# THE EFFECT OF NON-NORMALITY ON THE CUTSCORE OPERATING FUNCTION: ESTIMATION CORRECTNESS IN NON-NORMAL MONTE CARLO SIMULATIONS

By

Jesse R. Pace

Submitted to the graduate degree program in the Department of Educational Psychology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Education.

_____

Dr. John Poggio, Committee Chairperson

_____

Dr. Meagan Patterson, Committee Member

_____

Dr. Vicki Peyton, Committee Member

_____

Dr. Matthew Reynolds, Committee Member

_____

Dr. Suzanne Rice, Committee Member

Date Defended: 09/06/2019

The dissertation committee for Jesse Pace certifies that this is the approved version of the following dissertation:

THE EFFECT OF NON-NORMALITY ON THE CUTSCORE OPERATING FUNCTION: ESTIMATION CORRECTNESS IN NON-NORMAL MONTE CARLO SIMULATIONS

_____

Dr. John Poggio, Committee Chairperson

Date Approved: 09/10/2019

**Abstract**

Certification testing attempts to classify individuals into mutually exclusive categories, such as competent and non-competent. There is some potential for error whenever a classification decision is made as a result of a test score. The Grabovsky and Wainer cutscore operating function (GW-CSOF) is a recent addition to classification error estimates. This method allows for the prediction of error rates at all possible cutscore locations, but requires that certain assumptions about the examinee distribution are met. How the estimates made by the GW-CSOF compare to actual error values is currently unknown. Furthermore, the extent to which deviations from GW-CSOF assumptions impact error estimates is also unknown. The aim of this dissertation was to explore the extent to which non-normality of examinee true scores impacted the correctness of the GW-CSOF estimates. Monte Carlo methods were used to generate true score samples with systematically increased non-normality, and GW-CSOF estimates were compared to actual error rates. Findings indicated that GW-CSOF produced good estimates of error rates and optimal cutscore location in truly normal and minimally non-normal simulations. The degree to which GW-CSOF produced incorrect estimates was significantly correlated with the degree of non-normality. Specific guidelines for standard setting are discussed.

Acknowledgements

To Daryl Mellard, who taught me how to be a scientist. To Jonathan Templin, who taught me how to be a statistician. To Marianne Perie, who taught me how to be a psychometrician. And to John Poggio, who taught me valuable lessons in all of the above subjects, and who also taught me to not be "so damn naive". To those listed above, and to the remaining members of this committee: Drs. Rice, Reynolds, Peyton, and Patterson. You shared your knowledge with me, gave me opportunities to learn, and were available for questions and discussion. Thank you for embodying what academia is at its best.

To Irina Grabovsky, whose mentorship greatly improved my mathematics skills and confidence. I learned so much by working with you and the models you developed. Without your insightful mathematical formulations, this dissertation would not exist.

To my sisters and brothers, Misty, Elleni, Nik, and Bradford (Breeze). Thanks for the love, and for keeping life interesting.

To my friend and colleague Zhehan Jiang, who was an exemplar of how a student in our field should be, and a moving target of success to weigh myself against. I still haven't caught up; I doubt I ever will. But I will keep trying, and each time I improve, a bit of that improvement will be because of you and your example.

To my mother, Linda Pace, who was the only parent I ever needed; who saw to it that I had opportunities to succeed, even at great cost to herself. And who, perhaps most importantly of all, encouraged me to question everything, and by doing so, taught me how to truly learn. Mom - I went looking, but I still haven't found all the answers. Nor have I found a million dollars. I'll get back to you on both.

To the sun of my life, Kellie Amott, who showed me that learning anything worthwhile takes effort, and that learning what you really want to know is worth all the effort you can muster. You believed in me, even before I believed in myself. You are my best friend, my intellectual rival, and the best thing that ever happened to me. I have this degree, my knowledge of mathematics, and my sense of self because you challenged me to do something that I actually cared about. Thank you, and never stop challenging me.

Finally, to my other best friend, Bradley (Chet) Peterson. You said that math was for suckers. You encouraged me to play video games instead of study. You suggested that finishing my degree was less important than getting back home as quickly as possible. In many different ways, you told me that, no matter how different our paths may be, you will always be my best friend, and you will always want to spend time with me. I feel the same way about you. Thanks for that, it has meant the world to me.

Finally, for reals, to myself, and everyone who knows where I came from. Don’t be fooled by the doc that I got, I’m still (I’m still) Jeddy from the block.

# Table of Contents

## Chapter 1: Introduction

Classification is a process which involves making judgments in order to place observations into categories. There exist many statistical methods for classifying observations. Such methods arrive at a solution by attempting to minimize the number of incorrect classifications. When the true outcome of a classification decision is readily observable, such as whether or not a machine will work, logistic regression (e.g., Hastie, Tibshirani, and Friedman, 2009) is a suitable statistical classification method. Using training data from cases where the outcome was observed, alongside predictor variables, a model can be built by finding the parameters which maximize the likelihood of the observations. Thus, with an appropriate method such as logistic regression, various predictors of the outcome can be used to make a prediction about the classification status of a given observation prior to that status being directly observed. However, education is often concerned with latent constructs, where traits are not directly observable. There are methods which model latent variables as categorical variables, e.g., Diagnostic Classification Models (DCMs; Rupp, Templin, & Henson, 2010) resulting in classification decisions based on estimated probabilities. Another method commonly employed to classify individuals on latent variables is by using a continuous distribution for the latent variable, such as in Classical Test Theory or Item Response Theory (CTT and IRT, respectively; see Lord and Novick, 1968) and then selecting a cutscore along that continuum which divides the two categories. How this cutscore is set is arbitary, but one common way to do so is via standard setting. The standard setting process most often relies on subject matter experts (SMEs) to guide the identification of a reasonable point for a cutscore (e.g. Livingston & Zieky, 1989).

Whichever method is used to classify individuals, some amount of error is expected to occur. Such classification errors are of two forms: false positives (FP) and false negatives (FN).

Suppose a test is given to determine whether or not examinees are competent or non-competent in some academic subject. FPs occur when a non-competent examinee receives a passing score, while FNs occur when a competent examinee receives a failing score. The sum of the two errors provides an index of the total error present. In the case where a cutscore is used to divide a continuous latent distribution, any changes to that cutscore's location would likely change the total error as well. Thus, evaluating error rates across all possible cutscores allows one to determine the statistically optimal cut point (i.e., when the total error is minimized). Graphically speaking, this can be facilitated by plotting total error against all possible cutscores, and identifying the point where error is reaches its lowest point.

Determining the actual error after a test is given, and when there is a way to know the true status of an examinee, is straightforward and the calculation requires few assumptions. Other methods, via some strong statistical assumptions about the latent ability distribution, can estimate classification error. These methods are henceforth known as 'predictive' measures in this dissertation, in order to differentiate them from the actual error values as described above, and because they are, indeed, estimates of yet unknown actual values. One such predictive measure was recently developed by Grabovsky and Wainer (2017). This method, originally titled the "Cutscore Operating Function" (Grabovsky & Wainer, 2017a and b) and later as "the Grabovsky curve" (Wainer, 2017), is henceforth referred to as the GW-CSOF method to avoid confusion and to credit both of its original authors. This method estimates the error rate at all possible cutscore values, allowing one to estimate the statistically optimal cutscore. In order to provide this estimate, however, assumptions about the examinee population are required.

One of the key assumptions made by the GW-CSOF method is that examinee true scores are normally distributed. While this is a reasonable assumption in certain situations, it is not

always a valid approximation. Observed score distributions, which approximate the true score distribution to an extent depending on the size of the standard error of measurement, have been found to deviate notably from normality. Micceri (1989) found elements of non-normality in over 400 large-scale achievement tests. Among the problems found were distributions that were multimodal, tail weighted (kurtotic), or asymmetrical (skewed). Any deviation from the mathematical assumptions of a statistic can lead to errors in the estimates. To date, there has not been an exploration of how non-normality of true scores impacts the GW-CSOF method's estimates. In fact, there has not been an exploration of how the GW-CSOF method aligns with actual error values even when true scores are normally distributed. In other words, research has yet to answer the following questions: 1) Do GW-CSOF estimates of optimal cutscores match the actual location of the optimal cutscore, and does the match change as non-normality increases? 2) Do GW-CSOF estimates of error at the true cutscore location match actual error rates, and does the match change as non-normality increases?

These questions are important, as standard setting committees are often tasked with making important cutscore decisions, and being able to know how much error a certain cutscore might yield could be potent information. In order for such benefits to be realized, however, it is necessary to know how correct the method's estimates are. To answer the above research questions, a simulation design was proposed. Via simulation, normality can be systematically manipulated, and thus it is possible to determine the impact of increasing non-normality on the GW-CSOF method's estimates. The hypotheses of this dissertation were, I) The GW-CSOF method would produce error estimates close to actual error values when the normality assumptions of the true score distribution were met, II) Increased non-normality in the true score distribution would increase incorrectness in error estimates. For the above two hypotheses, the

comparisons were made at the location of the true cutscore, i.e., where the observed cutscore matched the value of the location on the true score that separated competent and non-competent examinees. In addition to the potential mismatch of error values at the true cutscore, the GW-CSOF method's estimated location of the optimal cutscore was also compared against the actual location. Thus, research hypothesis III, and IV paralleled I, and II. Hypothesis III) the GW-CSOF method would estimate a location for the optimal cutscore near the location of the actual optimal cutscore when normality assumptions of the true score distribution were met, IV) Increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore.

## Chapter 2: Literature Review

### Systematic Error and Validity

All measurements in the social and behavioral sciences, including education, are subject to measurement error. Classically speaking, measurement error takes two forms: systematic and random (Raykov & Marcoulides, 2011). Systematic errors occur when something inherent in a test or a testing situation depresses (or inflates) student scores. These sorts of errors are consistent: repeated testing would yield the same erroneous results. Systematic errors must be eliminated before a test can used (i.e., without eliminating systematic error, a given test is not validly assessing what it purports to measure). Random error, on the other hand, is expected to balance itself out over repeated testing. Random error is discussed further in this paper, but first attention is turned to the notion of what is meant by a valid assessment.

Validity refers to the appropriate inferences from an individual's test score. Historically, validity has often been discussed in terms of content, criterion, and construct validity (e.g., Helmstadter, 1964). Content validity refers to the extent to which a test completely and properly assesses the content it is purported to. This content is often referred to as the content domain, and a valid test, in terms of content validity, must contain items which span the entire content domain, and must allow examinees to respond in a way appropriate for what is being asked. Empirical validity, also known as criterion validity, refers to the extent to which a test can be shown to relate to the quality being assessed. Empirical validity takes two forms: concurrent and predictive (Thorndike, 1997, pg 143). Concurrent validity refers to the relationship between a test and another metric, such as a correlation between a test of mechanical ability and ratings of mechanical job performance. Predictive validity refers to a tests ability to predict relevant future outcomes, such as aptitude tests being used to predict later student success. Finally, construct validity refers to whether an assessment is actually measuring the latent construct that it purports

to measure. In the last three decades, consensus around what validity is and how best to document it have evolved.

Messick (1995) described validity as a "unified" concept. Thus, validity evidence, which historically were thought of as the above three separate forms of validity, come together to create a single pool of validity evidence for score interpretation. No one form of evidence is sufficient on its own, nor is any one form absolutely necessary: all that is required is that there is a, "compelling argument" that the evidence justifies the test interpretation. All of the evidence is thus unified under this argument.

Messick divided validity into a 2x2 table with four facets: interpretation, use, evidential basis and consequential basis. The evidential basis for test interpretation is essentially construct validity. The evidential basis of use is construct validity with the addition of some evidence supporting the relevance of a test to some use and some evidence from a cost/benefit (utility) point of view. The consequential basis of interpretation is also construct validity, but with the addition of value implications. Finally, the consequential basis of test use includes construct validity, relevance/utility, value implications, and the addition of social consequences – that is, what is the social impact of making a decision due to interpretation of a test score. This last notion is of particular importance in classification testing. If a license to practice a trade is withheld from an individual, that individual will lack job opportunities they might otherwise have had, which could have serious consequences on their livelihood. Likewise, if a license to practice a trade is given to an individual, then the public should be able to reasonably expect that the individual is competent in that trade. The consequences of giving a non-competent person a license to practice in such fields as aviation or medicine can be fatal. Thus, consequential validity is directly related to classification error. As indicated in the beginning of this section,

one piece of validity evidence is a lack of systematic error. Thus far, random error has not been discussed in this dissertation, but it too is directly related to classification error, as it is discussed in the next section.

**Random Error and Reliability**

In classical test theory, random errors of measurement affect an individual's score due to pure chance effects, and these effects are temporary (Raykov & Marcoulides, 2011). If a test is administered repeatedly, it is expected that random measurement error will balance itself out (i.e., that its expected value is zero). This error is part of the classical test theory equation, X=T+E, where X is a student's observed score, T is a student's true score, and E is the error of measurement (Lord and Novick, 1968). The standard deviation of this error is known as the standard error of measurement, and it is presented in further mathematical form in the next section of this dissertation. The degree to which a test produces the same results for the same inputs (i.e., same person at same time with the same knowledge and skill) is called test reliability (AERA, APA, NCME, 2014). If a test is entirely reliable, then it produces no random error, and every time this test is administered to an individual, the resulting score is the true score of that individual. When reliability is not perfect, random error is present. The relationship between random error, reliability, and true scores is quantified in classical test theory. Specifically, reliability is equal to the ratio of true score variance over observed score variance (Lord and Novick, 1968). Random errors can lead to incorrect classification decisions whenever a given examinee receives an observed score that differs from their true score. Determining the amount of error made by a classification test, then, is largely related to that test's random error component. This topic is further explicated in the methods section of this dissertation. First, the topic of classification accuracy warrants further attention.

## Classification Accuracy

Classification accuracy describes the extent to which classifications have been performed correctly. That is to say, that accuracy is the opposite of error: as classification errors go up, the accuracy of classification goes down. The topic of accuracy has been important in the literature surrounding classification decisions for a long time. The terms: false positive, true positive, false negative, and true negative abound in the literature, and have been present in medicine for at least the last 100 years (e.g., Solomon, 1920). A false positive occurs when a competent examinee is given a failing score. Conversely, a false negative occurs when a non-competent examinee is given a passing score. True positives and true negatives occur when competent examinees pass, and non-competent examinees fail, respectively. The sum of false positive and false negative errors can be termed 'total classification errors': as these two classifications represent all cases which have been misclassified. This metric is convenient because it consolidates error into a single metric. As an example, consider the case where an exam is administered to 10,000 students. Suppose that 100 students are incorrectly given passing scores, when they should have failed (false positive errors) and that 200 students are incorrectly given failing scores when they should have passed (false negative errors). The total error, then, is 300 of the total 10,000 examinees, which is a total error rate of .03, or 3%.

One method that can be used to determine a test's total error rate is to use a 'gold standard', which is believed to represent the knowable truth about examinee competency status (Feuerman, & Miller, 2008). Another method for determining error occurs when the latent value measured by a test becomes observable, such as Alzheimer's disease which can be definitively diagnosed post mortem (Dubois et al., 2010). When knowledge of the true state of examinee's classifications is knowable, it becomes possible to calculate false positive and false negative errors. These instances are termed 'actual' error values in this dissertation.

Several authors have expanded on the total error metric. Yerushalmy (1947) first defined the now ubiquitous terms: sensitivity and specificity. Sensitivity is defined as the ratio of examinees who are correctly classified as competent (true positives) to the total number of positive classifications (including both true and false positives). Specificity is the ratio of true negatives relative to the total number of examinees who were classified as a negative. Youden built upon Yerushalmy's work by combining sensitivity and specificity into a single metric: Youden's J. This metric is calculated as sensitivity + specificity – 1. Cohen (1960) described another metric, known today as Cohen's kappa, which is used to view overall agreement between two classifications. This method takes the additional step of removing an estimate of the chance agreement between the two methods. That is, Cohen's kappa calculates the probability that assessment A classifies an examine as competent, as well as the probability that assessment B classifies an examinee as competent, and takes their product, assuming rater's are independent, and removes that from the overall rate of agreement. Likewise, chance agreement where both assessments A and B agree that an examinee is non-competent are also removed. One of these assessments could be operationalized as the gold standard described earlier, and thus Cohen's kappa provides an index of classification error.

Using a metric such as total error, it is possible to determine the cutscore where error is minimized. For example, total error can be calculated at every potential cutscore along a test's scale range. The potential cutscore where the total error is smallest is the location with the least error. Such a point is henceforth termed the statistically optimal cutscore. While the statistically optimal cutscore is a defensible option for setting a cut-point, standard setting offers an alternative.

## Standard Setting

Standard setting is the process of establishing levels that separate examinees into different performance categories (Cizek, 2012). This process is often operationalized in the setting of cutscores, which represent the minimum score on a test necessary for classification in the category they represent. The standard setting procedure used to estimate this cutscore often involves the use of judges (e.g. Livingston & Zieky, 1989). In licensure testing, there is typically a single cutscore, as only two categories (e.g. competent vs. non-competent) are needed. Determining the cutscore that best separates examinees who are minimally competent from those who are non-competent, then, is the object of standard setting in licensure testing. Standard setting can be performed in a variety of ways, and these different methods can be classified into one of two categories: test centered and examinee centered (Jaeger, 1989 as cited in Kane, 1994):

**Test centered methods**. Test centered methods require judges to make judgments about the test content (Cizek, 2012). This often requires judges to review items and decide on the level of performance on each item necessary to be considered competent (e.g., Kane, 1994). There are several variations of the test centered method.

***Angoff.*** Angoff's 1971 chapter on scales (reprinted in 1984) was the first mention of this method. In short, the Angoff method utilizes experts to decide the number of items a minimally competent examinee (MCE) would be able to answer on a given test. Often, individual judges will decide on an ideal cutscore after viewing each test item and rating it as either a 1 or a 0 (or they will assign probabilities between 0 and 1). The average of the judges ratings will typically be the number established as the optimal cutscore (Hurtz & Auerbach, 2003). There are many different modifications to the Angoff method which have been developed over the decades since Angoff's 1971 paper, but all still follow the same general procedure of tasking judges to determine probabilities that the MCE could answer questions correctly (Plake & Cizek, 2012).

***Ebel's method.*** Ebel (1972, as cited in Kane, 1994) requires judges to categorize items by both difficulty and relevance to the construct being measured. The number of categories for each decision need to be decided upon beforehand (e.g., easy, medium, difficult; not relevant, somewhat, very). Judges then assign a value to each cell of the resulting difficulty by relevance matrix. This value represents the number of items a MCE would get right in that cell. These values are then converted to proportions and summed to give each judge's estimate of the cutscore.

***Nedelsky method.*** The Nedelsky (1956) method is purposed exclusively for multiple-choice items. Judges examine each item, and eliminate the response options that the MCE would be able to determine were incorrect. The reciprocal of the remaining number of choices is the probability that the MCE would get that item correct. The sum of all such probabilities over items on the test is the MCE's expected passing score: i.e., the specific judge's cutscore.

***Bookmark.*** The bookmark method is an item response theory-based procedure (Lewis, Mitzel, Mercado, & Schulz, 2012; Cizek & Bunch, 2007). This method utilizes item difficulties calculated during item calibration. Items are rank ordered from easiest to most difficult, and content experts are tasked with locating the position in the order where a MCE would have a high probability of success (this probability is often operationalized as .67). Once the item has been selected, the theta value corresponding to the selected probability and difficulty of the item can be calculated.

**Examinee centered methods.** These methods ask judges to use knowledge of actual students in order to determine the cutscore.

***Borderline-group method.*** The borderline-groups method (Livingston & Zieky, 1982) identifies actual students as 'borderline' examinees. This method requires a sample of students to

take the exam for which the cutscore is being set. Judges then identify which examinees they believe are borderline students, i.e., judges do not use the student's scores, but rather their personal knowledge of the students to categorize them. Then, the average of the scores of the students that have been categorized into the borderline category is used as the cutscore.

***Contrasting (Critrion)-groups method.*** The contrasting-group method (Livingston & Zieky, 1982) requires judges to assign a label of 'qualified' or 'not qualified' to each member of a sample of examinees. This assignment is not based on examinee scores, but instead on some separate criterion, such as judgments of their knowledge and skill. Once examinees have been divided into these groups, student test scores are analyzed using the resulting examinee cumulative distribution (using % qualified) to select the cutscore that corresponds to the .5 proportion of % qualified examinees. In cases where the sample is small, and sparsity of data is present, smoothing techniques are used to smooth the cumulative density function (CDF).

Standard setting techniques are commonly used to establish cutscores on education tests. While their presentation thus far in this paper has put them at odds with the statistically optimal cutscore, they are not truly opposites. Indeed, methods exist which attempt to utilize both statistical and standard setting information in determining the cutscore. Among these methods is the primary topic of this paper: the GW-CSOF method.

## The GW-CSOF Method

Grabovsky and Wainer (2017a & 2017b) described a method for estimating the optimal cutscore. The technique was created with the hope of aiding standard setting committees in choosing operational cutscores, by providing additional information standard setting committee members might not otherwise have had. As a starting point, the method requires the user (e.g., a standard setting committee) to specify where they believe the cutscore should be. The authors

recommend the standard setting techniques as methods to determine this point, as the point is conceptualized as the best guess of the 'true' cutscore: i.e., as the point that truly separates competent from non-competent examinees. The authors also noted that the location of the optimal cutscore may differ from the believed location of the true cutscore. The GW-CSOF method estimates classification error at the inputted value of the true cutscore, and also determines the accuracy at all other potential cutscores, in order to determine the statistically optimal cutscore. Thus, the GW-CSOF method is one which marries the concepts of standard setting and statistically optimal cutscores.

The GW-CSOF method utilizes classical test theory (CTT) conceptualizations of the relationship of observed scores and true scores in their mathematical explication, but could also be utilized with IRT methods. The CTT theory utilized in Grabovsky and Wainer (2017a) includes the properties of expected values of observed scores (being equal to true scores), which follow from the classic T= X - E assumption where the expected value of E is zero. The resulting derivations are also utilized: $\rho = \sigma_t^2/\sigma_x^2$, and $\sigma_x^2 = \sigma_T^2 + \sigma_E^2$. Thus, $\sigma_E^2 = \sigma_x^2 - \sigma_x^2 * \rho$, and it follows that, $\sigma_E^2 = \sigma_x^2(1-\rho)$. Where $\rho$ is the test reliability, $\sigma_t^2$ is true score variance, $\sigma_x^2$ is observed score variance, and $\sigma_E^2$is error variance, which is often termed the standard error of measurement (SEM; e.g., Harvill, 1991). Thus, error variance and true score variance can both be estimated using knowledge of reliability and observed score variance. It is also true that knowledge of true score variance and reliability leads to knowledge of standard error and observed score variance. The latter point is important for the methods developed in this study.

The GW-CSOF method is not the only predictive error measure, nor the first; other methods are described in the literature. Livingston (1993) described a method for estimating error that was found to be very close to actual values. Lee (2010) explicated a method that could

be used for such estimation with complicated item-response theory based assessments. Rudner (2001) described another model for estimating accuracy, but also went on to demonstrate how accuracy estimates change if the cutscore is altered. The GW-CSOF method incorporates this last idea, of looking at error at different possible points of the cutscore, into its mathematical model. GW-CSOF was the first method of predictive error estimation designed to identify the predicted error across all possible cutscores, allowing for the user to find the optimal error point.

In the GW-CSOF method, individual student ability, as indexed by true scores, is assumed to be normally distributed. The distribution can be estimated using error variance as the variance around a given examinee's true score.

The probability that a given examinee with observed score, x, and true score, τ, will obtain a score below a given cut point, c, using standard properties of a normally distributed variable, is:

$$p(x < c) = \mathrm{p}\left(\frac{x-\tau}{\sigma_E} < \frac{c-\tau}{\sigma_E}\right) = \mathrm{p}\left(\mathrm{z} < \frac{c-\tau}{\sigma_E}\right) =$$

$$\int_{-\infty}^{\frac{c-\tau}{\sigma_E^2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) dy \qquad (1)$$

Where z is a standard normal random variable. This follows from the fact that, for a given examinee, τ is the expected value of their observed score, and that individual's error variance (the variance between around the mean of their observed score) is estimatable by $\sigma_E^2$ (see Harvill, 1991).

The probability that an examinee will pass a given tests, then, is: $1 - p(x < c)$

If a student is selected at random from the true score population, the probability that the student will have τ less than the true cutscore, $\tau^*$, and pass the exam (i.e., a false positive error), is:

$$p(\text{FP}) = \int_{-\infty}^{\tau^*} [1 - p(x < c)] \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\tau - \bar{x}}{2\sigma_t^2}\right) d\tau \quad (2)$$

And the probability of a false negative error (i.e., that an examinee will have τ greater than the true cutscore and fail the exam) is:

$$p(\text{FN}) = \int_{\tau^*}^{\infty} p(x < c) \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\tau - \bar{x}}{2\sigma_t^2}\right) d\tau \quad (3)$$

Both p(FN) and p(FP) assume that the examinee true score distribution is normal, as indicated in their respective integrands. Deviation from this assumption may lead to incorrect estimates of classification accuracy.

True score distributions are rarely known in practice, and assumptions about their form may lead to incorrect results. This dissertation investigated the degree to which deviation from normality of the true score distribution impacts the correctness of the GW-CSOF method's estimates. This was accomplished via a simulation design in which a given simulated examinee's true score is known, as is their observed score, and the true cutscore is specified. Thus, actual accuracy is readily computable. The simulations varied the degree of non-normality, and the extent to which these deviations impact the GW-CSOF method's estimates, as compared to actual error values, was determined This research was important as the GW-CSOF method can provide valuable information to standard setting committees in their work to choose cutscores. However, it was crucial to know the extent to which violations to the method's assumptions might lead to incorrect estimates.

**Summary**

This review of the literature has expounded on the variables of interest of this dissertation. Most importantly, the GW-CSOF method is a recently conceived method for estimating optimal cutscores prior to test administration. GW-CSOF holds great potential utility as another source of information for standard setting committees, and serves as a method to integrate both standard setting and optimal cutscore methods of setting cutscores. However, the extent to which the GW-CSOF method matches actual error values remains unknown. In terms of application, misestimation of error and optimal cutscore location could have substantial impacts on standard setting. For example, in a situation in which GW-CSOF indicates an incorrect location for the optimal cutscore, the standard setting committee could very well end up choosing a point which differs from their theoretical cutscore, only to arrive at a point, which in reality, has even more error than if they had not altered the cutscore location at all. Reasons such as this were the motivation of this dissertation. It was hypothesized that GW-CSOF error would match actual error values when normality assumptions were not met, and that increased non-normality would lead to increased mismatch. Thus, one area of focus was on comparing GW-CSOF to actual error values in a truly normal simulation. Additionally, systematically increased non-normality was utilized to determine how GW-CSOF estimates matched actual error values under different normality manipulations. Normality manipulations were conducted to create skewed, bimodal, and kurtotic distributions. The overall purpose of the present dissertation was to explore the degree to which the GW-CSOF method corresponds to actual classification error rates under varying degrees of non-normality of the true score distribution. A simulation design was chosen to allow for a proper exploration of these unknowns.

## Chapter 3: Methods

The present dissertation sought to test the extent to which GW-CSOF estimates of error and optimal cutscore location match actual error rates and the actual optimal cutscore location. This work was important, as the GW-CSOF can potentially be used by standard setting panels as an additional source of information. Its use could then potentially lead a panel to choose a different cutscore than they might have otherwise. Cutscores are utilized in assessment (e.g., education, medicine) to make diagnostic decisions, such as whether a student is qualified or not, or whether a disease is present or not. Thus, use of the GW-CSOF has the potential to impact one of the most important decisions in assessment planning. Ideally, the GW-CSOF will provide useful information to standard setting panels, guiding them in choosing a cutscore which minimizes the rate of misclassifications: thus, reducing the number of examinees mishandled by the assessment process. In order to maximize the utility of the GW-CSOF, however, it is necessary to determine under which circumstances its estimates are valid. This was the focus of the current dissertation. This section explicates the methodology used to systematically manipulate non-normality of the true score distribution (i.e., the distribution of examinee's true abilities), in order to determine how such manipulations alter the correctness of the GW-CSOF estimates.

In order to investigate the research hypotheses, it was necessary to have knowledge of the true score distribution's shape as well as knowledge of the resulting properties of the observed score sample resulting from that true score distribution. A simulation design, in which the true score distribution was specified and used to generate true score observations, which were in turn used to generate observed score observations, provided a suitable basis for this research.

**Simulations**

This dissertation used Monte Carlo methods to simulate examinee test scores. A true score distribution was specified, as well as a true cutscore and test reliability. Using the specifications of the true score distribution, a sample of 10,000 true scores was generated. Using the true score simulations and test reliability, an observed score was simulated for each examinee, resulting in a simulated observed score sample. This observed score sample was used to calculate actual classification error: that is, each simulated examinee had a known true score and a known observed score, making for direct comparison. Four separate true cutscores were used in order to determine the differing effects of a true cutscore which was: i) extremely below the mean, ii) below the mean, iii) above the mean, or iv) extremely above the mean.

The mean of the true score distribution was fixed at 50 with a standard deviation 5. These values were chosen because they provided for a good fit to a 0 to 100 scale, which is a common scale in education (see Scale section below for more details on the scale). True cutscores were set at 45, 47.5, 52.5, and 55, in four separate analyses. These correspond to -1, -.5, +.5, and +1 standard deviations below and above the mean, respectively, representing the results when the cutscore is set near the mean, as well as when it set a considerable distance away, in either direction. This produced four different conditions that had straightforward connotations: a very easy test, where about 84% of students should pass, a somewhat easy test, where about 69% of students should pass, a somewhat hard test where about 31% of students should pass, and a very hard test where only about 16% of students should pass. A reliability of .8 is generally considered satisfactory in applied research (e.g., Raykov & Marcoulides, 2011), and this value was used in the simulations. From CTT we have that $\rho = \sigma_t^2/\sigma_x^2$, and, $\sigma_E^2 = \sigma_x^2 - \sigma_t^2$. Thus,

$$\sigma_E^2 = \frac{\sigma_t^2}{\rho} - \sigma_t^2$$

Where $\rho$ is the test reliability, $\sigma_t^2$ is true score variance, $\sigma_x^2$ is observed score variance, and $\sigma_E^2$is error variance, which is often termed the standard error of measurement (SEM; e.g., Harvill, 1991). It follows that, using knowledge of the true score variance and reliability we were able to calculate error variance,

$$\sigma_E^2 = \sigma_T^2\left(\frac{1}{\rho} - 1\right) \tag{4}$$

Thus, SEM was calculable using the known values of true score variance and reliability. Although the simulated true score sample variance differed slightly from the true score distribution variance, the differences were slight due to the large sample size (see Simulation Quality in Results section). Thus, the error variance that was used for all calculations was $25(1/.8 - 1) = 6.25$, and SEM was $\sqrt{6.25} = 2.5$.

The observed score sample was generated from the true score sample by use of the SEM. The SEM describes the standard deviation around a given examinee's true score (the mean of their observed scores), and this distribution was assumed to be normal. For each simulated examinee, then, their observed score is simply a random variable to be sampled from their individual observed score distribution, which was completely specified by their true score simulated value and the SEM. That is, using SEM and a given examinee's true score value, a random draw from a normal observed score distribution (~N(true score, SEM)) could be taken, which resulted in their simulated observed score.

Actual classification error was compared to estimates produced by the GW-CSOF method. Using the true cutscore, the reliability of the test, and the true score mean, the GW-CSOF method estimated the error for each possible value of the cutscore, and indicated the

location of the optimal cutscore. This method was compared directly with the actual error and actual optimal cutscore location.

**Scale**

While all of the simulation criteria were specifiable without need for a specific scale range, the GW-CSOF method requires a scale over which to search for the optimal cutscore. The scale range that was used in this study ranged from 0 to 100. This range fits nicely with the proposed true score distribution, and the probability of any score falling above 100 or below zero is infinitesimal. That is, the expected observed score standard deviation is $\sqrt{\frac{\sigma_T^2}{\rho}} = \sqrt{\frac{25}{.8}} = 5.59$ Thus, 99% of the sample in a truly normal case was within three standard deviations of 50, (33.23,66.77).

**Normality Manipulations**

Generally speaking, distributions need not be normal. Indeed, the normal distribution is but one of many continuous distributions one would find in an introductory text on mathematical statistics (e.g., Hogg, Tanis and Zimmerman, 2015). When determining whether or not a sample appears to have come from a normal distribution, introductory texts commonly present several criteria to look for. These criteria often include indices of skewness and kurtosis (e.g., Coladarci & Cobb, 2014) and multimodality (e.g., Glassnap & Poggio, 1985). All three of these criteria were observed by Micceri's (1989) research on violations to normality of observed score distributions. These three exceptions to normality were systematically manipulated in the present dissertation. Specifically, normality was increasingly distorted over the course of 50 simulations within each of the three conditions. A sample of 100 has been shown to be a sufficiently large sample to detect even small effects using Spearman's rho (Yue, Pilon, & Cavadias, 2002).

However, following the recommendation of the dissertation committee, this dissertation created 50 simulations per condition.

For example, in the skewness condition, each of the 50 simulations had increasingly more skewness. The total error (FP+FN) at the true cutscore was tabulated and compared between the GW-CSOF method and the simulated results, and the estimated optimal error location was compared to the actual location.

**True Normal Distribution**

A single truly normal distribution was simulated. This provided an opportunity to test the effectiveness of the GW-CSOF method against actual classification error rates, something not performed in the published GW-CSOF literature. This also provided a baseline for the rest of the comparisons, as all other simulations were intentionally violating the normality assumptions and thus, compared to the truely normal distribution, were expected to produce worse matches between GW-CSOF estimates and actual errors.

This manipulation was conducted to confirm or refute Research Hypotheses 1 and 2: The GW-CSOF method would produce error estimates close to actual error values when the normality assumptions of the true score distribution were met, and the GW-CSOF method would estimate a location for the optimal cutscore near the location of the actual optimal cutscore when normality assumptions of the true score distribution were met.

**Skewness**

Skewness values for normally distributed random variables generally range from ±3, with 0 indicating a symmetric distribution (Glassnap and Poggio, 1985). As an example of a skewed distribution, consider the case where most examinees do very well on an exam (scoring near the

maximum of the possible range), but a handful of students perform poorly (scoring near the minimum of the possible range). Such a distribution would be negatively skewed.

In order to generate a skewed examinee sample, it is necessary to sample from a distribution with a known skewness value. The exponentially modified normal distribution (e.g., Zabell, Foxworthy, Eaton, and Julian, 2014) describes the sum of two random variables, one from a normal distribution and the other from an exponential distribution. The resulting distribution has, as a function of $\lambda$ (introduced below), a normal density except for a positive skew.

Let S be a random variable, S is defined: $S = X + Y$

Let X be a normally distributed variable with variance $\sigma^2$ and mean μ, and Y be exponentially distributed with mean $\frac{1}{\lambda}$, with X and Y independent. As we transform from X,Y coordinates to S, Y coordinates, the resulting joint distribution, *f*(s,y), of S and Y, given by the transformation formula (Hogg, Mckean, & Craig, 2014) with Jacobian J=1 is:

$$f(\mathrm{s},\mathrm{y}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s-y-\mu}{2\sigma^2}\right) \lambda \exp(-\lambda y) * 1 \tag{5}$$

Marginalizing over this joint pdf with respect to y produces the marginal pdf of S, *f*(s):

$$f(\mathrm{s}) = \int_0^s \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s-x-\mu}{2\sigma^2}\right) \lambda \exp(-\lambda y)\, dy \tag{6}$$

Via independence, the mean is simply the sum of the means, and the variance is simply the sum of the variances. The skewness, as presented in Grushka (1972), is

$$\frac{2\theta^3}{(\sigma^2 + \theta^2)^{3/2}} \quad (7)$$

Where $\theta = \frac{1}{\lambda}$

Thus, the mean, variance, and skewness of the simulated examinee distributions could be specified. For consistency, skewness was increased while the mean and variance was held constant, via manipulation of the three input parameters (i.e., $\sigma^2$, μ, and $\lambda$). See Appendix A for derivation of the formulas used for this purpose. While the resulting simulated examinee data was slightly different than the distributions used to generate it, the size (10,000) yielded numbers that match very closely with the intended outcomes. The skewness of the exponentially modified normal distribution ranges from 0 to +2, and it is over this interval that the 50 simulations were conducted. Specifically, in order to have 50 total simulations of increasing skewness, [0, 2), steps of .04 were taken. The first simulation used skewness of 0, and the last used skewness of 1.96.

The skewness manipulations were conducted to confirm or refute Research Hypotheses II and IV: Increased non-normality in the true score distribution would cause increased incorrectness in error estimates, and increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore. Specifically, as skewness increased, density at the 45, 47.5, and 52.5 locations was expected to decrease. It was expected that the relative decrease in density near these points would lead to shifts in error estimates, as well as optimal cutscores. Because the skewness is increasing near the 55 location, there was no clear expectation about how error might change at that point.

**Bimodality**

A normal distribution has a single mode. The extent to which two modes are present is the extent to which a distribution can be termed ‘bimodal’. As an example of a bimodal

distribution, consider the case when examinee samples are drawn from different countries, i.e., such as an exam that was developed for use in the United States, such as a US college entrance exam, where many students from the US as well as from other countries participate in the examination. It may well be the case that US and non-US test takers differ in terms of their average test score. It may also be that both groups exhibit a relatively normal distribution about their respective means. In such a testing situation, if the separation of the two groups' modes is sufficiently large, the overall examinee distribution will be bimodal.

Bimodality was generated using a mixture distribution of two normals. In general, a mixture of two distributions is composed of their respective pdfs and a mixing probability (Hogg, Mckean, & Craig, 2014). Let Z be a mixture, and X and Y two independent random variables with their own distributions, with mixing probability w. Let I be an indicator function with I=1 with probability w, and I=0 with probability 1-w, then:

$$Z = IX + (1 - I)Y \tag{8}$$

And,

$$f(\mathrm{z}) = \mathrm{w}f(\mathrm{x}) + (1 - w)f(\mathrm{y}) \tag{9}$$

Mixtures of normals can be used to generate bimodal distributions (Rossi, 2014). This is accomplished by sampling from normals with different means. In the case where both distributions have the same mean and same variance, a single normal distribution is present, and thus there is no bimodality.

The mean of a mixture, $\mu_m$, of normals is given by Behboodian (1969) as

$$\mu_m = w\mu_1 + (w - 1)\mu_2 \tag{10}$$

Where $\mu_1$ and $\mu_2$ are the means of X and Y, respectively.

The variance, $\sigma_m^2$, is

$$\sigma_m^2 = w(\sigma_1^2 + {\mu_1}^2) + w(\sigma_2^2 + {\mu_2}^2) - {\mu_m}^2 \quad (11)$$

Let $\sigma_1^2 = \sigma_2^2$, so that both normals have the same variance. This also makes solving for the distance between modes algebraically possible; see Appendix B for derivation. Bimodal simulations were carried out using 50 successively more bimodal distributions. Following the definition used by Micceri (1989), a distribution where two modes were present and the modes had, "distances greater than two thirds (.667) of a distribution's standard deviation were defined as bimodal."

Here, the standard deviation and mean of the resulting mixture was held constant via manipulations of the input normals, maintaining the mixture mean of 50 and standard deviation of 5. A weight of .5 was used for w, resulting in equal probabilities that a given draw came from either of the two normal distributions. Let D be the distance between two means from two normal distributions, as specified above. It can be shown that the maximum distance D, while holding mixture variance and mean constant, as well as the variance of the two input normals equal, is strictly less than two times the standard deviation of the mixture standard deviation. That is, the input normals would each have to have variance of zero to produce the desired mixture variance of a D. In the current case, this means that D had to be strictly less than 10. Thus, the stopping condition was set at a distance of means up to but not including D=10. This point was just below two times the standard deviation of 5, and thus certainly met the criteria set forth in Micceri. In order to have 50 iterations of increasing bimodality between 0 and 10, steps of .2 was taken.

The bimodal manipulations were conducted to confirm or refute Research Hypotheses II and IV: Increased non-normality in the true score distribution would cause increased incorrectness in error estimates, and increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore. Specifically, as bimodality increased, density near the middle of the distribution was expected to move toward zero. Due to this decrease in density, shifts were expected in error estimates and optimal cutscores for the 47.5 and 52.5 true cutscores, as they were located near these shifting densities. Meanwhile, error was expected to increase for cutscores further out (at 45 and 55).

**Kurtosis**

Normal distributions typically have a kurtosis value of 3, with values less than 3 indicating a platykurtic distribution, and values above 3 indicating a leptokurtic distribution (Glassnap and Poggio, 1985). As an applied example of a kurtotic distribution, consider again the case where there is a US college entrance exam taken by both US students and non-US students. It is possible that both US and non-US students, on average, perform the same. That is, that the means of each group might have the same location. However, it also possible that the variability (i.e., the variance) around the mean of the non-US group might be larger than that around the US group. Should this prove to be the case, then the overall testing sample may take on positive kurtosis.

The mixture of two normals, each with the same mean, but different variances, produces a normal distribution with excess kurtosis (An and Ahmed, 2008). A mixture of two normals, as defined earlier in the bimodal section, is composed of two normal distributions and the probabilities of drawing from each of the two distributions. The mean and variance are given in the bimodal section.

The Kurtosis of the mixture is equal to

$$\frac{3(w_1\sigma_1^4 + w_2\sigma_2^4)}{(w_1\sigma_1^2 + w_2\sigma_2^2)^2} \quad (12)$$

The kurtosis is maximized when $w_1 = \sigma_2^2/(\sigma_1^2 + \sigma_{2)}^2$ and $w_2 = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ , which results in a maximum value of

$$\frac{3}{4}\left(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} + 2\right) \quad (13)$$

The kurtosis of the mixture of normal distributions ranges from 3 to 6. It is over this range that kurtosis was manipulated. Specifically, in order to have 50 total simulations of increasing kurtosis [3,6), steps of .06 was taken. The first simulation used kurtosis of 3, and the last used kurtosis of 5.94.

The kurtosis manipulations were conducted to confirm or refute Research Hypotheses II and IV: Increased non-normality in the true score distribution would cause increased incorrectness in error estimates, and increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore. Specifically, as kurtosis increased, density near the center of the distribution was expected to increase, while density further out was expected to decrease. Thus, it was expected that overall error would increase near the center, affecting the 47.5 and 52.5 conditions, and that error would shrink for the 45 and 55 conditions.

**Software and Hardware**

R software (R Core team, 2017) was used to generate the specified distributions. For the two mixture distributions, sampling was coerced to contain exactly 50% of each component of the mixture. This corresponds to the expected proportions (i.e., the average number of simulated values from each distribution if many simulations were generated using random Bernoulli

variables for each draw from a mixture), and allows for a more consistent true score distribution. For the convolution (skewness), two samples, one normal and one exponential, each of 10,000 cases, was simulated separately, and then summed along their index of creation. The GW-CSOF algorithm (equations 1, 2 and 3) was coded into R. The integrals were solved using R's built-in numeric integration functions, and steps of .1 of possible cutscores were taken between 0 and 100 (the possible score range) to determine the optimal cutscore. The code for the GW-CSOF, and the code for the skewness, bimodal, and kurtosis manipulations, when the true score was set to 45, are contained in appendix H. If the reader is interested in generating the 47.5, 52.5, or 55 conditions, universal replacement of values of 45 in the corresponding code will suffice to produce the intended result.

Hardware used for the simulations was a PC with an 8$^{th}$ generation Intel Core i-5 1.6 GHz quad-core processor, 8GB of ram, with an NVIDIA GeForce MX150 graphics processor, running Windows 10 for its operating system.

**Points of Comparison**

Each iteration of the non-normal simulations created two points of comparison between the GW-CSOF method and the actual error rates: the error rate at the true cutscore, and the location of the optimal cutscore. As non-normality increased, whether by increasing skewness, bimodality, or kurtosis, it was expected that the differences between GW-CSOF estimates and actual error values would increase. To test this assumption, a correlation between the difference in total error and the magnitude of non-normality was calculated. The correlation between the magnitude of non-normality and distance between the actual and estimated optimal cutscore was also be calculated. Spearman's rho (e.g., Hays, 1973, pg. 788) was chosen to calculate correlation. This technique is a non-parametric procedure, and is distributed asymptotically

normal (Ornstein & Lyhagen, 2016). There was no reason to expect that the distribution the increase in non-normality (on any of the three metrics) would follow any particular distribution, except that it was ordinal. The same was true of the differences between the actual and estimated error values. Thus, the proposed non-parametric procedure was used. When ties occurred, average rank was assigned to each member of the tie, as indicated in Hays, 1973, pg. 791. An alpha level of .01 was used in order to be conservative about conclusions of significant relationships. This is particularly important when using Spearman's rho with tie correction, as the significance levels are not determined exactly in the presence of ties, and thus it is helpful to err on the side of caution in interpreting resulting statistical significance.

These analyses will provide an indication of whether the incorrectness of the GW-CSOF method is related to the degree of non-normality.

**Summary**

This chapter has presented the methodology that was used in this dissertation. This dissertation used a simulation design to generate true scores from specified distributions with known characteristics, including non-normality. These simulated true score samples were then used to generate observed score samples which were used to calculate actual error values. This design determined the degree to which the GW-CSOF method's estimates matched actual error values in cases where true scores are normally distributed, as well as when true scores were non-normal.

# Chapter 4: Results

The present dissertation sought to investigate the correctness of the GW-CSOF under manipulations of normality of the examinee true score distribution. The GW-CSOF is a method for predicting classification error, and estimating the location of the optimal cutscore (i.e., where error is minimized). The GW-CSOF has potential utility for standard setting panels, enabling them to select cutscores which minimizes the number of misclassified examinees, as well as providing information about error at other potential cutscores. The present dissertation has sought to provide additional information for standard setting panels, particularly about how the GW-CSOF performs when its model assumptions are violated.

Based on the methodology proposed, there were 50 simulations per condition (i.e., skewness, bimodality, and kurtosis), and four subsets of simulation trials (for four different true cutscores), resulting in 600 total simulations. For all conditions, the 1$^{st}$ simulation was a truly normal simulation, and each subsequent simulation had increased non-normality. Comparisons were made between the GW-CSOF estimates of error at the true cutscore, as well as comparisons of the optimal error location.

Let $\Delta L$ be the difference in optimal cutscore location, and let $\Delta T$ be the difference in error rates when the observed cutscore is set equal to the true cutscore. Let $\Delta N$ be the change in non-normality in a given simulation. The proceeding results are divided into a section on the relationship between $\Delta L$ and $\Delta N$, and a section on the relationship between $\Delta T$ and $\Delta N$.

## Simulation Quality

As can be seen in Appendix D, the simulation results matched closely with desired values. For both the skewness and kurtosis manipulations, simulated true score sample values matched almost exactly with the desired distribution values, and the relationship between

increased distributional non-normality and simulated true score non-normality was almost monotonic.

The results of bimodal simulations were more complicated. Visual inspection of the 50 true score histograms was necessary to determine when bimodality became present. This inspection revealed that a truly bimodal distribution did not develop until the gap in the mixture means was greater than or equal to seven. Twenty images, beginning from a distribution mean gap of six, to the end of the simulations (gap of 9.8) are presented in Appendix F. As can be seen, true bimodality did not present until F21. Because bimodality did not present until the final 15 simulations, only those simulations were used to determine the relationship between increasing bimodality and the correctness of the GW-CSOF estimates.

**Truly Normal Case**

**ΔL results.** The top row of Appendices Appendix E, F and G all contain the same information: the results from a truly normal simulation. As can be seen by comparing Tables E1 and E3, the GW-CSOF estimates of the optimal cutscore were nearly identical to the actual values: 43.6 vs. 43.7 for true cutscore of 45, 46.8 vs. 46.8 for true cutscore of 47.5, 53.1 vs. 53 for true cutscore of 52.5, and 56.2 vs. 56.8 for true cutscore of 55.

**ΔT results.** By comparing Tables E2 and E4, it can be seen that, when the observed cutscore was set to be equal to the true cutscore, GW-CSOF estimates of error were nearly identical to actual values. Total error estimates vs. actual values were, for the 45, 47.5, 52.5, and 55 conditions, respectively: 0.095 vs. 0.088, 0.131 vs 0.134, 0.134 vs 0.131, and 0.099 vs 0.098. Thus, fewer than 1 in every 100 examinees would be classified differently between the GW-CSOF estimates and the actual values.

**Skewness**

Table 1: Skewness Results: Correlations Between ΔN, ΔL, and ΔT

| True Cut Location | Optimum Cutscore Location | | Error Rate at True Cutscore | |
|---|---|---|---|---|
| | Spearman's Rho | p | Spearman's Rho | p |
| 45 | 0.87 | <.001 | 0.97 | <.001 |
| 47.5 | 0.58 | <.001 | 0.98 | <.001 |
| 52.5 | 0.58 | <.001 | 0.93 | <.001 |
| 55 | -0.25 | 0.078 | 0.97 | <.001 |

**ΔL results.** Table 1 reports the correlations between ΔN, ΔL. As can be seen, the correlation was significant at an alpha level of .01 for all but the 55 condition.

***Cutscore at 45.*** As can be seen in table 1, there was a large correlation (.87) between ΔN and ΔL when the truecut score was set to 45.

As can be seen in Table E1 in Appendix E, the position chosen for the optimal error tended to the left as the skewness increased. This appeared to be due to the relative decrease in density to the left of the mean, as the skew pooled more and more examinees to the right. Thus, the probability of a false negative error (i.e., a student having a true score above 45, but an observed score below a certain cut) was relatively low just to the left of 45. By setting the observed cutscore to the left of 45, the optimal error location took advantage of a large decrease in FN for a modest increase in FP.

As can be seen by comparing Tables E1 and E2, i.e., comparing optimal error location with error at the (higher) observed cutscore location when set equal to the true cutscore location, by setting the cutscore to 43.6, the optimal point (rather than 45), in the first iteration, FP error was increased, while FN error was decreased. However, the relative decrease in FN outweighed the relative increase in FP. That is, as the observed cutscore was set more and more to the left of the true cutscore, FP increased, as there was an increased likelihood of a given examinee having a true score below the true cutscore, yet receiving an observed score above observed the cutscore.

As can be seen in Appendix E2, FN error at a fixed observed cutscore generally increased as the skew to the right increased. Thus, there were more and more candidates who have true scores above the true cutscore and thus there were more and more candidates that could potentially have been FN.

While the examinees in the extreme skew to the right should theoretically reduce the FN rate, they did not do so in this situation (with a true cutscore of 45). This was because those examinees with scores in the far right tail were already so beyond the true cutscore that the probability any of them would have been a FN, even in the truly normal case, was infinitesimal.

***Cutscore at 47.5.*** As can be seen in Table 1, there was a moderate correlation (.58) between $\Delta N$ and $\Delta L$ when the true cutscore was set to 47.5.

Similar to the 45 condition, in the beginning, the optimal cutscore was set just to the left of the true cutscore, taking advantage of a large decrease in FN for a modest increase in FP. As the skew to the right increases, however, the actual optimal location moved right, eventually moving to the right of the true cutscore in the most skewed simulations. Thus, the actual optimal cutscore moved in the same direction of the skew.

As can be seen in Figures E1 – E9, the location of the distribution mode moved left as the skew increased to the right. This occurred because the specified distribution maintained a mean of 50, requiring density to the left of that mean to balance out the skew in the extreme right tail. The true cutscore location of 47.5 became effectively in the middle of a normal distribution, as the individuals in the right skew were too far away to have any noticeable probability of being misclassified, and the left tail remained normal. Thus, at the true cutscore, FP and FN rates approached each other as the skew increased. The optimal error location reflected this trend, as the rates of FP and FN were practically equal in the last few iterations of the simulations. In

other words, the rates of FP and FN were so close to equal around the true cutscore that moving in either direction increased one error rate at least as quickly as it reduced the other, resulting in an optimal point that was effectively at the true cutscore.

***Cutscore at 52.5.*** There was a moderate correlation (.58) between ΔN and ΔL when the true cutscore was set to 52.5. The optimal cutscore for the initial simulation is just to the left of the true cutscore, at 53, taking advantage of a slight increase of FN for a larger decrease in FP. As the skew increases, the optimal cutscore generally moves further to the right, although it never moves far beyond 54.

***Cutscore at 55.*** There was no significant correlation between ΔN and ΔL for the 55 condition. Thus, increased skewness did not predict differences between GW-CSOF estimates of the optimal cutscore, and actual optimal cutscore. As can be seen in Table E1, the location of the optimal error remained relatively constant, and is very similar to the GW-CSOF estimated location, as can be seen in Table E3.

**ΔT results.** As can be seen in Table 1, there was a large significant correlation between the correctness of the GW-CSOF method at the true cutscore and ΔN for all four conditions.

As can be seen by comparing Tables E1 and E2, for the 52.5 and 55 conditions, as the skew to the right increases, fewer and fewer errors were actually made at a fixed point. The reverse was found for the 45 and 47.5 conditions, in which the error increased at a fixed point as the skew increased to the right. Thus, in the present manipulations, when the true cutscore was on the opposite side of the skew, error increased at a fixed point, and decreased at a fixed point when the true cutscore was on the same side as the skew. As can be seen by comparing Tables E2 and E4, the GW-CSOF underestimated error when the observed cutscore was set equal to a true cutscore on the opposite side of the skew, and overestimated error when the true cutscore

was on the same side as the skew. Furthermore, the degree of the over or underestimation was significantly predicted by the degree of skewness in the true score distribution.

**Bimodal**

Table 2: Bimodal Results (n=15 between D=7 and D=10) Correlations Between ΔN, ΔL, and ΔT

| | Optimum Cutscore Location | | Error Rate at True Cutscore | |
|---|---|---|---|---|
| True Cut Location | Spearman's Rho | p | Spearman's Rho | p |
| 45 | 0.59 | 0.021 | 0.87 | <.001 |
| 47.5 | 0.86 | <.001 | 0.75 | 0.001 |
| 52.5 | 0.85 | <.001 | 0.8 | <.001 |
| 55 | 0.43 | 0.106 | 0.93 | <.001 |

It was expected that the difference between GW-CSOF estimates and actual estimates would be roughly similar at the 45 and 55 condition, as the distribution should be approximately mirrored over the x axis. However, the results in Table 2 did not match this expectation. It seems likely that the small sample (15) may have led to anomalous results. Because of this, a second set of simulations was conducted which extended the number of simulations to 50 between D=7 and D=10, where D was the separation of mixture means. Table 3 presents these results.

Table 3: Bimodal Results (n=50 between D=7 and D=10) Correlations Between ΔN, ΔL, and ΔT

| | Optimum Cutscore Location | | Error Rate at True Cutscore | |
|---|---|---|---|---|
| True Cut Location | Spearman's Rho | p | Spearman's Rho | p |
| 45 | 0.4 | 0.004 | 0.88 | <.001 |
| 47.5 | 0.86 | <.001 | 0.83 | <.001 |
| 52.5 | 0.87 | <.001 | 0.87 | <.001 |
| 55 | 0.5 | <.001 | 0.9 | <.001 |

**ΔL results.** The results indicated that there was a significant correlation between ΔN and ΔL for all conditions.

***Cutscore at 45.*** As can be seen in Table 3, there was a significant moderate correlation between ΔN and ΔL. As can seen by comparing tables F1 and F2, the optimal cutscore began to the left of the true cutscore (44), but moved up and nearer to the cutscore as the degree of

bimodality increased, ending at 45 in the most extremely bimodal case (with true score D of 10.2).

***Cutscore at 47.5.*** As can be seen in Table 3, there was a significant large correlation between ΔN and ΔL. Similar to the 45 condition, the optimal cutscore moved right as the degree of bimodality increased. In this instance, however, the optimal cutscore began near the true cutscore (47.4) and reached a maximum of 51.1.

***Cutscore of 52.5.*** As can be seen in Table 3, there was a significant large correlation between ΔN and ΔL. As can be seen in Table F1, the optimal error began to the right of the true cutscore (53.3) and moved right as the bimodality increases, ending at a minimum of 49.7. Note, these results were roughly parallel to the 47.5 results, but in the opposite direction.

***Cutscore of 55.*** As can be seen in Table 3, there was a significant moderate correlation between ΔN and ΔL. As can seen in Table F2, the optimal cutscore began to the right of the true cutscore (56.2), but generally moved down and nearer to the true cutscore as the degree of bimodality increased, reaching a minimum of 44.6. Note, these results were roughly parallel to the 45 results, but in the opposite direction.

**ΔT results.** The results indicated that there was a large significant correlation between ΔN and ΔT for all conditions. As can be seen by viewing Table F2, error decreased for true cutscores near the center of the distribution, and increased for true cutscores further out from the center, as bimodality increased. As can be seen in Figures F1 – F50, as the bimodality increased, there was less and less density in the middle most part of the distribution, resulting in fewer FN and FP. Meanwhile, the density increased on either side of the true cutscores further out, resulting in more and more error as the bimodality increased. Accordingly, as can be seen in Table F4, when the observed cutscore is set equal to the true cutscore, GW-CSOF

underestimated error for true cutscores further out from the mean as bimodality increased, and overestimated error for true cutscores near the mean.

**Kurtosis**

Table 4: Kurtosis Results with n=50: Correlations Between ΔN, ΔL, and ΔT

| | Optimum Cutscore Location | | Error Rate at True Cutscore | |
|---|---|---|---|---|
| True Cut Location | Spearman's Rho | p | Spearman's Rho | p |
| 45 | 0.35 | 0.011 | 0.96 | <.001 |
| 47.5 | 0.89 | <.001 | 0.49 | <.001 |
| 52.5 | 0.9 | <.001 | 0.55 | <.001 |
| 55 | 0.22 | 0.118 | 0.96 | <.001 |

**ΔL results.** As can be seen in Table 4, only the 47.5 and 52.5 conditions were significant at the .01 alpha level of significance.

***Cutscore at 45.*** There was no significant correlation between ΔN and ΔL for the 45 condition. Thus, increased kurtosis did not predict differences between GW-CSOF estimates of the optimal cutscore, and actual optimal cutscore. As can be seen in Table G1, the location of the optimal error remained relatively constant, and ends almost exactly where it begins, thus it never departed far from the GW-CSOF estimated location (Table G3).

***Cutscore at 47.5.*** There was a significant large correlation between ΔN and ΔL for the 47.5 condition. As can be seen in tables G1 and G2, as kurtosis increased, FN went up and FP went down at a fixed cutscore location. This was due to the increased density just to the right of the true cutscore, resulting in an increase in the FN rate. The optimal cutscore moved to the left as the kurtosis increased, seeking a location to alleviate the increase in FN while also keeping FP low, which was made possible by the lower FP rates to the left as kurtosis increases.

***Cutscore at 52.5.*** There was a significant large correlation between ΔN and ΔL for the 52.5 condition. As can be seen in Tables E1 and E2, this was a similar of the 47.5 condition, with

FP going up and FN going down at a fixed cutscore location as kurtosis increased. Similarly, the optimal cutscore moved right as the kurtosis increased.

***Cutscore at 55.*** There was no significant correlation between ΔN and ΔL for the 55 condition. This was effectively a mirroring of the 45 condition. Thus, increased kurtosis did not predict differences between GW-CSOF estimates of the optimal cutscore, and actual optimal cutscore. As can be seen in Table G1, the location of the optimal error remained relatively constant, and ends almost exactly where it began, thus it never moved far away from the GW-CSOF estimated location (Table G3).

**ΔT results.** There was a significant large correlation for both the 45 and 55 conditions, and a significant moderate correlation for both the 47.5 and 52.5 conditions.

As can be seen in Table G2, total error remained similar for true cutscores near the center of the distribution, resulting in only a moderate correlation between ΔN and ΔT. This was due to small decreases in FP with only slightly larger increases in FN as the kurtosis increased for 47.5, and the reverse effect for 52.5. For 45 and 55, the large correlation between ΔN and ΔT was visibly due to the near monotonic decrease in both types of error as kurtosis increased: FP and FN error each fall to near half their starting values by the end of kurtosis manipulations for both conditions.

**Summary**

The results indicated that for the truly normal condition, GW-CSOF error estimates and actual error values, as well as GW-CSOF estimates of the optimal cutscore and actual location of the optimal cutscore, were nearly identical. For the non-normal manipulations, some conditions were significant and others were not. For skewness, the 45, 47.5, and 52.5 conditions all showed a significant correlation between increased non-normality and ΔL. The correlations were large

for a true cutscore of 45, and moderate for 47.5 and 52.5. The correlation between non-normality and ΔT were large and significant for all conditions.

For bimodality, there was a significant correlation for all conditions between non-normality and ΔL. The correlations were moderate for 45 and 55, and large for 47.5 and 52.5. The correlation between non-normality and ΔT were large and significant for all conditions.

Finally, for kurtosis, there was a significant and large correlation for the 47.5 and 52.5 conditions, and no significant correlation for the 45 and 55 conditions, between non-normality and ΔL. The correlation between non-normality and ΔT was large and significant for 45 and 55, and moderate and significant for 47.5 and 52.5.

## Chapter 5: Discussion

The GW-CSOF method, also known in the literature as the cutscore operating function, or the 'Grabovsky curve', is a method for estimating classification error rates, as well as for finding the optimal cutscore (i.e., the observed cutscore where classification error is minimized). Among the assumptions of the GW-CSOF is that examinee true scores are distributed normally. The present research has investigated the extent to which the GW-CSOF estimates are correct using Monte Carlo simulations. These simulations systematically manipulated normality (i.e., how Gaussian the distribution was) of examinee true score distributions in three different ways (skewness, bimodality, and kurtosis), and the correctness of the GW-CSOF method was checked at each progressive step away from normality. Answers to specific research questions were sought, which would confirm or refute specific research hypotheses.

**Do GW-CSOF estimates of optimal cutscores match the actual location of the optimal cutscore, and does the match change as non-normality increases?**

The GW-CSOF estimates of optimal cutscore location were found to be almost identical to the location of the actual cutscore in the truly normal case. This provides good evidence that the GW-CSOF estimates are valuable predictors of the true optimal cutscore when model assumptions are met.

There was a significant relationship between the degree of non-normality and the degree to which the GW-CSOF estimated optimal cutscore differed from the actual optimal cutscore location for most of the non-normality manipulations. For three of the four skewness conditions, for all the bimodal conditions, and for two of the four kurtosis conditions, increased non-normality predicted increased incorrectness in the GW-CSOF estimates.

Thus, research hypothesis I, that *the GW-CSOF method would estimate a location for the optimal cutscore near the location of the actual optimal cutscore when normality assumptions of the true score distribution were met*, was supported: the GW-CSOF method produced optimal cutscore locations close to actual optimal cutscore locations when the normality assumptions of the true score distribution were met.

Research hypothesis II, *increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore*, was partially supported. Increased non-normality in the true score distribution caused increased incorrectness in optimal cutscore location estimation, but not in all conditions. In the specific manipulations conducted in this dissertation, increased skewness predicted increased incorrectness in GW-CSOF estimation of optimal cutscore location when the true cutscore was on the opposite side of the skew, and when it was near the middle of the distribution on either side, but not when it was located a standard deviation away from the center, near to the tail containing the increasing skewness. This was likely due to the relatively constant density just above and below the true cutscore (55 in this case). That is to say that the proportion of simulated examinees that were just below the minimally proficient threshold, relative to the proportion of simulated examinees just above that threshold, remained relatively constant throughout the manipulation. While there were more and more examinees in the extreme right of the tail, those examinees had an infinitesimal chance of being FN errors, meanwhile the corresponding decrease in the center of the distribution did not alleviate FP rates as those examinees were too far below the true cutscore to have been likely FP errors already. This maintained relatively constant density near the true cutscore.

There was a significant relationship between increased bimodality and increased error in GW-CSOF estimation of optimal cutscore location for all bimodal manipulations. For the outer true cutscores (45 and 55) the optimal location for observed cutscores in the truly normal case began just a little toward the tail from the true cutscore. As bimodality increased, the two modes effectively became 45 and 55, placing the true cutscore essentially at the center of a normal distribution. Thus, there was roughly an equal cost to moving in either direction away from the center (i.e., the rate of FP increased at roughly the same rate FN decreased), and the optimal cutscore came to settle at approximately the same location as the mode. For those manipulations near the center, there was no place near that the optimal cutscore could move toward where error (FP for 47.5 and FN for 52.5) could be minimized, without a dramatic increase in the respective error.

In the kurtotic manipulations, increased (positive) kurtosis predicted increased incorrectness of GW-CSOF estimation of optimal cutscore locations when the true cutscore was near the center of the distribution, but not when it was located a standard deviation away from the mean in either direction. The density near the cutscores in these locations remained relatively constant, while the center of the distribution took on more and more of the density. Thus, there was a dramatic shift in the optimal cutscore location near the center, but no significant relationship between kurtosis and the location of the optimal cutscore location further out.

**Do GW-CSOF estimates of error at the true cutscore location match actual error rates, and does the match change as non-normality increases?**

The GW-CSOF estimates of total error at the true cutscore location (as well FP and FN) were found to be almost identical to the actual error rates in the truly normal case. This provides good evidence that the GW-CSOF estimates of error are valuable predictors of the classification

error when model assumptions are met. There was also a significant relationship between the degree of non-normality and the degree to which the GW-CSOF estimated error rates differed from the actual error rates for all of the non-normality manipulations.

Thus, research hypothesis III, that *the GW-CSOF method would produce error estimates close to actual error values when the normality assumptions of the true score distribution were met*, was supported. The GW-CSOF method produced error estimates close to actual error values when the normality assumptions of the true score distribution were met.

Hypothesis IV, that *increased non-normality in the true score distribution would cause increased incorrectness in error estimates*, was also supported. Increased non-normality caused increased incorrectness in GW-CSOF error estimates for all manipulations tested.

For the skewness manipulations, when the true cutscore was on the opposite side as the skew, error increased at a fixed point as the skewness increased. For true cutscores on the same side as the skew, at a fixed point, error decreased as the skewness increased. Thus, GW-CSOF underestimated error at the true cutscore when the true cutscore location was on the opposite side of the distribution from the skew, and GW-CSOF overestimated error when the true cutscore was on the same side as the skew.

For the bimodality manipulations, when the true cutscore was near the center of the distribution, there was less and less FN and FP error as bimodality increased, as the density at the center shrank. For true cutscores further away from the center, error increased as the density under the true cutscore increased. Thus, as bimodality increased, GW-CSOF underestimated error for true cutscores a standard deviation away from the mean, and overestimated error for true cutscores nearer to the mean.

Finally, for the kurtosis manipulations, both FP and FN error for true cutscores one standard deviation away from the mean decreased as kurtosis increased. Thus, GW-CSOF overestimated error for both cutscores one standard deviation from the mean. For true cutscores nearer to the mean, the relationship was more complicated. For the 47.5 condition, FP error decreased, while FN error increased, as kurtosis increased. This resulted in an overall error that was relatively steady throughout, resulting in only a moderate correlation between changes in normality and total error. The reverse was found for the 52.5 condition, where FN decreased while FP increased proportionally. Again, the total error was roughly the same, resulting in a moderate correlation between changes in normality and total error. Thus, GW-CSOF underestimated FN error for the true cutscore manipulation just below the mean, while overestimating FP error, as kurtosis increased. Similarly, GW-CSOF underestimated FP error for the true cutscore just above the mean, while overestimating FN error, as kurtosis increased.

**Are the Differences Meaningful?**

The present paper has demonstrated and discussed the statistical relationship between increases in non-normality and the correctness of GW-CSOF estimates, but thus far it has not discussed the magnitude of that incorrectness. In order to be useful to a standard setting committee, it is necessary to know how different G&W estimates are from actual values. Discussion proceeds by breaking non-normality manipulations into three categories: the 17$^{th}$, the 35$^{th}$, and the 50$^{th}$ manipulations. These corresponded to true score skewness of .63, 1.32, and 1.97, true score bimodality D of 6.1, 8.9, and 9.9, and true score kurtosis of 3.9, 5.1, and 6.03, respectively. These three divisions are henceforth referred to as the ‘minutely non-normal’, the ‘moderately non-normal’, and the ‘largely non-normal’. For each manipulation that was found to have a significant effect, the meaningfulness of the GW-CSOF methods misestimation is

discussed. First, differences between error at the actual optimal cutscore was compared to the actual error rates at the GW-CSOF estimated optimal cutscore. This allowed for comparison of how much more or less error would be made by using GW-CSOF estimated optimal cutscores rather than the actual optimal cutscores.

Additionally, comparison was made between actual error and GW-CSOF estimates of error for observed cutscores set to be equal to true cutscores. This information might be useful to standard setting committees who wish to use GW-CSOF to determine the error present at a given true cutscore.

**Optimal cutscore.** Table 5 provides information on the differences between error rates at the actual optimal cutscore and the GW-CSOF estimated cutscore. For the minutely non-normal manipulations, the largest difference in total error was -.01, indicating that for every 100 examinees tested, one additional examinee would be misclassified beyond what GW-CSOF estimated at its estimate of the optimal cutscore. Based on the present manipulations, it appears that small amounts of skewness, bimodality, or kurtosis do not have a meaningful impact on the utility of the GW-CSOF methods estimation of the optimal cutscore location.

For medium skewness and kurtosis, there was a similarly trivial increase in error. For the bimodal manipulations, however, when the true cutscore was located near the center of the distribution, an additional 3 examinees for every 100 tested would be misclassified. Thus, it appears that moderate non-normality did have a small but meaningful impact on GW-CSOF's estimation of the optimal cutscore, particularly for bimodality when the true cutscore is near the mean.

Finally, in the largely non-normal manipulation, several conditions appear to have meaningfully increased error rates. For the skewness condition, when the true cutscore was on

the opposite side of the distribution from the skew, an additional 5 of every 100 examinees would be misclassified. For the kurtosis condition, for true cutscores near the center of the distribution, an additional 3 out of every 100 examinees would be misclassified. Finally, for the bimodal conditions with true cutscores near the mean, an additional 9 out of every 100 examinees would be misclassified. Thus, it appears that large non-normality had a meaningful impact on GW-CSOF's estimation of the optimal cutscore for all three normality manipulations.

In short, it appears that GW-CSOF's estimates of the optimal cutscore location provide a close approximation (in terms of minimal error location) when normality is only slightly violated. In the case of moderate or large violations of normality, caution should be taken when using GW-CSOF estimates of optimal error location.

**True cutscore.** Table 6 provides useful information for the event that a standard setting panel uses the GW-CSOF estimates to get a sense of the error rates if the observed cutscore is set equal to the true cutscore. Based on the manipulations in this dissertation, it appears that minute non-normality lead to relatively minor misestimation of the error at the true cutscore. In the worst case, with minute kurtosis and a true cutscore of 47.5, two additional students per every 100 would be misclassified beyond what was estimated by the GW-CSOF method.

For moderate skewness, GW-CSOF underestimated error for both conditions on the opposite side of the distribution from the skew. For the worst of these, an additional four out of every 100 examinees would have been misclassified in addition to the GW-CSOF method estimates. GW-CSOF overestimated error for cutscores on the same side of the distribution as the skew. For moderate non-normality, the bimodal conditions were fairly closely estimated by

Table 5: Difference between actual error at actual optimal cutscore & actual error at estimated optimal cutscore

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Minute** | Δ FP | Δ FN | ΔTot. | Δ FP | Δ FN | ΔTot. | Δ FP | Δ FN | ΔTot. | Δ FP | Δ FN | ΔTot. |
| Skew | 0.00 | 0.00 | 0.00 | -0.01 | 0.00 | 0.00 | -0.01 | 0.01 | 0.00 | NA | NA | NA |
| Bimodal | -0.01 | 0.00 | 0.00 | -0.03 | 0.02 | -0.01 | 0.02 | -0.03 | -0.01 | 0.00 | -0.01 | 0.00 |
| Kurtosis | NA | NA | NA | 0.02 | -0.02 | 0.00 | -0.01 | 0.01 | 0.00 | NA | NA | NA |
| **Moderate** | | | | | | | | | | | | |
| Skew | 0.02 | -0.02 | 0.00 | -0.02 | 0.02 | 0.00 | -0.01 | 0.01 | 0.00 | NA | NA | NA |
| Bimodal | -0.02 | 0.02 | 0.00 | -0.05 | 0.02 | -0.03 | 0.04 | -0.06 | -0.02 | 0.02 | -0.02 | 0.00 |
| Kurtosis | NA | NA | NA | 0.01 | -0.02 | -0.01 | -0.04 | 0.03 | -0.01 | NA | NA | NA |
| **Large** | | | | | | | | | | | | |
| Skew | 0.02 | -0.07 | -0.05 | -0.05 | 0.04 | -0.01 | -0.02 | 0.01 | -0.01 | NA | NA | NA |
| Bimodal | 0.02 | -0.02 | 0.00 | -0.10 | 0.01 | -0.09 | 0.02 | -0.11 | -0.09 | 0.06 | -0.06 | -0.01 |
| Kurtosis | NA | NA | NA | 0.02 | -0.05 | -0.03 | -0.06 | 0.03 | -0.03 | NA | NA | NA |

*Note: Difference is actual optimal error - actual error at the GW-CSOF estimated optimal location. NA's denote non-significant results.

Table 6: Difference between actual and GW-CSOF estimate of error at true cutscore

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Minute** | Δ FP | Δ FN | ΔTot. | Δ FP | Δ FN | ΔTot. | Δ FP | Δ FN | ΔTot. | Δ FP | Δ FN | ΔTot. |
| Skew | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | -0.01 | -0.01 | 0.00 | -0.01 | -0.01 |
| Bimodal | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Kurtosis | -0.01 | 0.00 | -0.01 | 0.00 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | -0.01 | -0.01 |
| **Moderate** | | | | | | | | | | | | |
| Skew | 0.00 | 0.03 | 0.03 | 0.02 | 0.02 | 0.04 | 0.00 | -0.02 | -0.02 | -0.01 | -0.01 | -0.02 |
| Bimodal | 0.01 | 0.00 | 0.01 | 0.00 | -0.01 | -0.01 | -0.01 | 0.00 | -0.01 | 0.00 | 0.00 | 0.01 |
| Kurtosis | -0.02 | -0.01 | -0.03 | -0.02 | 0.03 | 0.01 | 0.03 | -0.02 | 0.01 | -0.01 | -0.02 | -0.03 |
| **Large** | | | | | | | | | | | | |
| Skew | -0.02 | 0.06 | 0.05 | 0.06 | 0.02 | 0.07 | -0.01 | -0.02 | -0.03 | -0.02 | -0.01 | -0.04 |
| Bimodal | 0.05 | 0.04 | 0.09 | 0.03 | -0.07 | -0.04 | -0.07 | 0.03 | -0.04 | 0.03 | 0.05 | 0.09 |
| Kurtosis | -0.02 | -0.02 | -0.04 | -0.03 | 0.03 | 0.00 | 0.03 | -0.03 | 0.00 | -0.02 | -0.01 | -0.04 |

*Note: Difference is actual error at true cutscore - GW-CSOF estimate of error at true cutscore

GW-CSOF, at worst being slight overestimates for 47.5 and 52.5, and slight underestimates for 45 and 55. For moderate kurtosis, GW-CSOF underestimated error for 45 and 55. An additional three out of every 100 examinees would have been misclassified at these cutscores than were estimated by GW-CSOF. For those cutscores nearer the mean (47.5 and 52.5) GW-CSOF slightly overestimated error (one per 100).

For largely non-normal true score distributions, substantial misestimation occurred. As many as 9 additional examinees out of every 100 would have been misclassified in the bimodal conditions 45 and 55, and as many as 7 additional examinees would have been misclassified for skewness of 45 and 47.5. Kurtosis did not appear to substantially impact GW-CSOF estimation of error for moderate and large kurtosis when true cutscores were set near the mean of the distribution.

The direction of the misclassifications discussed above is important to consider. Underestimating error is likely to be more problematic than overestimation, as it may provide a false and potentially harmful belief in the accuracy of classification results. In the interest of minimizing classification error, and thus minimizing the number of examinees who are mishandled by the classification processes, underestimation is a greater problem than is underestimation. Thus, the most problematic conditions are those which have large positive values in Table 6. Particularly problematic were moderate skewness with true cutscore 47.5, large skewness with true cutscores 45 and 47.5, and large bimodality with true cutscores 45 and 55.

**Limitations**

Larger simulations, or non-random simulations, would have been better. Simulations of 10,000 produced problematic volatility between runs, and made generalizations from this study

more tenuous than they could have been with larger samples. Random variation due to random sampling produced somewhat inconsistent results in the true score distributions. A more controlled and elegant approach would have been to directly specify the true score distribution using intentional sampling (rather than random) at regular increments of the respective distributions. I.e., in the skewness manipulation, for a given skewness value, the actual distribution is completely specified as indicated in the methods section. It would have been possible to use that specified distribution, and to create from it a distribution which represented it systematically.

Kendall's tau may be been a better statistic to use than Spearman's rho. Spearman's rho was chosen because it had been shown to have good power to detect small differences. However, statistical significance of Spearman's rho is not exactly specified when there are ties, making it necessary to be conservative in interpreting significant results. Kendall's tau would have alleviated this complication.

**Future Research Recommendations**

The results and discussion of this dissertation presented information that indicates that, in certain conditions, the GW-CSOF overestimates error, and underestimates error in other situations. Additional research should be conducted to replicate these findings, particularly with respect to the magnitude of differences between the GW-CSOF and actual error values/locations of optimal cutscores. This work could then be used to provide explicit correction guidelines for standard setting panels who observe specific amounts of non-normality in their examinee samples. Ideally, these correction factor guidelines would span the range of all possible cutscores, in order for standard setting panels to be able to consider potential cutscores with well researched correction factors, thus providing as much information to standard setting panels as

possible. Ideally, these correction factors would be built into software for standard setting (e.g., Runyon & Grabovsky, 2018; Pace & Grabovsky, 2019) to maximize their ease-of-use

This dissertation has demonstrated that non-normality, in the form of skewness, bimodality, or kurtosis, has an effect on the accuracy of the GW-CSOF. It would be useful for future research to explore the impact of combined forms of non-normality. That is, to investigate the degree to which combined non-normality, such as simultaneously increased skewness and kurtosis, yields inaccuracies in the GW-CSOF estimates.

**Conclusions**

The GW-CSOF method has great potential for use in standard setting. It offers standard setting committees the ability to predict classification error rates at all possible observed scores, as well as to determine the location of the optimal cutscore (i.e., where error is minimized). This dissertation sought to determine how well GW-CSOF estimates matched actual values of error and optimal cutscores when true score normality assumptions were met, as well as when those assumptions were systematically violated. Generally, this dissertation supports the use of GW-CSOF estimates for normally distributed true scores, as well as for true score distributions which are only slightly non-normal. Caution is advised for standard setting panels who might wish to use GW-CSOF estimates with substantially non-normal examinee data. Particularly, standard setting panels should be advised that GW-CSOF may indicate a sub-optimal cutscore location, i.e., choosing a position where error is not in fact minimized. In the worst of these cases, simulation results showed that for large skewness, when the true cutscore was set to the opposite side of the skew, an additional 5 of every 100 examinees would be misclassified due to selecting the GW-CSOF estimated optimal cutscore rather than the actual optimal cutscore. For large kurtosis, when the true cutscore is near the center of the distribution, an additional 3 out of every

100 examinees would have been misclassified, and for large bimodality, with true cutscores near the mean, an additional 9 out of every 100 examinees would be misclassified. Thus, using GW-CSOF estimates of optimal error should be done with caution when working with moderately or largely non-normal data.

Furthermore, standard setting panels might also want to use GW-CSOF to estimate error if the observed cutscore is set equal to the true cutscore. Caution should be taken for heavily non-normal examinee distributions in this situation as well. Particularly, the panel should be cautious of skewness which exists on the opposite side of the examinee distribution from their true cutscore, and also of bimodality in which either of the separate modes are near to the true cutscore, as simulation results showed that as many an additional 9 out of every 100 examinees would be misclassified over and above what was estimated by GW-CSOF. Thus, in such situations, GW-CSOF should be thought of as a lower estimate of error, and only used with extreme caution.

From a practical and applied standpoint, standard setting panel members should consider the following. As was observed in this simulation study, observed score non-normality may well be lower than the corresponding true score non-normality. The true shape of the true score distribution is always unknown in practice, but the information found in the present study can provide some approximate guidelines. This dissertation demonstrated that, at least as manipulated in this study, observed score skewness values below .46, bimodality (when equally distanced from the mean) of less than about 1.66 times the standard deviation (i.e., 8.3/5 = 1.66), and kurtosis of less than 3.5, all appeared to produce relatively correct GW-CSOF estimates of the optimal cutscore. Panelists should exercise caution when they observed score histograms which indicate that the observed score distribution may have skewness, bimodality, or kurtosis

above these points, and note that they may be receiving over-estimates or underestimates (depending on the condition, see earlier discussion) if they proceed to use the GW-CSOF with such non-normal distributions.

To bring the discussion thus far into context, it is worth returning to an example from the literature review portion of this dissertation, to refresh the notions on which this dissertation was written, specifically, about what false positive and false negatives mean in context of an actual exam. Suppose that an exam is given to 10,000 examinees, and the error rate is 3%. That would mean that 300 examinees are being wrongly classified by the given exam. Suppose that, through the use of the GW-CSOF, the standard setting panel is able to find the optimal cutscore location which decreases the error rate by 2%. This would then result in 200 examinees receiving the correct classification that otherwise would have been handled incorrectly by the exam. Now, suppose that the GW-CSOF provides an incorrect estimate of the optimal cutscore, which leads a standard setting panel to choose a score which is 1% less optimal than the actual optimal (which, as was said, decreased the error rate by 2%). That would mean that an additional 100 of every 10,000 examinees is misclassified over what the actual optimal cutscore would have yielded. These principals will be useful to bear in mind when considering the information that has been provided above.

Overall, GW-CSOF is a useful and powerful tool for standard setting. This dissertation will hopefully contribute to the research base for the GW-CSOF method, and inform potential users of when they should proceed cautiously, and what they might expect in various non-normal situations.

## References


An, L., & Ahmed, S. E. (2008). Improving the performance of kurtosis estimator. *Computational Statistics & Data Analysis, 52*(5), 2669-2681.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME; 2014). *The standards for educational and psychological testing.* Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME); 2014.

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193-212.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.

Azzalini, A. (2013). *The skew-normal and related families* (Vol. 3). Cambridge University Press.

Behboodian, J. (1970). On a mixture of normal distributions. *Biometrika*, 57 (1), 215-217.

Buckendahl, C. W., Smith, R. W., Impara, J. C., & Plake, B. S. (2002). A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational Measurement, 39*(3), 253-263.

Cizek, G. J., & Bunch, M. B. (2007). The Bookmark Method. In *Standard setting: A guide to establishing and evaluating performance standards on tests.*, 155-192, SAGE Publications Ltd.

Cizek, G. J. (2012). An Introduction to Contemporary Standard Setting: Concepts, Characteristics, and Concepts. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations (2nd ed., pp 3-13). New York, NY: Routledge.*

Coladarci, T., Cobb, C. D., (2014). *Fundamentals of statistical reasoning in education, 4th edition*. Wiley.

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N., Galasko, D., Gauthier, S., Hampel, H., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Sarazin, M., de Souza, L. C., Stern, Y., Visser, P., Scheltens, P. (2010). Revising the definition of Alzheimer's disease: a new lexicon. *The Lancet Neurology, 9*(11), 1118-1127.

Hurtz, G. M., & Auerbach, M. A. (2003). A meta-analysis of the effects of modifications to the Angoff method on cutoff scores and judgment consensus. *Educational and Psychological Measurement*, *63*(4), 584-601.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*(3), 425-461.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

Feuerman, M., & Miller, A. R. (2008). Relationships between statistical measures of agreement: sensitivity, specificity and kappa. *Journal of evaluation in clinical practice*, *14*(5), 930-933.

Glasnapp, D., & Poggio, J. (1985). *Essentials of Statistical Analysis for the Behavioral Sciences, 1985.* Columbus,OH. Merrill Publishing Co.

Grabovsky, I., & Wainer, H. (2017a). The cutscore operating function: A new tool to aid in standard setting. *Journal of Educational and Behavioral Statistics*, 42, 251-263.

Grabovsky, I., & Wainer, H. (2017b). A Guide for setting the cutscores to minimize weighted classification errors in test batteries. *Journal of Educational and Behavioral Statistics*, *42*, 264-281.

Greiner, M., Pfeiffer, D., & Smith, R. D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive veterinary medicine*, *45*(1-2), 23-41.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction.* New York, NY. Springer Series in Statistics.

Harvill, L. M. (1991). Standard Error of Measurement: an NCME Instructional Module on. *Educational Measurement: issues and practice*, *10*(2), 33-41.

Helmstadter, G. C. (1964). *Principles of psychological measurement.* New York, NY. Appleton-Century-Crofts.

Hogg, R., McKean, J., & Craig, A. (2013*). Introduction to Mathematical Statistics, 7th edition.* Boston,MA. Pearson.

Hogg, R. V., Tanis, E. A. , Zimmerman, D.L. (2015). *Probability and statistical inference (7th ed.)*. Pearson Education Inc.

Livingston, S. A., & Zieky, M. J. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests.

Livingston, S. A., & Zieky, M. J. (1989). A comparative study of standard-setting methods. *Applied Measurement in Education*, *2*(2), 121-141.

Lee, W. C. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement*, 47, 1-17.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark Standard Setting Procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations (2nd ed., pp. 225 - 253). New York, NY: Routledge*.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Ornstein, P., & Lyhagen, J. (2016). Asymptotic properties of Spearman's rank correlation for variables with finite support. *PloS one*, *11*(1).

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, *50*, 741.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological bulletin*, *105*(1), 156.

Pace, J.R., Grabovsky, I. (April 2019). *Minimizing classification errors when the true cut score is not known, with software for standard setting*. Presentation given at the annual meeting of the National Council on Measurement in Education, Toronto, ON, Canada.

Plake, B. S., & Cizek, G. J. (2012). Variations on a theme: The Modified Angoff. Extended Angoff. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp 181-199). New York, NY: Routledge.

R Core Team. (2017). *R: a language and environment for statistical computing. R Foundation for Statistical Computing.* Vienna, Austria. Retrieved from http://www. R-project.org/

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Rossi, P. (2014). *Bayesian non-and semi-parametric methods and applications*. Princeton University Press.

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment, Research & Evaluation*, 7, 1-8.

Runyon, C., Grabovsky, I. (April 2018). *Cutscore: a shiny app for the cut-score operating function.* Presentation given at the annual meeting of the National Council on Measurement in Education, New York, NY.

Solomon, H. C. (1920). Agreement in results of the Wassermann reaction: study of tests performed by two laboratories in three thousand successive hospital admissions. Journal of the American Medical Association, 74, 788-790.

Thorndike, R. M. (1997). Measurement and evaluation in psychology and education. Uppe Saddle River, NJ: Prentice Hall, Inc.

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports (1896-1970)*, 1432-1449.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32-35.

Wainer, H. (2017). Visual Revelations: The Grabovsky Curve. *CHANCE, 30*, 44-48.

Yue, S., Pilon, P., & Cavadias, G. (2002). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of hydrology*, 259(1-4), 254-271.

Zabell, A. P., Foxworthy, T., Eaton, K. N., & Julian, R. K. (2014). Diagnostic application of the exponentially modified Gaussian model for peak quality and quantitation in high-throughput liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A*, *1369*, 92-97.

## Appendix A: Systematic increase of skewness with fixed mean and variance

Using the exponentially modified normal distribution, with S=X+Y, the mean and variance are given by the sum of the means and variances of x and y, and skewness is given by equation 7. As stated earlier, the mean of S will be fixed to 50, and variance of S will be 25. We have three equations and three unknowns. That is,

Let μ be the mean of the normal distribution, X, and let $\theta$ be the mean of the exponential distribution Y. Let $\sigma^2$ be the variance of X, and $\theta^2$ be the variance of Y. Let s be the fixed (i.e., constant) skewness of a given manipulation. Then, using the above specifications:

$$\mu + \theta = 50 \quad \text{(i)}$$

$$\sigma^2 + \theta^2 = 25 \quad \text{(ii)}$$

$$\frac{2\theta^3}{(\sigma^2 + \theta^2)^{3/2}} = s \quad \text{(iii)}$$

Where s is the fixed skewness. We have three unknowns and 3 equations, which can be solved algebraically. Using equations ii and iii

$$\theta^2 = 25 - \sigma^2$$

And,

$$\sigma^2 = \left(\frac{2\theta^3}{s}\right)^{\frac{2}{3}} - \theta^2$$

Thus,

$$\theta^2 = \frac{25(s)^{2/3}}{2^{2/3}} \quad \text{(iv)}$$

Once s is chosen, $\theta$, $\sigma^2$ and μ are determined.

## Appendix B: Systematic increase of bimodality with fixed mean and variance

Using a mixture of normals with equal weight (i.e., mixing weight $w$ = ½) the mean and variance are given by equations 10 and 11. Let Z be the mixture, then $Z = IX + (1 - I)Y$, where X and Y are the respective normal random variables, I is an indicator with I=1 having probability w. As stated earlier, the mean of Z was fixed to 50, and variance of Z was fixed to 25. Let $\mu_{\boldsymbol{m}}$ be the mixture mean, and let $\sigma_m^2$ be the mixture variance. We establish that $\mu_1$ and $\mu_2$ are the means of the first input normal, and the second, respectively, and $\sigma_1^2$ and $\sigma_2^2$ are their respective variances. We define D as the distance between $\mu_1$ and $\mu_2$, and let $\mu_1 > \mu_2$ for all values, thus $D = |\mu_1 - \mu_2| = \mu_1 - \mu_2$. Using the equations for the mean and variance of the mixture, given by equations 10 and 11, we now have two equations and two unknowns. That is:

$$\mu_m = w\mu_1 + (w - 1)\mu_2$$

With $\mu_m = 50$ and w=1/2 this simplifies to

$$50 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 \qquad \text{(i)}$$

The variance also simplifies by setting $\sigma_1^2$ and $\sigma_2^2$ to be equal, which we call $v$.

$$\sigma_m^2 = 25 = \frac{1}{2}(v + {\mu_1}^2) + w(v + {\mu_2}^2) - 50^2$$

Which, replacing $\mu_1$ and $\mu_2$ by their respective relationship to D, which determines each, and solving algebraically, we can rewrite as

$$v = 25 - \frac{D^2}{4} \qquad \text{(ii)}$$

Using equations i and ii, for any specified D, the values of $v$, $\mu_1, \mu_2$, are specified.

## Appendix C: Systematic increase of kurtosis with fixed mean and variance

Using a mixture of normals with equal weight (i.e., mixing weight $w$ = ½) the mean and variance are given by equations 10 and 11, and the kurtosis of the mixture is given by equation 12 Let Z be the mixture, then $Z = IX + (1 - I)Y$, where X and Y are the respective normal random variables, I is an indicator with I=1 having probability w. As stated earlier, the mean of Z was fixed to 50, and variance of Z were fixed to 25. Let $\mu_m$ be the mixture mean, and let $\sigma_m^2$ be the mixture variance. We establish that $\mu_1$ and $\mu_2$ are equal, as stated in the Methods section, and we let $\sigma_1^2$ and $\sigma_2^2$ be the variances of the two mixtures. Using the equations for the variance and kurtosis of the mixture, given by equations 11 and 12, we now have two equations and two unknowns. That is:

Each input normal has a mean of 50, and the resulting mixture does as well.

Using this result, $\sigma_2^2$ can be expressed in terms of $\sigma_1^2$, i.e.:

$$\sigma_2^2 = 50 - \sigma_1^2 \qquad \text{(i)}$$

Finally, substituting equation i into equation 12, we find the result that for a fixed kurtosis, k, $\sigma_1^2$ is equal to:

$$\sigma_1^2 = 25 + \sqrt{\frac{k * 625}{3} - 625} \qquad \text{(ii)}$$

Thus, with $\mu_1$ and $\mu_2$ fixed, $\sigma_1^2$ and $\sigma_2^2$ are determined by the fixed kurtosis value.

## Appendix D: Simulation Results

### Skewness

| Distribution | True | Observed | Mean True | Mean Obs | Var True | Var Obs |
|---|---|---|---|---|---|---|
| 0.00 | -0.03 | -0.03 | 50.08 | 50.12 | 24.95 | 31.55 |
| 0.04 | 0.04 | 0.01 | 49.97 | 49.99 | 24.70 | 30.63 |
| 0.08 | 0.09 | 0.06 | 50.01 | 50.06 | 24.99 | 31.37 |
| 0.12 | 0.11 | 0.07 | 50.05 | 50.04 | 25.31 | 31.42 |
| 0.16 | 0.12 | 0.08 | 49.94 | 49.99 | 25.21 | 32.20 |
| 0.20 | 0.28 | 0.19 | 49.94 | 49.94 | 24.87 | 30.98 |
| 0.24 | 0.28 | 0.20 | 50.03 | 50.03 | 24.40 | 30.41 |
| 0.28 | 0.28 | 0.18 | 50.08 | 50.10 | 25.09 | 31.27 |
| 0.32 | 0.31 | 0.25 | 49.99 | 49.99 | 25.04 | 30.69 |
| 0.36 | 0.36 | 0.27 | 49.96 | 49.94 | 24.54 | 30.49 |
| 0.40 | 0.39 | 0.29 | 49.91 | 49.88 | 24.73 | 30.67 |
| 0.44 | 0.41 | 0.27 | 50.06 | 50.06 | 24.38 | 30.31 |
| 0.48 | 0.50 | 0.33 | 49.99 | 50.01 | 25.15 | 31.64 |
| 0.52 | 0.53 | 0.35 | 50.06 | 50.05 | 25.22 | 31.56 |
| 0.56 | 0.51 | 0.35 | 49.92 | 49.97 | 24.58 | 30.11 |
| 0.60 | 0.62 | 0.44 | 49.99 | 50.00 | 25.19 | 31.38 |
| 0.64 | 0.63 | 0.46 | 50.00 | 49.99 | 25.15 | 31.29 |
| 0.68 | 0.74 | 0.55 | 50.13 | 50.12 | 25.48 | 32.04 |
| 0.72 | 0.78 | 0.55 | 49.94 | 49.92 | 25.61 | 32.16 |
| 0.76 | 0.86 | 0.59 | 49.98 | 50.03 | 25.41 | 32.19 |
| 0.80 | 0.72 | 0.49 | 49.99 | 50.02 | 23.96 | 29.94 |
| 0.84 | 0.82 | 0.60 | 49.96 | 49.94 | 24.72 | 31.19 |
| 0.88 | 0.87 | 0.60 | 50.03 | 50.05 | 25.44 | 31.85 |
| 0.92 | 0.85 | 0.59 | 49.93 | 49.96 | 23.61 | 29.76 |
| 0.96 | 0.92 | 0.68 | 49.97 | 49.98 | 24.91 | 30.72 |
| 1.00 | 0.96 | 0.71 | 50.01 | 49.98 | 24.63 | 30.64 |
| 1.04 | 1.00 | 0.74 | 49.95 | 49.96 | 24.76 | 31.02 |
| 1.08 | 1.06 | 0.78 | 50.07 | 50.09 | 24.82 | 30.85 |
| 1.12 | 1.15 | 0.82 | 50.10 | 50.14 | 25.87 | 32.18 |
| 1.16 | 1.26 | 0.88 | 50.00 | 50.02 | 24.90 | 31.23 |
| 1.20 | 1.29 | 0.91 | 50.04 | 50.03 | 26.12 | 32.20 |
| 1.24 | 1.20 | 0.87 | 50.02 | 50.05 | 25.04 | 31.15 |
| 1.28 | 1.24 | 0.91 | 49.97 | 49.97 | 24.14 | 30.00 |
| 1.32 | 1.32 | 0.97 | 50.01 | 49.97 | 25.56 | 31.68 |
| 1.36 | 1.32 | 0.94 | 50.00 | 50.04 | 24.71 | 30.92 |
| 1.40 | 1.39 | 1.00 | 49.95 | 49.95 | 25.09 | 30.89 |
| 1.44 | 1.37 | 0.94 | 49.95 | 49.92 | 24.37 | 30.57 |
| 1.48 | 1.53 | 1.13 | 49.98 | 49.99 | 25.39 | 31.36 |
| 1.52 | 1.47 | 1.08 | 49.94 | 49.96 | 24.40 | 30.61 |
| 1.56 | 1.44 | 1.03 | 49.97 | 50.01 | 24.24 | 30.59 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1.60 | 1.61 | 1.14 | 50.11 | 50.10 | 26.03 | 32.29 |
| 1.64 | 1.72 | 1.27 | 49.98 | 49.96 | 25.72 | 32.06 |
| 1.68 | 1.77 | 1.28 | 49.99 | 50.01 | 25.48 | 31.92 |
| 1.72 | 1.68 | 1.20 | 49.98 | 49.97 | 24.58 | 31.29 |
| 1.76 | 1.66 | 1.21 | 50.11 | 50.12 | 25.29 | 31.60 |
| 1.80 | 1.70 | 1.22 | 49.98 | 49.99 | 24.93 | 30.99 |
| 1.84 | 1.79 | 1.27 | 50.08 | 50.05 | 25.32 | 32.01 |
| 1.88 | 2.07 | 1.51 | 50.00 | 50.04 | 26.06 | 32.30 |
| 1.92 | 1.85 | 1.30 | 49.97 | 49.92 | 24.12 | 30.37 |
| 1.96 | 1.97 | 1.40 | 49.98 | 49.99 | 24.77 | 30.96 |

**Bimodality**

| Distribution | True | Observed | Mean True | Mean Obs | Var True | Var Obs |
|---|---|---|---|---|---|---|
| 7.00 | 1.50 | 1.10 | 49.97 | 49.93 | 24.68 | 30.77 |
| 7.20 | 3.70 | 2.40 | 49.99 | 49.97 | 25.30 | 31.09 |
| 7.40 | 2.80 | 2.90 | 49.94 | 49.94 | 24.56 | 30.53 |
| 7.60 | 4.70 | 3.60 | 50.01 | 50.02 | 24.88 | 31.56 |
| 7.80 | 7.40 | 2.30 | 50.01 | 50.02 | 25.35 | 32.03 |
| 8.00 | 7.90 | 5.70 | 49.99 | 50.00 | 25.08 | 31.17 |
| 8.20 | 7.80 | 5.20 | 50.03 | 50.01 | 24.61 | 31.34 |
| 8.40 | 8.60 | 10.30 | 50.06 | 50.07 | 25.25 | 31.58 |
| 8.60 | 9.20 | 8.30 | 49.97 | 49.97 | 25.00 | 31.29 |
| 8.80 | 9.50 | 5.60 | 49.98 | 50.02 | 25.10 | 30.97 |
| 9.00 | 8.70 | 8.50 | 49.99 | 49.99 | 24.93 | 31.12 |
| 9.20 | 8.50 | 9.60 | 50.01 | 50.02 | 25.26 | 31.78 |
| 9.40 | 9.20 | 7.50 | 50.00 | 49.95 | 24.81 | 30.44 |
| 9.60 | 10.10 | 9.70 | 50.00 | 49.99 | 24.97 | 31.91 |
| 9.80 | 10.10 | 8.70 | 50.00 | 49.98 | 24.94 | 31.30 |

**Kurtosis**

| Distribution | True | Observed | Mean True | Mean Obs | Var True | Var Obs |
|---|---|---|---|---|---|---|
| 3.00 | 3.03 | 2.92 | 50.08 | 50.12 | 24.95 | 31.55 |
| 3.06 | 3.14 | 3.07 | 49.97 | 49.97 | 24.70 | 30.84 |
| 3.12 | 3.11 | 3.06 | 49.94 | 49.90 | 24.41 | 30.83 |
| 3.18 | 3.12 | 3.07 | 50.10 | 50.15 | 24.83 | 31.19 |
| 3.24 | 3.31 | 3.21 | 49.99 | 49.94 | 25.04 | 31.26 |
| 3.30 | 3.28 | 3.17 | 50.03 | 50.03 | 24.58 | 30.97 |
| 3.36 | 3.31 | 3.26 | 50.06 | 50.03 | 25.26 | 31.44 |
| 3.42 | 3.39 | 3.20 | 50.03 | 50.04 | 25.10 | 31.18 |
| 3.48 | 3.39 | 3.32 | 50.02 | 50.04 | 24.13 | 30.24 |
| 3.54 | 3.53 | 3.37 | 49.96 | 50.00 | 24.39 | 30.54 |
| 3.60 | 3.49 | 3.32 | 49.97 | 49.98 | 25.15 | 31.61 |
| 3.66 | 3.61 | 3.49 | 50.01 | 49.98 | 25.57 | 31.64 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3.72 | 3.58 | 3.39 | 50.06 | 50.09 | 25.29 | 31.24 |
| 3.78 | 3.82 | 3.48 | 50.00 | 50.02 | 25.43 | 32.04 |
| 3.84 | 3.67 | 3.48 | 50.08 | 50.06 | 24.83 | 30.62 |
| 3.90 | 3.90 | 3.52 | 50.00 | 50.02 | 25.54 | 31.53 |
| 3.96 | 3.90 | 3.54 | 49.96 | 49.98 | 24.09 | 30.50 |
| 4.02 | 4.08 | 3.57 | 50.03 | 49.97 | 25.01 | 31.49 |
| 4.08 | 4.01 | 3.49 | 49.97 | 49.94 | 24.05 | 30.23 |
| 4.14 | 4.15 | 3.88 | 49.99 | 49.96 | 25.21 | 31.28 |
| 4.20 | 4.08 | 3.64 | 49.92 | 49.96 | 25.31 | 32.04 |
| 4.26 | 4.21 | 3.72 | 49.96 | 49.93 | 24.51 | 30.63 |
| 4.32 | 4.63 | 4.14 | 50.05 | 50.10 | 24.70 | 31.26 |
| 4.38 | 4.39 | 3.86 | 50.07 | 50.07 | 25.51 | 31.75 |
| 4.44 | 4.47 | 3.95 | 50.00 | 49.96 | 25.47 | 32.10 |
| 4.50 | 4.54 | 4.10 | 50.09 | 50.05 | 25.70 | 31.82 |
| 4.56 | 4.69 | 4.06 | 50.04 | 50.01 | 24.24 | 30.43 |
| 4.62 | 4.59 | 4.02 | 50.02 | 50.05 | 25.68 | 31.88 |
| 4.68 | 4.98 | 4.30 | 50.08 | 50.08 | 24.54 | 30.77 |
| 4.74 | 5.02 | 4.31 | 50.06 | 50.08 | 25.02 | 31.57 |
| 4.80 | 4.81 | 4.19 | 49.95 | 49.93 | 25.53 | 31.38 |
| 4.86 | 4.80 | 4.19 | 50.01 | 50.06 | 26.15 | 32.30 |
| 4.92 | 4.86 | 4.16 | 49.97 | 50.00 | 24.58 | 30.71 |
| 4.98 | 4.85 | 4.15 | 50.08 | 50.05 | 25.30 | 32.01 |
| 5.04 | 4.97 | 4.19 | 49.97 | 49.97 | 24.74 | 30.70 |
| 5.10 | 5.10 | 4.41 | 49.98 | 49.94 | 24.57 | 30.61 |
| 5.16 | 5.24 | 4.38 | 50.01 | 50.00 | 24.78 | 30.63 |
| 5.22 | 5.14 | 4.50 | 49.87 | 49.87 | 25.23 | 31.55 |
| 5.28 | 5.44 | 4.49 | 50.00 | 50.01 | 25.49 | 31.72 |
| 5.34 | 5.31 | 4.48 | 50.04 | 50.04 | 25.85 | 32.49 |
| 5.40 | 5.40 | 4.58 | 50.01 | 50.03 | 24.48 | 30.69 |
| 5.46 | 5.26 | 4.40 | 50.01 | 50.00 | 23.83 | 30.79 |
| 5.52 | 5.57 | 4.59 | 50.13 | 50.13 | 24.80 | 31.13 |
| 5.58 | 5.82 | 4.93 | 49.96 | 49.96 | 25.06 | 31.50 |
| 5.64 | 5.68 | 4.76 | 49.99 | 50.03 | 24.72 | 30.59 |
| 5.70 | 6.08 | 4.93 | 49.99 | 49.99 | 24.10 | 30.40 |
| 5.76 | 5.67 | 4.68 | 50.07 | 50.08 | 24.40 | 30.95 |
| 5.82 | 5.96 | 4.88 | 49.97 | 49.92 | 24.66 | 30.51 |
| 5.88 | 5.74 | 4.72 | 49.99 | 49.98 | 25.10 | 31.37 |
| 5.94 | 6.03 | 4.93 | 49.99 | 49.97 | 24.68 | 31.20 |

## Appendix E: Skewness Results

Table E1:Error Rates and Location of Actual Optimal Cutscore with Increasing Skewness

| Truescore | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Skewness | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| -0.03 | 0.052 | 0.024 | 43.6 | 0.076 | 0.054 | 46.8 | 0.059 | 0.069 | 53 | 0.021 | 0.063 | 56.8 |
| 0.04 | 0.06 | 0.026 | 43.6 | 0.081 | 0.046 | 46.7 | 0.056 | 0.077 | 53.2 | 0.025 | 0.062 | 56.5 |
| 0.09 | 0.052 | 0.03 | 43.9 | 0.073 | 0.059 | 47 | 0.055 | 0.074 | 53.2 | 0.029 | 0.053 | 56.2 |
| 0.11 | 0.055 | 0.028 | 43.8 | 0.077 | 0.053 | 46.8 | 0.054 | 0.071 | 53 | 0.031 | 0.052 | 56.1 |
| 0.12 | 0.06 | 0.028 | 43.6 | 0.077 | 0.057 | 46.9 | 0.051 | 0.07 | 53.3 | 0.027 | 0.056 | 56.5 |
| 0.28 | 0.058 | 0.034 | 43.9 | 0.071 | 0.069 | 47.3 | 0.04 | 0.084 | 53.6 | 0.019 | 0.057 | 56.7 |
| 0.28 | 0.057 | 0.027 | 43.6 | 0.082 | 0.05 | 46.8 | 0.053 | 0.074 | 53.2 | 0.023 | 0.054 | 56.5 |
| 0.28 | 0.065 | 0.02 | 43.2 | 0.075 | 0.059 | 47 | 0.048 | 0.074 | 53.4 | 0.026 | 0.054 | 56.4 |
| 0.31 | 0.064 | 0.028 | 43.6 | 0.07 | 0.061 | 47.1 | 0.05 | 0.077 | 53.2 | 0.033 | 0.049 | 56 |
| 0.36 | 0.06 | 0.025 | 43.5 | 0.084 | 0.054 | 46.7 | 0.048 | 0.073 | 53.3 | 0.024 | 0.05 | 56.4 |
| 0.39 | 0.071 | 0.026 | 43.3 | 0.073 | 0.07 | 47.1 | 0.04 | 0.08 | 53.6 | 0.028 | 0.044 | 56.1 |
| 0.41 | 0.063 | 0.027 | 43.5 | 0.085 | 0.059 | 46.8 | 0.049 | 0.077 | 53.3 | 0.031 | 0.043 | 56 |
| 0.5 | 0.057 | 0.032 | 43.7 | 0.081 | 0.062 | 47 | 0.039 | 0.086 | 53.9 | 0.021 | 0.053 | 56.7 |
| 0.53 | 0.058 | 0.027 | 43.6 | 0.076 | 0.065 | 47 | 0.045 | 0.07 | 53.5 | 0.018 | 0.056 | 56.7 |
| 0.51 | 0.064 | 0.029 | 43.5 | 0.092 | 0.052 | 46.7 | 0.043 | 0.077 | 53.6 | 0.028 | 0.044 | 56.1 |
| 0.62 | 0.061 | 0.033 | 43.6 | 0.082 | 0.06 | 47 | 0.039 | 0.075 | 53.6 | 0.018 | 0.051 | 56.6 |
| 0.63 | 0.062 | 0.03 | 43.5 | 0.077 | 0.064 | 47.1 | 0.045 | 0.07 | 53.4 | 0.024 | 0.052 | 56.5 |
| 0.74 | 0.065 | 0.026 | 43.4 | 0.082 | 0.059 | 46.8 | 0.036 | 0.084 | 53.9 | 0.024 | 0.046 | 56.3 |
| 0.78 | 0.069 | 0.025 | 43.3 | 0.088 | 0.059 | 46.8 | 0.041 | 0.07 | 53.5 | 0.018 | 0.046 | 56.7 |
| 0.86 | 0.061 | 0.031 | 43.5 | 0.098 | 0.05 | 46.6 | 0.032 | 0.075 | 54 | 0.02 | 0.045 | 56.5 |
| 0.72 | 0.053 | 0.032 | 43.7 | 0.086 | 0.069 | 47.1 | 0.038 | 0.077 | 53.7 | 0.019 | 0.046 | 56.5 |
| 0.82 | 0.064 | 0.032 | 43.5 | 0.082 | 0.073 | 47.1 | 0.048 | 0.065 | 53.3 | 0.027 | 0.039 | 56.1 |
| 0.87 | 0.069 | 0.024 | 43.2 | 0.09 | 0.06 | 46.8 | 0.053 | 0.062 | 53.2 | 0.02 | 0.047 | 56.5 |
| 0.85 | 0.064 | 0.028 | 43.3 | 0.101 | 0.056 | 46.8 | 0.034 | 0.078 | 53.9 | 0.017 | 0.047 | 56.6 |
| 0.92 | 0.062 | 0.036 | 43.8 | 0.091 | 0.07 | 47 | 0.043 | 0.068 | 53.5 | 0.017 | 0.047 | 56.7 |
| 0.96 | 0.078 | 0.015 | 42.5 | 0.108 | 0.054 | 46.6 | 0.043 | 0.07 | 53.5 | 0.019 | 0.042 | 56.4 |
| 1 | 0.075 | 0.022 | 43 | 0.092 | 0.068 | 47 | 0.037 | 0.072 | 53.7 | 0.017 | 0.046 | 56.6 |
| 1.06 | 0.068 | 0.026 | 43.3 | 0.09 | 0.062 | 46.9 | 0.053 | 0.059 | 53.1 | 0.019 | 0.043 | 56.6 |
| 1.15 | 0.064 | 0.03 | 43.4 | 0.096 | 0.063 | 46.9 | 0.041 | 0.067 | 53.6 | 0.024 | 0.044 | 56.3 |
| 1.26 | 0.076 | 0.019 | 42.7 | 0.09 | 0.073 | 47.1 | 0.04 | 0.069 | 53.6 | 0.029 | 0.034 | 55.9 |
| 1.29 | 0.07 | 0.025 | 43.1 | 0.107 | 0.056 | 46.5 | 0.032 | 0.071 | 54 | 0.018 | 0.04 | 56.4 |
| 1.2 | 0.064 | 0.026 | 43.2 | 0.092 | 0.076 | 47.2 | 0.04 | 0.062 | 53.5 | 0.021 | 0.04 | 56.3 |
| 1.24 | 0.065 | 0.027 | 43.3 | 0.106 | 0.062 | 46.8 | 0.026 | 0.075 | 54.2 | 0.018 | 0.041 | 56.4 |
| 1.32 | 0.068 | 0.023 | 43 | 0.102 | 0.065 | 46.8 | 0.035 | 0.067 | 53.7 | 0.018 | 0.04 | 56.4 |
| 1.32 | 0.078 | 0.017 | 42.4 | 0.081 | 0.091 | 47.5 | 0.041 | 0.063 | 53.6 | 0.015 | 0.048 | 56.9 |
| 1.39 | 0.081 | 0.016 | 42.4 | 0.11 | 0.064 | 46.8 | 0.03 | 0.073 | 54 | 0.023 | 0.036 | 56.2 |
| 1.37 | 0.073 | 0.022 | 42.7 | 0.093 | 0.076 | 47.1 | 0.03 | 0.067 | 53.8 | 0.017 | 0.038 | 56.5 |
| 1.53 | 0.078 | 0.016 | 42.4 | 0.105 | 0.074 | 47.1 | 0.023 | 0.078 | 54.3 | 0.019 | 0.037 | 56.4 |
| 1.47 | 0.075 | 0.014 | 42.4 | 0.108 | 0.075 | 47 | 0.041 | 0.057 | 53.5 | 0.021 | 0.034 | 56.2 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.44 | 0.07 | 0.018 | 42.5 | 0.102 | 0.084 | 47.2 | 0.041 | 0.053 | 53.5 | 0.012 | 0.045 | 57 |
| 1.61 | 0.069 | 0.014 | 42.2 | 0.107 | 0.08 | 47.1 | 0.043 | 0.052 | 53.3 | 0.018 | 0.037 | 56.4 |
| 1.72 | 0.074 | 0.01 | 41.7 | 0.105 | 0.088 | 47.3 | 0.031 | 0.061 | 53.8 | 0.022 | 0.033 | 56 |
| 1.77 | 0.072 | 0.011 | 41.8 | 0.099 | 0.091 | 47.4 | 0.032 | 0.058 | 53.8 | 0.018 | 0.037 | 56.4 |
| 1.68 | 0.066 | 0.017 | 42.2 | 0.086 | 0.107 | 47.7 | 0.025 | 0.065 | 54.1 | 0.011 | 0.043 | 56.8 |
| 1.66 | 0.073 | 0.004 | 40.7 | 0.108 | 0.087 | 47.3 | 0.035 | 0.058 | 53.5 | 0.022 | 0.035 | 56.2 |
| 1.7 | 0.077 | 0.004 | 40.3 | 0.122 | 0.074 | 47.1 | 0.023 | 0.064 | 54.2 | 0.017 | 0.035 | 56.3 |
| 1.79 | 0.07 | 0.001 | 39.1 | 0.096 | 0.104 | 47.6 | 0.029 | 0.059 | 53.8 | 0.02 | 0.032 | 56.2 |
| 2.07 | 0.061 | 0.002 | 39.9 | 0.092 | 0.106 | 47.8 | 0.035 | 0.053 | 53.6 | 0.019 | 0.031 | 56.2 |
| 1.85 | 0.054 | <.001 | 38.6 | 0.097 | 0.104 | 47.7 | 0.028 | 0.056 | 53.8 | 0.016 | 0.039 | 56.5 |
| 1.97 | 0.039 | <.001 | 38.1 | 0.104 | 0.101 | 47.7 | 0.028 | 0.059 | 53.8 | 0.018 | 0.033 | 56.4 |

*Note: Skew, as presented in this table, is the skewness of the simulated truescores

Table E2: Actual Error Rates at True Cutscore Location with Increasing Skewness

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| -0.03 | 0.03 | 0.058 | 45 | 0.057 | 0.077 | 47.5 | 0.076 | 0.055 | 52.5 | 0.065 | 0.033 | 55 |
| 0.04 | 0.035 | 0.06 | 45 | 0.057 | 0.073 | 47.5 | 0.082 | 0.057 | 52.5 | 0.062 | 0.033 | 55 |
| 0.09 | 0.035 | 0.062 | 45 | 0.059 | 0.077 | 47.5 | 0.08 | 0.054 | 52.5 | 0.063 | 0.034 | 55 |
| 0.11 | 0.033 | 0.062 | 45 | 0.058 | 0.076 | 47.5 | 0.074 | 0.055 | 52.5 | 0.06 | 0.033 | 55 |
| 0.12 | 0.035 | 0.064 | 45 | 0.059 | 0.079 | 47.5 | 0.076 | 0.049 | 52.5 | 0.064 | 0.032 | 55 |
| 0.28 | 0.037 | 0.064 | 45 | 0.065 | 0.077 | 47.5 | 0.075 | 0.054 | 52.5 | 0.059 | 0.03 | 55 |
| 0.28 | 0.034 | 0.069 | 45 | 0.064 | 0.073 | 47.5 | 0.076 | 0.055 | 52.5 | 0.059 | 0.03 | 55 |
| 0.28 | 0.035 | 0.065 | 45 | 0.06 | 0.076 | 47.5 | 0.08 | 0.05 | 52.5 | 0.062 | 0.032 | 55 |
| 0.31 | 0.04 | 0.063 | 45 | 0.06 | 0.076 | 47.5 | 0.073 | 0.055 | 52.5 | 0.058 | 0.033 | 55 |
| 0.36 | 0.033 | 0.066 | 45 | 0.061 | 0.086 | 47.5 | 0.077 | 0.051 | 52.5 | 0.058 | 0.029 | 55 |
| 0.39 | 0.037 | 0.07 | 45 | 0.061 | 0.086 | 47.5 | 0.076 | 0.052 | 52.5 | 0.056 | 0.029 | 55 |
| 0.41 | 0.037 | 0.067 | 45 | 0.065 | 0.084 | 47.5 | 0.076 | 0.056 | 52.5 | 0.058 | 0.028 | 55 |
| 0.5 | 0.033 | 0.067 | 45 | 0.067 | 0.08 | 47.5 | 0.083 | 0.051 | 52.5 | 0.056 | 0.029 | 55 |
| 0.53 | 0.035 | 0.066 | 45 | 0.062 | 0.085 | 47.5 | 0.079 | 0.046 | 52.5 | 0.057 | 0.028 | 55 |
| 0.51 | 0.037 | 0.071 | 45 | 0.067 | 0.081 | 47.5 | 0.082 | 0.047 | 52.5 | 0.056 | 0.028 | 55 |
| 0.62 | 0.037 | 0.071 | 45 | 0.067 | 0.078 | 47.5 | 0.076 | 0.049 | 52.5 | 0.054 | 0.028 | 55 |
| 0.63 | 0.036 | 0.071 | 45 | 0.066 | 0.08 | 47.5 | 0.072 | 0.049 | 52.5 | 0.057 | 0.029 | 55 |
| 0.74 | 0.034 | 0.074 | 45 | 0.061 | 0.085 | 47.5 | 0.08 | 0.049 | 52.5 | 0.058 | 0.028 | 55 |
| 0.78 | 0.037 | 0.073 | 45 | 0.065 | 0.086 | 47.5 | 0.077 | 0.047 | 52.5 | 0.052 | 0.024 | 55 |
| 0.86 | 0.037 | 0.076 | 45 | 0.072 | 0.083 | 47.5 | 0.078 | 0.044 | 52.5 | 0.052 | 0.025 | 55 |
| 0.72 | 0.033 | 0.07 | 45 | 0.072 | 0.086 | 47.5 | 0.079 | 0.048 | 52.5 | 0.053 | 0.027 | 55 |
| 0.82 | 0.038 | 0.077 | 45 | 0.07 | 0.089 | 47.5 | 0.075 | 0.046 | 52.5 | 0.049 | 0.025 | 55 |
| 0.87 | 0.037 | 0.074 | 45 | 0.068 | 0.086 | 47.5 | 0.079 | 0.045 | 52.5 | 0.052 | 0.026 | 55 |
| 0.85 | 0.036 | 0.077 | 45 | 0.075 | 0.086 | 47.5 | 0.08 | 0.044 | 52.5 | 0.053 | 0.025 | 55 |
| 0.92 | 0.036 | 0.074 | 45 | 0.072 | 0.092 | 47.5 | 0.077 | 0.044 | 52.5 | 0.051 | 0.025 | 55 |
| 0.96 | 0.035 | 0.078 | 45 | 0.076 | 0.089 | 47.5 | 0.075 | 0.048 | 52.5 | 0.048 | 0.024 | 55 |
| 1 | 0.04 | 0.078 | 45 | 0.075 | 0.089 | 47.5 | 0.075 | 0.046 | 52.5 | 0.046 | 0.025 | 55 |
| 1.06 | 0.037 | 0.076 | 45 | 0.072 | 0.086 | 47.5 | 0.073 | 0.046 | 52.5 | 0.054 | 0.025 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.15 | 0.037 | 0.078 | 45 | 0.076 | 0.086 | 47.5 | 0.074 | 0.042 | 52.5 | 0.051 | 0.025 | 55 |
| 1.26 | 0.036 | 0.081 | 45 | 0.077 | 0.088 | 47.5 | 0.079 | 0.046 | 52.5 | 0.049 | 0.025 | 55 |
| 1.29 | 0.035 | 0.084 | 45 | 0.075 | 0.094 | 47.5 | 0.078 | 0.037 | 52.5 | 0.046 | 0.024 | 55 |
| 1.2 | 0.034 | 0.083 | 45 | 0.08 | 0.089 | 47.5 | 0.071 | 0.04 | 52.5 | 0.046 | 0.025 | 55 |
| 1.24 | 0.037 | 0.083 | 45 | 0.08 | 0.092 | 47.5 | 0.075 | 0.04 | 52.5 | 0.047 | 0.024 | 55 |
| 1.32 | 0.033 | 0.08 | 45 | 0.077 | 0.094 | 47.5 | 0.07 | 0.042 | 52.5 | 0.042 | 0.024 | 55 |
| 1.32 | 0.039 | 0.089 | 45 | 0.081 | 0.091 | 47.5 | 0.073 | 0.04 | 52.5 | 0.049 | 0.025 | 55 |
| 1.39 | 0.039 | 0.091 | 45 | 0.085 | 0.092 | 47.5 | 0.071 | 0.043 | 52.5 | 0.044 | 0.022 | 55 |
| 1.37 | 0.034 | 0.09 | 45 | 0.078 | 0.095 | 47.5 | 0.07 | 0.039 | 52.5 | 0.046 | 0.021 | 55 |
| 1.53 | 0.037 | 0.09 | 45 | 0.089 | 0.093 | 47.5 | 0.071 | 0.041 | 52.5 | 0.043 | 0.023 | 55 |
| 1.47 | 0.037 | 0.09 | 45 | 0.089 | 0.1 | 47.5 | 0.071 | 0.038 | 52.5 | 0.046 | 0.022 | 55 |
| 1.44 | 0.034 | 0.094 | 45 | 0.092 | 0.099 | 47.5 | 0.07 | 0.034 | 52.5 | 0.048 | 0.022 | 55 |
| 1.61 | 0.031 | 0.091 | 45 | 0.091 | 0.098 | 47.5 | 0.068 | 0.038 | 52.5 | 0.041 | 0.022 | 55 |
| 1.72 | 0.032 | 0.097 | 45 | 0.098 | 0.097 | 47.5 | 0.07 | 0.036 | 52.5 | 0.04 | 0.023 | 55 |
| 1.77 | 0.033 | 0.102 | 45 | 0.095 | 0.096 | 47.5 | 0.071 | 0.036 | 52.5 | 0.044 | 0.021 | 55 |
| 1.68 | 0.032 | 0.106 | 45 | 0.095 | 0.098 | 47.5 | 0.068 | 0.037 | 52.5 | 0.04 | 0.021 | 55 |
| 1.66 | 0.029 | 0.103 | 45 | 0.099 | 0.098 | 47.5 | 0.068 | 0.038 | 52.5 | 0.044 | 0.021 | 55 |
| 1.7 | 0.033 | 0.105 | 45 | 0.104 | 0.092 | 47.5 | 0.065 | 0.035 | 52.5 | 0.039 | 0.02 | 55 |
| 1.79 | 0.029 | 0.114 | 45 | 0.101 | 0.1 | 47.5 | 0.064 | 0.033 | 52.5 | 0.038 | 0.021 | 55 |
| 2.07 | 0.026 | 0.115 | 45 | 0.106 | 0.093 | 47.5 | 0.066 | 0.033 | 52.5 | 0.04 | 0.019 | 55 |
| 1.85 | 0.025 | 0.127 | 45 | 0.108 | 0.094 | 47.5 | 0.061 | 0.034 | 52.5 | 0.042 | 0.022 | 55 |
| 1.97 | 0.017 | 0.126 | 45 | 0.114 | 0.093 | 47.5 | 0.064 | 0.036 | 52.5 | 0.04 | 0.02 | 55 |

*Note: Skew, as presented in this table, is the skewness of the simulated truescores

Table E3: GW-CSOF Estimate of Location of & Error at Optimal Cutscore with Increasing Skewness

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| -0.03 | 0.056 | 0.028 | 43.7 | 0.076 | 0.052 | 46.8 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 0.04 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 0.09 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.056 | 56.2 |
| 0.11 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 0.12 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 0.28 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 0.28 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 0.28 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 0.31 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 0.36 | 0.056 | 0.03 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 0.39 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.051 | 0.075 | 53.2 | 0.028 | 0.055 | 56.3 |
| 0.41 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.056 | 56.2 |
| 0.5 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 0.53 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 0.51 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.62 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 0.63 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 0.74 | 0.056 | 0.029 | 43.7 | 0.076 | 0.052 | 46.8 | 0.056 | 0.075 | 53.1 | 0.032 | 0.058 | 56.2 |
| 0.78 | 0.057 | 0.032 | 43.8 | 0.075 | 0.056 | 46.9 | 0.055 | 0.073 | 53.1 | 0.029 | 0.057 | 56.3 |
| 0.86 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.057 | 56.2 |
| 0.72 | 0.056 | 0.028 | 43.7 | 0.073 | 0.054 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 0.82 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 0.87 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 0.85 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.054 | 0.073 | 53.1 | 0.027 | 0.056 | 56.3 |
| 0.92 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 0.96 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 1 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.057 | 56.3 |
| 1.06 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 1.15 | 0.056 | 0.029 | 43.7 | 0.075 | 0.052 | 46.8 | 0.056 | 0.076 | 53.1 | 0.032 | 0.058 | 56.2 |
| 1.26 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 1.29 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.057 | 56.2 |
| 1.2 | 0.057 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.056 | 56.2 |
| 1.24 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.054 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 1.32 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 1.32 | 0.057 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 1.39 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 1.37 | 0.056 | 0.03 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 1.53 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 1.47 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 1.44 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 1.61 | 0.057 | 0.029 | 43.7 | 0.076 | 0.052 | 46.8 | 0.056 | 0.075 | 53.1 | 0.032 | 0.058 | 56.2 |
| 1.72 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 1.77 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 1.68 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 1.66 | 0.056 | 0.028 | 43.7 | 0.076 | 0.052 | 46.8 | 0.055 | 0.075 | 53.1 | 0.031 | 0.058 | 56.2 |
| 1.7 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 1.79 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 2.07 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.056 | 0.075 | 53.1 | 0.032 | 0.057 | 56.2 |
| 1.85 | 0.056 | 0.03 | 43.8 | 0.075 | 0.055 | 46.9 | 0.054 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 1.97 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |

*Note: Skew, as presented in this table, is the skewness of the simulated truescores

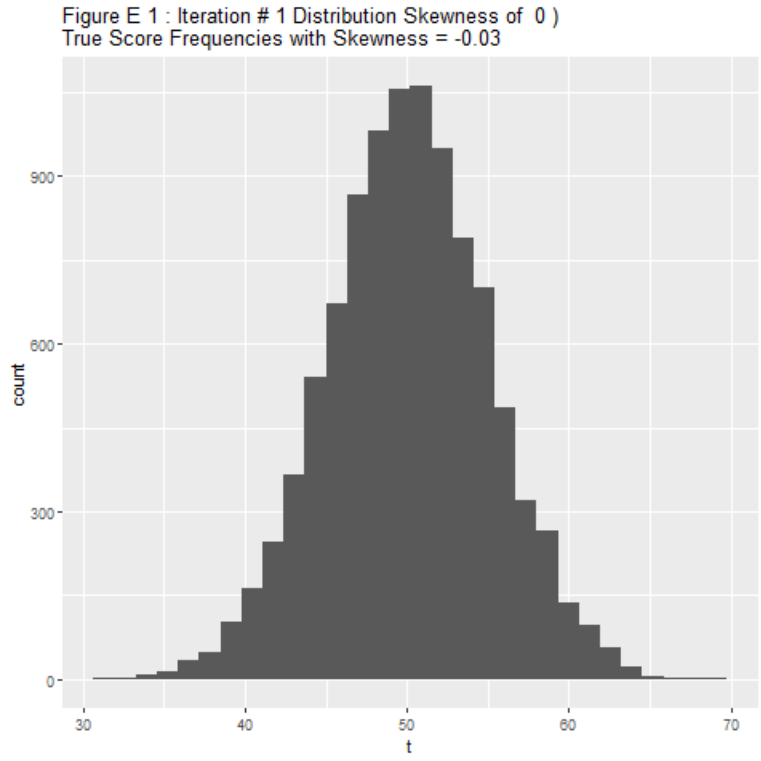
Table E4: GW-CSOF Error Rates at True Cutscore Location with Increasing Skewness

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| -0.03 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 0.04 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 0.09 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 0.11 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.12 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 0.28 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 0.28 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 0.28 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 0.31 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 0.36 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 0.39 | 0.035 | 0.062 | 45 | 0.058 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 0.41 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 0.5 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 0.53 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 0.51 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 0.62 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 0.63 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 0.74 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 0.78 | 0.036 | 0.063 | 45 | 0.058 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 0.86 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 0.72 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 0.82 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 0.87 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 0.85 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 0.92 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 0.96 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 1 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 1.06 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 1.15 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 1.26 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 1.29 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 1.2 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 1.24 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 1.32 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 1.32 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 1.39 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 1.37 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 1.53 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 1.47 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 1.44 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 1.61 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 1.72 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 1.77 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 1.68 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 1.66 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 1.7 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 1.79 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 2.07 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.036 | 55 |
| 1.85 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.97 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |

*Note: Skew, as presented in this table, is the skewness of the simulated truescores



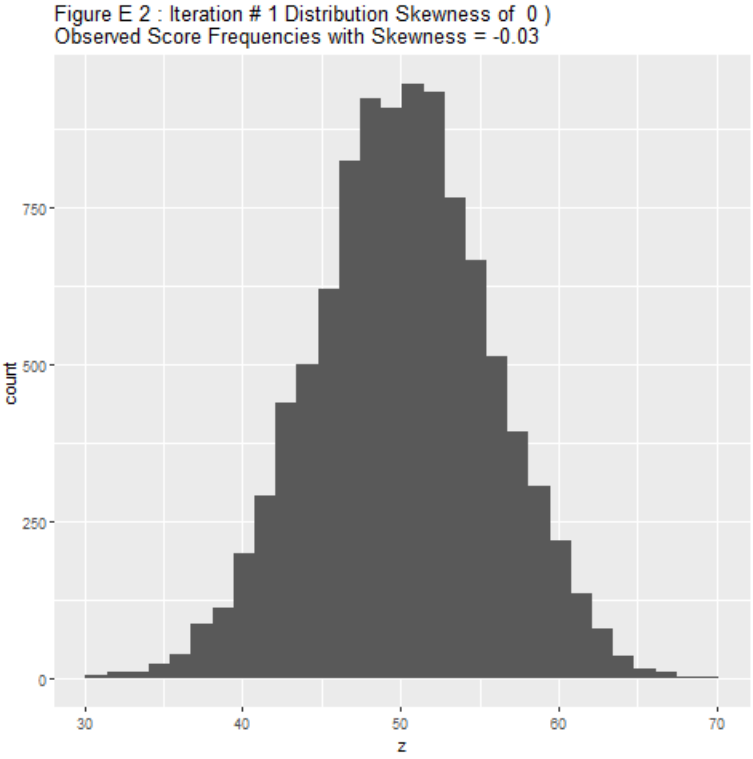



Figure E 3 : Iteration # 10 Distribution Skewness of 0.36 )
True Score Frequencies with Skewness = 0.36

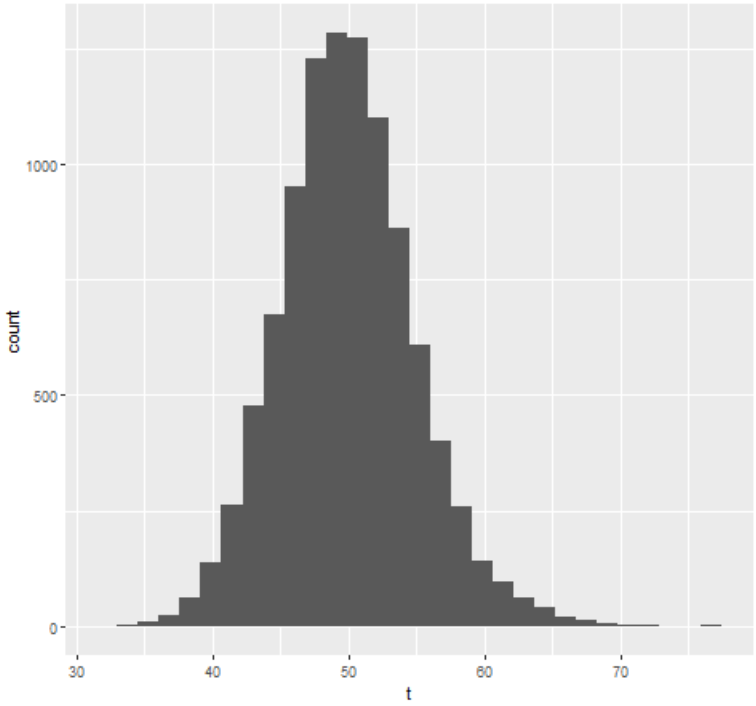



Figure E 4 : Iteration # 10 Distribution Skewness of 0.36 )
Observed Score Frequencies with Skewness = 0.27

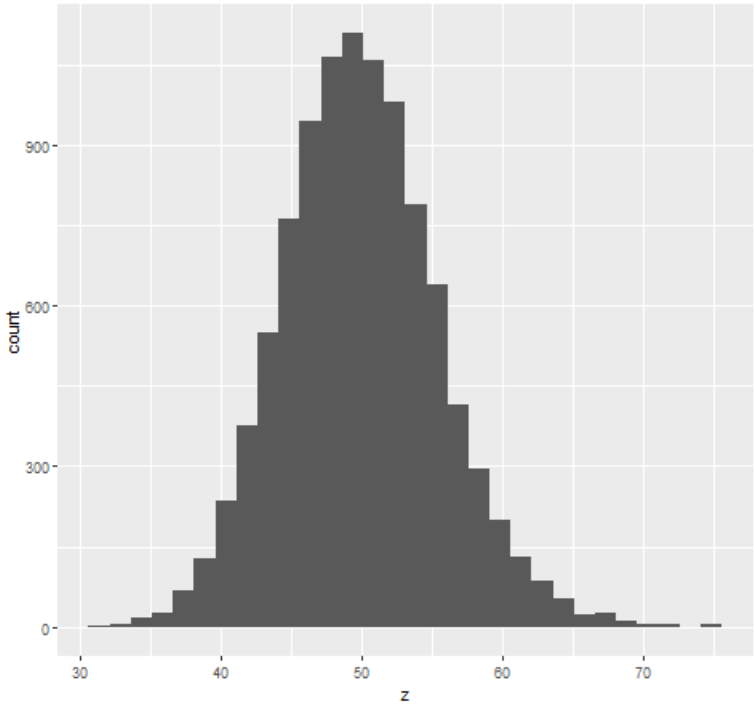


Figure E 5 : Iteration # 20 Distribution Skewness of 0.76 )
True Score Frequencies with Skewness = 0.86

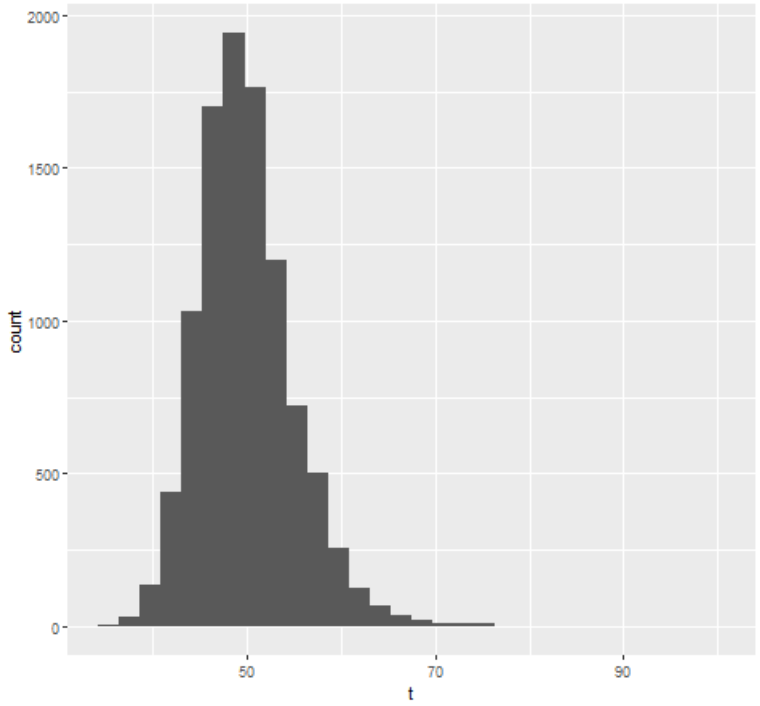



Figure E 6 : Iteration # 20 Distribution Skewness of 0.76 )
Observed Score Frequencies with Skewness = 0.59

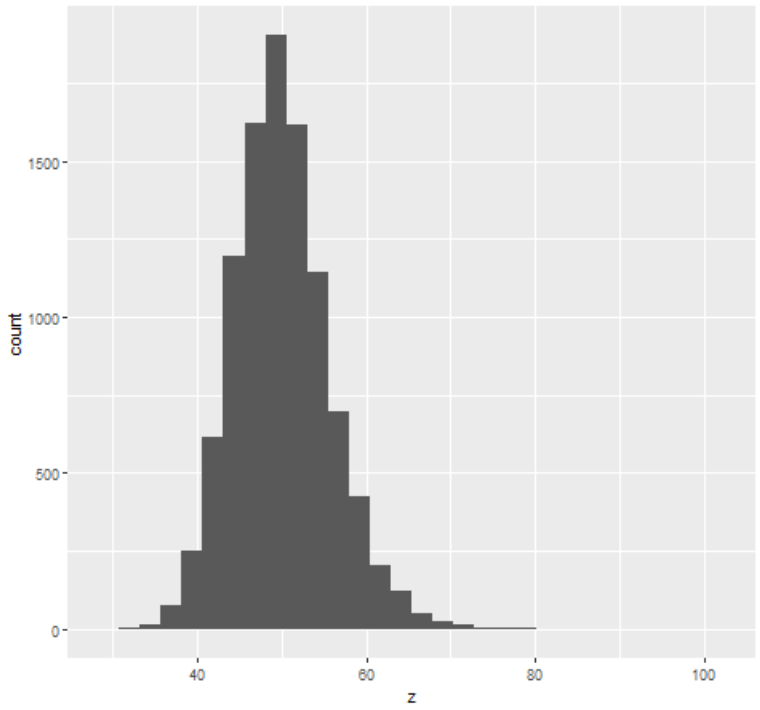


Figure E 7 : Iteration # 30 Distribution Skewness of 1.16 )
True Score Frequencies with Skewness = 1.26

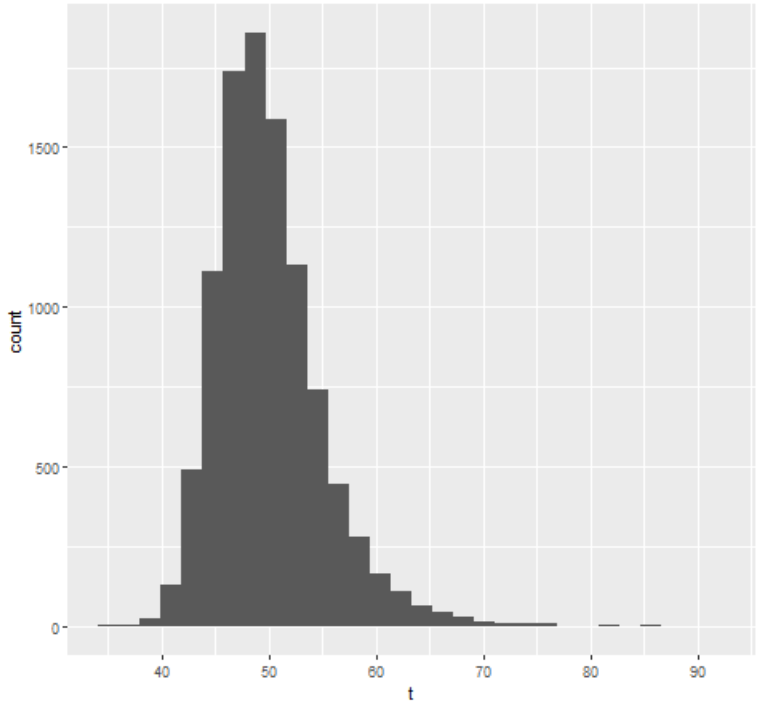



Figure E 8 : Iteration # 30 Distribution Skewness of 1.16 )
Observed Score Frequencies with Skewness = 0.88

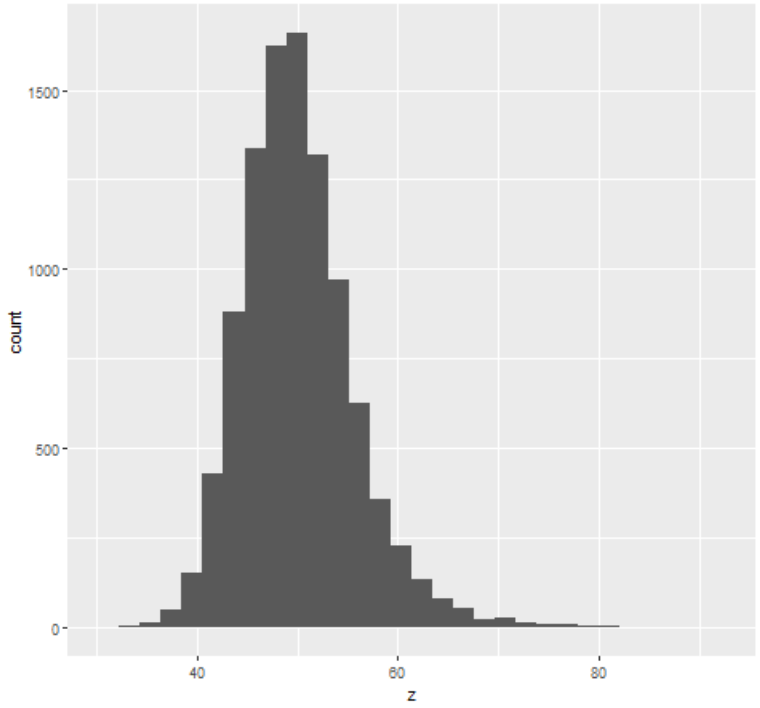


Figure E 9 : Iteration # 40 Distribution Skewness of 1.56 )
True Score Frequencies with Skewness = 1.44

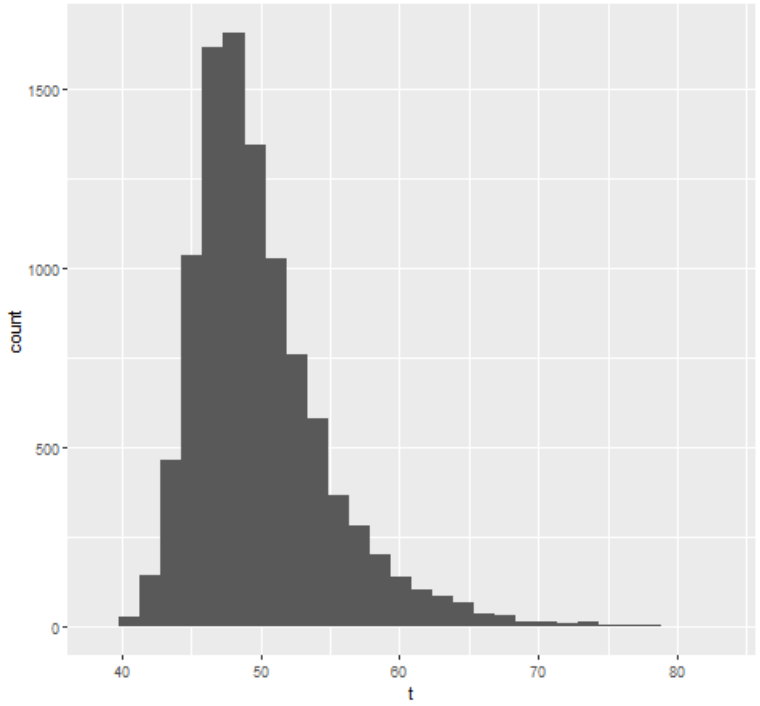



Figure E 10 : Iteration # 40 Distribution Skewness of 1.56 )
Observed Score Frequencies with Skewness = 1.03

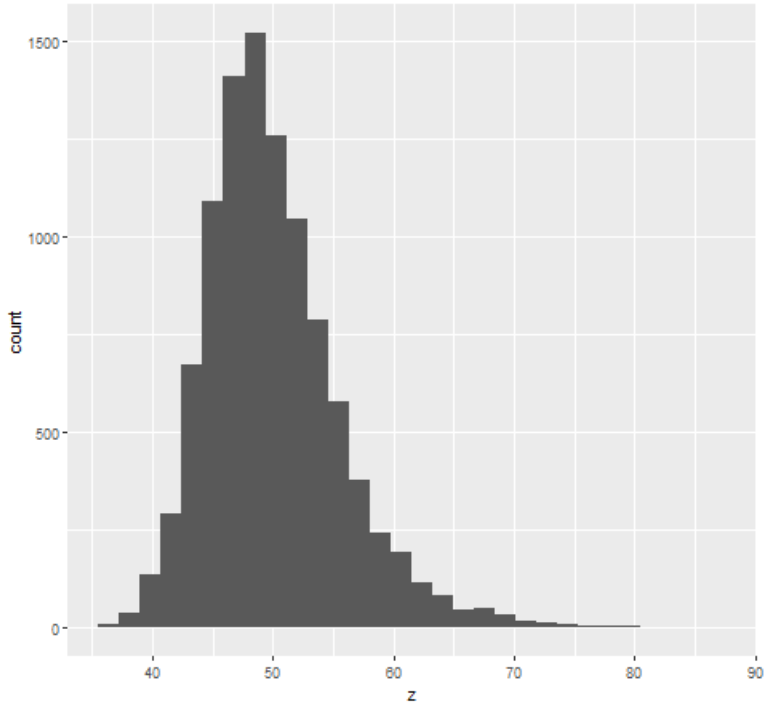


Figure E 11 : Iteration # 50 Distribution Skewness of 1.96 )
True Score Frequencies with Skewness = 1.97

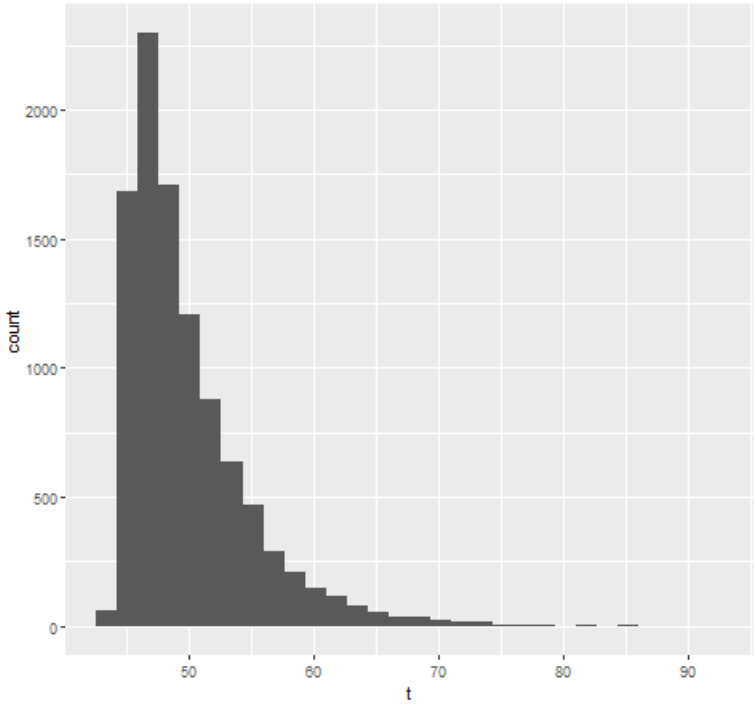
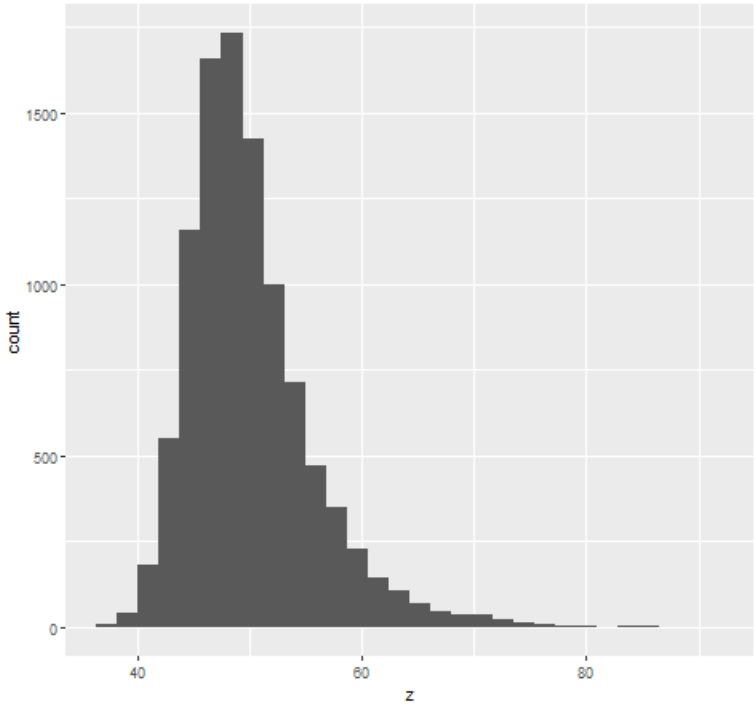



Figure E 12 : Iteration # 50 Distribution Skewness of 1.96 )
Observed Score Frequencies with Skewness = 1.4



## Appendix F: Bimodal Results

Table F1:Error Rates and Location of Actual Optimal Cutscore with Increasing Skewness

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TS D | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 0.8 | 0.06 | 0.036 | 44 | 0.061 | 0.063 | 47.4 | 0.046 | 0.084 | 53.3 | 0.033 | 0.063 | 56.2 |
| 2.8 | 0.06 | 0.036 | 44 | 0.072 | 0.055 | 47.1 | 0.066 | 0.06 | 52.6 | 0.035 | 0.054 | 56 |
| 3.1 | 0.063 | 0.034 | 43.8 | 0.064 | 0.061 | 47.3 | 0.053 | 0.071 | 52.8 | 0.026 | 0.066 | 56.4 |
| 3 | 0.062 | 0.031 | 43.9 | 0.061 | 0.066 | 47.5 | 0.058 | 0.065 | 52.8 | 0.04 | 0.061 | 56 |
| 4.9 | 0.06 | 0.036 | 44 | 0.068 | 0.056 | 47.1 | 0.047 | 0.075 | 53.1 | 0.044 | 0.047 | 55.5 |
| 6.9 | 0.066 | 0.027 | 43.7 | 0.065 | 0.06 | 47.4 | 0.066 | 0.064 | 52.6 | 0.032 | 0.066 | 56.2 |
| 4.4 | 0.066 | 0.028 | 43.7 | 0.069 | 0.054 | 47.2 | 0.053 | 0.073 | 52.9 | 0.031 | 0.064 | 56.1 |
| 5.2 | 0.062 | 0.038 | 44.1 | 0.064 | 0.061 | 47.4 | 0.051 | 0.067 | 52.9 | 0.034 | 0.062 | 56.1 |
| 5.7 | 0.056 | 0.046 | 44.4 | 0.07 | 0.055 | 47.2 | 0.065 | 0.061 | 52.5 | 0.04 | 0.065 | 56 |
| 6 | 0.067 | 0.036 | 44 | 0.066 | 0.052 | 47.4 | 0.054 | 0.073 | 52.9 | 0.034 | 0.061 | 56 |
| 3.3 | 0.061 | 0.044 | 44.3 | 0.064 | 0.063 | 47.4 | 0.052 | 0.075 | 53 | 0.045 | 0.055 | 55.6 |
| 5.9 | 0.062 | 0.039 | 44.1 | 0.057 | 0.066 | 47.7 | 0.053 | 0.072 | 52.7 | 0.043 | 0.062 | 55.8 |
| 7.7 | 0.056 | 0.044 | 44.4 | 0.074 | 0.052 | 47.2 | 0.056 | 0.066 | 52.8 | 0.039 | 0.06 | 55.9 |
| 5.6 | 0.059 | 0.043 | 44.3 | 0.066 | 0.057 | 47.3 | 0.06 | 0.062 | 52.5 | 0.034 | 0.064 | 56 |
| 5.4 | 0.073 | 0.029 | 43.6 | 0.05 | 0.07 | 47.9 | 0.049 | 0.077 | 53 | 0.026 | 0.077 | 56.5 |
| 8.2 | 0.066 | 0.038 | 44 | 0.057 | 0.067 | 47.7 | 0.053 | 0.07 | 52.8 | 0.04 | 0.064 | 55.9 |
| 8.5 | 0.055 | 0.051 | 44.5 | 0.053 | 0.067 | 47.8 | 0.064 | 0.059 | 52.3 | 0.037 | 0.065 | 55.9 |
| 8.1 | 0.057 | 0.05 | 44.4 | 0.06 | 0.057 | 47.5 | 0.061 | 0.065 | 52.4 | 0.04 | 0.07 | 56 |
| 7.6 | 0.064 | 0.038 | 44 | 0.069 | 0.048 | 47.3 | 0.062 | 0.06 | 52.3 | 0.046 | 0.06 | 55.6 |
| 8.1 | 0.076 | 0.039 | 43.9 | 0.06 | 0.059 | 47.6 | 0.064 | 0.056 | 52.2 | 0.036 | 0.069 | 56.1 |
| 7.9 | 0.061 | 0.041 | 44.2 | 0.056 | 0.063 | 47.8 | 0.048 | 0.068 | 52.8 | 0.058 | 0.051 | 55.2 |
| 7.3 | 0.069 | 0.041 | 44.1 | 0.063 | 0.06 | 47.6 | 0.063 | 0.061 | 52.2 | 0.047 | 0.064 | 55.6 |
| 9.3 | 0.076 | 0.036 | 43.9 | 0.053 | 0.066 | 48 | 0.06 | 0.059 | 52.3 | 0.054 | 0.06 | 55.5 |
| 7.3 | 0.064 | 0.051 | 44.4 | 0.047 | 0.069 | 48.2 | 0.064 | 0.061 | 52.2 | 0.035 | 0.079 | 56.2 |
| 8.1 | 0.079 | 0.039 | 43.9 | 0.044 | 0.07 | 48.2 | 0.062 | 0.052 | 51.9 | 0.038 | 0.078 | 56.1 |
| 9.4 | 0.078 | 0.039 | 43.9 | 0.058 | 0.061 | 47.8 | 0.06 | 0.052 | 52 | 0.042 | 0.074 | 56 |
| 8.4 | 0.067 | 0.047 | 44.2 | 0.052 | 0.061 | 48 | 0.059 | 0.057 | 52 | 0.06 | 0.057 | 55.3 |
| 8.4 | 0.076 | 0.043 | 44.2 | 0.053 | 0.064 | 48 | 0.064 | 0.052 | 51.9 | 0.039 | 0.083 | 56 |
| 8.4 | 0.066 | 0.056 | 44.5 | 0.055 | 0.059 | 48 | 0.056 | 0.055 | 52.1 | 0.047 | 0.069 | 55.8 |
| 7.7 | 0.076 | 0.045 | 44.2 | 0.044 | 0.069 | 48.5 | 0.065 | 0.047 | 51.7 | 0.055 | 0.067 | 55.5 |
| 8.1 | 0.081 | 0.045 | 44 | 0.059 | 0.051 | 47.8 | 0.068 | 0.051 | 51.7 | 0.047 | 0.077 | 55.7 |
| 7.7 | 0.071 | 0.056 | 44.4 | 0.056 | 0.056 | 48.2 | 0.06 | 0.05 | 51.9 | 0.051 | 0.074 | 55.7 |
| 8.3 | 0.067 | 0.062 | 44.8 | 0.045 | 0.061 | 48.5 | 0.061 | 0.042 | 51.5 | 0.063 | 0.066 | 55.3 |
| 8.7 | 0.083 | 0.046 | 44 | 0.031 | 0.073 | 48.9 | 0.06 | 0.041 | 51.5 | 0.048 | 0.08 | 55.7 |
| 8.1 | 0.075 | 0.058 | 44.5 | 0.049 | 0.049 | 48.2 | 0.064 | 0.041 | 51.4 | 0.056 | 0.075 | 55.6 |
| 8.9 | 0.079 | 0.058 | 44.3 | 0.036 | 0.059 | 48.8 | 0.058 | 0.035 | 51.1 | 0.061 | 0.076 | 55.5 |
| 9.1 | 0.091 | 0.05 | 44 | 0.039 | 0.058 | 48.8 | 0.058 | 0.033 | 51.1 | 0.057 | 0.08 | 55.6 |
| 8.3 | 0.087 | 0.06 | 44.3 | 0.038 | 0.058 | 48.9 | 0.058 | 0.039 | 51.1 | 0.077 | 0.069 | 55 |
| 10 | 0.078 | 0.059 | 44.4 | 0.03 | 0.058 | 49.2 | 0.05 | 0.038 | 51.2 | 0.061 | 0.077 | 55.5 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9.1 | 0.086 | 0.058 | 44.3 | 0.037 | 0.047 | 48.9 | 0.052 | 0.031 | 50.8 | 0.057 | 0.089 | 55.8 |
| 10.2 | 0.07 | 0.083 | 45 | 0.027 | 0.052 | 49.4 | 0.058 | 0.022 | 50.3 | 0.045 | 0.105 | 56.2 |
| 9.5 | 0.108 | 0.051 | 43.8 | 0.023 | 0.056 | 49.7 | 0.047 | 0.026 | 50.5 | 0.066 | 0.087 | 55.6 |
| 10 | 0.085 | 0.075 | 44.6 | 0.026 | 0.041 | 49.5 | 0.046 | 0.023 | 50.4 | 0.06 | 0.091 | 55.8 |
| 9.7 | 0.098 | 0.064 | 44.3 | 0.025 | 0.036 | 49.6 | 0.037 | 0.023 | 50.3 | 0.077 | 0.087 | 55.4 |
| 9.9 | 0.094 | 0.076 | 44.7 | 0.017 | 0.036 | 49.9 | 0.034 | 0.018 | 50.1 | 0.08 | 0.087 | 55.2 |
| 9.7 | 0.112 | 0.065 | 44.2 | 0.016 | 0.03 | 50 | 0.028 | 0.017 | 50 | 0.084 | 0.093 | 55.4 |
| 10.1 | 0.092 | 0.091 | 44.8 | 0.02 | 0.022 | 49.8 | 0.024 | 0.016 | 49.9 | 0.067 | 0.113 | 55.8 |
| 10.1 | 0.099 | 0.093 | 44.8 | 0.018 | 0.016 | 49.7 | 0.022 | 0.014 | 49.7 | 0.069 | 0.127 | 55.9 |
| 10 | 0.141 | 0.062 | 43.7 | 0.012 | 0.015 | 50.1 | 0.012 | 0.014 | 50.1 | 0.096 | 0.104 | 55.2 |
| 9.9 | 0.165 | 0.047 | 43.2 | 0.016 | 0.009 | 49.7 | 0.016 | 0.009 | 49.7 | 0.123 | 0.088 | 54.6 |

Table F2: Actual Error Rates at True Cutscore Location with Increasing Bimodality

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TS D | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 0.8 | 0.04 | 0.061 | 45 | 0.058 | 0.067 | 47.5 | 0.072 | 0.061 | 52.5 | 0.064 | 0.039 | 55 |
| 2.8 | 0.039 | 0.064 | 45 | 0.061 | 0.069 | 47.5 | 0.07 | 0.056 | 52.5 | 0.062 | 0.036 | 55 |
| 3.1 | 0.041 | 0.065 | 45 | 0.058 | 0.068 | 47.5 | 0.064 | 0.062 | 52.5 | 0.063 | 0.041 | 55 |
| 3 | 0.04 | 0.063 | 45 | 0.061 | 0.066 | 47.5 | 0.069 | 0.056 | 52.5 | 0.068 | 0.042 | 55 |
| 4.9 | 0.04 | 0.062 | 45 | 0.056 | 0.07 | 47.5 | 0.068 | 0.058 | 52.5 | 0.058 | 0.038 | 55 |
| 6.9 | 0.041 | 0.061 | 45 | 0.063 | 0.064 | 47.5 | 0.07 | 0.061 | 52.5 | 0.064 | 0.042 | 55 |
| 4.4 | 0.042 | 0.061 | 45 | 0.061 | 0.064 | 47.5 | 0.067 | 0.06 | 52.5 | 0.062 | 0.041 | 55 |
| 5.2 | 0.044 | 0.063 | 45 | 0.061 | 0.064 | 47.5 | 0.065 | 0.057 | 52.5 | 0.064 | 0.041 | 55 |
| 5.7 | 0.044 | 0.062 | 45 | 0.062 | 0.066 | 47.5 | 0.065 | 0.061 | 52.5 | 0.065 | 0.044 | 55 |
| 6 | 0.044 | 0.062 | 45 | 0.062 | 0.056 | 47.5 | 0.066 | 0.062 | 52.5 | 0.06 | 0.042 | 55 |
| 3.3 | 0.047 | 0.064 | 45 | 0.061 | 0.066 | 47.5 | 0.068 | 0.061 | 52.5 | 0.064 | 0.043 | 55 |
| 5.9 | 0.044 | 0.062 | 45 | 0.063 | 0.062 | 47.5 | 0.061 | 0.066 | 52.5 | 0.065 | 0.046 | 55 |
| 7.7 | 0.044 | 0.061 | 45 | 0.065 | 0.064 | 47.5 | 0.066 | 0.059 | 52.5 | 0.061 | 0.044 | 55 |
| 5.6 | 0.042 | 0.063 | 45 | 0.06 | 0.063 | 47.5 | 0.06 | 0.062 | 52.5 | 0.065 | 0.045 | 55 |
| 5.4 | 0.043 | 0.067 | 45 | 0.063 | 0.058 | 47.5 | 0.065 | 0.064 | 52.5 | 0.063 | 0.045 | 55 |
| 8.2 | 0.045 | 0.066 | 45 | 0.063 | 0.062 | 47.5 | 0.062 | 0.063 | 52.5 | 0.065 | 0.045 | 55 |
| 8.5 | 0.045 | 0.066 | 45 | 0.062 | 0.059 | 47.5 | 0.059 | 0.065 | 52.5 | 0.062 | 0.045 | 55 |
| 8.1 | 0.043 | 0.068 | 45 | 0.06 | 0.057 | 47.5 | 0.059 | 0.069 | 52.5 | 0.067 | 0.048 | 55 |
| 7.6 | 0.042 | 0.065 | 45 | 0.062 | 0.056 | 47.5 | 0.057 | 0.066 | 52.5 | 0.066 | 0.045 | 55 |
| 8.1 | 0.048 | 0.071 | 45 | 0.063 | 0.057 | 47.5 | 0.057 | 0.066 | 52.5 | 0.063 | 0.045 | 55 |
| 7.9 | 0.045 | 0.067 | 45 | 0.065 | 0.055 | 47.5 | 0.059 | 0.06 | 52.5 | 0.065 | 0.046 | 55 |
| 7.3 | 0.046 | 0.068 | 45 | 0.066 | 0.058 | 47.5 | 0.054 | 0.07 | 52.5 | 0.064 | 0.051 | 55 |
| 9.3 | 0.05 | 0.064 | 45 | 0.07 | 0.054 | 47.5 | 0.055 | 0.065 | 52.5 | 0.07 | 0.047 | 55 |
| 7.3 | 0.049 | 0.068 | 45 | 0.069 | 0.051 | 47.5 | 0.056 | 0.07 | 52.5 | 0.068 | 0.05 | 55 |
| 8.1 | 0.05 | 0.07 | 45 | 0.068 | 0.051 | 47.5 | 0.048 | 0.07 | 52.5 | 0.071 | 0.051 | 55 |
| 9.4 | 0.05 | 0.075 | 45 | 0.067 | 0.053 | 47.5 | 0.048 | 0.068 | 52.5 | 0.072 | 0.05 | 55 |
| 8.4 | 0.049 | 0.072 | 45 | 0.068 | 0.049 | 47.5 | 0.046 | 0.073 | 52.5 | 0.071 | 0.049 | 55 |
| 8.4 | 0.057 | 0.069 | 45 | 0.07 | 0.051 | 47.5 | 0.05 | 0.071 | 52.5 | 0.068 | 0.057 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8.4 | 0.053 | 0.074 | 45 | 0.071 | 0.048 | 47.5 | 0.045 | 0.07 | 52.5 | 0.073 | 0.049 | 55 |
| 7.7 | 0.056 | 0.07 | 45 | 0.072 | 0.046 | 47.5 | 0.045 | 0.074 | 52.5 | 0.071 | 0.055 | 55 |
| 8.1 | 0.057 | 0.076 | 45 | 0.07 | 0.045 | 47.5 | 0.048 | 0.075 | 52.5 | 0.07 | 0.058 | 55 |
| 7.7 | 0.056 | 0.076 | 45 | 0.079 | 0.043 | 47.5 | 0.046 | 0.07 | 52.5 | 0.073 | 0.056 | 55 |
| 8.3 | 0.062 | 0.069 | 45 | 0.079 | 0.04 | 47.5 | 0.039 | 0.077 | 52.5 | 0.073 | 0.058 | 55 |
| 8.7 | 0.057 | 0.079 | 45 | 0.074 | 0.04 | 47.5 | 0.037 | 0.075 | 52.5 | 0.07 | 0.06 | 55 |
| 8.1 | 0.061 | 0.072 | 45 | 0.075 | 0.035 | 47.5 | 0.04 | 0.075 | 52.5 | 0.078 | 0.059 | 55 |
| 8.9 | 0.06 | 0.079 | 45 | 0.079 | 0.032 | 47.5 | 0.031 | 0.08 | 52.5 | 0.076 | 0.062 | 55 |
| 9.1 | 0.062 | 0.082 | 45 | 0.079 | 0.033 | 47.5 | 0.031 | 0.085 | 52.5 | 0.077 | 0.062 | 55 |
| 8.3 | 0.066 | 0.084 | 45 | 0.079 | 0.033 | 47.5 | 0.033 | 0.084 | 52.5 | 0.077 | 0.069 | 55 |
| 10 | 0.063 | 0.08 | 45 | 0.084 | 0.027 | 47.5 | 0.027 | 0.082 | 52.5 | 0.079 | 0.064 | 55 |
| 9.1 | 0.067 | 0.084 | 45 | 0.085 | 0.025 | 47.5 | 0.026 | 0.085 | 52.5 | 0.085 | 0.064 | 55 |
| 10.2 | 0.07 | 0.083 | 45 | 0.086 | 0.022 | 47.5 | 0.023 | 0.087 | 52.5 | 0.086 | 0.068 | 55 |
| 9.5 | 0.072 | 0.091 | 45 | 0.09 | 0.022 | 47.5 | 0.02 | 0.09 | 52.5 | 0.088 | 0.07 | 55 |
| 10 | 0.073 | 0.088 | 45 | 0.089 | 0.017 | 47.5 | 0.019 | 0.088 | 52.5 | 0.091 | 0.067 | 55 |
| 9.7 | 0.078 | 0.09 | 45 | 0.093 | 0.012 | 47.5 | 0.012 | 0.089 | 52.5 | 0.091 | 0.076 | 55 |
| 9.9 | 0.085 | 0.087 | 45 | 0.093 | 0.011 | 47.5 | 0.011 | 0.09 | 52.5 | 0.088 | 0.081 | 55 |
| 9.7 | 0.086 | 0.094 | 45 | 0.091 | 0.007 | 47.5 | 0.006 | 0.093 | 52.5 | 0.1 | 0.081 | 55 |
| 10.1 | 0.086 | 0.1 | 45 | 0.089 | 0.005 | 47.5 | 0.005 | 0.09 | 52.5 | 0.099 | 0.088 | 55 |
| 10.1 | 0.091 | 0.102 | 45 | 0.086 | 0.002 | 47.5 | 0.002 | 0.095 | 52.5 | 0.104 | 0.097 | 55 |
| 10 | 0.095 | 0.113 | 45 | 0.083 | 0.001 | 47.5 | 0.001 | 0.086 | 52.5 | 0.104 | 0.097 | 55 |
| 9.9 | 0.103 | 0.115 | 45 | 0.083 | 0.001 | 47.5 | 0.001 | 0.082 | 52.5 | 0.109 | 0.103 | 55 |

Table F3: GW-CSOF Estimate of Error at & Location of Optimal Cutscore with Increasing Bimodality

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TS D | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 0.8 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 2.8 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 3.1 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 3 | 0.056 | 0.028 | 43.7 | 0.075 | 0.052 | 46.8 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.9 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.03 | 0.055 | 56.2 |
| 6.9 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.4 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.056 | 56.3 |
| 5.2 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 5.7 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 6 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 3.3 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 5.9 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 7.7 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.056 | 56.2 |
| 5.6 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 5.4 | 0.057 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8.2 | 0.057 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 8.5 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 8.1 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.029 | 0.058 | 56.3 |
| 7.6 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 8.1 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.03 | 0.055 | 56.2 |
| 7.9 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 7.3 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.031 | 0.056 | 56.2 |
| 9.3 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.029 | 0.057 | 56.3 |
| 7.3 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 8.1 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.031 | 0.056 | 56.2 |
| 9.4 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 8.4 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 8.4 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 8.4 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 7.7 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.029 | 0.057 | 56.3 |
| 8.1 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 7.7 | 0.057 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.056 | 56.2 |
| 8.3 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.056 | 56.2 |
| 8.7 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 8.1 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 8.9 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.029 | 0.058 | 56.3 |
| 9.1 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 8.3 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.031 | 0.056 | 56.2 |
| 10 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 9.1 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 10.2 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 9.5 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 10 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 9.7 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 9.9 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 9.7 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 10.1 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 10.1 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.029 | 0.057 | 56.3 |
| 10 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.056 | 56.2 |
| 9.9 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |

Table F4: GW-CSOF Error Rates at True Cutscore Location with Increasing Bimodality

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 0.8 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 2.8 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.1 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 3 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4.9 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 6.9 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 4.4 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.06 | 0.033 | 55 |
| 5.2 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 5.7 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 6 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.3 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 5.9 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 7.7 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 5.6 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 5.4 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.2 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.5 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.1 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 7.6 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 8.1 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 7.9 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 7.3 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 9.3 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 7.3 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 8.1 | 0.036 | 0.062 | 45 | 0.058 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 9.4 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 8.4 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.4 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.4 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 7.7 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.1 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.062 | 0.035 | 55 |
| 7.7 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 8.3 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.7 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 8.1 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 8.9 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.062 | 0.035 | 55 |
| 9.1 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 8.3 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.035 | 55 |
| 10 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 9.1 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 10.2 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 9.5 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 10 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 9.7 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 9.9 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 9.7 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 10.1 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 10.1 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.035 | 55 |
| 10 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9.9 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |

Figure F 1 : Iteration # 26 (distance of in mixture means= 5 )
True Score Frequencies with D undefined

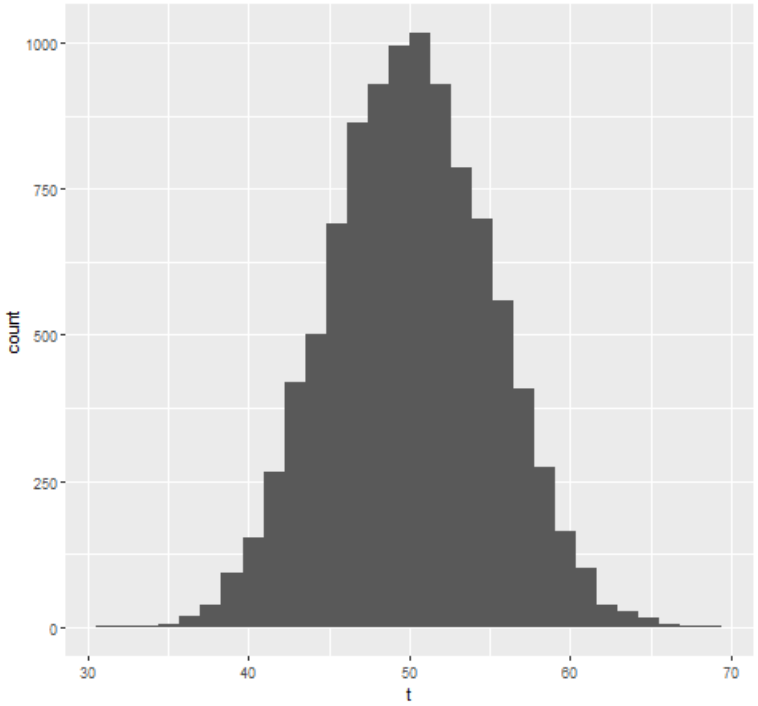



Figure F 2 : Iteration # 26 (distance of in mixture means= 5 )
Observed Score Frequencies with D undefined

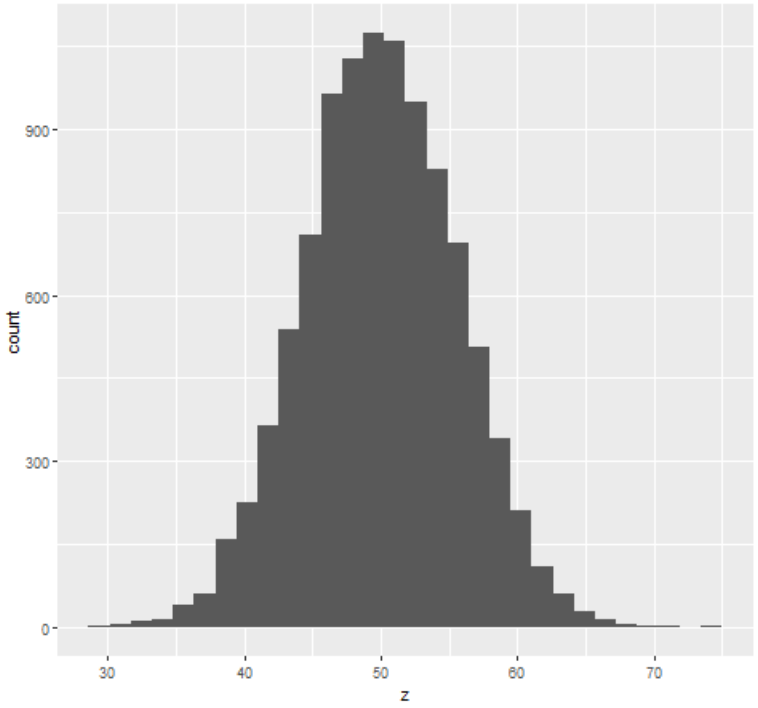


Figure F 3 : Iteration # 27 (distance of in mixture means= 5.2 )
True Score Frequencies with D undefined

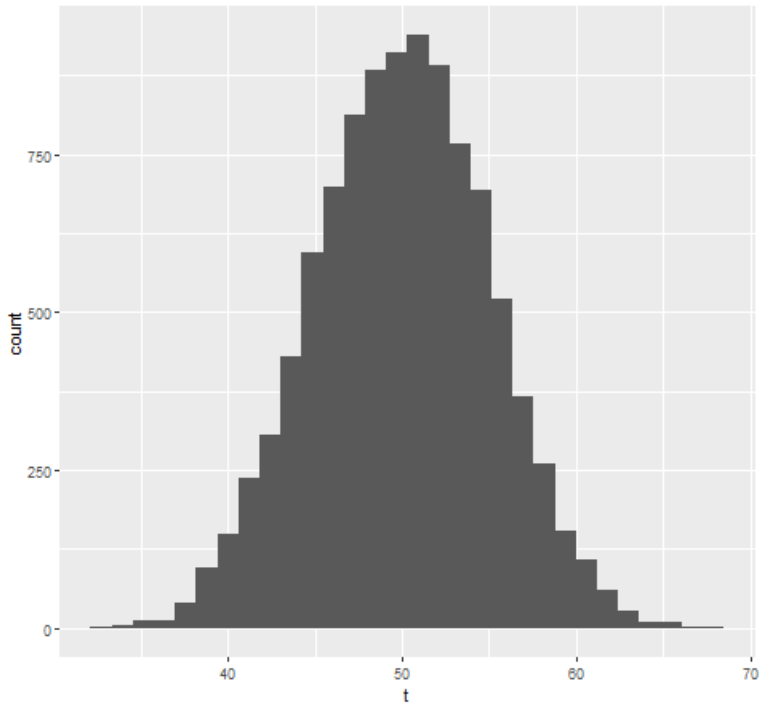



Figure F 4 : Iteration # 27 (distance of in mixture means= 5.2 )
Observed Score Frequencies with D undefined

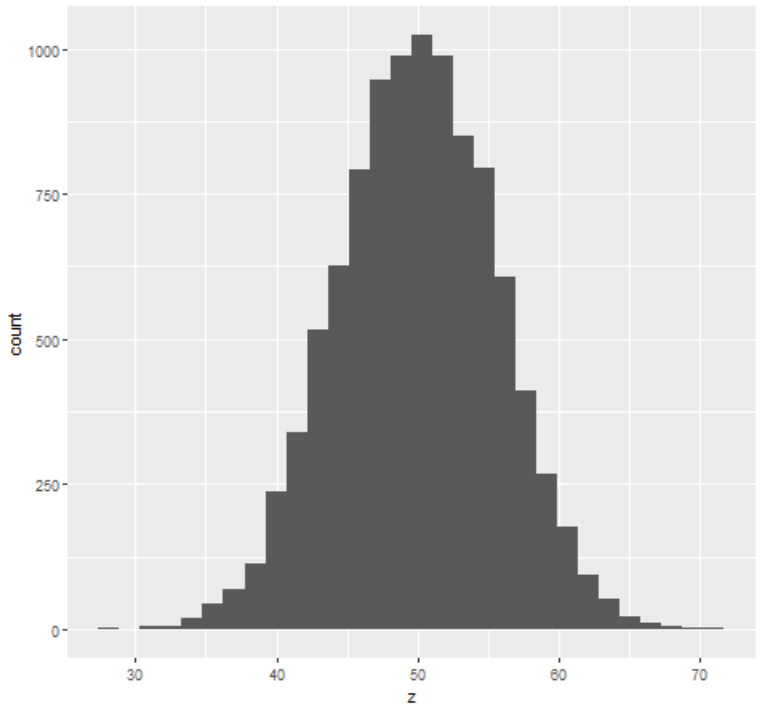


Figure F 5 : Iteration # 28 (distance of in mixture means= 5.4 )
True Score Frequencies with D undefined

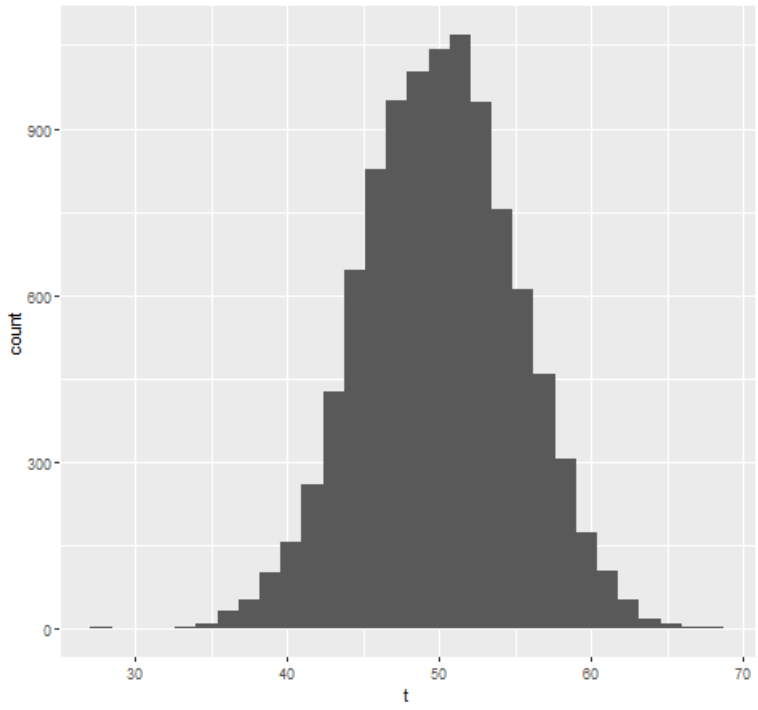



Figure F 6 : Iteration # 28 (distance of in mixture means= 5.4 )
Observed Score Frequencies with D undefined

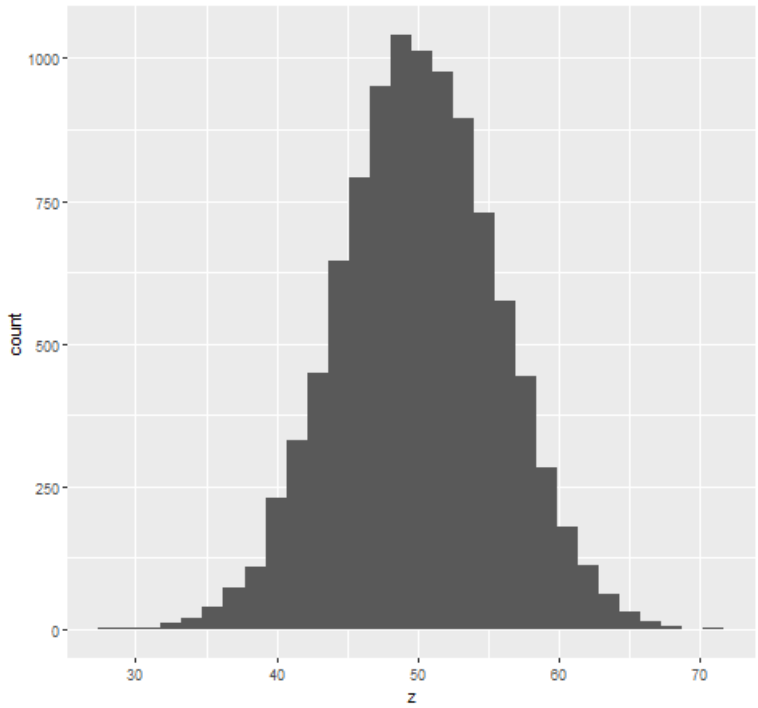


Figure F 7 : Iteration # 29 (distance of in mixture means= 5.6 )
True Score Frequencies with D undefined

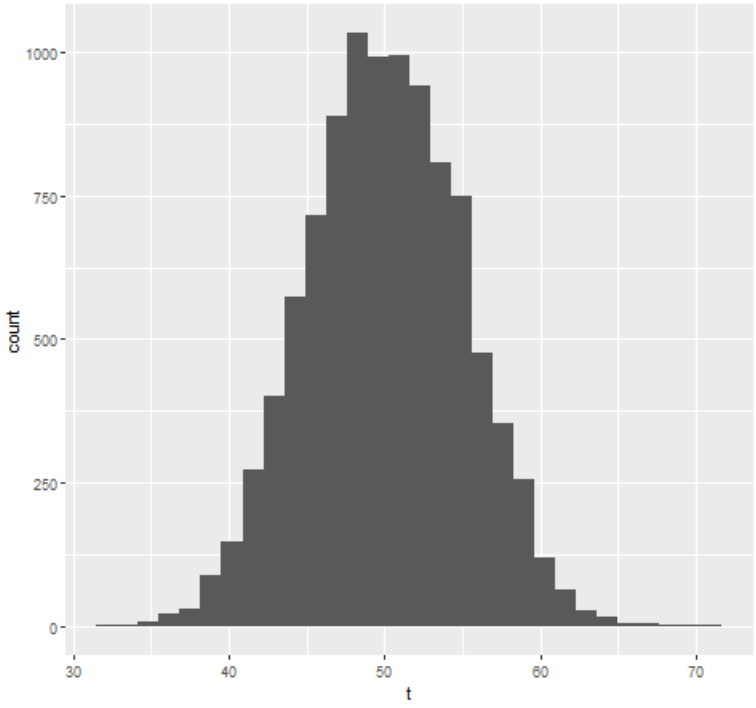



Figure F 8 : Iteration # 29 (distance of in mixture means= 5.6 )
Observed Score Frequencies with D undefined

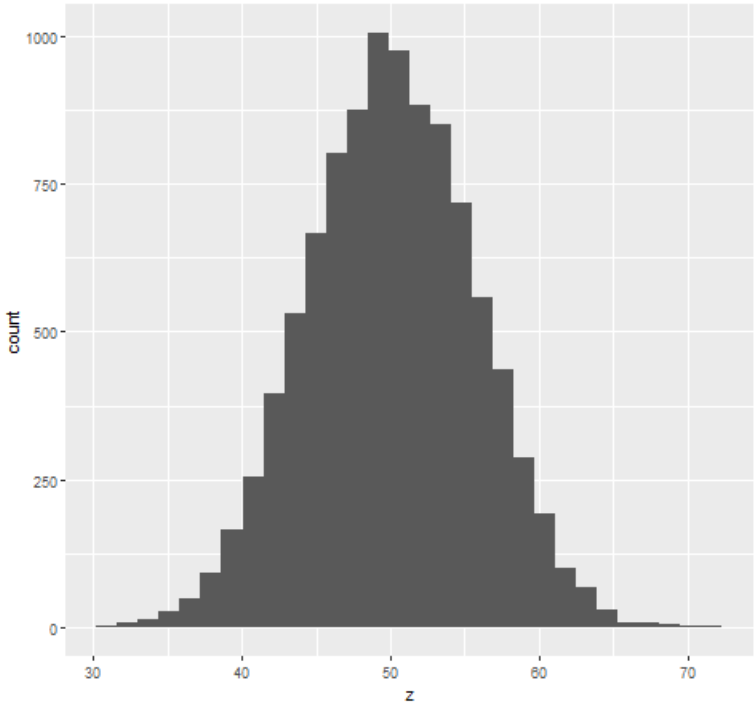


Figure F 9 : Iteration # 30 (distance of in mixture means= 5.8 )
True Score Frequencies with D undefined

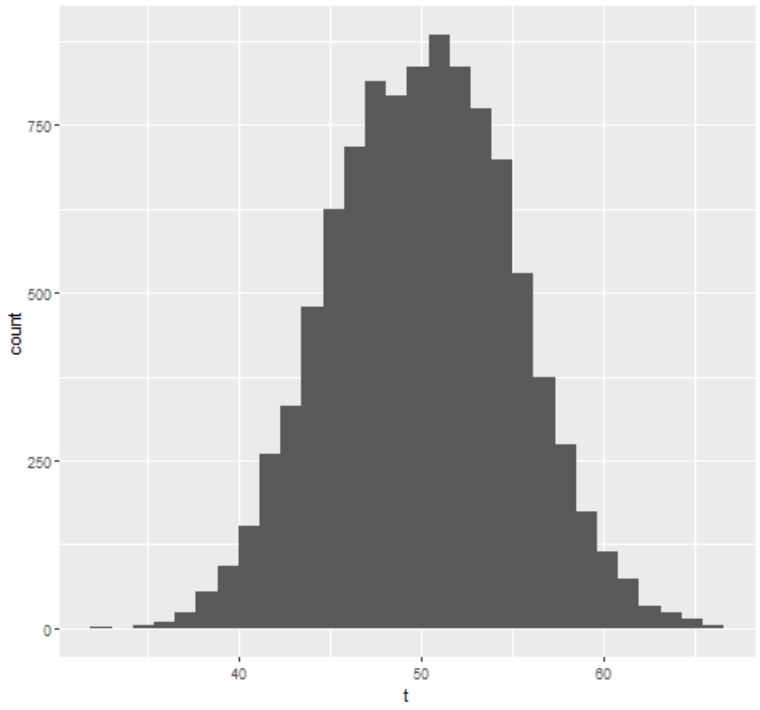



Figure F 10 : Iteration # 30 (distance of in mixture means= 5.8 )
Observed Score Frequencies with D undefined

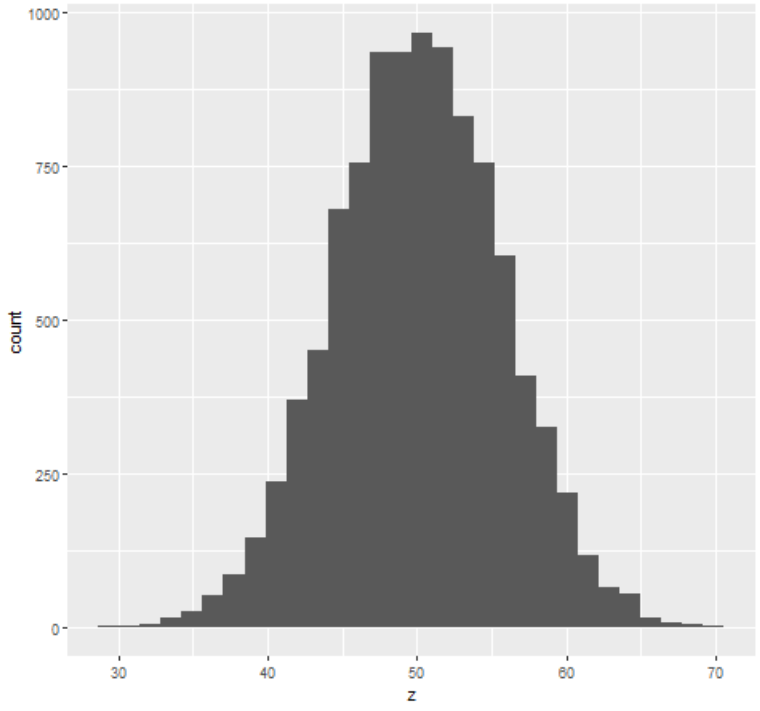


Figure F 11 : Iteration # 31 (distance of in mixture means= 6 )
True Score Frequencies with D undefined

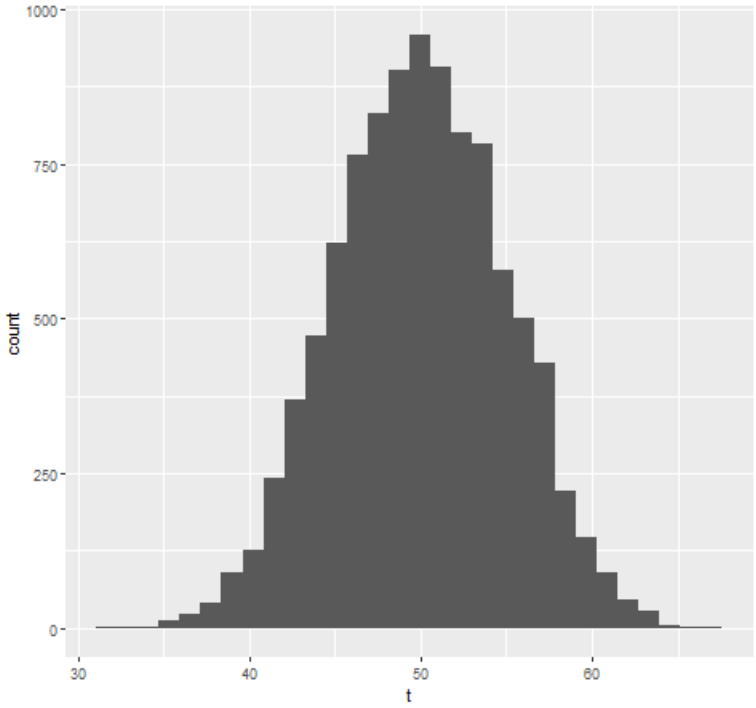



Figure F 12 : Iteration # 31 (distance of in mixture means= 6 )
Observed Score Frequencies with D undefined

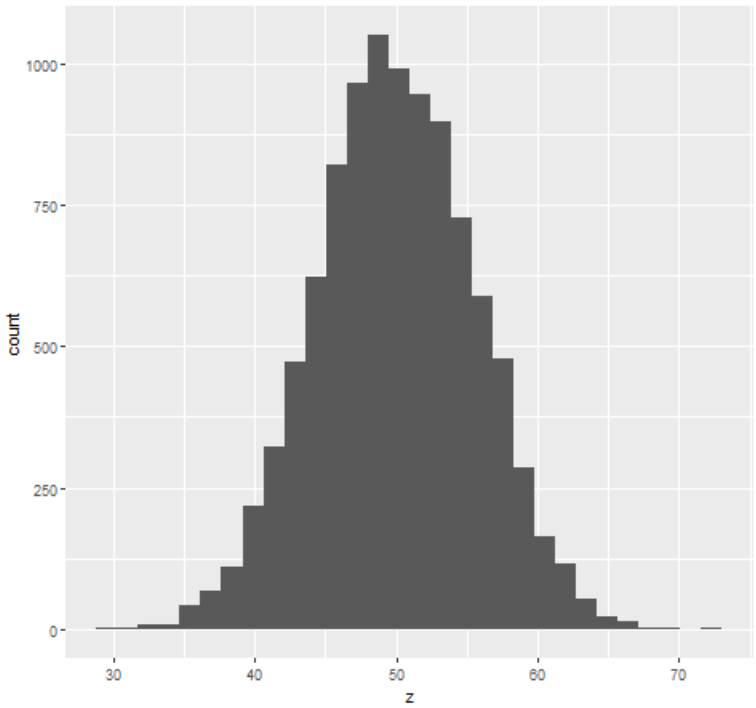


Figure F 13 : Iteration # 32 (distance of in mixture means= 6.2 )
True Score Frequencies with D undefined

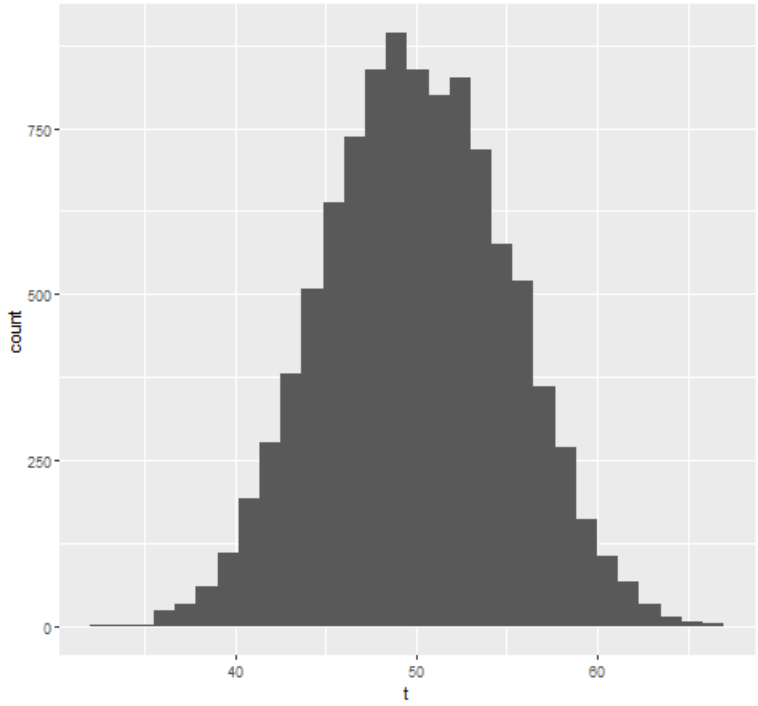



Figure F 14 : Iteration # 32 (distance of in mixture means= 6.2 )
Observed Score Frequencies with D undefined

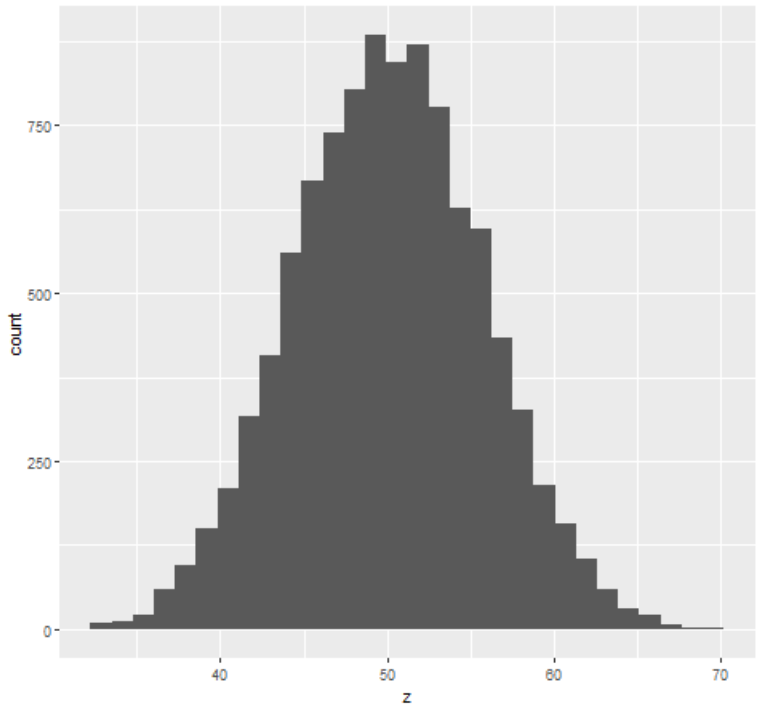


Figure F 15 : Iteration # 33 (distance of in mixture means= 6.4 )
True Score Frequencies with D undefined

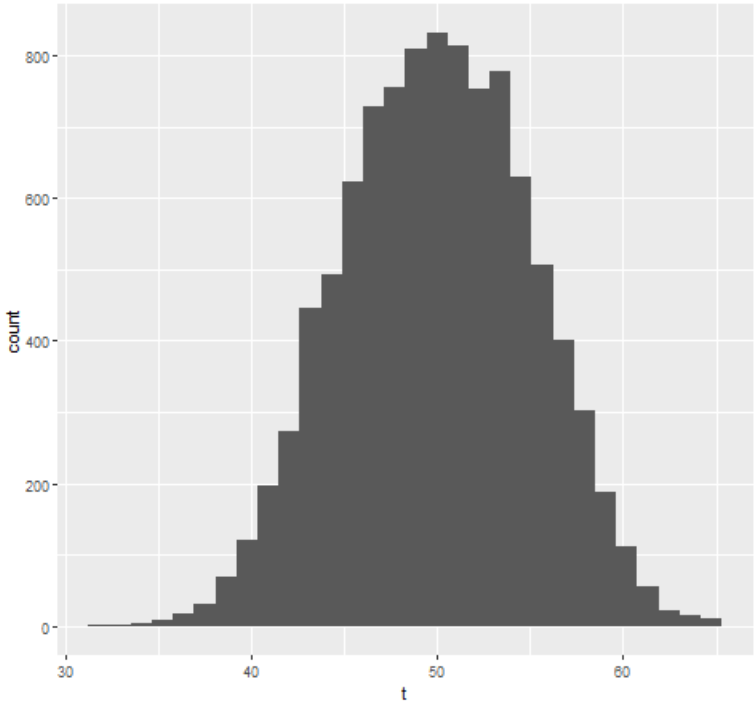



Figure F 16 : Iteration # 33 (distance of in mixture means= 6.4 )
Observed Score Frequencies with D undefined

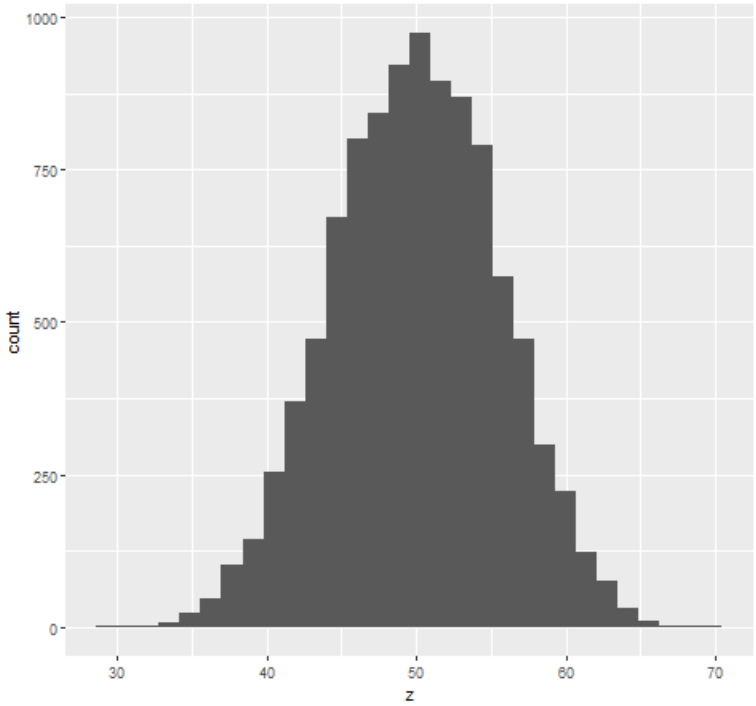


Figure F 17 : Iteration # 34 (distance of in mixture means= 6.6 )
True Score Frequencies with D undefined

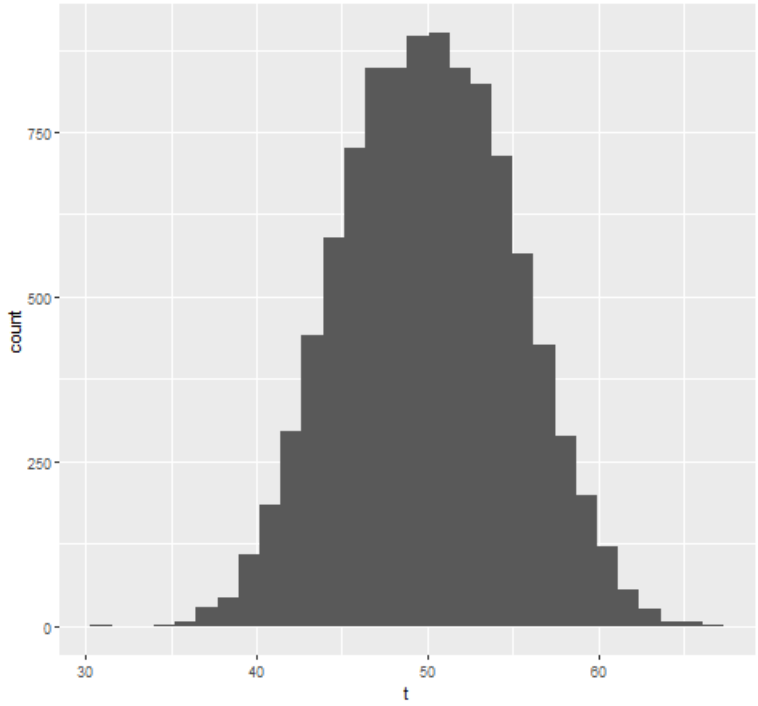



Figure F 18 : Iteration # 34 (distance of in mixture means= 6.6 )
Observed Score Frequencies with D undefined

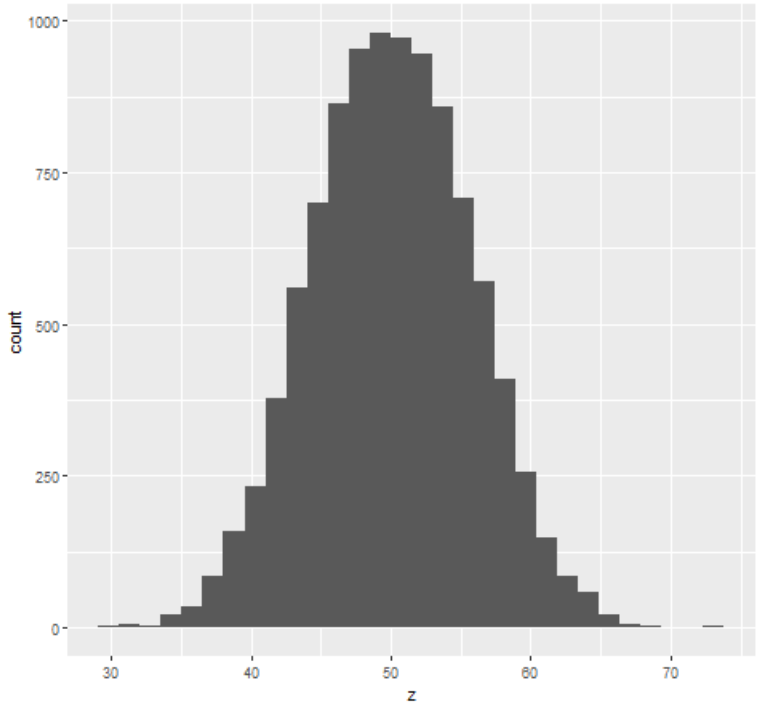


Figure F 19 : Iteration # 35 (distance of in mixture means= 6.8 )
True Score Frequencies with D undefined

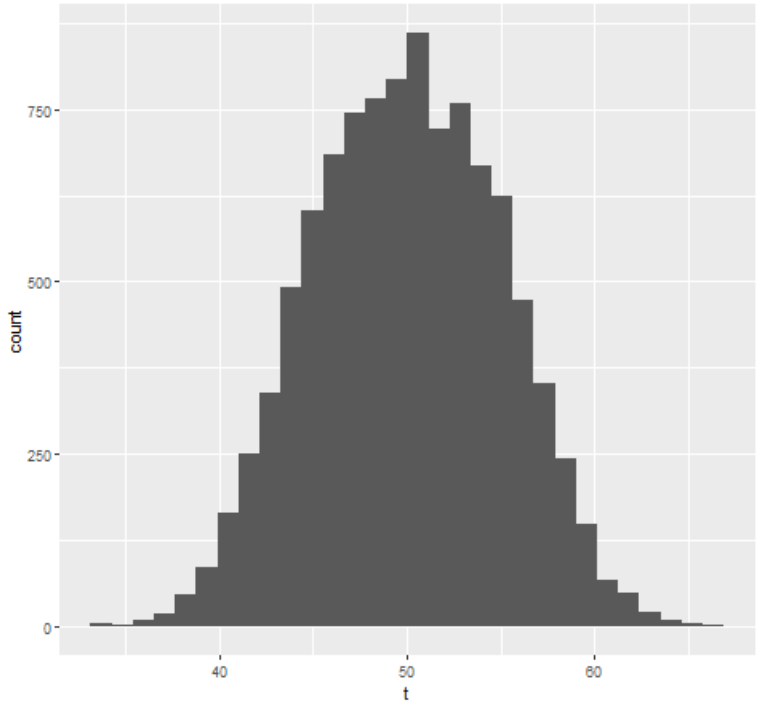



Figure F 20 : Iteration # 35 (distance of in mixture means= 6.8 )
Observed Score Frequencies with D undefined

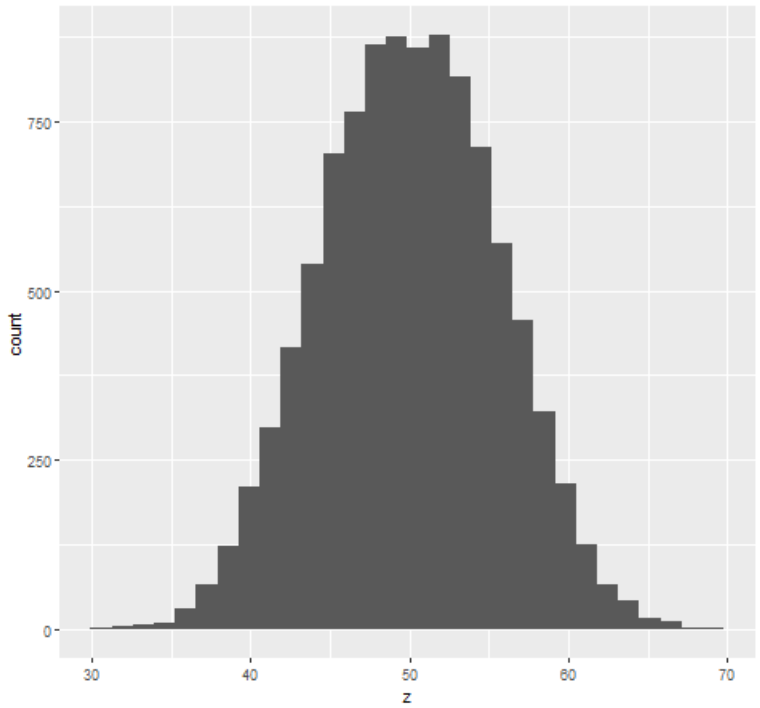


Figure F 21 : Iteration # 36 (distance of mixture means= 7 )
True Score Frequencies with D = 1.5

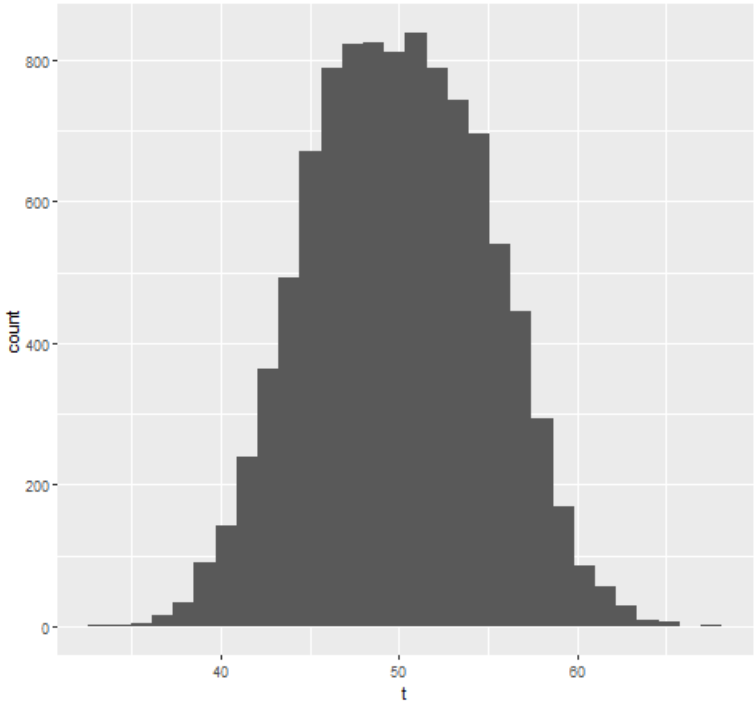



Figure F 22 : Iteration # 36 (distance of in mixture means= 7 )
Observed Score Frequencies with D = 1.1

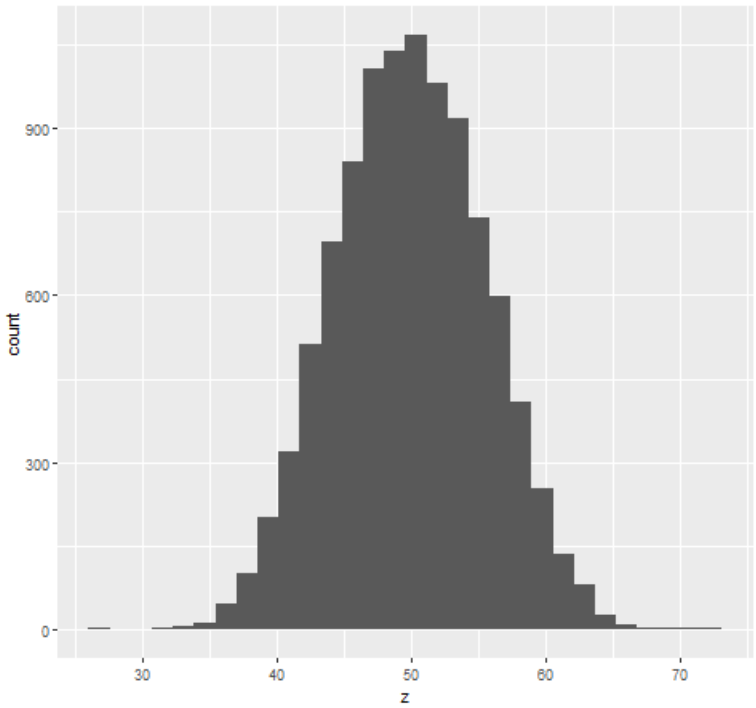


Figure F 23 : Iteration # 37 (distance of mixture means= 7.2 )
True Score Frequencies with D = 3.7

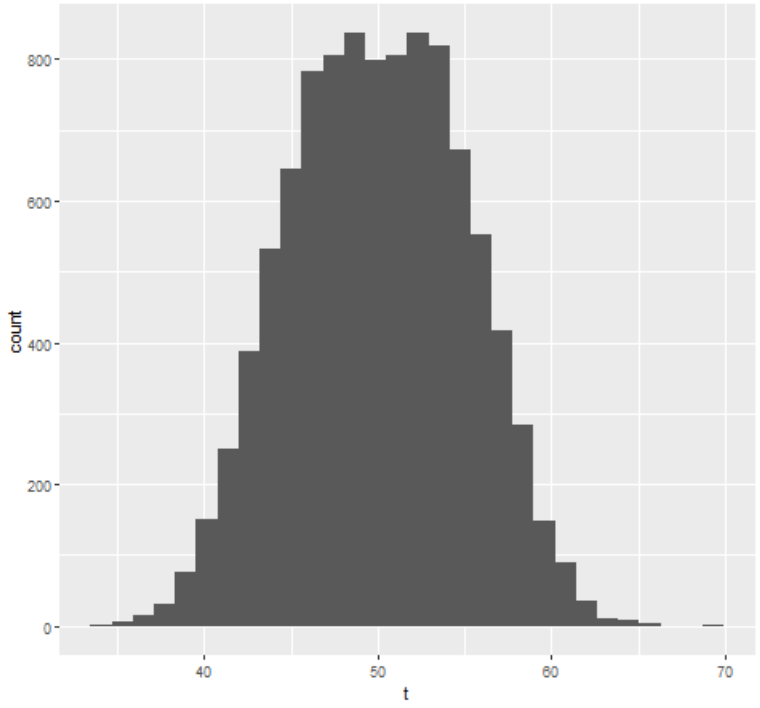



Figure F 24 : Iteration # 37 (distance of in mixture means= 7.2 )
Observed Score Frequencies with D = 2.4

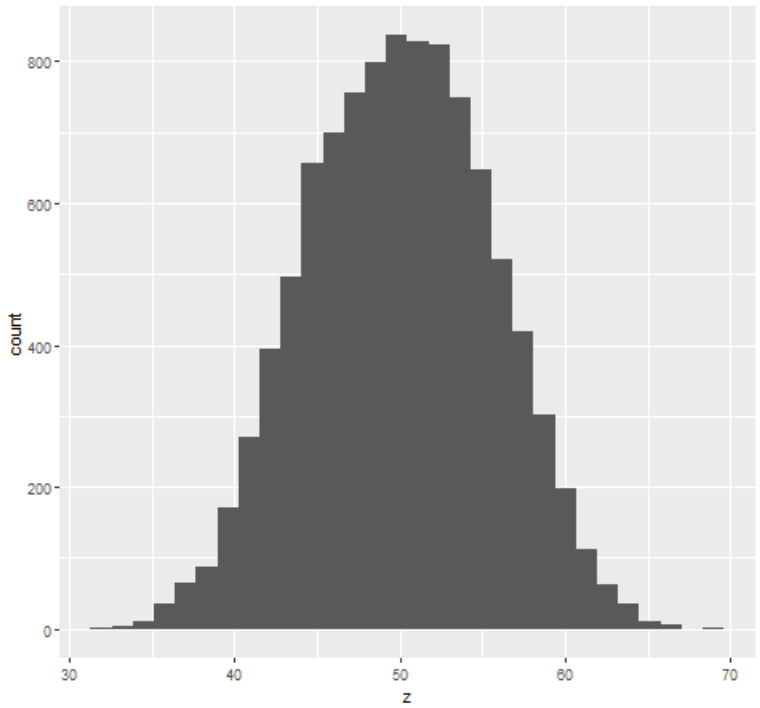


Figure F 25 : Iteration # 38 (distance of mixture means= 7.4 )
True Score Frequencies with D = 2.8

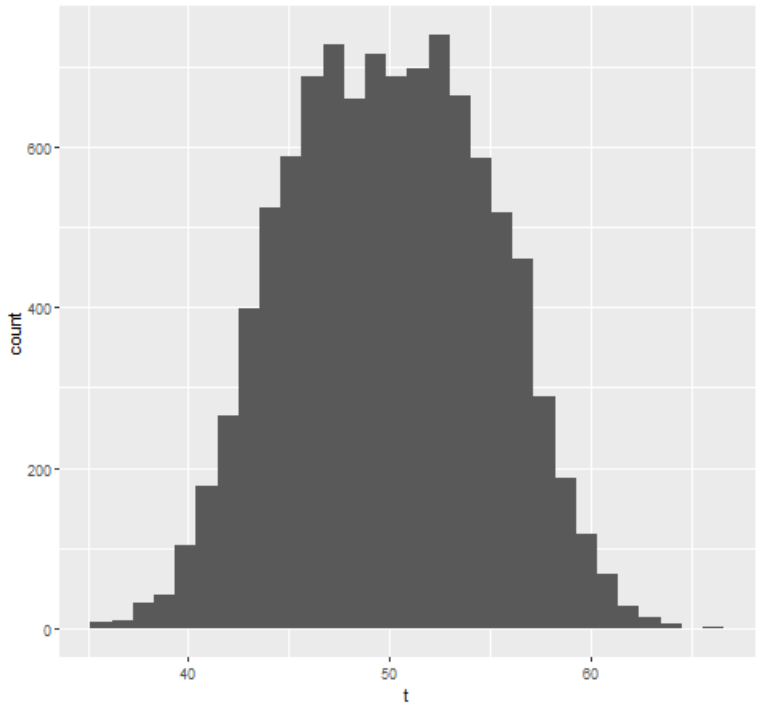



Figure F 26 : Iteration # 38 (distance of in mixture means= 7.4 )
Observed Score Frequencies with D = 2.9

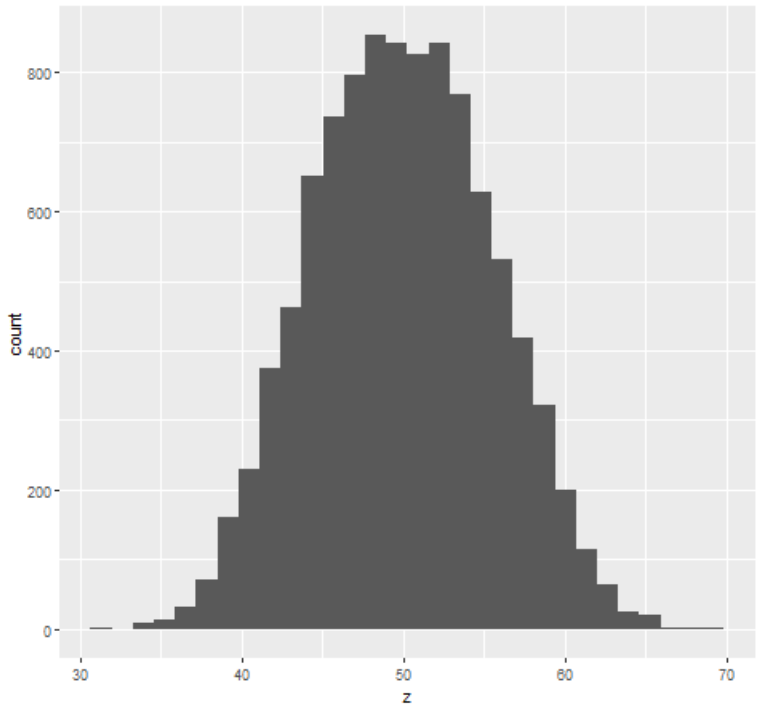


Figure F 27 : Iteration # 39 (distance of mixture means= 7.6 )
True Score Frequencies with D = 4.7

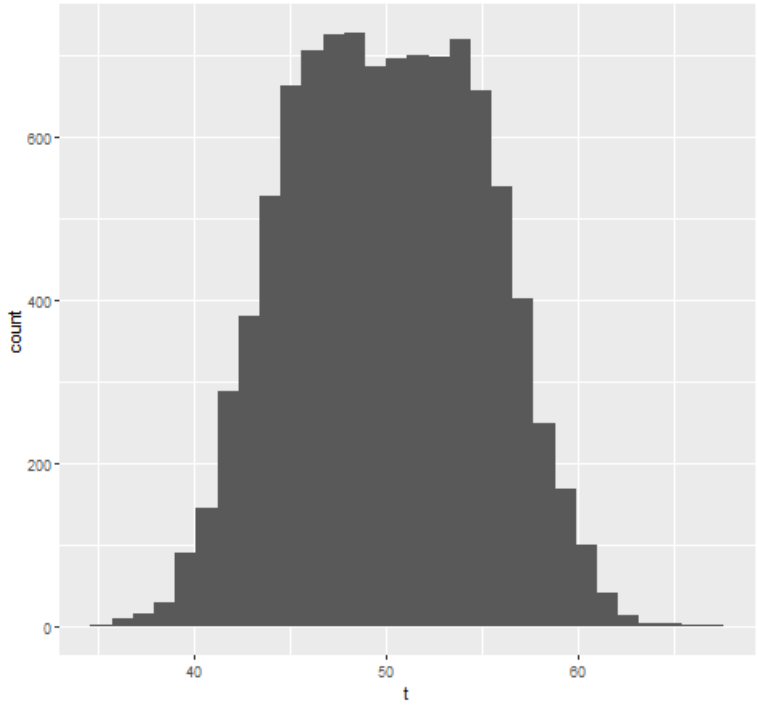



Figure F 28 : Iteration # 39 (distance of in mixture means= 7.6 )
Observed Score Frequencies with D = 3.6

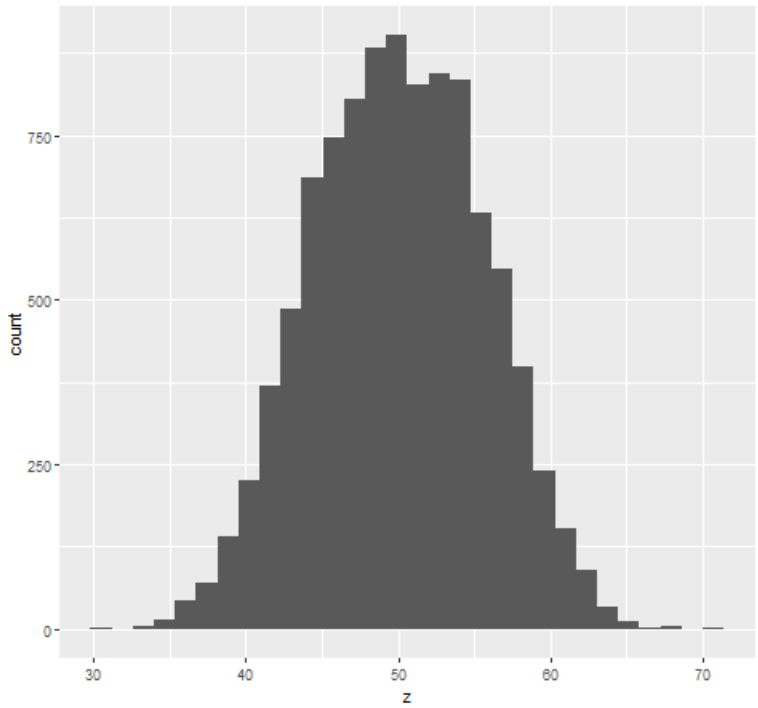


Figure F 29 : Iteration # 40 (distance of mixture means= 7.8 )
True Score Frequencies with D = 7.4

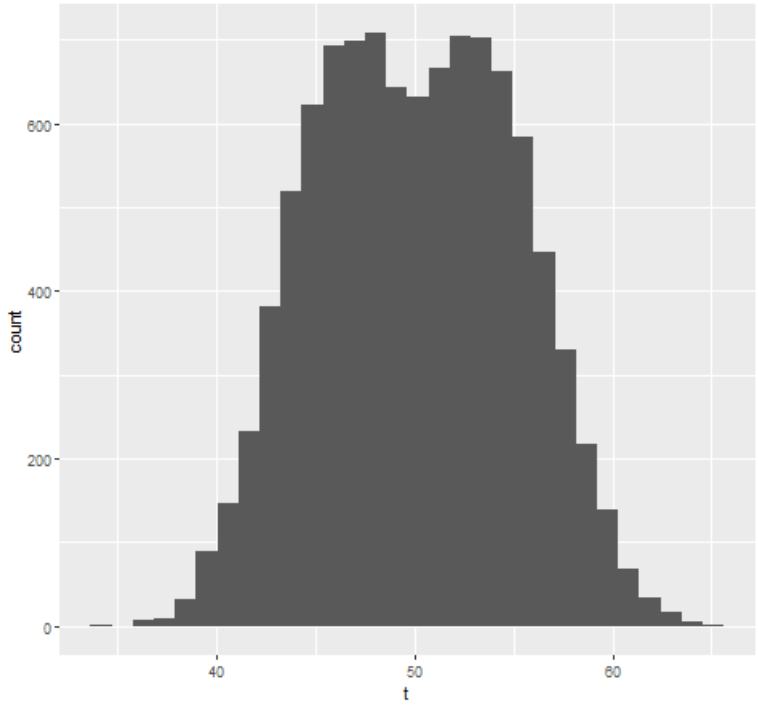



Figure F 30 : Iteration # 40 (distance of in mixture means= 7.8 )
Observed Score Frequencies with D = 2.3

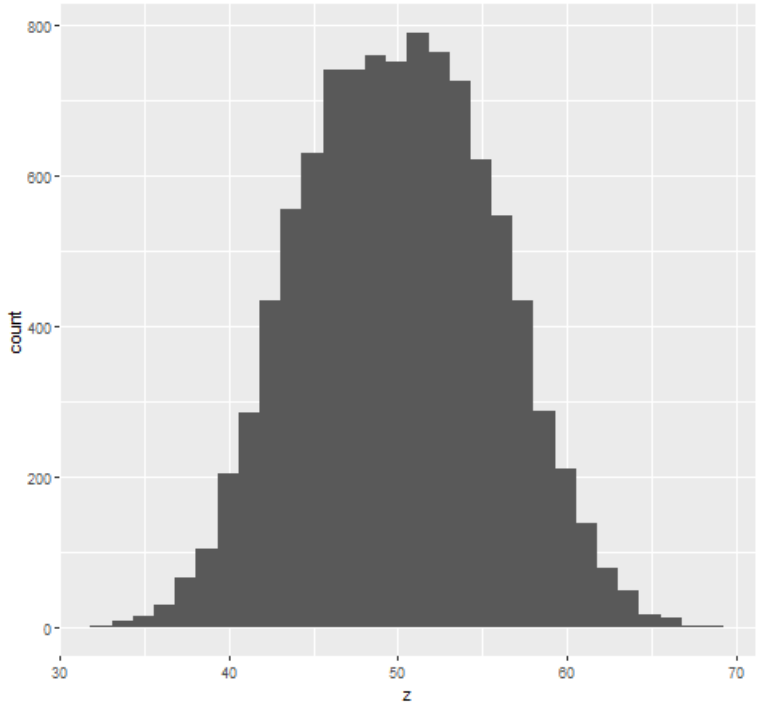


Figure F 31 : Iteration # 41 (distance of mixture means= 8 )
True Score Frequencies with D = 7.9

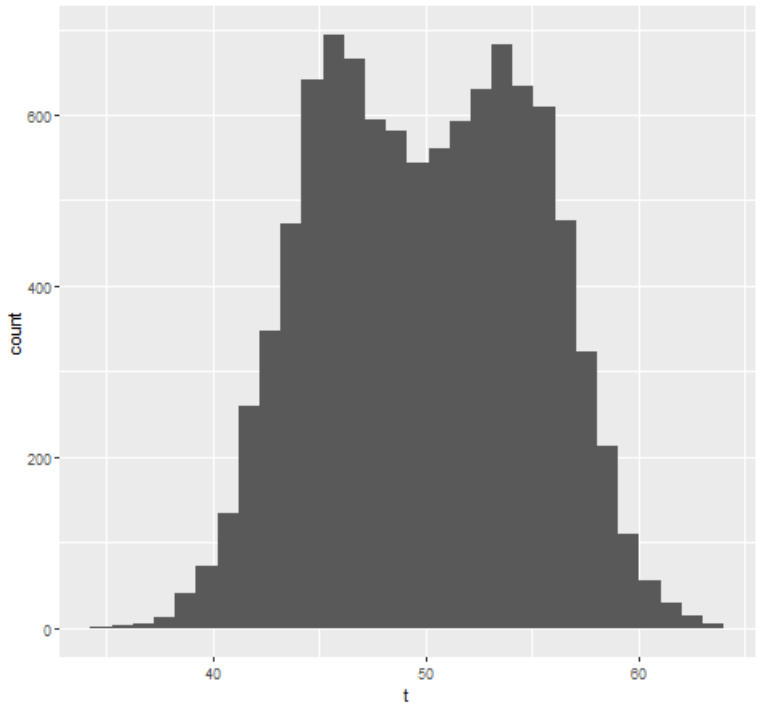



Figure F 32 : Iteration # 41 (distance of in mixture means= 8 )
Observed Score Frequencies with D = 5.7

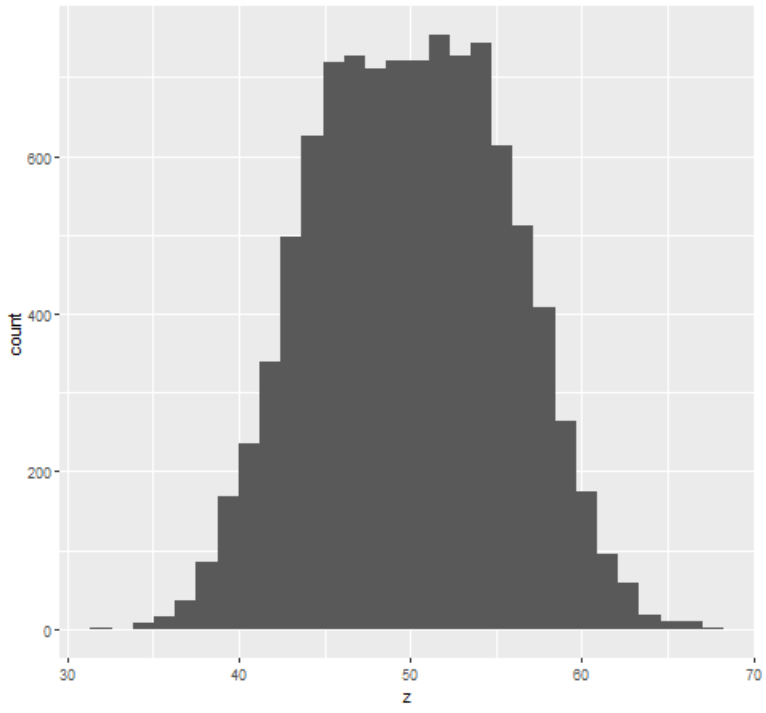


Figure F 33 : Iteration # 42 (distance of mixture means= 8.2 )
True Score Frequencies with D = 7.8

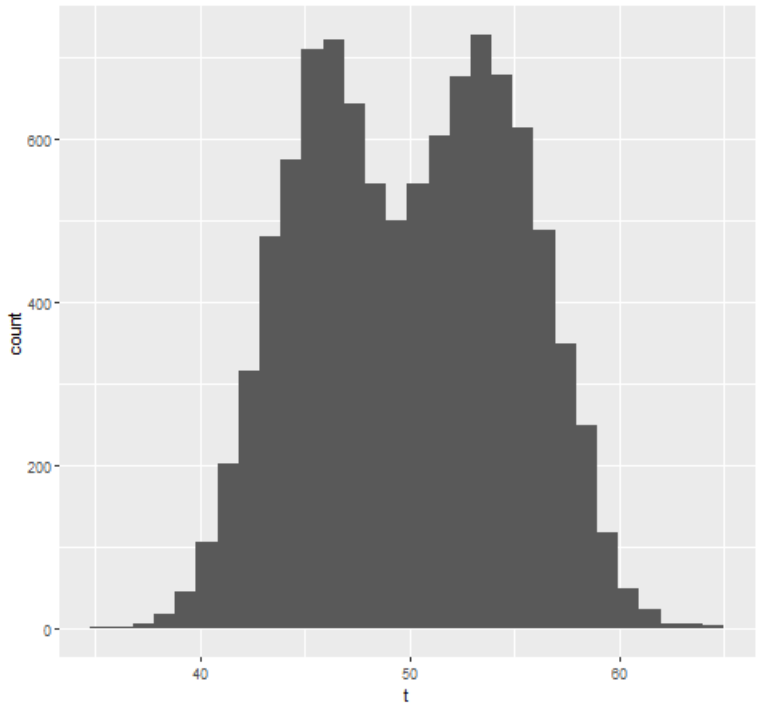



Figure F 34 : Iteration # 42 (distance of in mixture means= 8.2 )
Observed Score Frequencies with D = 5.2

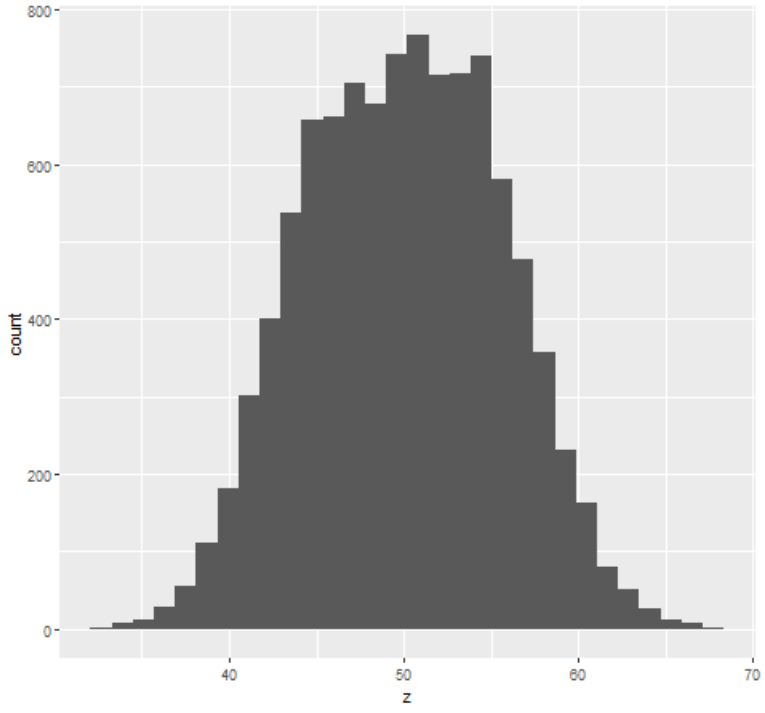


Figure F 35 : Iteration # 43 (distance of mixture means= 8.4 )
True Score Frequencies with D = 8.6

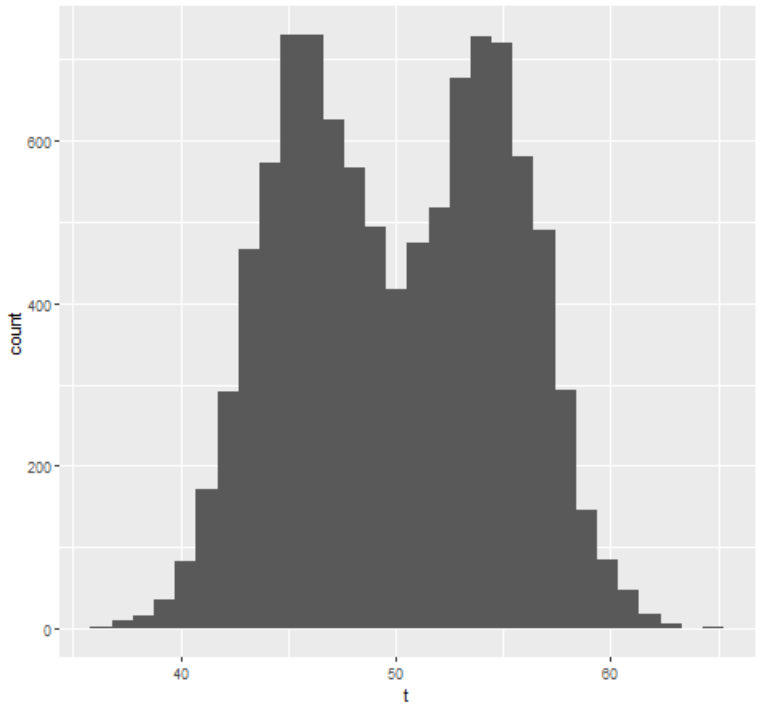



Figure F 36 : Iteration # 43 (distance of in mixture means= 8.4 )
Observed Score Frequencies with D = 10.3

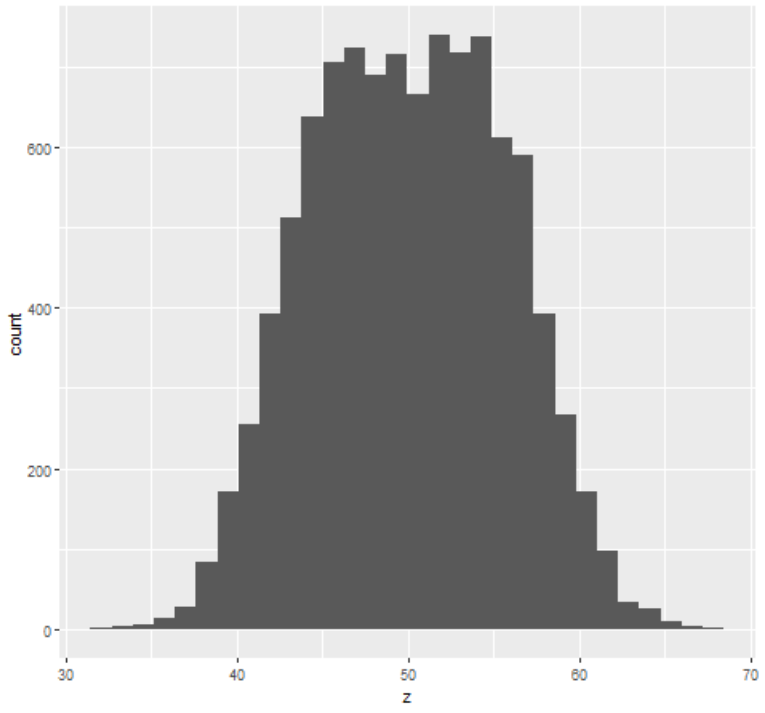


Figure F 37 : Iteration # 44 (distance of mixture means= 8.6 )
True Score Frequencies with D = 9.2

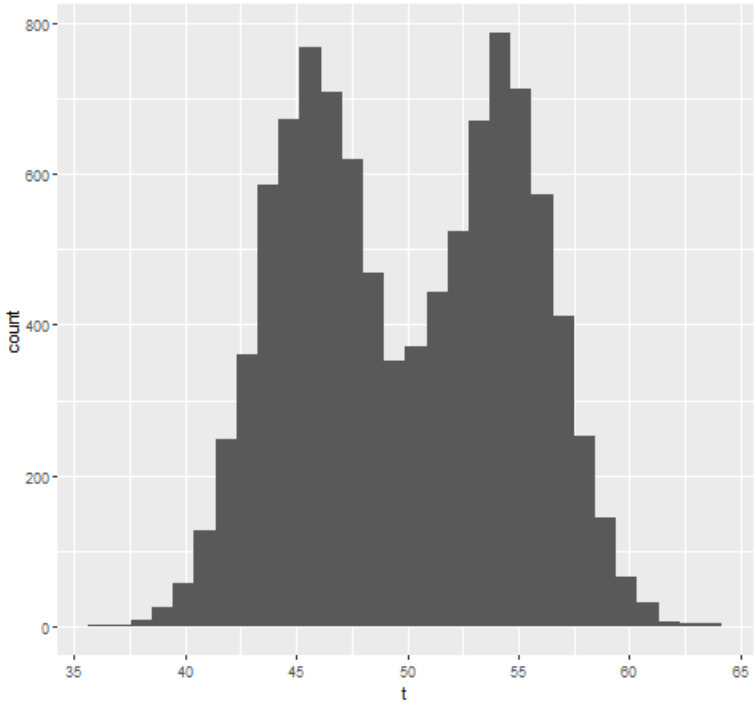



Figure F 38 : Iteration # 44 (distance of in mixture means= 8.6 )
Observed Score Frequencies with D = 8.3

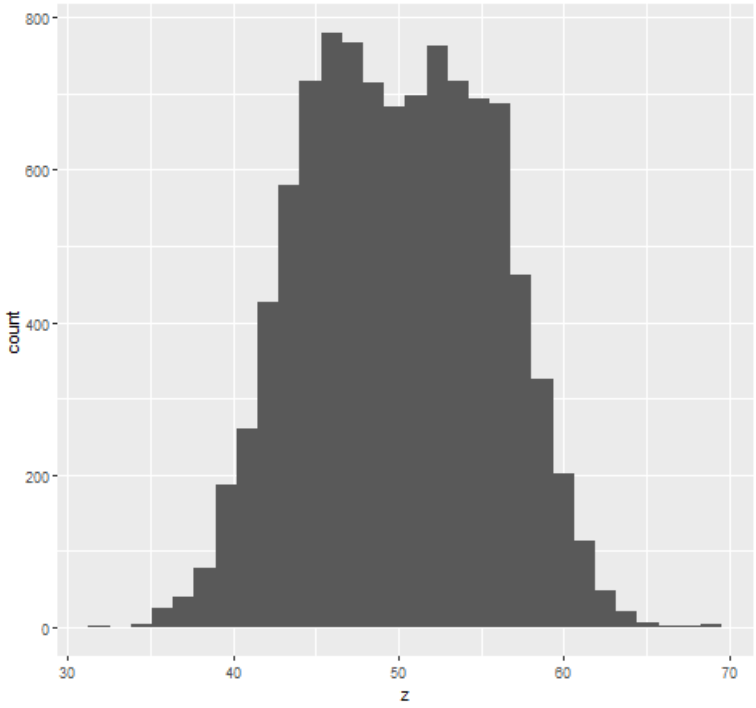


Figure F 39 : Iteration # 45 (distance of mixture means= 8.8 )
True Score Frequencies with D = 9.5

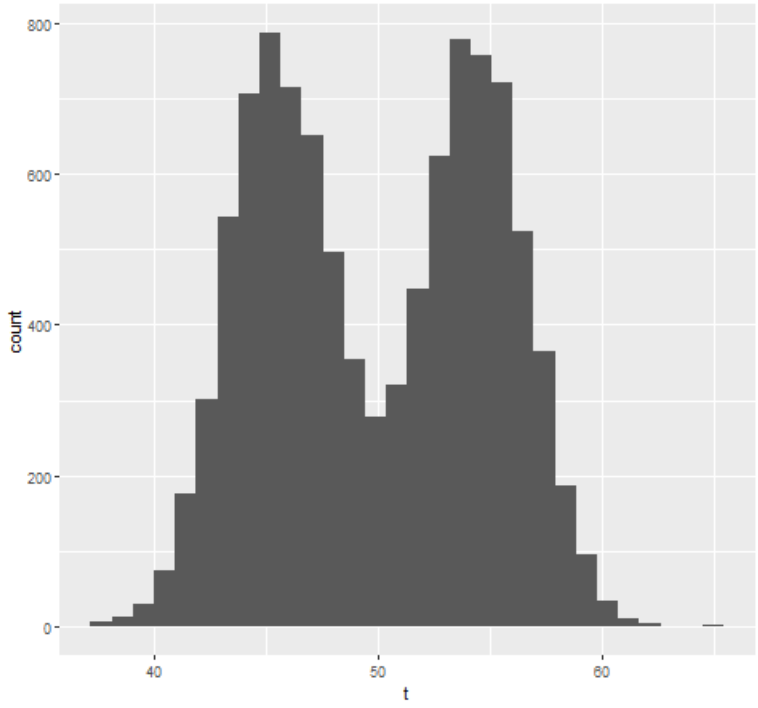



Figure F 40 : Iteration # 45 (distance of in mixture means= 8.8 )
Observed Score Frequencies with D = 5.6

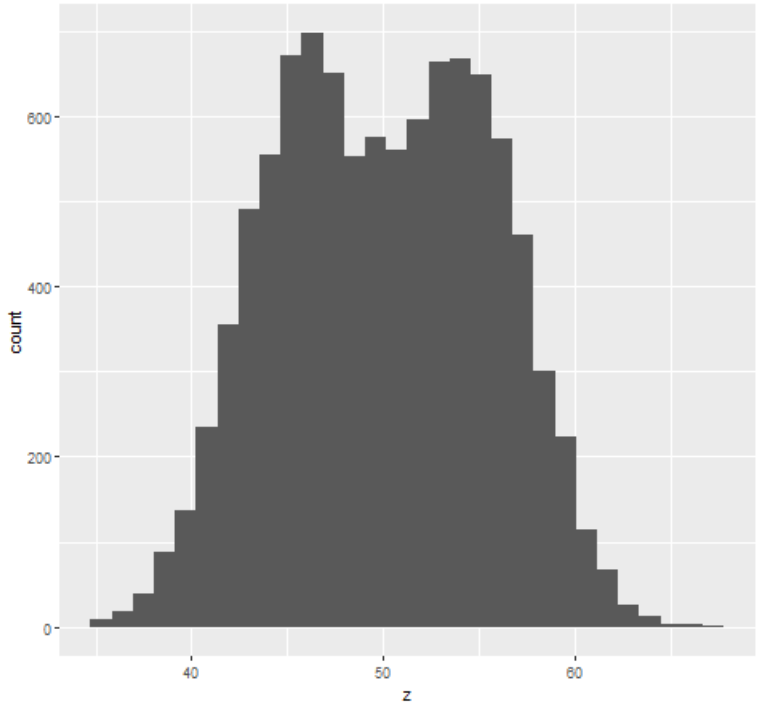


Figure F 41 : Iteration # 46 (distance of mixture means= 9 )
True Score Frequencies with D = 8.7

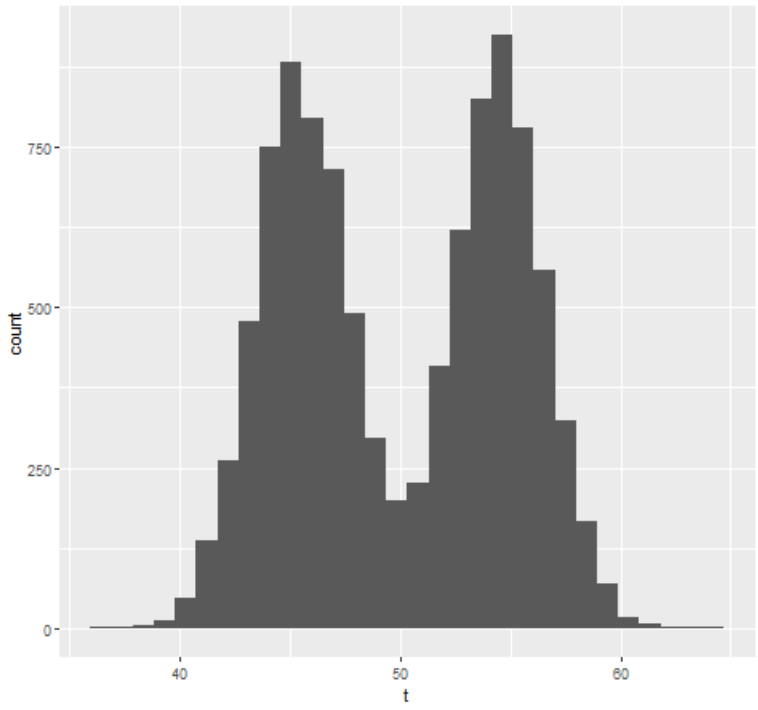



Figure F 42 : Iteration # 46 (distance of in mixture means= 9 )
Observed Score Frequencies with D = 8.5

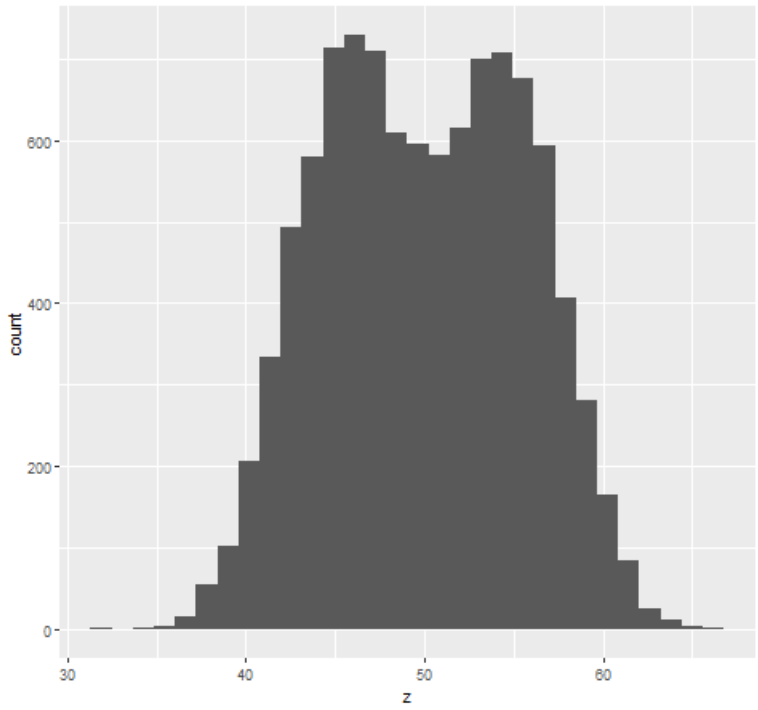


Figure F 43 : Iteration # 47 (distance of mixture means= 9.2 )
True Score Frequencies with D = 8.5

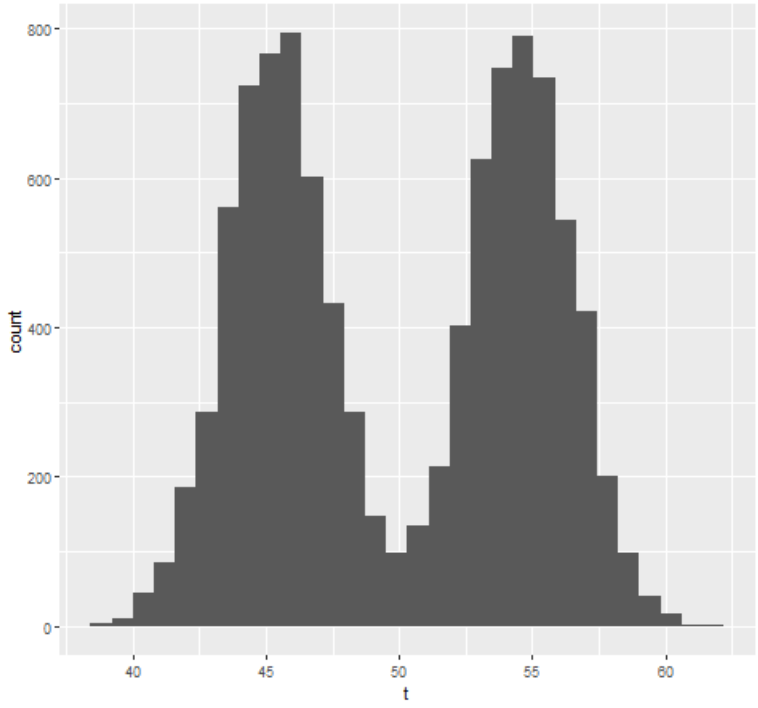



Figure F 44 : Iteration # 47 (distance of in mixture means= 9.2 )
Observed Score Frequencies with D = 9.6

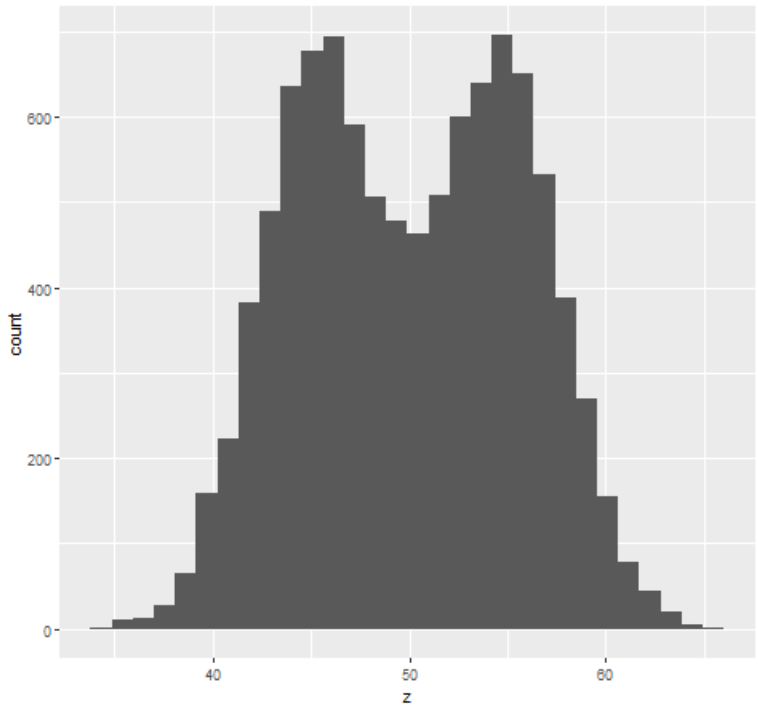


Figure F 45 : Iteration # 48 (distance of mixture means= 9.4 )
True Score Frequencies with D = 9.2

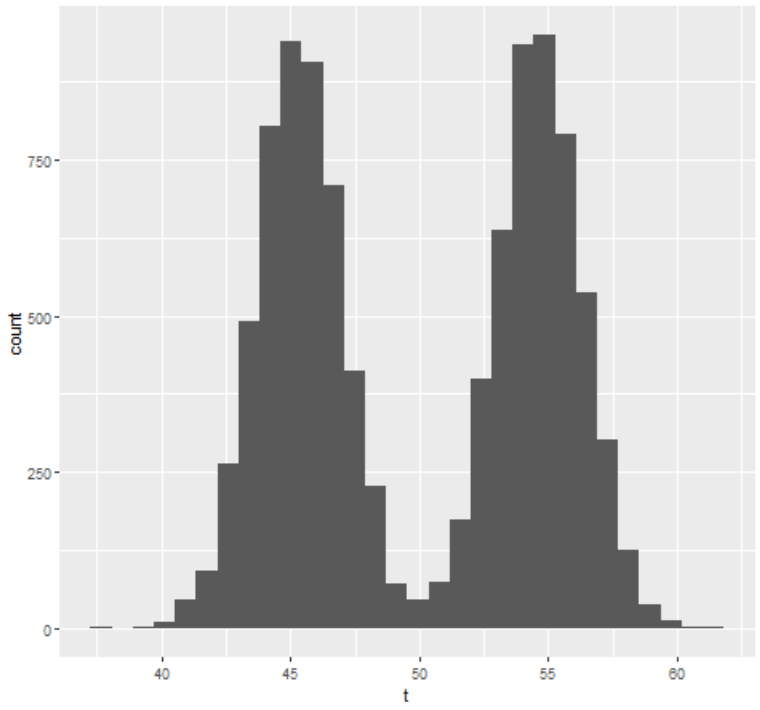



Figure F 46 : Iteration # 48 (distance of in mixture means= 9.4 )
Observed Score Frequencies with D = 7.5

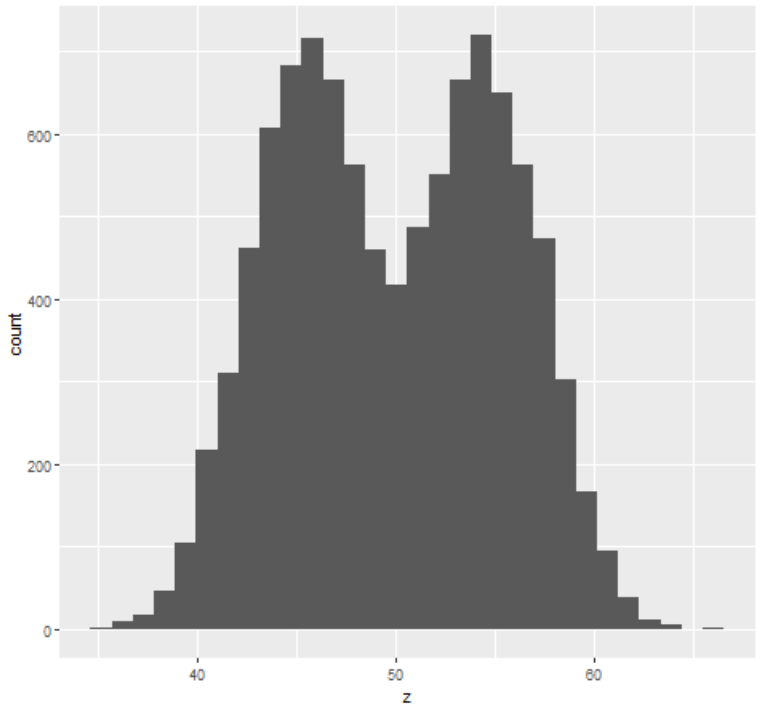


Figure F 47 : Iteration # 49 (distance of mixture means= 9.6 )
True Score Frequencies with D = 10.1

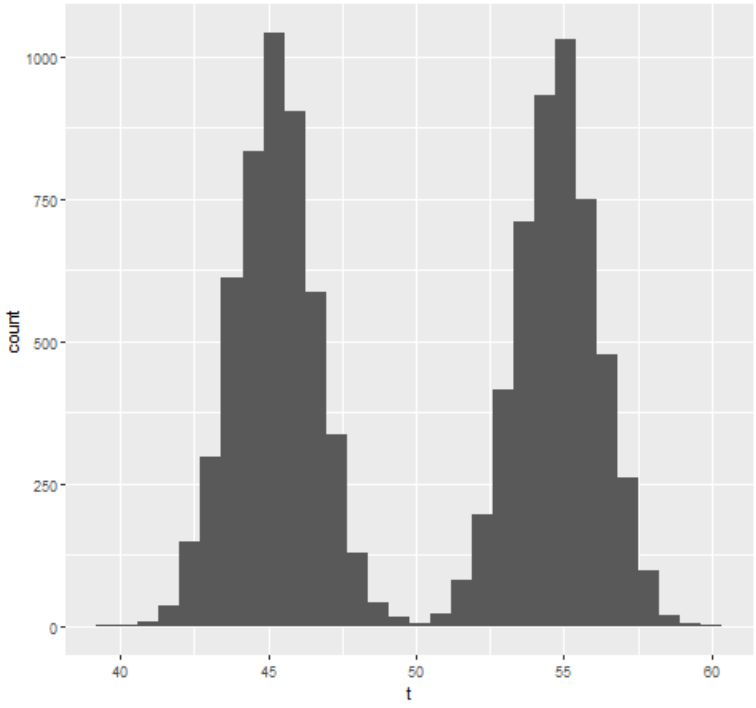



Figure F 48 : Iteration # 49 (distance of in mixture means= 9.6 )
Observed Score Frequencies with D = 9.7

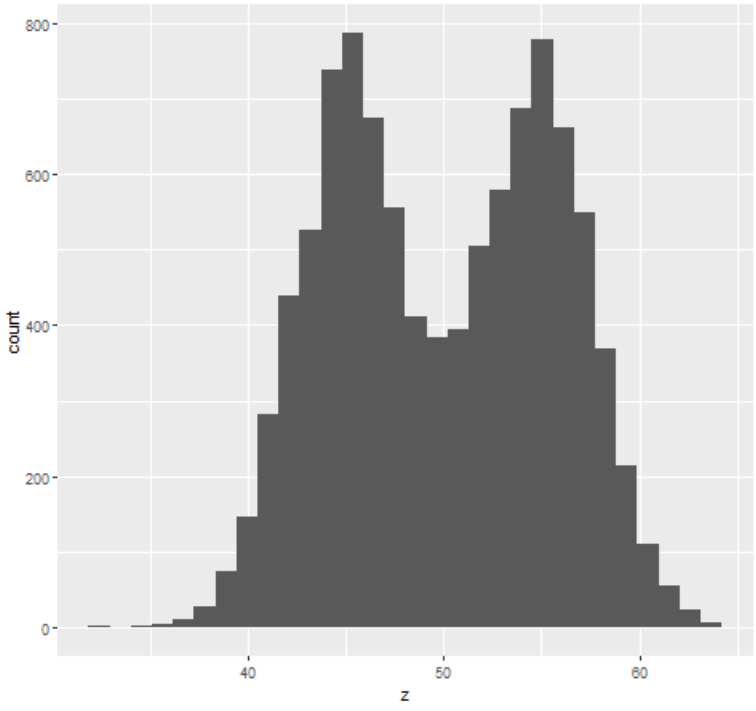


Figure F 49 : Iteration # 50 (distance of mixture means= 9.8 )
True Score Frequencies with D = 10.1

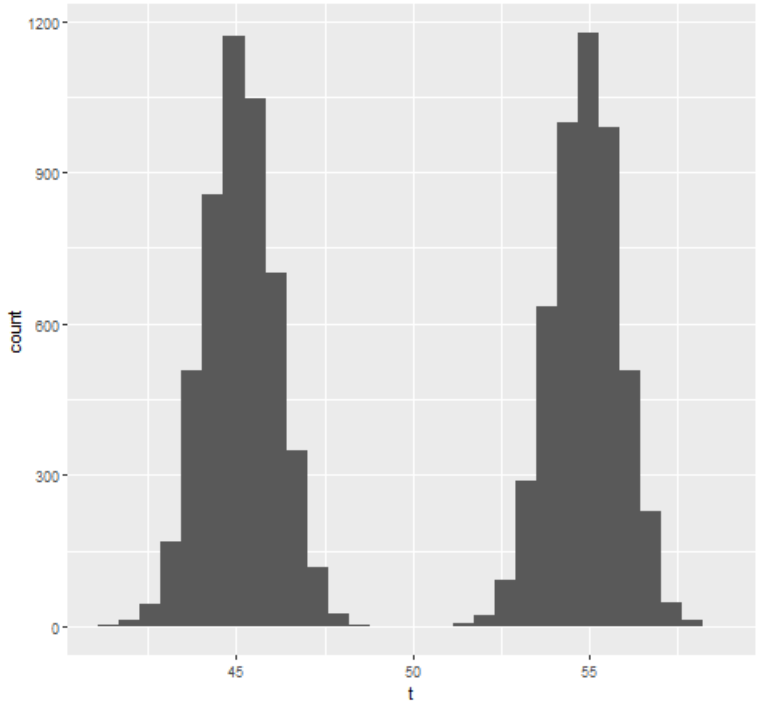
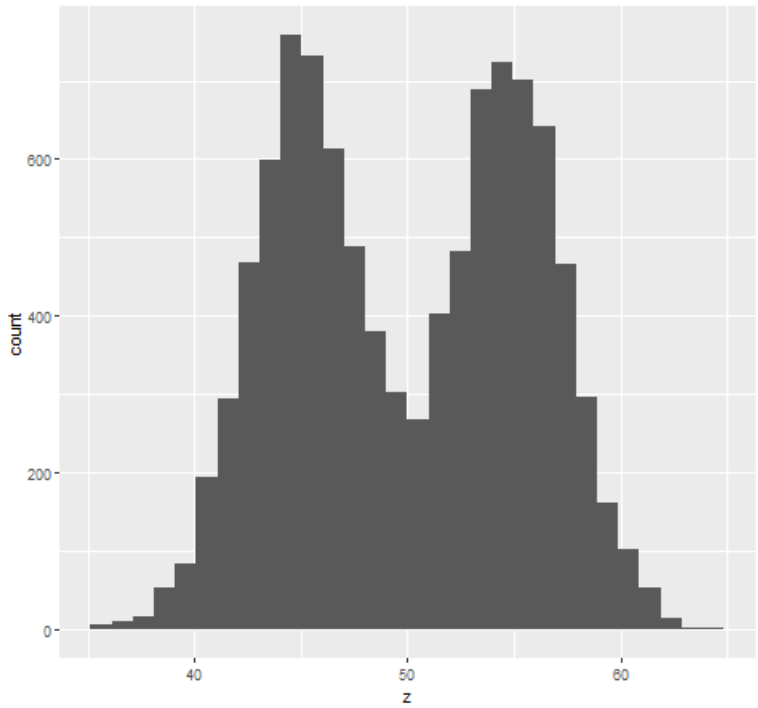



Figure F 50 : Iteration # 50 (distance of in mixture means= 9.8 )
Observed Score Frequencies with D = 8.7



## Appendix G: Kurtosis Results

Table G1: Error Rates and Location of Actual Optimal Cutscore with Increasing Kurtosis

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 3.03 | 0.052 | 0.024 | 43.6 | 0.076 | 0.054 | 46.8 | 0.059 | 0.069 | 53 | 0.021 | 0.063 | 56.8 |
| 3.14 | 0.056 | 0.03 | 43.8 | 0.076 | 0.054 | 46.8 | 0.061 | 0.065 | 52.9 | 0.03 | 0.054 | 56.2 |
| 3.11 | 0.063 | 0.025 | 43.4 | 0.08 | 0.052 | 46.7 | 0.053 | 0.075 | 53.1 | 0.023 | 0.06 | 56.6 |
| 3.12 | 0.053 | 0.029 | 43.8 | 0.078 | 0.047 | 46.7 | 0.058 | 0.076 | 53.2 | 0.03 | 0.061 | 56.4 |
| 3.31 | 0.06 | 0.025 | 43.4 | 0.074 | 0.054 | 46.9 | 0.046 | 0.079 | 53.3 | 0.021 | 0.061 | 56.6 |
| 3.28 | 0.056 | 0.019 | 43.2 | 0.078 | 0.049 | 46.7 | 0.052 | 0.074 | 53.2 | 0.033 | 0.055 | 56 |
| 3.31 | 0.058 | 0.028 | 43.6 | 0.088 | 0.04 | 46.3 | 0.056 | 0.072 | 53.1 | 0.031 | 0.055 | 56.2 |
| 3.39 | 0.059 | 0.025 | 43.4 | 0.068 | 0.062 | 47.1 | 0.052 | 0.08 | 53.3 | 0.028 | 0.052 | 56.3 |
| 3.39 | 0.058 | 0.021 | 43.4 | 0.087 | 0.048 | 46.4 | 0.051 | 0.081 | 53.3 | 0.023 | 0.058 | 56.6 |
| 3.53 | 0.053 | 0.029 | 43.7 | 0.076 | 0.058 | 46.9 | 0.044 | 0.088 | 53.6 | 0.022 | 0.057 | 56.7 |
| 3.49 | 0.056 | 0.025 | 43.5 | 0.09 | 0.041 | 46.4 | 0.053 | 0.078 | 53.3 | 0.024 | 0.056 | 56.5 |
| 3.61 | 0.049 | 0.028 | 43.7 | 0.088 | 0.045 | 46.4 | 0.059 | 0.07 | 53 | 0.028 | 0.053 | 56.2 |
| 3.58 | 0.054 | 0.023 | 43.6 | 0.085 | 0.045 | 46.5 | 0.063 | 0.073 | 53 | 0.027 | 0.054 | 56.4 |
| 3.82 | 0.042 | 0.033 | 44 | 0.079 | 0.054 | 46.7 | 0.06 | 0.069 | 53.2 | 0.021 | 0.051 | 56.6 |
| 3.67 | 0.06 | 0.016 | 43.1 | 0.073 | 0.057 | 46.9 | 0.055 | 0.076 | 53.2 | 0.026 | 0.055 | 56.4 |
| 3.9 | 0.048 | 0.026 | 43.7 | 0.083 | 0.047 | 46.4 | 0.046 | 0.084 | 53.6 | 0.02 | 0.052 | 56.8 |
| 3.9 | 0.05 | 0.021 | 43.5 | 0.092 | 0.046 | 46.2 | 0.051 | 0.077 | 53.4 | 0.018 | 0.053 | 56.9 |
| 4.08 | 0.047 | 0.026 | 43.6 | 0.087 | 0.044 | 46.2 | 0.056 | 0.075 | 53.2 | 0.021 | 0.052 | 56.7 |
| 4.01 | 0.056 | 0.017 | 43.1 | 0.101 | 0.034 | 45.8 | 0.044 | 0.079 | 53.6 | 0.017 | 0.055 | 57 |
| 4.15 | 0.046 | 0.025 | 43.6 | 0.094 | 0.041 | 46.2 | 0.038 | 0.091 | 53.8 | 0.016 | 0.053 | 56.8 |
| 4.08 | 0.048 | 0.026 | 43.5 | 0.081 | 0.047 | 46.4 | 0.049 | 0.074 | 53.5 | 0.021 | 0.05 | 56.8 |
| 4.21 | 0.051 | 0.02 | 43.3 | 0.087 | 0.046 | 46.3 | 0.047 | 0.082 | 53.5 | 0.017 | 0.049 | 57 |
| 4.63 | 0.046 | 0.019 | 43.3 | 0.079 | 0.047 | 46.4 | 0.065 | 0.069 | 53.2 | 0.015 | 0.053 | 57.1 |
| 4.39 | 0.048 | 0.016 | 43.1 | 0.097 | 0.034 | 45.9 | 0.038 | 0.092 | 53.9 | 0.017 | 0.053 | 57 |
| 4.47 | 0.041 | 0.027 | 43.7 | 0.081 | 0.054 | 46.4 | 0.046 | 0.078 | 53.6 | 0.018 | 0.045 | 56.7 |
| 4.54 | 0.049 | 0.016 | 42.9 | 0.089 | 0.037 | 46.1 | 0.045 | 0.083 | 53.7 | 0.021 | 0.047 | 56.6 |
| 4.69 | 0.051 | 0.014 | 42.8 | 0.081 | 0.055 | 46.4 | 0.045 | 0.079 | 53.7 | 0.018 | 0.042 | 56.7 |
| 4.59 | 0.042 | 0.018 | 43.4 | 0.094 | 0.033 | 45.7 | 0.054 | 0.071 | 53.4 | 0.022 | 0.043 | 56.6 |
| 4.98 | 0.041 | 0.017 | 43.2 | 0.088 | 0.038 | 46 | 0.048 | 0.079 | 53.6 | 0.012 | 0.049 | 57.3 |
| 5.02 | 0.038 | 0.016 | 43.2 | 0.086 | 0.036 | 45.9 | 0.041 | 0.084 | 53.9 | 0.02 | 0.04 | 56.8 |
| 4.81 | 0.042 | 0.024 | 43.5 | 0.08 | 0.048 | 46.3 | 0.052 | 0.071 | 53.6 | 0.024 | 0.038 | 56.3 |
| 4.8 | 0.04 | 0.02 | 43.5 | 0.076 | 0.046 | 46.3 | 0.043 | 0.087 | 54.1 | 0.018 | 0.043 | 56.8 |
| 4.86 | 0.042 | 0.016 | 43.2 | 0.079 | 0.039 | 46 | 0.032 | 0.092 | 54.5 | 0.017 | 0.041 | 56.7 |
| 4.85 | 0.039 | 0.016 | 43.1 | 0.089 | 0.027 | 45.4 | 0.047 | 0.078 | 53.8 | 0.018 | 0.041 | 56.7 |
| 4.97 | 0.04 | 0.021 | 43.6 | 0.067 | 0.049 | 46.3 | 0.028 | 0.092 | 54.5 | 0.015 | 0.042 | 57 |
| 5.1 | 0.041 | 0.013 | 43 | 0.086 | 0.034 | 45.6 | 0.039 | 0.077 | 54.1 | 0.014 | 0.039 | 56.8 |
| 5.24 | 0.035 | 0.02 | 43.4 | 0.084 | 0.034 | 45.7 | 0.032 | 0.082 | 54.4 | 0.013 | 0.04 | 56.9 |
| 5.14 | 0.036 | 0.017 | 43.5 | 0.086 | 0.032 | 45.6 | 0.032 | 0.084 | 54.5 | 0.01 | 0.046 | 57.3 |
| 5.44 | 0.039 | 0.014 | 43.1 | 0.079 | 0.026 | 45.3 | 0.028 | 0.087 | 54.7 | 0.016 | 0.036 | 56.6 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5.31 | 0.039 | 0.016 | 43.2 | 0.079 | 0.03 | 45.5 | 0.03 | 0.075 | 54.5 | 0.019 | 0.033 | 56.6 |
| 5.4 | 0.034 | 0.017 | 43.5 | 0.077 | 0.028 | 45.3 | 0.025 | 0.084 | 54.9 | 0.009 | 0.042 | 57.3 |
| 5.26 | 0.036 | 0.015 | 43.3 | 0.071 | 0.026 | 45.2 | 0.02 | 0.084 | 55.1 | 0.016 | 0.035 | 56.6 |
| 5.57 | 0.034 | 0.012 | 43.2 | 0.071 | 0.026 | 45.2 | 0.027 | 0.074 | 54.7 | 0.015 | 0.034 | 56.6 |
| 5.82 | 0.035 | 0.011 | 43.1 | 0.07 | 0.021 | 45 | 0.022 | 0.069 | 55 | 0.016 | 0.032 | 56.4 |
| 5.68 | 0.031 | 0.018 | 43.8 | 0.071 | 0.023 | 45 | 0.028 | 0.065 | 54.6 | 0.01 | 0.038 | 56.9 |
| 6.08 | 0.034 | 0.012 | 43.3 | 0.058 | 0.032 | 45.5 | 0.025 | 0.06 | 54.8 | 0.012 | 0.036 | 56.7 |
| 5.67 | 0.032 | 0.019 | 43.8 | 0.064 | 0.021 | 45 | 0.028 | 0.058 | 54.7 | 0.014 | 0.033 | 56.5 |
| 5.96 | 0.036 | 0.011 | 43.2 | 0.062 | 0.022 | 45.2 | 0.021 | 0.059 | 54.8 | 0.011 | 0.033 | 56.5 |
| 5.74 | 0.036 | 0.011 | 43.2 | 0.061 | 0.019 | 45 | 0.017 | 0.058 | 55.1 | 0.01 | 0.035 | 56.8 |
| 6.03 | 0.027 | 0.015 | 43.7 | 0.052 | 0.023 | 45.3 | 0.022 | 0.059 | 54.7 | 0.015 | 0.033 | 56.2 |

Table G2: Actual Error Rates at True Cutscore Location with Increasing Kurtosis

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 3.03 | 0.03 | 0.058 | 45 | 0.057 | 0.077 | 47.5 | 0.076 | 0.055 | 52.5 | 0.065 | 0.033 | 55 |
| 3.14 | 0.035 | 0.061 | 45 | 0.056 | 0.079 | 47.5 | 0.074 | 0.054 | 52.5 | 0.065 | 0.035 | 55 |
| 3.11 | 0.034 | 0.064 | 45 | 0.059 | 0.078 | 47.5 | 0.074 | 0.057 | 52.5 | 0.062 | 0.033 | 55 |
| 3.12 | 0.033 | 0.062 | 45 | 0.056 | 0.075 | 47.5 | 0.081 | 0.056 | 52.5 | 0.066 | 0.034 | 55 |
| 3.31 | 0.034 | 0.063 | 45 | 0.058 | 0.075 | 47.5 | 0.075 | 0.055 | 52.5 | 0.061 | 0.036 | 55 |
| 3.28 | 0.03 | 0.063 | 45 | 0.056 | 0.078 | 47.5 | 0.075 | 0.054 | 52.5 | 0.06 | 0.035 | 55 |
| 3.31 | 0.035 | 0.062 | 45 | 0.054 | 0.078 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 3.39 | 0.034 | 0.061 | 45 | 0.057 | 0.075 | 47.5 | 0.081 | 0.057 | 52.5 | 0.061 | 0.031 | 55 |
| 3.39 | 0.03 | 0.059 | 45 | 0.06 | 0.086 | 47.5 | 0.082 | 0.058 | 52.5 | 0.062 | 0.033 | 55 |
| 3.53 | 0.032 | 0.063 | 45 | 0.059 | 0.08 | 47.5 | 0.081 | 0.055 | 52.5 | 0.063 | 0.031 | 55 |
| 3.49 | 0.032 | 0.066 | 45 | 0.059 | 0.079 | 47.5 | 0.081 | 0.056 | 52.5 | 0.063 | 0.03 | 55 |
| 3.61 | 0.03 | 0.062 | 45 | 0.056 | 0.085 | 47.5 | 0.078 | 0.057 | 52.5 | 0.059 | 0.035 | 55 |
| 3.58 | 0.03 | 0.057 | 45 | 0.057 | 0.08 | 47.5 | 0.084 | 0.059 | 52.5 | 0.062 | 0.033 | 55 |
| 3.82 | 0.028 | 0.061 | 45 | 0.058 | 0.082 | 47.5 | 0.086 | 0.053 | 52.5 | 0.063 | 0.027 | 55 |
| 3.67 | 0.031 | 0.06 | 45 | 0.057 | 0.082 | 47.5 | 0.078 | 0.056 | 52.5 | 0.063 | 0.034 | 55 |
| 3.9 | 0.028 | 0.06 | 45 | 0.052 | 0.086 | 47.5 | 0.084 | 0.056 | 52.5 | 0.064 | 0.025 | 55 |
| 3.9 | 0.028 | 0.056 | 45 | 0.056 | 0.094 | 47.5 | 0.083 | 0.051 | 52.5 | 0.06 | 0.026 | 55 |
| 4.08 | 0.026 | 0.063 | 45 | 0.051 | 0.092 | 47.5 | 0.081 | 0.057 | 52.5 | 0.062 | 0.03 | 55 |
| 4.01 | 0.028 | 0.066 | 45 | 0.053 | 0.094 | 47.5 | 0.085 | 0.051 | 52.5 | 0.057 | 0.026 | 55 |
| 4.15 | 0.027 | 0.06 | 45 | 0.059 | 0.087 | 47.5 | 0.085 | 0.056 | 52.5 | 0.052 | 0.028 | 55 |
| 4.08 | 0.028 | 0.062 | 45 | 0.05 | 0.085 | 47.5 | 0.089 | 0.049 | 52.5 | 0.063 | 0.024 | 55 |
| 4.21 | 0.027 | 0.062 | 45 | 0.053 | 0.091 | 47.5 | 0.086 | 0.056 | 52.5 | 0.06 | 0.025 | 55 |
| 4.63 | 0.024 | 0.056 | 45 | 0.051 | 0.09 | 47.5 | 0.092 | 0.052 | 52.5 | 0.063 | 0.025 | 55 |
| 4.39 | 0.024 | 0.058 | 45 | 0.052 | 0.091 | 47.5 | 0.089 | 0.053 | 52.5 | 0.061 | 0.027 | 55 |
| 4.47 | 0.024 | 0.063 | 45 | 0.054 | 0.095 | 47.5 | 0.088 | 0.049 | 52.5 | 0.06 | 0.023 | 55 |
| 4.54 | 0.023 | 0.06 | 45 | 0.053 | 0.087 | 47.5 | 0.09 | 0.052 | 52.5 | 0.057 | 0.025 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4.69 | 0.023 | 0.062 | 45 | 0.052 | 0.095 | 47.5 | 0.091 | 0.05 | 52.5 | 0.06 | 0.022 | 55 |
| 4.59 | 0.022 | 0.056 | 45 | 0.052 | 0.094 | 47.5 | 0.09 | 0.051 | 52.5 | 0.058 | 0.023 | 55 |
| 4.98 | 0.02 | 0.054 | 45 | 0.051 | 0.095 | 47.5 | 0.093 | 0.052 | 52.5 | 0.062 | 0.023 | 55 |
| 5.02 | 0.019 | 0.057 | 45 | 0.047 | 0.098 | 47.5 | 0.098 | 0.049 | 52.5 | 0.061 | 0.02 | 55 |
| 4.81 | 0.025 | 0.058 | 45 | 0.05 | 0.095 | 47.5 | 0.091 | 0.047 | 52.5 | 0.052 | 0.023 | 55 |
| 4.8 | 0.021 | 0.054 | 45 | 0.048 | 0.092 | 47.5 | 0.105 | 0.049 | 52.5 | 0.06 | 0.021 | 55 |
| 4.86 | 0.019 | 0.056 | 45 | 0.045 | 0.095 | 47.5 | 0.102 | 0.046 | 52.5 | 0.054 | 0.021 | 55 |
| 4.85 | 0.018 | 0.056 | 45 | 0.043 | 0.1 | 47.5 | 0.101 | 0.047 | 52.5 | 0.055 | 0.022 | 55 |
| 4.97 | 0.023 | 0.054 | 45 | 0.04 | 0.098 | 47.5 | 0.098 | 0.044 | 52.5 | 0.055 | 0.019 | 55 |
| 5.1 | 0.021 | 0.054 | 45 | 0.043 | 0.105 | 47.5 | 0.103 | 0.041 | 52.5 | 0.051 | 0.021 | 55 |
| 5.24 | 0.02 | 0.052 | 45 | 0.043 | 0.106 | 47.5 | 0.101 | 0.044 | 52.5 | 0.052 | 0.019 | 55 |
| 5.14 | 0.02 | 0.049 | 45 | 0.045 | 0.105 | 47.5 | 0.103 | 0.043 | 52.5 | 0.05 | 0.023 | 55 |
| 5.44 | 0.019 | 0.051 | 45 | 0.038 | 0.107 | 47.5 | 0.108 | 0.039 | 52.5 | 0.051 | 0.018 | 55 |
| 5.31 | 0.019 | 0.052 | 45 | 0.041 | 0.107 | 47.5 | 0.109 | 0.037 | 52.5 | 0.05 | 0.018 | 55 |
| 5.4 | 0.016 | 0.046 | 45 | 0.036 | 0.11 | 47.5 | 0.109 | 0.038 | 52.5 | 0.05 | 0.019 | 55 |
| 5.26 | 0.017 | 0.05 | 45 | 0.032 | 0.117 | 47.5 | 0.113 | 0.035 | 52.5 | 0.046 | 0.018 | 55 |
| 5.57 | 0.016 | 0.045 | 45 | 0.032 | 0.115 | 47.5 | 0.113 | 0.034 | 52.5 | 0.046 | 0.018 | 55 |
| 5.82 | 0.016 | 0.044 | 45 | 0.029 | 0.11 | 47.5 | 0.117 | 0.03 | 52.5 | 0.043 | 0.017 | 55 |
| 5.68 | 0.019 | 0.043 | 45 | 0.028 | 0.11 | 47.5 | 0.113 | 0.029 | 52.5 | 0.043 | 0.018 | 55 |
| 6.08 | 0.018 | 0.043 | 45 | 0.028 | 0.117 | 47.5 | 0.117 | 0.027 | 52.5 | 0.041 | 0.017 | 55 |
| 5.67 | 0.02 | 0.041 | 45 | 0.025 | 0.118 | 47.5 | 0.111 | 0.023 | 52.5 | 0.044 | 0.018 | 55 |
| 5.96 | 0.018 | 0.038 | 45 | 0.026 | 0.116 | 47.5 | 0.109 | 0.025 | 52.5 | 0.035 | 0.019 | 55 |
| 5.74 | 0.019 | 0.038 | 45 | 0.025 | 0.116 | 47.5 | 0.109 | 0.021 | 52.5 | 0.04 | 0.018 | 55 |
| 6.03 | 0.015 | 0.039 | 45 | 0.023 | 0.108 | 47.5 | 0.11 | 0.024 | 52.5 | 0.036 | 0.02 | 55 |

Table G3: GW-CSOF Estimate of Location of & Error at Optimal Cutscore with Increasing Kurtosis

| | 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |
| 3.03 | 0.056 | 0.028 | 43.7 | 0.076 | 0.052 | 46.8 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 3.14 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 3.11 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 3.12 | 0.055 | 0.028 | 43.7 | 0.075 | 0.052 | 46.8 | 0.055 | 0.076 | 53.1 | 0.031 | 0.058 | 56.2 |
| 3.31 | 0.056 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.057 | 56.3 |
| 3.28 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 3.31 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 3.39 | 0.057 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 3.39 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 3.53 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 3.49 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.61 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 3.58 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 3.82 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 3.67 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.03 | 0.056 | 56.2 |
| 3.9 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 3.9 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 4.08 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 4.01 | 0.056 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.054 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 4.15 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 4.08 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 4.21 | 0.056 | 0.03 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 4.63 | 0.056 | 0.028 | 43.7 | 0.076 | 0.052 | 46.8 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.39 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.47 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.058 | 56.3 |
| 4.54 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.69 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 4.59 | 0.057 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.98 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.056 | 56.2 |
| 5.02 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.81 | 0.057 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.057 | 56.3 |
| 4.8 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.056 | 0.075 | 53.1 | 0.032 | 0.057 | 56.2 |
| 4.86 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 4.85 | 0.057 | 0.029 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 4.97 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 5.1 | 0.056 | 0.03 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 5.24 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 5.14 | 0.058 | 0.031 | 43.8 | 0.075 | 0.055 | 46.9 | 0.052 | 0.075 | 53.2 | 0.028 | 0.056 | 56.3 |
| 5.44 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.031 | 0.056 | 56.2 |
| 5.31 | 0.058 | 0.029 | 43.7 | 0.074 | 0.055 | 46.9 | 0.056 | 0.075 | 53.1 | 0.032 | 0.057 | 56.2 |
| 5.4 | 0.057 | 0.028 | 43.7 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.056 | 56.2 |
| 5.26 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 5.57 | 0.055 | 0.028 | 43.7 | 0.075 | 0.052 | 46.8 | 0.055 | 0.075 | 53.1 | 0.031 | 0.057 | 56.2 |
| 5.82 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 5.68 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.03 | 0.055 | 56.2 |
| 6.08 | 0.055 | 0.03 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.028 | 0.057 | 56.3 |
| 5.67 | 0.056 | 0.028 | 43.7 | 0.073 | 0.055 | 46.9 | 0.055 | 0.075 | 53.1 | 0.031 | 0.056 | 56.2 |
| 5.96 | 0.056 | 0.03 | 43.8 | 0.075 | 0.055 | 46.9 | 0.055 | 0.073 | 53.1 | 0.028 | 0.056 | 56.3 |
| 5.74 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |
| 6.03 | 0.056 | 0.031 | 43.8 | 0.074 | 0.055 | 46.9 | 0.055 | 0.074 | 53.1 | 0.029 | 0.057 | 56.3 |

Table G4: GW-CSOF Error Rates at True Cutscore Location with Increasing Kurtosis

| 45 | | | 47.5 | | | 52.5 | | | 55 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| fp | fn | loc | fp | fn | loc | fp | fn | loc | fp | fn | loc |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.03 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 3.14 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 3.11 | 0.035 | 0.062 | 45 | 0.058 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 3.12 | 0.033 | 0.06 | 45 | 0.055 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 3.31 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 3.28 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 3.31 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.39 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.39 | 0.034 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 3.53 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 3.49 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.035 | 55 |
| 3.61 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.58 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 3.82 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.67 | 0.034 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.9 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 3.9 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 4.08 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 4.01 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 4.15 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 4.08 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 4.21 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.034 | 55 |
| 4.63 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 4.39 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 4.47 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 4.54 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 4.69 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 4.59 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 4.98 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 5.02 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 4.81 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 4.8 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.058 | 52.5 | 0.063 | 0.036 | 55 |
| 4.86 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 4.85 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.035 | 55 |
| 4.97 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 5.1 | 0.035 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 5.24 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 5.14 | 0.036 | 0.063 | 45 | 0.058 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 5.44 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 5.31 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.036 | 55 |
| 5.4 | 0.034 | 0.061 | 45 | 0.056 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 5.26 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 5.57 | 0.033 | 0.06 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.058 | 52.5 | 0.062 | 0.036 | 55 |
| 5.82 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 5.68 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6.08 | 0.034 | 0.061 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.056 | 52.5 | 0.061 | 0.034 | 55 |
| 5.67 | 0.034 | 0.061 | 45 | 0.056 | 0.075 | 47.5 | 0.076 | 0.057 | 52.5 | 0.062 | 0.035 | 55 |
| 5.96 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.075 | 0.056 | 52.5 | 0.06 | 0.033 | 55 |
| 5.74 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |
| 6.03 | 0.035 | 0.062 | 45 | 0.057 | 0.076 | 47.5 | 0.076 | 0.057 | 52.5 | 0.061 | 0.034 | 55 |

Figure G 1 : Iteration # 1 (Distribution Kurtosis of 3 )
True Score Frequencies with Kurtosis = 3.03

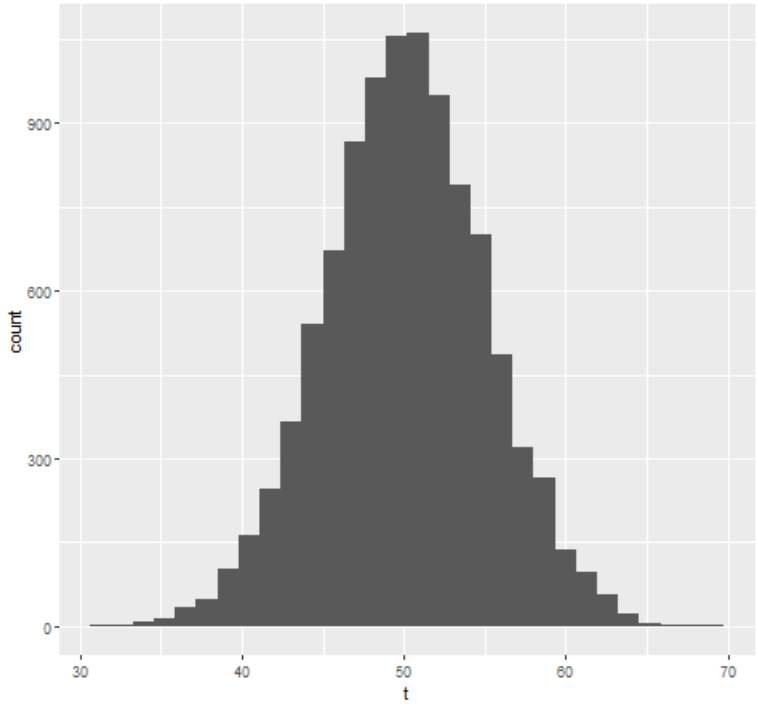



Figure G 2 : Iteration # 1 (Distribution Kurtosis of 3 )
Observed Score Frequencies with Kurtosis = 2.92

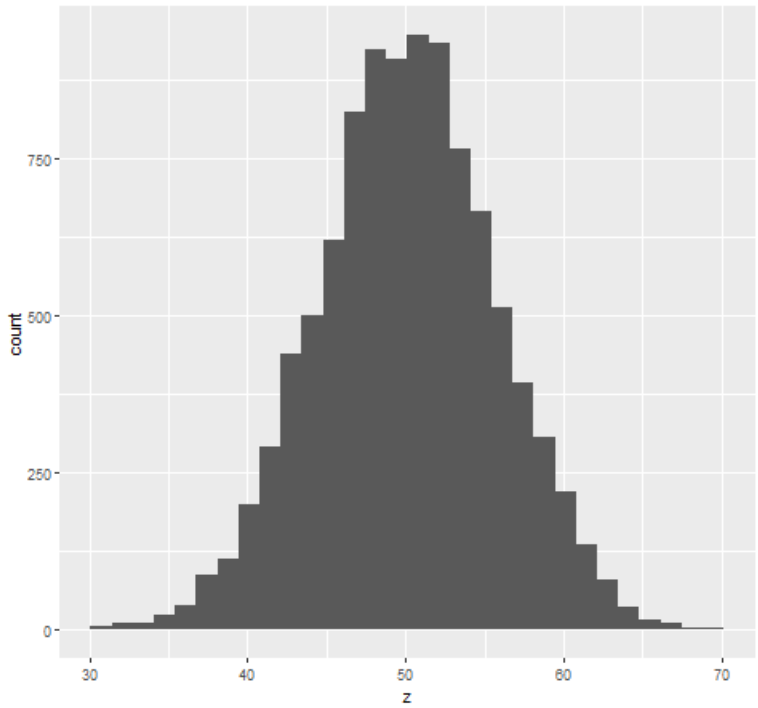


Figure G 3 : Iteration # 10 (Distribution Kurtosis of 3.54 )
True Score Frequencies with Kurtosis = 3.53

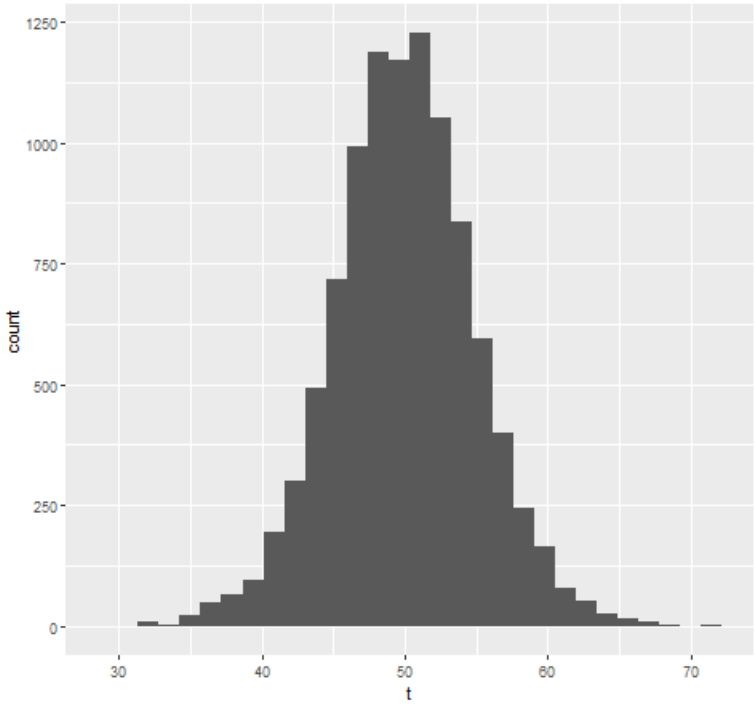



Figure G 4 : Iteration # 10 (Distribution Kurtosis of 3.54 )
Observed Score Frequencies with Kurtosis = 3.37

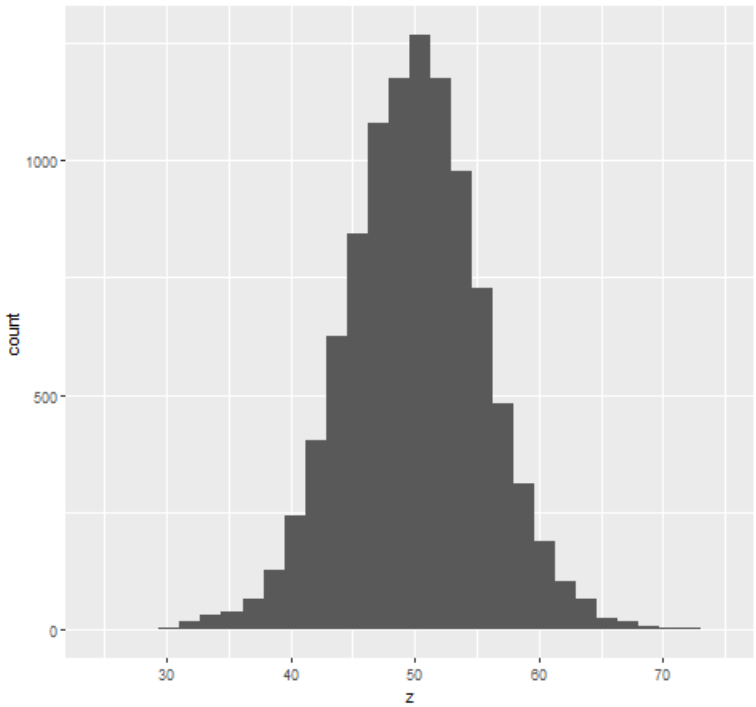


Figure G 5 : Iteration # 20 (Distribution Kurtosis of 4.14 )
True Score Frequencies with Kurtosis = 4.15

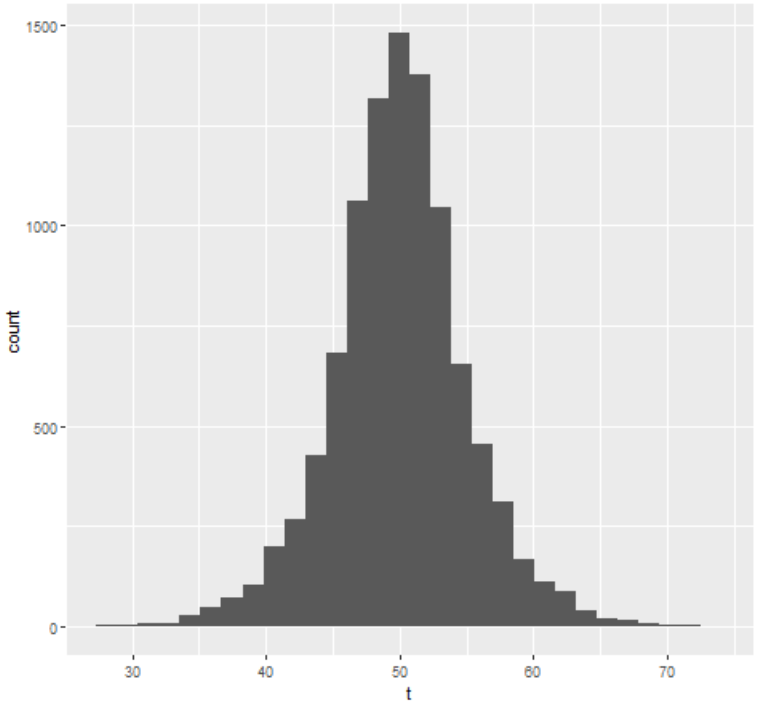



Figure G 6 : Iteration # 20 (Distribution Kurtosis of 4.14 )
Observed Score Frequencies with Kurtosis = 3.88

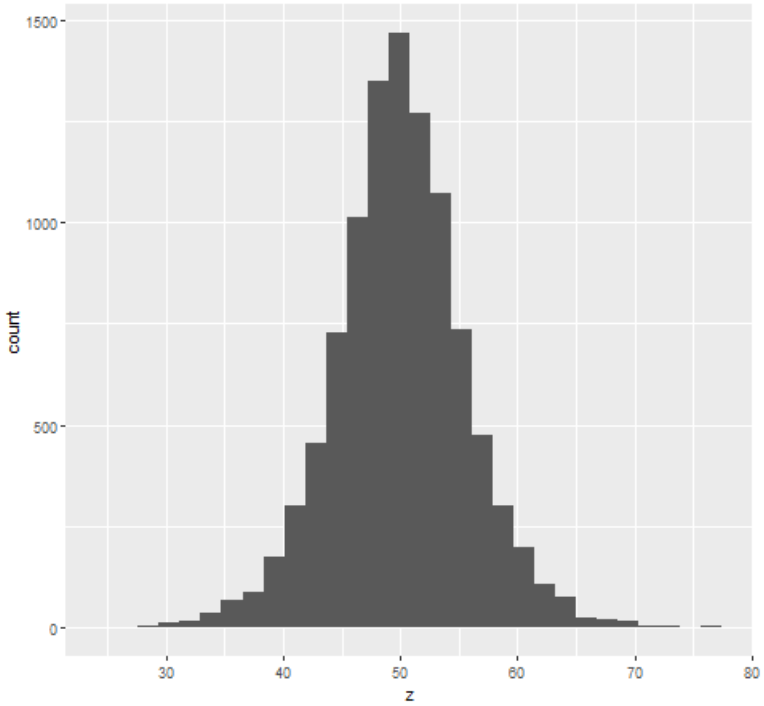


Figure G 7 : Iteration # 30 (Distribution Kurtosis of 4.74 )
True Score Frequencies with Kurtosis = 5.02

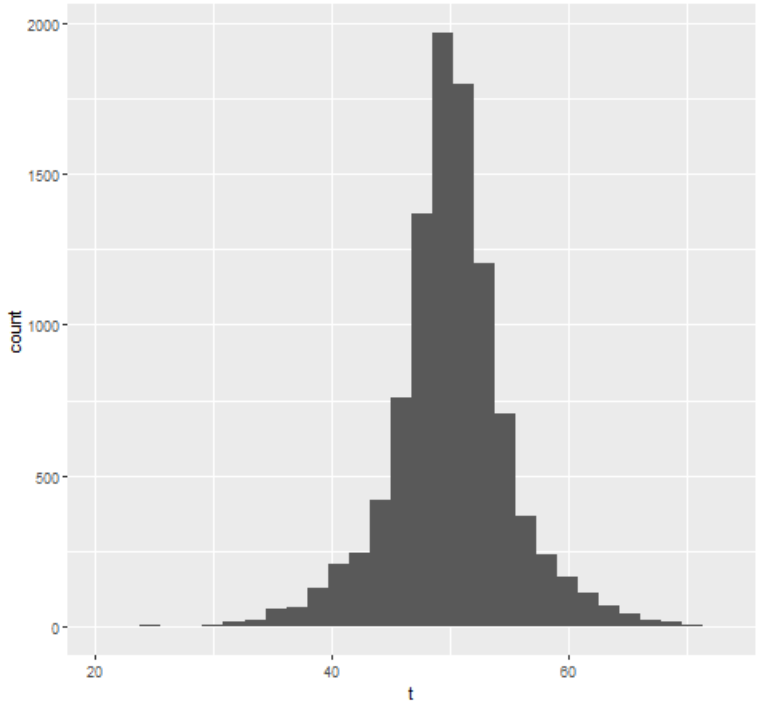



Figure G 8 : Iteration # 30 (Distribution Kurtosis of 4.74 )
Observed Score Frequencies with Kurtosis = 4.31

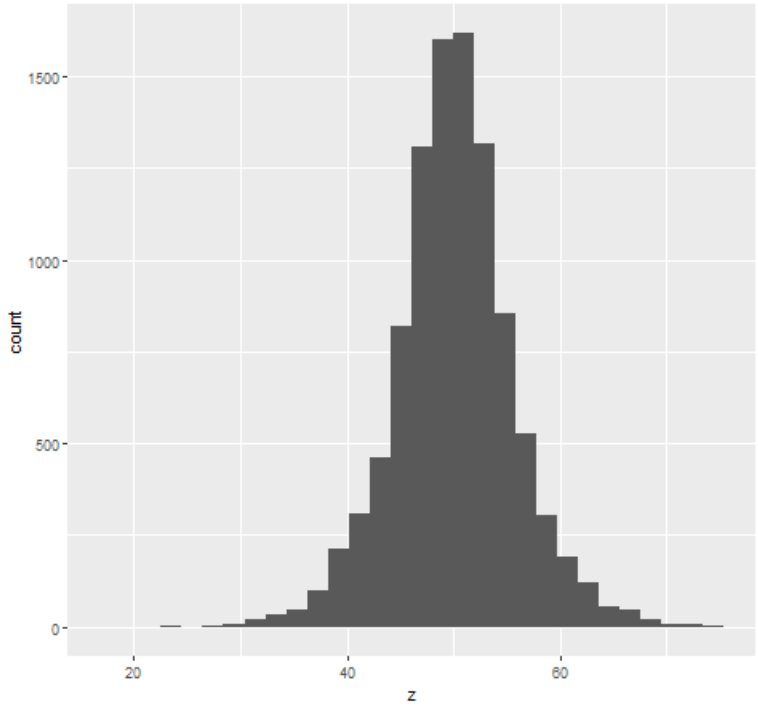


Figure G 9 : Iteration # 40 (Distribution Kurtosis of 5.34 )
True Score Frequencies with Kurtosis = 5.31

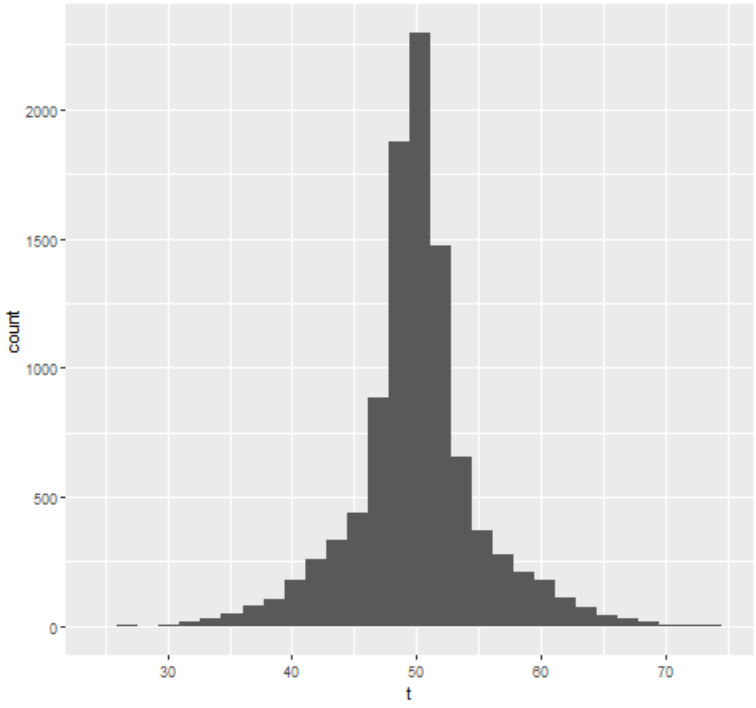



Figure G 10 : Iteration # 40 (Distribution Kurtosis of 5.34 )
Observed Score Frequencies with Kurtosis = 4.48

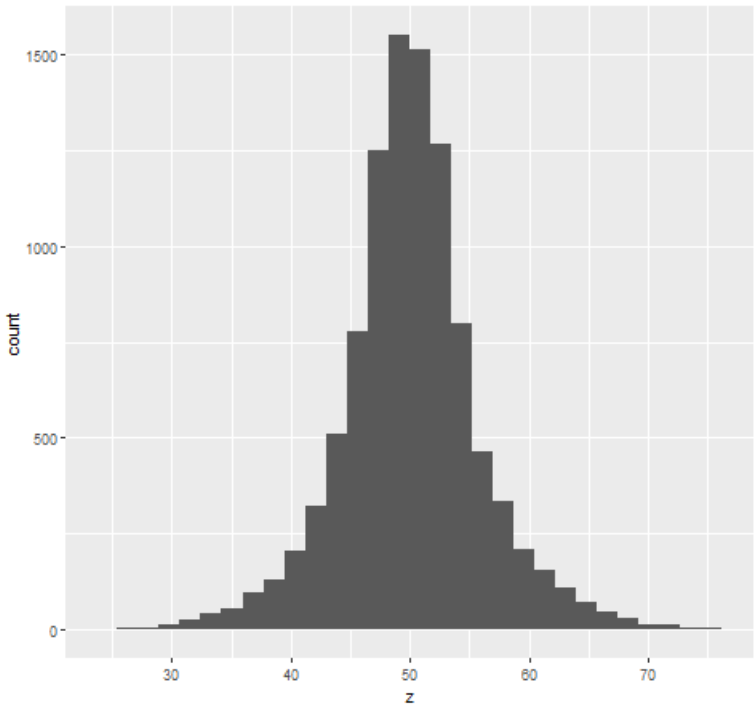


Figure G 11 : Iteration # 50 (Distribution Kurtosis of 5.94 )
True Score Frequencies with Kurtosis = 6.03

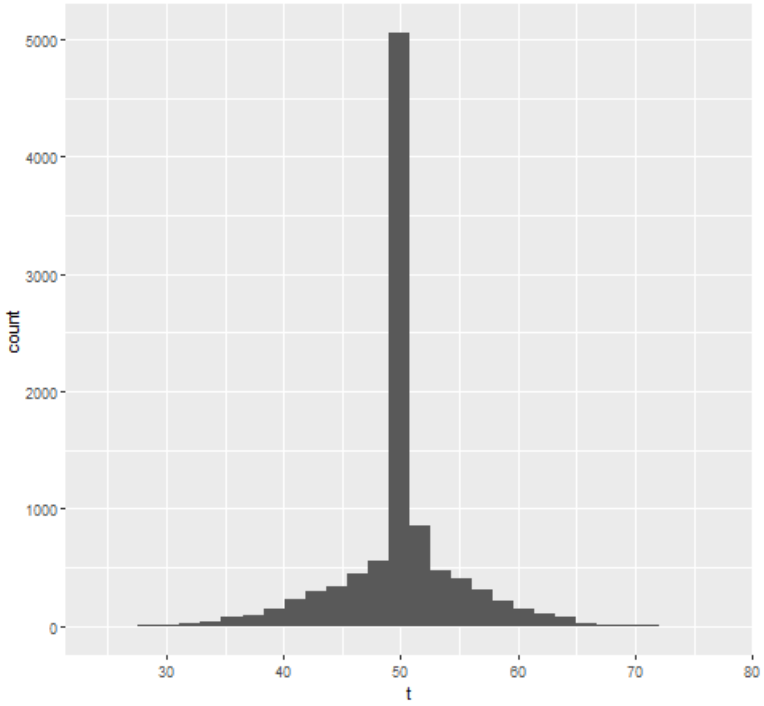
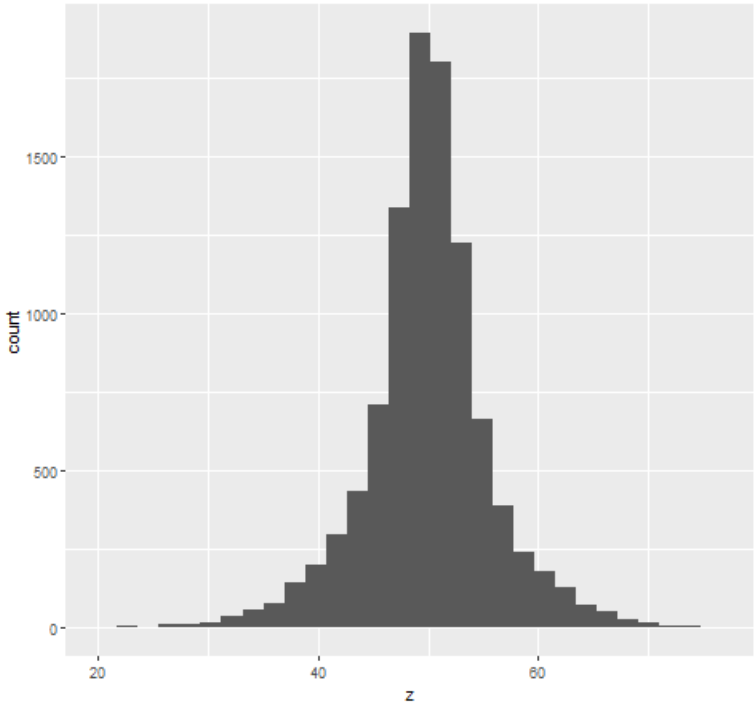



Figure G 12 : Iteration # 50 (Distribution Kurtosis of 5.94 )
Observed Score Frequencies with Kurtosis = 4.93



## Appendix H: R Code

### *H1: Code for the Callable GW Function*

```
#/*
callable GW function
#*/

if(require(moments)==FALSE){install.packages('moments')}
if(require(ggplot2)==FALSE){install.packages('ggplot2')}
library(ggplot2)
library(moments)
gandw=function(obsmean,obsvar,truecut)
{
 #GW method as a callable function for tabling
 #######Functions (run these first)
 phi=function(x)
 {(1/2)*(1+erf(x/sqrt(2)))}
 falseposfunc2=function(theta)
 {
  (1-phi(((c-theta)/sderr)))*dnorm(theta,obsscoremean,truesd)#exp(-(theta^2)/2)
 }
 falsenegfunc2=function(theta)
 {
  (1-(1-phi(((c-theta)/sderr))))*dnorm(theta,obsscoremean,truesd)
 }
 library(pracma)
 obsscoremean=obsmean
```

```
sdobs=sqrt(obsvar)
rel=.8
w=1
sderr=sqrt(1-rel)*sdobs
truevar=rel*(sdobs)^2
truesd=sqrt(truevar)
newmat2=matrix(nrow=100000,ncol=7)
colnames(newmat2)=c("theta","wce","FP","FN","wcec","FPc","FNc")
fpvec2=Vectorize(falseposfunc2,'theta')
fnvec2=Vectorize(falsenegfunc2,'theta')
pshouldfail=pnorm((truecut-obsscoremean)/truesd,0,1)
pshouldpass=1-pshouldfail
w=1
c=20
# end=4*sdobs+truecut
i=1
while(c<=90)
{
  fp2=integrate(fpvec2,-Inf,truecut)
  fn2=integrate(fnvec2,truecut,Inf)
  fpc=fp2$value/pshouldfail
  fnc=fn2$value/pshouldpass
  wce=(w*fp2$value+fn2$value)
  newmat2[i,1]=c
  newmat2[i,2]=wce
```

```
  newmat2[i,3]=fp2$value
  newmat2[i,4]=fn2$value
  newmat2[i,5]=(1/2)*(fpc+fnc)
  newmat2[i,6]=fpc
  newmat2[i,7]=fnc
  i=i+1
  c=c+.1
 }
 outmat=subset(newmat2,newmat2[,2]>0)
#optimal error point
loc=which(outmat[,2]==min(outmat[,2]))
minerrC=outmat[loc,1]
minerrFP=outmat[loc,3]
minerrFN=outmat[loc,4]
#now I need error at the true cut
a=round(as.vector(outmat[,1]),digits=1)
loc=which(a[]==truecut)
truecutFP=outmat[loc,3]
truecutFN=outmat[loc,4]
outvector=as.matrix(c(truecutFP,truecutFN,minerrC,minerrFP,minerrFN))
rownames(outvector)=c('truecutFP','truecutFN','minerrC','minerrFP','minerrFN')
return(outvector)
}#end g&w function
```

***H2: Code for Skewness simulations***

```
library(rstudioapi)

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

sourcepath='../'

source(file.path(sourcepath,'callable GW function.r'))

Finaloutput=matrix(nrow=10000,ncol=21)

colnames(Finaloutput)=c('distribution.degree.of.nonnorm','truescore.degree.of.nonnorm','obs.degree.o
f.nonnorm','mean.true','mean.obs','var.true','var.obs','actual.fp.at.truecut','actual.fn.at.truecut',

'actual.tot.err.truecut','actual.opt.cut','fp.err.at.opt','fn.err.at.opt','toterr.at.opt',

'est.fp.at.truecut','est.fn.at.truecut','est.opt.cut','fp.est.at.opt','fn.est.atopt','toterr.est.attrue','toterr.est.
atopt')

library(ggplot2)

library(moments)

set.seed(9999)

#kurtosis

#we want to loop from skew=0 to 1.96

k=0.00

i=1 # i iterates with k, but needs to be integers increasing by 1

while(k<=1.96)

{

#mixture kurtosis

skew=k

theta = sqrt(25*(skew^(2/3))/(2^(2/3)))

var=25-theta^2

mew=50-theta

x=rnorm(10000,mew,sqrt(var))

y=rexp(10000,1/theta)
```

```
t=x+y
skewness(t)
#let y be the vector of our observed scores, recall that 2.5 is our std err
z=vector(length=10000)
n=1
while(n<=10000)
{
z[n]=rnorm(1,t[n],2.5)
n=n+1
}
trueandobs=cbind(t,z) #true scores on left, obs scores on right
#okay, now we'll search over score ranges, taking steps of .1
findat=NULL
#we'll build this row by row, so start with a vector
c=20
while(c<=80)
{
tableofdata=vector(length=4)
shouldpass=subset(trueandobs,trueandobs[,1]>=45)
falsenegcount=nrow(subset(shouldpass,shouldpass[,2]<c))
shouldfail=subset(trueandobs,trueandobs[,1]<45)
falseposcount=nrow(subset(shouldfail,shouldfail[,2]>=c))
tableofdata[1]=c
tableofdata[3]=falseposcount/10000
tableofdata[4]=falsenegcount/10000
```

```
tableofdata[2]=tableofdata[3]+tableofdata[4] this is column 2, but has to be written after 3 and 4

findat=rbind(findat,tableofdata)

c=c+.1

}

#need to locate point of optimal error

#and also line where truecut=45

loc45=which(round(findat[,1],digits=1)==45)

locmin=min(which(findat[,2]==min(findat[,2])))  # i take the min of the location, cause sometimes we'll get two points with same error/etc.

#in those rare instances, I'll take the left most (min on horizontal scale) location

Finaloutput[i,1]=k  #the non-normality parameter (skew, kurt, or distance between modes)

Finaloutput[i,2]=skewness(t)

Finaloutput[i,'obs.degree.of.nonnorm']=skewness(z)

Finaloutput[i,'mean.true']=mean(t)

Finaloutput[i,'mean.obs']=mean(z)

Finaloutput[i,'var.true']=var(t)

Finaloutput[i,'var.obs']=var(z)

Finaloutput[i,'actual.fp.at.truecut']=findat[loc45,3]

Finaloutput[i,'actual.fn.at.truecut']=findat[loc45,4]

Finaloutput[i,'actual.tot.err.truecut']=Finaloutput[i,'actual.fn.at.truecut']+Finaloutput[i,'actual.fp.at.truecut']

Finaloutput[i,'actual.opt.cut']=findat[locmin,1]

Finaloutput[i,'fp.err.at.opt']=findat[locmin,3]

Finaloutput[i,'fn.err.at.opt']=findat[locmin,4]

Finaloutput[i,'toterr.at.opt']=Finaloutput[i,'fp.err.at.opt']+Finaloutput[i,'fn.err.at.opt']

#need to call these after the var has been written
```

```
Finaloutput[i,15:19]=gandw(Finaloutput[i,'mean.obs'],Finaloutput[i,'var.obs'],45)

Finaloutput[i,'toterr.est.attrue'] =Finaloutput[i,'est.fp.at.truecut']+Finaloutput[i,'est.fn.at.truecut']

Finaloutput[i,'toterr.est.atopt']= Finaloutput[i,'fp.est.at.opt']+Finaloutput[i,'fn.est.atopt']


i=i+1

k=round(k+.04,digits=3)

}# END OF skewness LOOP

output=Finaloutput[1:50,]

write.csv(output,file='skewness45.csv')
```

***H3: Code for Bimodal simulations***

```
library(rstudioapi)

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

sourcepath='../../'

source(file.path(sourcepath,'callable GW function.r'))  #file path does our paste work for us

modeless50 <- function(v2) {

u=round(v2,digits=1)

v=subset(u,u<=49.8)

u=round(u,digits=0)

uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

modemore50 <- function(v2) {

u=round(v2,digits=1)

v=subset(u,u>=50.2)
```

```
uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

Finaloutput=matrix(nrow=10000,ncol=21)

colnames(Finaloutput)=c('distribution.degree.of.nonnorm','truescore.degree.of.nonnorm','obs.degree.of.nonnorm','mean.true','mean.obs','var.true','var.obs','actual.fp.at.truecut','actual.fn.at.truecut',

'actual.tot.err.truecut','actual.opt.cut','fp.err.at.opt','fn.err.at.opt','toterr.at.opt',

'est.fp.at.truecut','est.fn.at.truecut','est.opt.cut','fp.est.at.opt','fn.est.atopt','toterr.est.attrue','toterr.est.atopt')

library(moments)

set.seed(9999)

#we want to loop from k=7 to k=9.94. do 50 total. I have (effectively) 3/50 = a delta x of .06

k=7.00

i=1 # i iterates with k, but needs to be integers increasing by 1

while(k<=9.94)

{

#mixture bimodality

#let

d=k #then

mew2=(100+d)/2

mew1=100-mew2

v=25+(50)^2-2500-(d^2)/4

x=rnorm(5000,mew2,sqrt(v) )

y=rnorm(5000,mew1,sqrt(v))

t=c(x,y)

var(t)
```

```
#let y be the vector of our observed scores, recall that 2.5 is our std err

z=vector(length=10000)

n=1

while(n<=10000)

{

z[n]=rnorm(1,t[n],2.5)

n=n+1

}

trueandobs=cbind(t,z) #true scores on left, obs scores on right

#okay, now we'll search over score ranges, taking steps of .5

findat=NULL

#we'll build this row by row, so start with a vector

c=20

while(c<=80)

{

tableofdata=vector(length=4)

shouldpass=subset(trueandobs,trueandobs[,1]>=45)

falsenegcount=nrow(subset(shouldpass,shouldpass[,2]<c))

shouldfail=subset(trueandobs,trueandobs[,1]<45)

falseposcount=nrow(subset(shouldfail,shouldfail[,2]>=c))

tableofdata[1]=c

tableofdata[3]=falseposcount/10000

tableofdata[4]=falsenegcount/10000

tableofdata[2]=tableofdata[3]+tableofdata[4] # are you watching? this is column 2, but has to be written after 3 and 4

findat=rbind(findat,tableofdata)
```

```
c=c+.1

}

#need to locate point of optimal error

#and also line where truecut=45

loc45=which(round(findat[,1],digits=1)==45)

locmin=min(which(findat[,2]==min(findat[,2])))  # i take the min of the location, cause sometimes we'll get two points with same error/etc.

#in those rare instances, I'll take the left most (min on horizontal scale) location

Finaloutput[i,1]=k  #the non-normality parameter (skew, kurt, or distance between modes)

Finaloutput[i,2]=modemore50(t)-modeless50(t)

Finaloutput[i,'obs.degree.of.nonnorm']=modemore50(z)-modeless50(z)

Finaloutput[i,'mean.true']=mean(t)

Finaloutput[i,'mean.obs']=mean(z)

Finaloutput[i,'var.true']=var(t)

Finaloutput[i,'var.obs']=var(z)

Finaloutput[i,'actual.fp.at.truecut']=findat[loc45,3]

Finaloutput[i,'actual.fn.at.truecut']=findat[loc45,4]

Finaloutput[i,'actual.tot.err.truecut']=Finaloutput[i,'actual.fn.at.truecut']+Finaloutput[i,'actual.fp.at.truecut']

Finaloutput[i,'actual.opt.cut']=findat[locmin,1]

Finaloutput[i,'fp.err.at.opt']=findat[locmin,3]

Finaloutput[i,'fn.err.at.opt']=findat[locmin,4]

Finaloutput[i,'toterr.at.opt']=Finaloutput[i,'fp.err.at.opt']+Finaloutput[i,'fn.err.at.opt']

#need to call these after the var has been written

Finaloutput[i,15:19]=gandw(Finaloutput[i,'mean.obs'],Finaloutput[i,'var.obs'],45)

Finaloutput[i,'toterr.est.attrue'] =Finaloutput[i,'est.fp.at.truecut']+Finaloutput[i,'est.fn.at.truecut']
```

```
Finaloutput[i,'toterr.est.atopt']= Finaloutput[i,'fp.est.at.opt']+Finaloutput[i,'fn.est.atopt']


i=i+1

k=round(k+.06,digits=3)

}# end bimodiality loop

output=Finaloutput[1:50,]

write.csv(output,file='bimodality45.csv')
```

***H4: Code for Kurtosis simulations***

```
library(rstudioapi)

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

sourcepath='../'

source(file.path(sourcepath,'callable GW function.r'))

Finaloutput=matrix(nrow=10000,ncol=21)

colnames(Finaloutput)=c('distribution.degree.of.nonnorm','truescore.degree.of.nonnorm','obs.degree.of.nonnorm','mean.true','mean.obs','var.true','var.obs','actual.fp.at.truecut','actual.fn.at.truecut',

'actual.tot.err.truecut','actual.opt.cut','fp.err.at.opt','fn.err.at.opt','toterr.at.opt',
'est.fp.at.truecut','est.fn.at.truecut','est.opt.cut','fp.est.at.opt','fn.est.atopt','toterr.est.attrue','toterr.est.atopt')

library(moments)

set.seed(9999)

#kurtosis

#we want to loop from k=3 to k=5.99. I proposed to do 50 total. I have (effectively) 3/50 = a delta x of .06

k=3.00

i=1 # i iterates with k, but needs to be integers increasing by 1

while(k<=5.999)
```

```
{       #BEGIN KURTOSIS LOOP
#mixture kurtosis

 var2=sqrt(((k*625)/3)-1250+625)+25
var1 = (25-.5*(var2))/.5

 x=rnorm(5000,50,sqrt(var2))
 y=rnorm(5000,50,sqrt(var1))
t=c(x,y)

#let y be the vector of our observed scores, recall that 2.5 is our std err
z=vector(length=10000)
n=1
while(n<=10000)
{
 z[n]=rnorm(1,t[n],2.5)
 n=n+1
}
trueandobs=cbind(t,z) #true scores on left, obs scores on right
#okay, now we'll search over score ranges, taking steps of .1
findat=NULL#vector(length=4)
#we'll build this row by row, so start with a vector
c=20
while(c<=80)
{
```

```
 tableofdata=vector(length=4)
 shouldpass=subset(trueandobs,trueandobs[,1]>=45)
falsenegcount=nrow(subset(shouldpass,shouldpass[,2]<c))
shouldfail=subset(trueandobs,trueandobs[,1]<45)
falseposcount=nrow(subset(shouldfail,shouldfail[,2]>=c))
tableofdata[1]=c
 tableofdata[3]=falseposcount/10000
 tableofdata[4]=falsenegcount/10000
 tableofdata[2]=tableofdata[3]+tableofdata[4] # this is column 2, but has to be written after 3 and 4
findat=rbind(findat,tableofdata)
c=c+.1
}
#need to locate point of optimal error
#and also line where truecut=45
loc45=which(round(findat[,1],digits=1)==45)
locmin=min(which(findat[,2]==min(findat[,2])))  # i take the min of the location, cause sometimes we'll get two points with same error/etc.
                                    #in those rare instances, I'll take the left most (min on horizontal scale) location
Finaloutput[i,1]=k  #the non-normality parameter (skew, kurt, or distance between modes)
Finaloutput[i,2]=kurtosis(t)
Finaloutput[i,'obs.degree.of.nonnorm']=kurtosis(z)
Finaloutput[i,'mean.true']=mean(t)
Finaloutput[i,'mean.obs']=mean(z)
Finaloutput[i,'var.true']=var(t)
Finaloutput[i,'var.obs']=var(z)
```

```
Finaloutput[i,'actual.fp.at.truecut']=findat[loc45,3]

Finaloutput[i,'actual.fn.at.truecut']=findat[loc45,4]

Finaloutput[i,'actual.tot.err.truecut']=Finaloutput[i,'actual.fn.at.truecut']+Finaloutput[i,'actual.fp.at.true
cut']

Finaloutput[i,'actual.opt.cut']=findat[locmin,1]

Finaloutput[i,'fp.err.at.opt']=findat[locmin,3]

Finaloutput[i,'fn.err.at.opt']=findat[locmin,4]

Finaloutput[i,'toterr.at.opt']=Finaloutput[i,'fp.err.at.opt']+Finaloutput[i,'fn.err.at.opt']

#need to call these after the var has been written

Finaloutput[i,15:19]=gandw(Finaloutput[i,'mean.obs'],Finaloutput[i,'var.obs'],45)

Finaloutput[i,'toterr.est.attrue'] =Finaloutput[i,'est.fp.at.truecut']+Finaloutput[i,'est.fn.at.truecut']

Finaloutput[i,'toterr.est.atopt']= Finaloutput[i,'fp.est.at.opt']+Finaloutput[i,'fn.est.atopt']

i=i+1

k=round(k+.06,digits=3)

}# END OF KURTOSIS LOOP

output=Finaloutput[1:50,]

write.csv(output,file='kurtosis45.csv')
```