2010-03-09

Parameter Estimation and Hypothesis Testing for the Truncated Normal Distribution with Applications to Introductory Statistics Grades

James T. Hattaway
Brigham Young University - Provo

Parameter Estimation and Hypothesis Testing for the Truncated Normal
Distribution with Applications to Introductory Statistics Grades

James Taylor Hattaway

A project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Scott D. Grimshaw, Chair
E. Shannon Neeley
Bruce J. Collings

Department of Statistics

Brigham Young University

April 2010

ABSTRACT

Parameter Estimation and Hypothesis Testing for the Truncated Normal
Distribution with Applications to Introductory Statistics Grades

James Taylor Hattaway

Department of Statistics

Master of Science

The normal distribution is a commonly seen distribution in nature, education, and business. Data that are mounded or bell shaped are easily found across various fields of study. Although there is high utility with the normal distribution; often the full range can not be observed. The truncated normal distribution accounts for the inability to observe the full range and allows for inferring back to the original population. Depending on the amount of truncation, the truncated normal has several distinct shapes. A simulation study evaluating the performance of the maximum likelihood estimators and method of moment estimators is conducted and a comparison of performance is made. The level α Likelihood Ratio Test (LRT) is derived for testing the null hypothesis of equal population means for truncated normal data. A simulation study evaluating the power of the LRT to detect absolute standardized differences between the two population means with small sample size was conducted and the power curves were approximated. Another simulation study evaluating the power of the LRT to detect absolute differences for testing the hypothesis with large unequal sample sizes was conducted. The level α LRT was extended to a k population hypothesis test for equal population means. A simulation study examining the power of the k population LRT for detecting absolute standardized differences when one of the population means is different than the others was conducted and the power curve approximated. Stat 221 is the largest introductory statistics course at BYU serving about 4,500 students a year. Every section of Stat 221 shares common homework assignments and tests. This controls for confounding when making comparisons between sections. Historically grades have been thought to be bell shaped, but with grade inflation and other factors, the upper tail is lost because of the truncation at 100. It is reasonable to assume that grades follow a truncated normal distribution. Inference using the final grades should be done recognizing the truncation. Performance of the different Stat 221 sections was evaluated using the level α LRTs derived.

Keywords: maximum likelihood estimators, method of moments estimators, likelihood ratio test, assessment of student learning, Stat 221

ACKNOWLEDGMENTS

CONTENTS

INTRODUCTION

Stat 221 at Brigham Young University (BYU) is an introductory statistics course taught to about 4,500 students each year. Due to the high number of participants, Stat 221 is offered through various instructional methods. It is taught during the day as well as through evening classes. There are courses available on the main BYU campus, at the Salt Lake Center Annex, and through BYU Independent Study. Each year there are about 15 different instructors teaching the course.

Often we want to compare different sections of the same course in the assessment of student learning to determine if there are differences between the instructors. This assessment typically is based on the final grade, which is a weighted average of the course elements. Comparing grades between the various sections of Stat 221, the inference needs to recognize the truncated nature of the distribution of grades. The final grades can be considered to be normally distributed because they are an average of various measures of a student's learning, but they are truncated at 100, cutting off the upper tail of the distribution.

1.1 TRUNCATED DISTRIBUTIONS

A truncated distribution is a *conditional distribution* resulting when the domain of the parent distribution is restricted to a smaller region. A truncated distribution occurs when there is no ability to know about or record events occurring above or below a set threshold or outside a certain range.

Truncated data is an acceptable commonplace occurrence in the field of reliability, when the variable of interest is related to failure rates of items. Truncation is different than censoring. With censoring, knowledge of items outside the restricted range is retained, but the full information is unable to be recorded. With truncation, knowledge of items outside

the restricted range cannot be obtained. An example of truncation from manufacturing occurs when a sample of items is selected to be studied from a population that has already had items removed due to a failure to meet the set requirements. Another example of truncation is the population of standardized test scores. Standardized tests, such as the SAT, are designed to be normally distributed with a known mean and variance. Universities and colleges have established minimum requirements for admittance, creating a population of SAT scores with part of the lower tail of the distribution missing.

Let X be a random variable from a distribution with a probability density function, $f(x)$, a cumulative distribution function, $F(x)$, and the range of the support $(-\infty, \infty)$. The density function of X given the restriction that $a \leq X < b$ is

$$f(x|a \leq X < b) = \begin{cases} \frac{f(x)}{F(b)-F(a)} & a \leq x < b \\ \\ 0 & \text{otherwise} \end{cases} . \tag{1.1}$$

Because $f(x|a \leq X < b)$ is scaled up to account for the probability of being in the restricted support, $f(x|a \leq X < b)$ is a density function. The restriction can occur either on a single side or on both sides of the range. Truncation that occurs on a single side of the range is called *singly truncated* and on both sides of the range is called *doubly truncated*. When the truncation occurs at the upper (or right) end of the support range, this is called *truncation from above*. When the truncation occurs at the lower (or left) end of the support range, this is called *truncation from below*.

1.2 TRUNCATED NORMAL DISTRIBUTION

The normal distribution is a commonly seen distribution in nature, education, and business. Data, which are mounded or have a bell shape curve, are easily found across various fields of study. This is often due to the Central Limit Effect, where a measurement is the mean of a collection of random effects. Johnson and Thomopoulos (2002) indicate although normal data have high utility, situations occur where the infinite range, $X \in (-\infty, \infty)$, can cause problems when estimating and inferring back to the population.

Applying (1.1) to the normal density function, the truncated normal density function is

$$f(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2} \cdot P(Y \in (a,b)|\mu, \sigma^2)}, \tag{1.2}$$

where $P(Y \in (a,b)|\mu, \sigma^2) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} dy$ scales the distribution. This scaling factor can be interpreted as the probability of being in the restricted range (a,b) for a given μ and σ^2. The value $1 - P(Y \in (a,b)|\mu, \sigma^2)$ is the percent (%) truncation of the distribution. If the distribution function of the standard normal distribution, denoted $\Phi(\cdot)$, is available then $P(Y \in (a,b)|\mu, \sigma^2) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{b-\mu}{\sigma}\right)$.

Johnson and Kotz (1970) provide the expected value of a truncated normal random variable X as

$$E(X) = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma, \tag{1.3}$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. The expected value is equal to μ plus an adjustment for the trunation on the distribution. This adjustment shifts the $E(X)$ into the appropriate tail based on the truncation. For example, when there is more truncation on the lower portion of the domain than the upper, then $E(X)$ shift into the upper tail.

Johnson and Kotz (1970) provide the variance of a truncated normal random variable X as

$$\text{Var}(X) = \left[1 + \frac{\left(\frac{a-\mu}{\sigma}\right)\phi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right)\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)\right]\sigma^2. \tag{1.4}$$

Like the expected value, the variance of X is adjusted for the truncation.

When there is symmetric truncation, that is $a - \mu = -(b - \mu) = -\delta$, the expected value and variance of a truncated normal random variable X are

$$E(X) = \mu$$

$$Var(X) = \left(1 + \frac{2\delta\phi(\delta)}{2\Phi(\delta) - 1}\right)\sigma^2.$$

Often the goal is to make inference back to the original population and not on the truncated population that is sampled. This means that the inference is made on μ and not $E(X)$.

_____

PARAMETER ESTIMATION

This document considers the maximum likelihood estimators (MLE) and the method of moments (MoM) estimators for the truncated normal distribution. The case of a left truncated normal distribution with the point of truncation at $a = 0$ will be used in the simulation study. This distribution is also referred to as a "positive" normal distribution. Without loss of generality, the symmetric nature of the normal distribution permits changing truncation to match a left truncated distribution. For a right truncated distribution, a new variable would need to be defined as $X' = T^* - X$, where T^* is the point of truncation and $\mu^* = T^* - \mu$ will be the new parameter of interest.

Figures 2.1 and 2.2 show the graphs of density functions from left truncated distributions with different values of μ and σ and truncation at $a = 0$. Figure 2.1 illustrates that when distance $|T^* - \mu|$ is large relative to σ, the distribution appears to be normally shaped. Johnson and Kotz (1970) indicate when $\frac{|T^* - \mu|}{\sigma} \geq 3.5$, the effect of truncation is essentially gone and parameter estimation can be done using the normal distribution. Figure 2.2 shows with larger values of σ the distribution appears to flatten.

2.1 Maximum Likelihood Estimators

The likelihood for the truncated normal distribution is

$$
\begin{aligned}
L(\mu, \sigma^2) &= \prod_{i=1}^{n} f(x) = \left(\frac{1}{P(Y \in (a,b)|\mu, \sigma^2)} \right)^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{\frac{-\sum (x_i - \mu)^2}{2\sigma^2}} \\
&= \left(P(Y \in (a,b)|\mu, \sigma^2) \right)^{-n} \left(\sqrt{2\pi\sigma^2} \right)^{-n} + e^{\frac{-\sum (x_i - \mu)^2}{2\sigma^2}}.
\end{aligned}
\tag{2.1}
$$

$$\tag{2.2}$$

Different Density Functions of Left Truncated Normal Distributions

σ = 1

Figure 2.1: Graph of various density functions of a left truncated normal distribution with truncation point $a = 0$ and $\sigma = 1$. The shape of the distribution changes as different values for μ are considered.

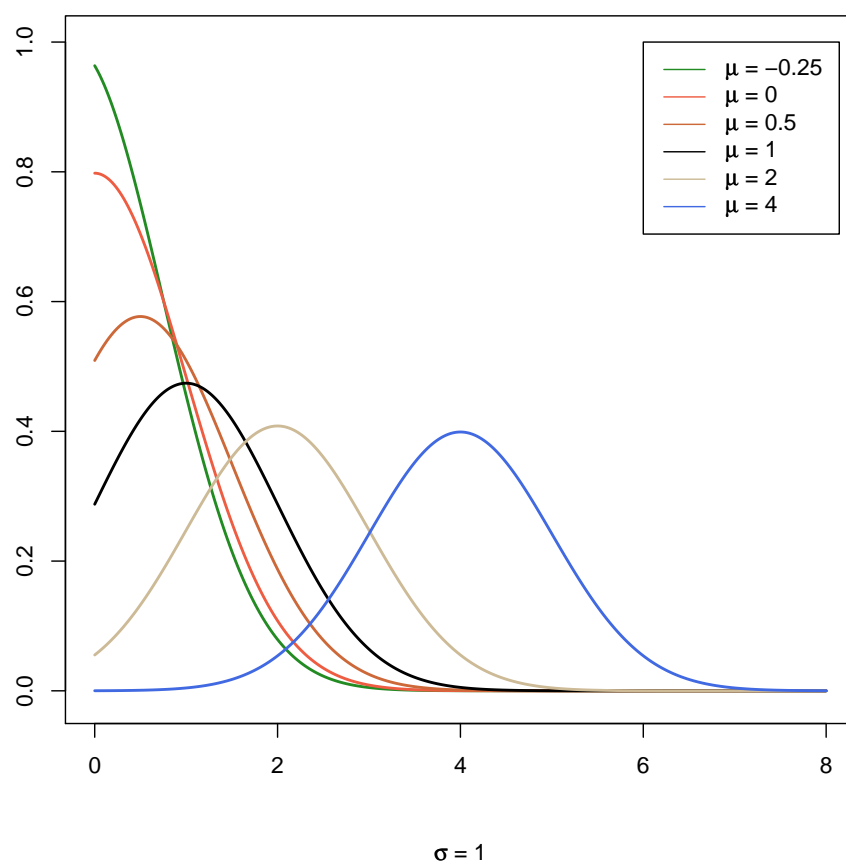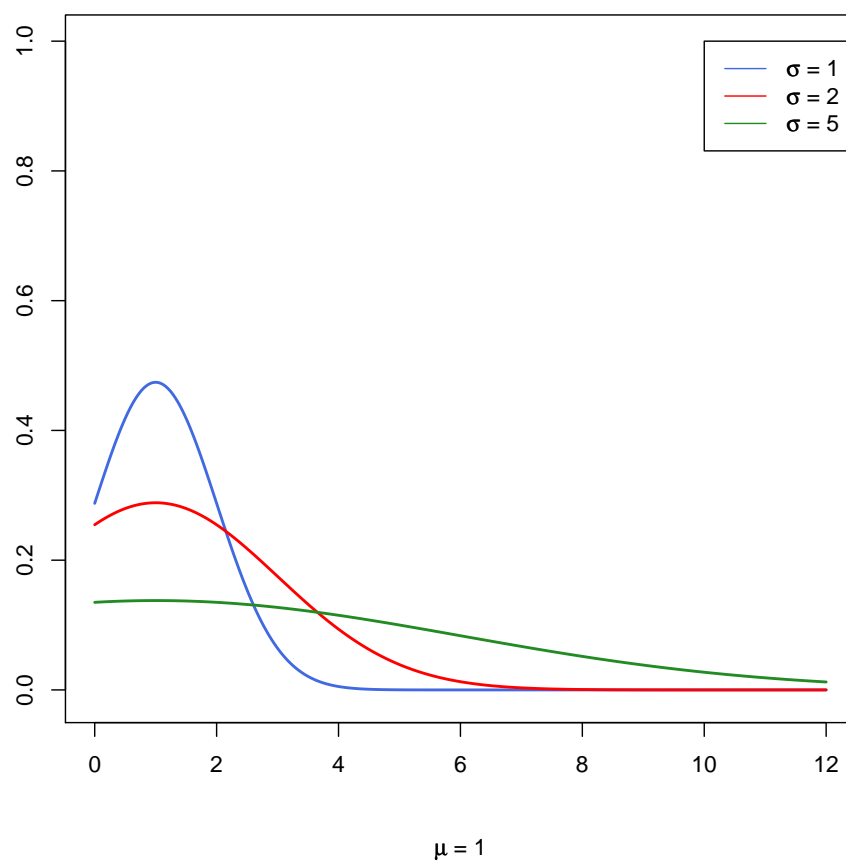**Different Density Functions of Left Truncated Normal Distributions**

Figure 2.2: Graph of various density functions of a left truncated normal distribution with truncation point $a = 0$ and $\mu = 1$. The distribution flattens when larger values for σ are considered.

The log-likelihood is

$$
\begin{aligned}
\ell \quad &= \ln L(\mu, \sigma^2) \\
&= -n \ln \left(P(Y \in (a,b) | \mu, \sigma^2) \right) - n \ln \left(\sqrt{2\pi\sigma^2} \right) - \frac{\sum (x_i - \mu)^2}{2\sigma^2}. \tag{2.3}
\end{aligned}
$$

Under normal circumstances, maximum likelihood and method of moment estimation is straightforward. However, when dealing with a truncated normal, the $P(Y \in (a,b) | \mu, \sigma^2)$ term is part of the estimation. This probability is defined in (1.2) as

$$
\psi(\mu, \sigma) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, dy, \tag{2.4}
$$

which is equivalent to $\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$. The Lebesgue's Dominated Convergence Theorem states that

$$
\frac{\partial}{\partial \theta} \int_a^b \phi(z) \, dz. = \int_a^b \left[\frac{\partial}{\partial \theta} \phi(z) \right] dz
$$

under the condition a dominating function $g(x)$ exists for $\phi(z)$, which converges to a finite value. Interchanging differentiation and integration, the first derivates of (2.4) with respect to the μ and σ are

$$
\begin{aligned}
\psi'_\mu(\mu, \sigma) \quad &= \frac{\partial \psi}{\partial \mu} \quad = \int_a^b \exp \left\{ \frac{-(y-\mu)^2}{2\sigma^2} \right\} \times \frac{(y-\mu)}{\sigma^3 \sqrt{2\pi}} \, dy, \quad \text{and} \\
\psi'_\sigma(\mu, \sigma) \quad &= \frac{\partial \psi}{\partial \sigma} \quad = \int_a^b \exp \left\{ \frac{-(y-\mu)^2}{2\sigma^2} \right\} \times \left(\frac{(y-\mu)^2}{\sigma^4 \sqrt{2\pi}} - \frac{1}{\sigma^2 \sqrt{2\pi}} \right) dy.
\end{aligned}
$$

Using the derivatives above, the gradient ($\mathbf{G}$) or the first partial derivative vector of (2.3) with respect to the parameters is

$$
\mathbf{G} = \begin{bmatrix} \frac{\partial \ell}{\partial \mu} \\ \frac{\partial \ell}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} -n\frac{\psi'_\mu}{\psi} - \frac{1}{\sigma^2}(n\mu - \sum x_i). \\ -n\frac{\psi'_{\sigma^2}}{\psi} - \frac{n}{2\sigma^2} + \frac{2\sum(x_i-\mu)^2}{(2\sigma^2)^2}. \end{bmatrix}. \tag{2.5}
$$

As Cohen (1949) points out, the problem of no closed-form solution to (2.5) means that solving $\mathbf{G} = 0$ must be performed iteratively. Newton-Raphson is suggested by Cohen (1950), but Halperin (1952) points out that Cohen relies on rough estimates for the starting values, which can cause problems with the estimates converging too quickly in the Newton-Raphson method.

The second derivatives of (2.4) needed for the Hessian are

$$\psi''_\mu(\mu,\sigma) = \frac{\partial^2\psi}{\partial^2\mu} = \int_a^b \exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \times \left(\frac{(y-\mu)^2}{\sigma^5\sqrt{2\pi}} - \frac{1}{\sigma^3\sqrt{2\pi}}\right) dy,$$

$$\psi''_\sigma(\mu,\sigma) = \frac{\partial^2\psi}{\partial^2\sigma} = \int_a^b \exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \times \left(\frac{2}{\sigma^3\sqrt{2\pi}} - \frac{5(y-\mu)^2}{\sigma^5\sqrt{2\pi}} + \frac{(y-\mu)^4}{\sigma^7\sqrt{2\pi}}\right) dy, \quad \text{and}$$

$$\psi''_{\mu,\sigma}(\mu,\sigma) = \frac{\partial^2\psi}{\partial\mu\partial\sigma} = \int_a^b \exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\} \times \left(\frac{(y-\mu)^3}{\sigma^6\sqrt{2\pi}} - \frac{3(y-\mu)}{\sigma^4\sqrt{2\pi}}\right) dy.$$

Using the equations above and taking the derivatives of (2.5), the Hessian, or second derivative matrix of (2.3), is

$$\mathbf{H} = \begin{bmatrix} \frac{\partial g_1}{\partial\mu} & \frac{\partial g_1}{\partial\sigma} \\[6pt] \frac{\partial g_2}{\partial\mu} & \frac{\partial g_2}{\partial\sigma} \end{bmatrix} = \begin{bmatrix} -n\frac{\psi\psi''_\mu - (\psi_{\mu'})^2}{\psi^2} - \frac{n}{\sigma^2} & -n\frac{\psi\psi''_{\mu|\sigma^2} - \psi'_\mu\psi'_{\sigma^2}}{\psi^2} + \frac{(n\mu - \sum x_i)}{(\sigma^2)^2} \\[8pt] -n\frac{\psi\psi''_{\mu|\sigma^2} - \psi'_\mu\psi'_{\sigma^2}}{\psi^2} + \frac{(n\mu - \sum x_i)}{(\sigma^2)^2} & -n\frac{\psi\psi''_{\sigma^2} - (\psi'_{\sigma^2})^2}{\psi^2} + \frac{2n}{(2\sigma^2)^2} - \frac{4\sum(x_i-\mu)^2}{(2\sigma^2)^3} \end{bmatrix}.$$
$$(2.6)$$

Iterating to find a solution is one of the problems with the MLE. The other problem is that numerical integration is required to evaluate $\psi, \psi_\mu,$ and ψ_σ at each step of the iteration. Tiku (1989) suggests a *work around* using *standardization* and using approximations to the standard normal density function ϕ and distribution function Φ.

The coding for the MLE, found in Appendix A.1, has several integrals to be evaluated at each iteration of the Newton-Raphson (NR) method. These integrals rely on the integrate command in R. The NR optimization is done using the nlminb function in R. This function allows for user-defined gradient and Hessian, instead of relying on numerical approximations to them. The funtion nlminb also has an option to add bounds for the parameter estimates, preventing invalid parameter estimates. The downside to the nlminb function is that it is slower than other optimization functions in R, but the other built-in functions do not allow for user-defined gradient, Hessian, and bound on the parameters. Attempts using them resulted in invalid parameter estimates with negative variances estimates, or the optimization function never deviated from the initial values.

Cohen (1950) suggests the initial values for the NR optimization should be the sample moments. Halperin (1952) cautions against using sample moments because the NR opti-

mization tended to stay at the starting values. The estimates with more than one iteration through the NR optimization process were considered.

2.2 METHOD OF MOMENTS ESTIMATORS

Method of moments (MoM) estimation for the truncated normal distribution is done by Pearson and Lee (1908), based on estimating μ and σ from a random sample of normal data where the number of observations and values in the truncated region were known. Cohen (1949) illustrates that these estimators are similar to the MLE and share the same problem of relying on iterative solutions to find the estimates. Pearson and Lee (1908) state that the MoM estimates are

$$\hat{\mu} = x_0 - h'\hat{\sigma}, \tag{2.7}$$

$$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i \cdot \psi_2\right), \quad \text{and} \tag{2.8}$$

$$\psi_1 = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{\left(\sum x_i\right)^2}, \tag{2.9}$$

where $h' = \frac{x_o - \hat{\mu}}{\hat{\sigma}}$ is the point of truncation measured in standard units of the population and ψ_1 and ψ_2 are moment functions of h'. Here ψ_2 is dependant on tables and the calcuated values of h' and ψ_1. Cohen (1949) indicates the MoM equations derived by Pearson and Lee (1908) are equivalent to the MLE equations derived by Fisher (1930).

Applying the suggestions by Tiku (1989) for the MLE, the MoM estimates are modified to be

$$\hat{\mu} = \mu + \sqrt{\sigma^2}\lambda(\alpha), \text{ and} \tag{2.10}$$

$$\hat{\sigma}^2 = \sigma^2[1 - \delta(\alpha)], \tag{2.11}$$

where $\alpha = \frac{a-\mu}{\sigma}$, $\lambda(\alpha) = \frac{\phi(\alpha)}{1-\Phi(\alpha)}$, and $\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$. Solving (2.10) and (2.11) using the sample moments, the MoM estimates are

$$
\begin{aligned}
E(X) =& \mu + \sigma\lambda(\alpha) = \frac{\sum(x_i)}{n} \\
=& \mu + \sigma\lambda(\alpha) - \bar{x} \text{ and} \\
Var(X) =& E(X^2) - [E(X)]^2 \\
=& \sigma^2[1 - \delta(\alpha)] + \mu^2 + 2\mu\sigma\lambda(\alpha) + \sigma^2\lambda^2(\alpha) - \frac{\sum(x_i)^2}{n}.
\end{aligned}
$$

(2.12)

(2.13)

The lack of a closed-form solution means that the MoM estimators must be found using an iterative approach. The code for this, found in Appendix A.2, uses Newton-Raphson to find the estimators. The code minimizes the parameters while solving (2.12) and (2.13) simultaneously. Like with the MLE code, the sample moments were used as the initial estimates for Newton-Raphson.

2.3 SIMULATION STUDY

The effect of truncation was studied. Different parameter combinations were used and the performance of the MLE and MoM estimators were evaluated based on bias and mean square error (MSE). The simulation study used a left truncated normal distribution with truncation point $a = 0$. From the literature, σ is known to be negatively biased and dependent on the $P(Y \in (a, b)|\mu, \sigma)$.

The different μ and σ combinations shown in Figure 2.3 provide a visual representation of the distributions of interest. Combinations with larger values of σ spread out and flatten. The flattening effect makes distinguishing distributions with different values of μ and the same values of σ hard. Running a pilot study, any of the three parameter combinations with $\sigma = 5$ had a low rate of convergence. Further exploration found these three combinations were hard to distinguish one from another. Based on the results from the pilot study, these three parameter combinations were dropped. Combinations with smaller values of σ preserve the shape of the normal distribution. From Figure 2.3 the remaining parameter

combinations reflect differences in the amount of truncation and not on particular values of μ and σ. None of the parameter combinations were chosen to have more than 50% truncation in this study.

There are five parameter combinations of interest in the simulation study. The first combination is $\mu = 0$, $\sigma = 1$ and represents a 50% truncation. At 50% truncation, the distribution maintains the appearance of a half normal distribution. The next two combinations are $\mu = 1$, $\sigma = 1$ and $\mu = 2$, $\sigma = 2$ with each representing 15.87% truncation. Because they represent the same level of truncation, the expectation is that they will have a similar performance in estimating parameters. The next combination is $\mu = 2$, $\sigma = 1$ and represents 2.27% truncation. The final combination is $\mu = 4$, $\sigma = 1$ and represents almost no truncation. This combination should perform like a regular estimation of the normal distribution and acts as a control on the performance of the code.

To generate from the truncated normal distribution, a function taking the parameters $(\mu, \sigma, n, t_1, t_2)$ generates values from a normal distribution(μ, σ). When the value generated is outside the range (t_1, t_2), a new value is sampled until a value inside the range is obtained. The function returns the sample of size n.

Analysis of Simulation

Harter and Moore (1966) illustrated, with sampling experiments, a negative bias for the MLE of $\hat{\mu}$ and $\hat{\sigma}$, which increased based on the degree of truncation. Figure 2.4 shows that the bias for $\hat{\mu}$ shrinks as the sample size increases. The negative bias persists among the different parameter combinations, but the negative bias is no longer apparent and is essentially zero for the parameter combination $\mu = 4$, $\sigma = 1$ and is virtually a "complete" normal distribution. All of the MLE biases converge to 0 as n increases. The MLEs for the 15.87% truncation both had larger biases for the small n but quickly converged to 0. The parameter combation $\mu = 4$, $\sigma = 1$ performed as expected with the MLE and MoM estimators. The only other

MoM estimator with bias that converged to 0 was the $\mu = 2$, $\sigma = 1$ combination. The other three combinations each converged to a positive bias.

Figure 2.5 shows the MSE for $\hat{\mu}$. The performance of the $\mu = 4$, $\sigma = 1$ combination is as expected for both the MLE and MoM estimators. The MLE for the $\mu = 0$, $\sigma = 1$ combination had a relatively small MSE, but the MoM estimator had a larger MSE that failed to converge to 0 with an increasing sample size. The only other estimator that failed to converge to 0 was the MoM estimator for the $\mu = 2$, $\sigma = 2$ combination. The other estimators had MSE's that converged to 0. The 15.87% truncated parameter combinations had larger variability for the MLE at the small n, but as n increased, the MSE converged to 0. Even though the $\mu = 2$, $\sigma = 1$ combination had near zero bias for $\hat{\mu}$ for the different sample sizes of n, the MSE was large for the small sample sizes.

Figure 2.6 depicts the bias for $\hat{\sigma}$ from the simulation study. The MLEs for $\hat{\sigma}$ had a negative bias as expected, but they perform remarkably well in each of the 5 cases studied having a near 0 bias. The MoM estimator for $\hat{\sigma}$ for the $\mu = 4$, $\sigma = 1$ combination again performed as expected, but the other MoM estimators had significant negative biases. The MoM estimator for the $\mu = 2$, $\sigma = 1$ combination had the smallest bias, converging to about -0.05.

Figure 2.7 shows the MSE for $\hat{\sigma}$. The MLEs behave in a similar fashion as did the MLEs for $\hat{\mu}$. There is less variability in the estimators, but the higher MSE is present for the two 15.87% truncation parameter combinations. The MoM estimators for $\hat{\sigma}$ behave like those for $\hat{\mu}$ when the percent truncation is small, but at higher levels of truncation, they have larger MSE.

The table of results for the simulation study can be found in Appendix A.3. The MLEs performed best in estimating the parameters of a truncated normal distribution. The MoM estimators exhibited biases when there was a high level of truncation. When inference is done with data from a truncated normal population, the MLEs can be used.

Different Density Functions of Left Truncated Normal Distributions
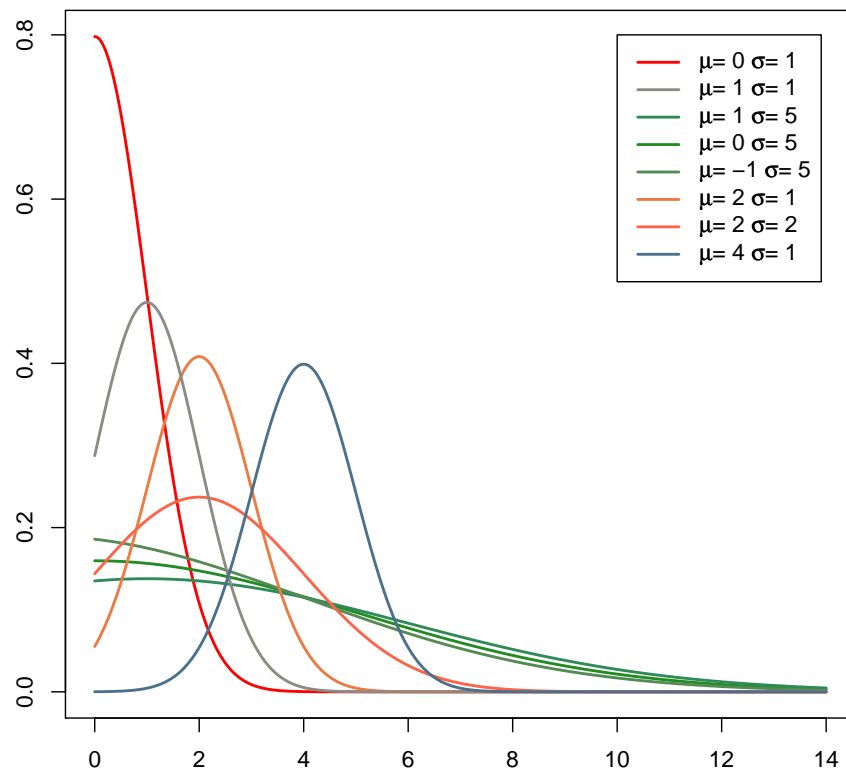
Figure 2.3: Various parameter combinations with a truncation point at 0.

Figure 2.4: Graph of bias for μ for the different parameter combinations. The percent truncation for each parameter combination is found in parentheses next to the parameter combination.

Figure 2.5: Graph of MSE for μ for the different parameter combinations. The percent truncation for each parameter combination is found in parentheses next to the parameter combination.

16

Figure 2.6: Graph of bias for σ for the different parameter combinations. The percent truncation for each parameter combination is found in parentheses next to the parameter combination.

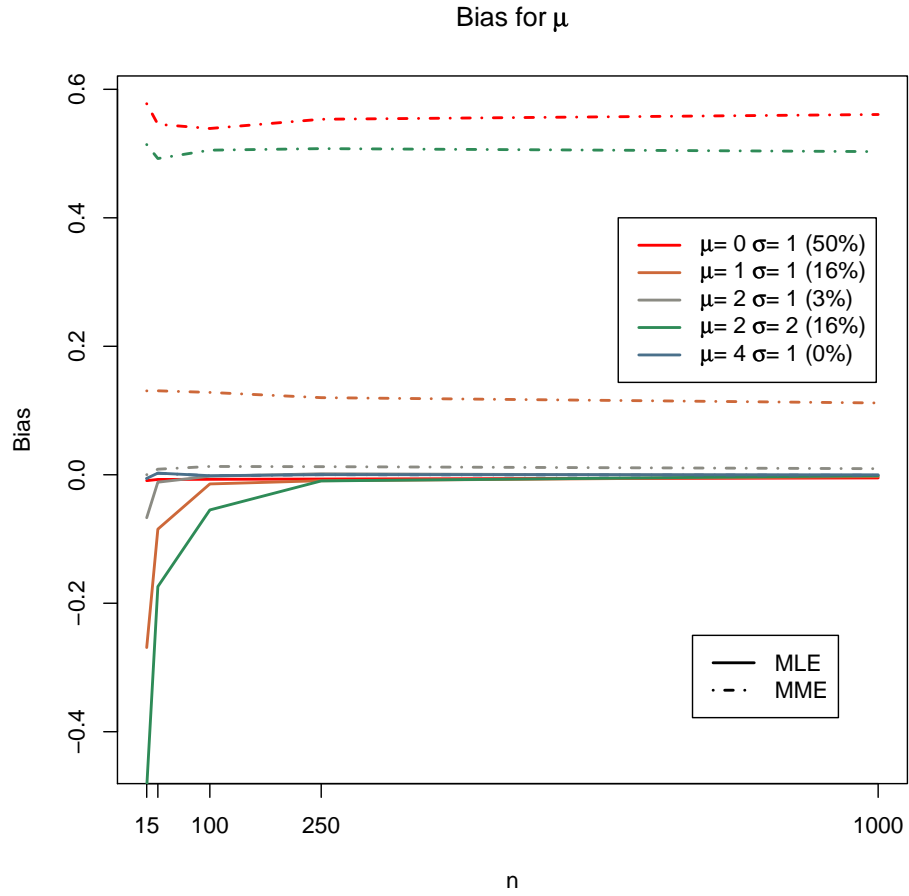Figure 2.7: Graph of MSE for σ for the different parameter combinations. The percent truncation for each parameter combination is found in parentheses next to the parameter combination.
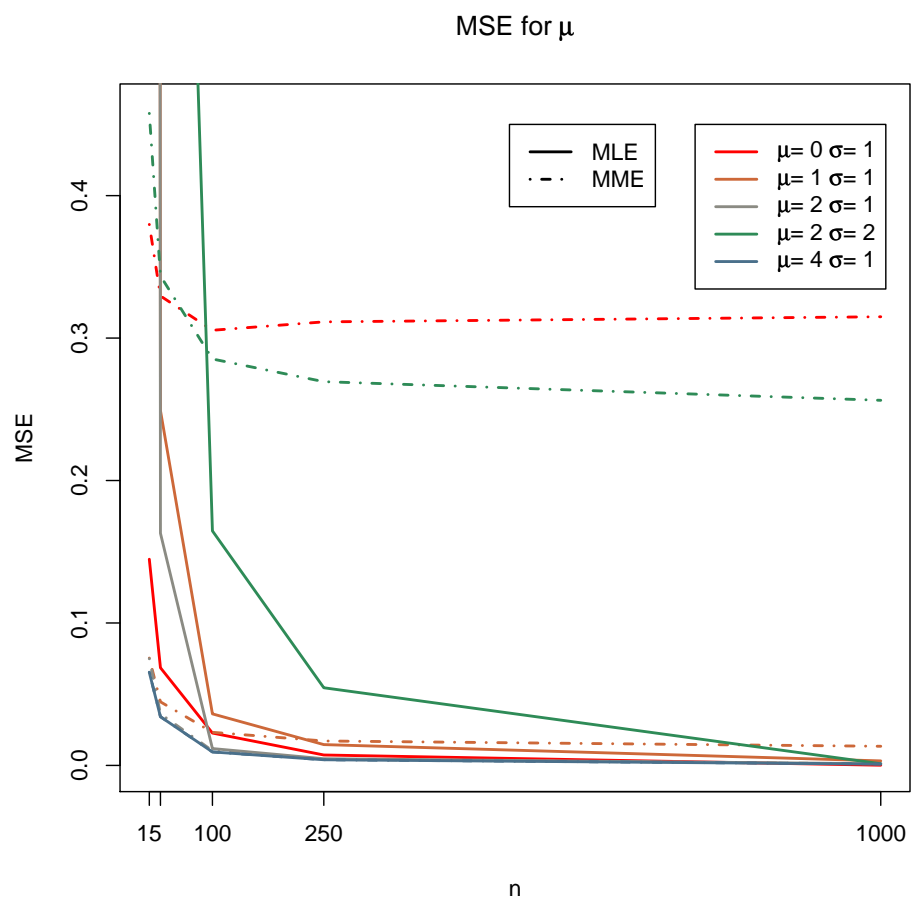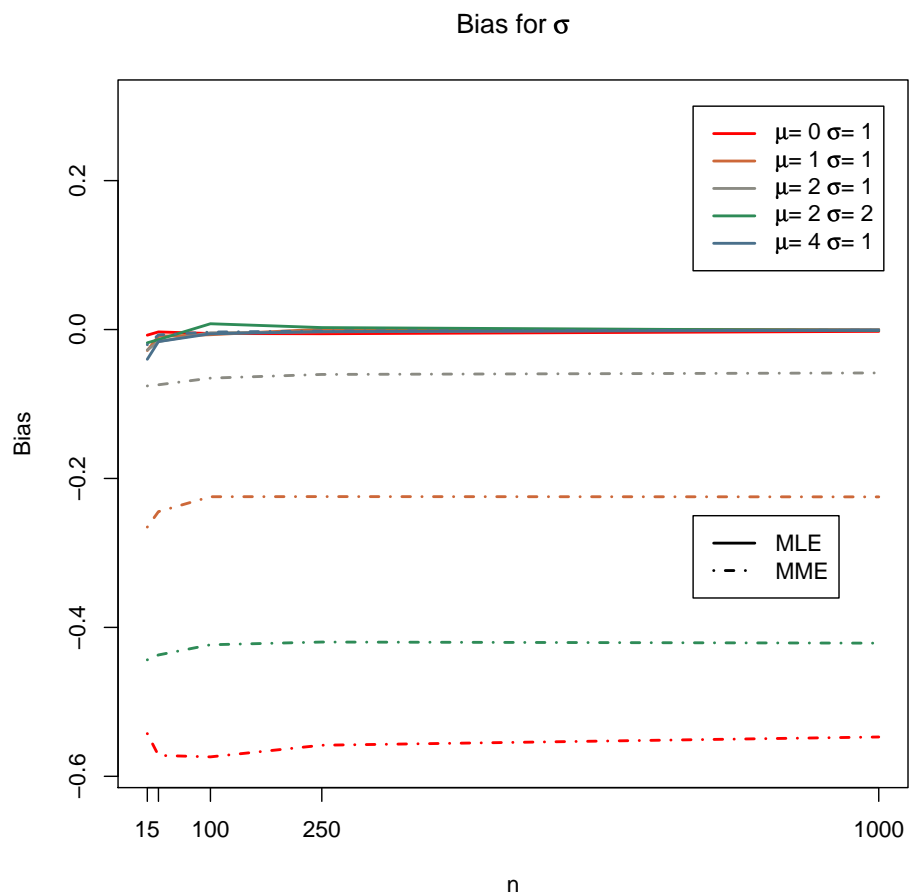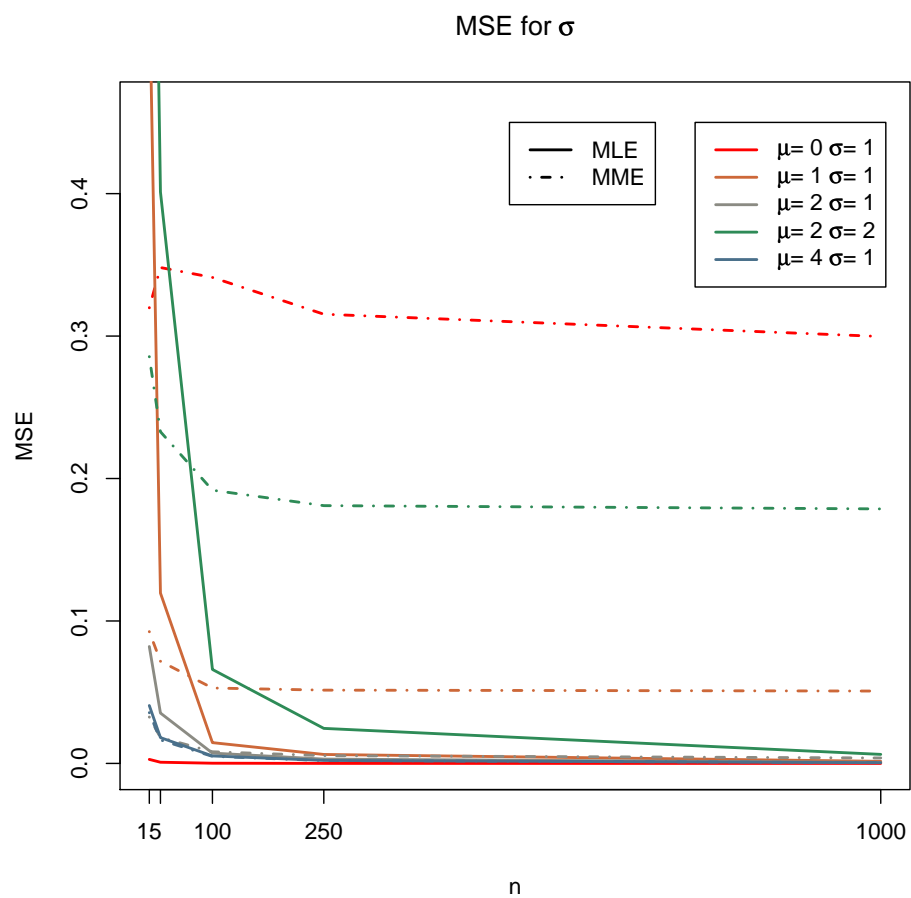
LIKELIHOOD RATIO TEST

Let $X_{11}, \ldots, X_{1n_1}$ be a random sample from a normal population (μ_1, σ) and let $X_{21}, \ldots, X_{2n_2}$ be an independent random sample from a normal population (μ_2, σ). The Likelihood Ratio Test (LRT) for testing the hypothesis $H_o : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ simplifies to the well known two-sample t-test with test statistic,

$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \tag{3.1}$$

which is compared to a critical value from a $t^*_{n_1+n_2-2,\alpha}$. Figure 3.1 depicts the densities of two normal populations with different μ_is but the same value of σ. Under this condition the two-sample t-test would be used to test the hypothesis $H_o : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$.

Let T^* be a point of truncation that restricts the domain of the two populations from Figure 3.1. Figure 3.2 represents the change in the density of the two populations given upper truncation point T^*. The density of the population further from T^* changes by a small amount, but the density of the population closer to T^* shows the upscaling on the density due to truncation.

Let $X_{11}, \ldots, X_{1n_1}$ be a random sample from a truncated normal distribution (μ_1, σ, a, b) and let $X_{21}, \ldots, X_{2n_2}$ be an independent random sample from a truncated normal distribution (μ_2, σ, a, b). The two populations share the same truncation points, (a, b), and variance. The LRT for testing the hypothesis $H_o : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ no longer simplifies to the two-sample t-test in (3.1). Consider the test statistic

$$\lambda(\mathbf{X_1}, \mathbf{X_2}) = \frac{\sup_{\Theta_0} L(\tilde{\mu}, \tilde{\sigma} | \mathbf{X_1}, \mathbf{X_2})}{\sup_{\Theta} L(\hat{\mu}_{X_1}, \hat{\mu}_{X_2}, \hat{\sigma} | \mathbf{X_1}, \mathbf{X_2})}, \tag{3.2}$$
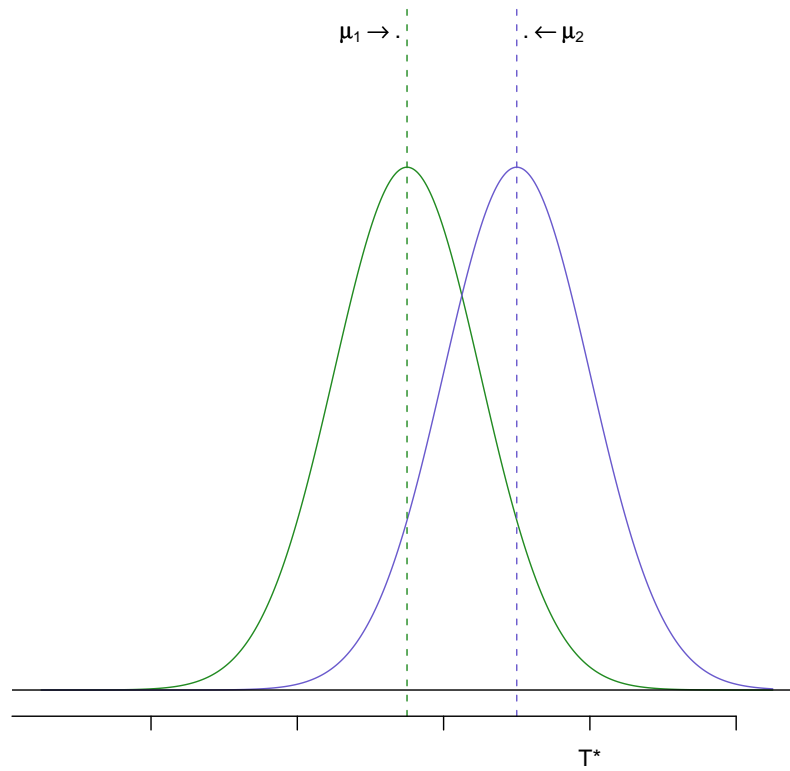
Figure 3.1: Graph of two normally distributed populations with the same σ^2 but different values of μ.
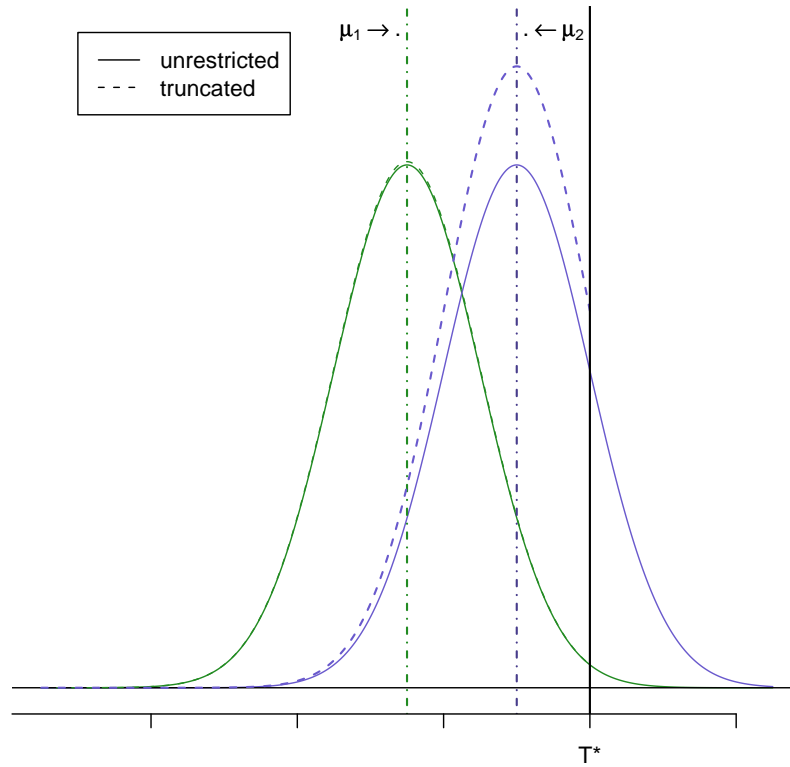
Figure 3.2: Graph of two normally distributed populations with the same σ^2 but different values of μ. Overlayed is a point of truncation T^* and how the densities for the two populations change to reflect the truncation.

where Θ_o is the parameter space, assuming the null hypothesis is true, and Θ is the unconstrained parameter space. The numerator is the supremum, or maximum value, of the likelihood under the null hypothesis, and the demoninator is the supremum of the likelihood under the alternative hypothesis. Computationally, the MLEs are found under the contrained and unconstrained space, and then the likelihood is evaluated at these estimates. The numerator is the ML under the hypothesis $H_0 : \mu_1 = \mu_2$ and is found treating the data as if they were sampled from one population. The denominator is the ML with no constraint on the μ_is and is found treating the data as if they were sampled from two different populations sharing the same σ.

The LRT is a test statistic, but the distribution of the test statistc when H_0 is true must be known. Instead of the ratio in (3.2), consider the test statistic, $-2\log(\lambda(\mathbf{x}, \mathbf{y}))$, which converges asymptotically to the χ^2 distribtion with 1 degree of freedom as the n increases. This test statistic can be found by taking the log of the ratio of likelihoods and multiplying by -2. This test statistics is compared to a χ_1^2 distribution to obtain the p-value. The test statistic of interest is $-2\log(\lambda(\mathbf{X_1}, \mathbf{X_2}))$, because the distribution is known when H_0 is true. When referencing the LRT in the document, this is the test statistic used.

The numerator in (3.2) is the ML under the constraint $\mathbf{X_{11}}, \cdots, \mathbf{X_{1n_1}} \sim TN(\mu, \sigma, a, b)$ and $\mathbf{X_{21}}, \cdots, \mathbf{X_{2n_2}} \sim TN(\mu, \sigma, a, b)$ are two random samples from the same truncated normal distribution. The numerator ML is found as described in Section 2.1. The denominator in (3.2) is the ML in the unconstrained space. This is an extension of the single population likelihood to a two population likelihood. The log-likelihood for the two population is

$$\ell = -n_1 \log(\psi(\mu_1, \sigma)) - \frac{n_1}{2} \log(2\pi) - n_1 \log(\sigma) - \frac{\sum(X_{1j} - \mu_1)^2}{2\sigma}$$
$$- n_2 \log(\psi(\mu_2, \sigma)) - \frac{n_2}{2} \log(2\pi) - n_2 \log(\sigma) - \frac{\sum(X_{2j} - \mu_2)^2}{2\sigma}, \qquad (3.3)$$

22

and the gradient for (3.3) is

$$
\mathbf{G} = \begin{bmatrix} \mu_1 - \bar{X}_1 + \frac{\psi'_\mu(\mu_1,\sigma)}{\psi(\mu_1,\sigma)} \\ \mu_2 - \bar{X}_2 + \frac{\psi'_\mu(\mu_2,\sigma)}{\psi(\mu_2,\sigma)} \\ \sigma - \frac{\sum(X_{1j}-\mu_1)^2}{n_1} - \frac{\sum(X_{2j}-\mu_2)^2}{n_2} + \frac{\psi'_\sigma(\mu_1,\sigma)}{\psi(\mu_1,\sigma)} + \frac{\psi'_\sigma(\mu_2,\sigma)}{\psi(\mu_2,\sigma)} \end{bmatrix}. \tag{3.4}
$$

Like the single population MLEs, the two population MLEs must be found using an iterative approach. The Newton-Raphson algorithm optimizes (3.3) utilizing (3.4) so a high number of the numerical approximations can be avoided.

The LRT of interest is written as

$$
-2\log(\lambda(\mathbf{X_1}, \mathbf{X_2})) = -2(\ell(\tilde{\mu}, \tilde{\sigma}) - \ell(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma})), \tag{3.5}
$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ are the MLEs under the null hypothesis and $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\sigma}$ are the MLEs in the unconstrained space.

3.1 POWER STUDY

The power of the two-sample LRT for $H_0 : \mu_1 = \mu_2$ was studied using simulation. The purpose of the study was to determine the power the LRT in detecting differences of $\Delta = \frac{|\mu_1-\mu_2|}{\sigma}$. The simulation studied truncated normal distributions with bounds $a = 0$ and $b = 100$ and a common $\sigma = 10$. Without the loss of generality, $\mu = 80$. Various Δs were selected to explore and approximate the power curve.

At each value of Δ selected, a random sample was taken from a truncated normal (μ, σ, a, b), an independent sample was taken from a truncated normal $(\mu + \Delta\sigma, \sigma, a, b)$. The LRT was calculated and compared to a critical value of χ^2 with 1 degree of freedom. The study uses $\alpha = 0.05$ so the critical value is $\chi^2_{1,\alpha} = 3.8415$. In the study 10,000 simulations were performed at each Δ.

Power study of LRT with $n_1 = n_2$ and small sample sizes

The first simulation study looked at various $\Delta = \frac{|\mu_1-\mu_2|}{\sigma}$ values to determine what the power curve was for $n_1 = n_2 = \{15, 30, 100\}$. In this study Δ represents absolute standardized

differences between μ_1 and μ_2. Figure 3.3 shows the results of the simulation study. The study set $\mu_1 = 80$ and $\sigma = 10$ with μ_2 shifting by Δ. The values of Δ were selected to provide a reasonable approximation of the power curve. Nine values of Δ were between 0 and 1 standardized differences, and four more values were selected between 1 and 2 standardized differences. The power when $\Delta = 0$ is α when n is sufficently large, but the LRT for small samples has P(Type I error) slightly larger than $\alpha = 0.05$ since α is about 0.06 for the three sample sizes. Table B.2 contains the results of the simulation. For samples of $n_1 = n_2 = 100$, 90% power is achieved at about $\Delta = 0.5$. For samples of $n_1 = n_2 = 30$, 90% power is achieved at about $\Delta = 0.875$. For samples of $n_1 = n_2 = 15$, 90% power is achieved at about $\Delta = 1.25$.

Power study of LRT with $n_1 \neq n_2$ and large sample sizes

The second simulation study examined the power for detecting the absolute difference, $\Delta = |\mu_1 - \mu_2|$, when $m \neq n$ and n_1 and n_2 are both large. In the study $n_1 = 1000$ and $n_2 = 250$. The first simulation study provided when $n_1 = n_2 = 100$, there was 90% power when the absolute difference was greater than a half a standard deviation. The focus of this simulation study was on values of $\Delta \leq 4$. The study used set $\mu_1 = 80$ and $\sigma = 10$ and let μ_2 represent a shift of Δ from μ_1. Figure 3.4 shows the power curve from the simulation study. In this study with the larger sample sizes, the power does equal α with $\Delta = 0$. The simulation study indicates that to detect differences between μ_1 and μ_2 with 90% power the observed difference in the MLEs must be at about 2.5.

Power study of LRT with $n_1 = n_2$ and small sample sizes

The last simulation study performed examined the power curve for detecting the absolute difference between μ_1 and μ_2 for small sample sizes $n_1 = n_2 = \{15, 30\}$.

Power Curve

Figure 3.3: The power curve of the LRT testing the hypothesis of $H_0 : \mu_1 = \mu_2$, where the sample sizes are the sample. The sample size considered were $n_1 = n_2 = \{15, 30, 100\}$. The detectable difference is measured on absolute standardized differences.
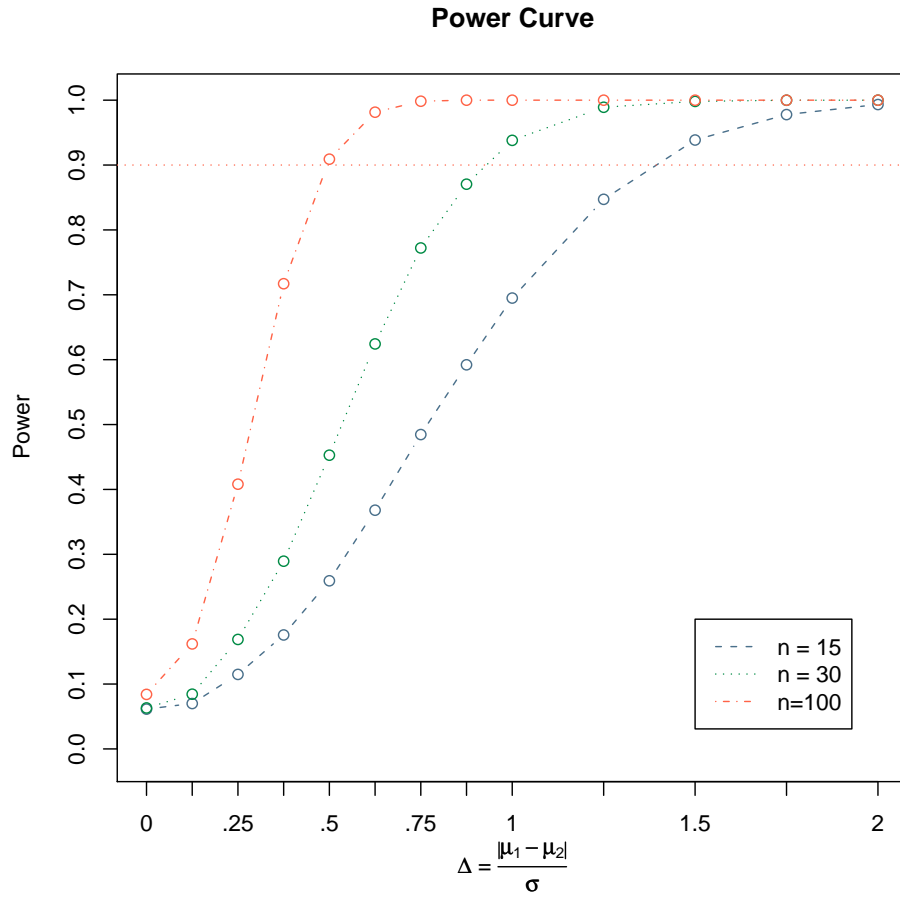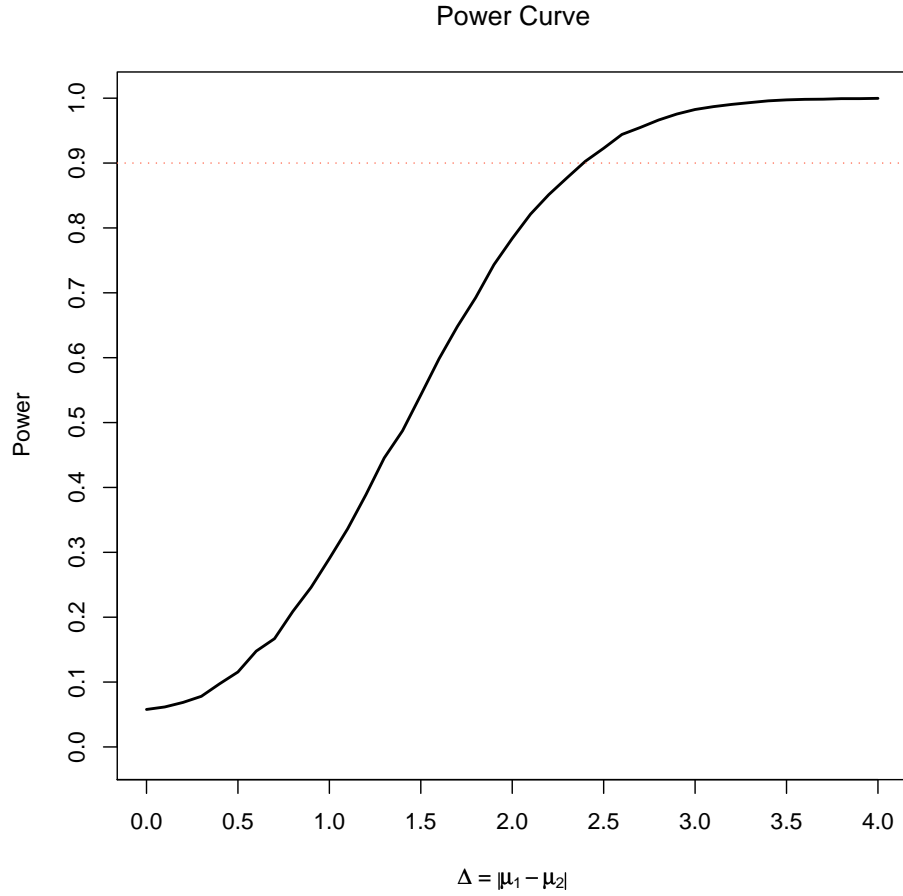
Power Curve

Figure 3.4: The power curve of the LRT testing the hypothesis of $H_0 : \mu_1 = \mu_2$, where the samples sizes are different. This curve considers $n_1 \neq n_2$, where $n_1 = 1000$ and $n_2 = 250$. The detectable difference is measured in $|\mu_1 - \mu_2|$.

Figure 3.5 shows the power curves. The study set $\mu_1 = 80$ and $\sigma = 10$ with $\mu_2 = \mu_1 + \Delta$. The simulation study shows with $n_1 = n_2 = 15$, 90% power is achieved at about $\Delta = 13$, and with $n_1 = n_2 = 30$, 90% power is achieved at about $\Delta = 9$. These results indicate that in order to detect differences between μ_1 and μ_2 with high power when $n_1 = n_2 = \{15, 30\}$, the absolute difference $|\mu_1 - \mu_2|$ must be at least 10. This results parallels the findings seen on the absolute standardized differences that showed one standard deviation was needed to have 90% power.

3.2 k POPULATION LRT

Let $X_{i1}, \ldots, X_{in_i}$ be independent random samples from truncated normal distributions (μ_i, σ, a,b), where $i = 1, \cdots, k$, the n_is are the sample sizes, and the k populations share the same truncation points, (a, b), and variance. The k population LRT tests the hypothesis $H_o : \mu_1 = \cdots = \mu_k$ versus H_a : at least one of the μ_is is different. The k population likelihood ratio is

$$\lambda(\mathbf{x_1}, \cdots, \mathbf{x_k}) = \frac{\sup_{\Theta_0} L(\tilde{\mu}, \tilde{\sigma} | \mathbf{x_1}, \cdots, \mathbf{x_k})}{\sup_{\Theta} L(\hat{\mu}_1, \cdots, \hat{\mu}_k, \hat{\sigma} | \mathbf{x_1}, \cdots, \mathbf{x_k})}, \tag{3.6}$$

where Θ_o is the parameter space, assuming the null hypothesis, and Θ is the unconstrained parameter space. The numerator is the supremum, or maximum value, of the likelihood under the null hypothesis, and the denominator is the supremum of the likelihood in the unrestricted space. Computationally, the MLEs are found under the contrained and uncon-strained space, and then the likelihood is evaluated at the estimates. The numerator is the ML under the hypothesis $H_o : \mu_1 = \cdots = \mu_k$ and is found treating the data as if they were sampled from one population. The denominator is the ML in the unrestricted space and is found treating the data as if they were sampled from k different populations sharing the same σ.

The numerator is found using the ML approach implemented for the single population found in Section 2.1. The ML in the unrestriced space is an extension to a k population

27

likelihood. The log-likelihood for the k population is

$$\ell = -\sum_{i=1}^{k} \left(n_i \log(\psi(\mu_i, \sigma)) + \frac{n_i}{2} \log(2\pi) + n_i \log(\sigma) + \frac{\sum_{j=1}^{n_i}(X_{ij} - \mu_i)^2}{2\sigma} \right), \qquad (3.7)$$

and the gradient for (3.7) is

$$\mathbf{G} = \begin{bmatrix} \mu_1 - \bar{X}_1 + \frac{\psi'_\mu(\mu_1, \sigma)}{\psi(\mu_1, \sigma)} \\ \mu_2 - \bar{X}_2 + \frac{\psi'_\mu(\mu_2, \sigma)}{\psi(\mu_2, \sigma)} \\ \vdots \\ \mu_k - \bar{X}_k + \frac{\psi'_\mu(\mu_k, \sigma)}{\psi(\mu_k, \sigma)} \\ \sigma - \sum_{i=1}^{k} \left(\frac{\sum_{j=1}^{n_i}(X_{ij} - \mu_i)^2}{n_i} + \frac{\psi'_\sigma(\mu_i, \sigma)}{\psi(\mu_i, \sigma)} \right) \end{bmatrix}. \qquad (3.8)$$

Similar to the single population MLE, the k population MLE has no closed-form solution and must be found using an iterative approach. The same considerations are made for the k population case. The Newton-Raphson algorithm optimizes (3.7) utilizing (3.8), so some numerical approximations can be avoided.

The LRT of interest is written as

$$-2\log(\lambda(\mathbf{X_1}, \cdots, \mathbf{X_k})) = -2(\ell(\tilde{\mu}, \tilde{\sigma}) - \ell(\hat{\mu}_1, \cdots, \hat{\mu}_k, \hat{\sigma})), \qquad (3.9)$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ are the MLEs under the null hypothesis and $\hat{\mu}_1, \cdots, \hat{\mu}_k$ and $\hat{\sigma}$ are the unrestricted MLEs. The LRT converges asymptotically to the χ^2 distribtion with $k-1$ degree of freedom as the n_is increase. This test statistic is compared to a χ^2_{k-1} distribution to obtain the p-value. The degrees of freedom for the χ^2 distribution are found by determining the difference in the number of parameters estimated. Under the null hypotheses only two parameters are estimated, and in the unrestricted case, $k+1$ parameters are estimated. The difference of $k-1$ represents the degrees of freedom available for the test.

Power Study

A power study for the $k = 5$ population LRT was performed similarly to the two population cases studied. The purpose of the study was to determine how much power the LRT had

to detect differences in $\Delta = \frac{|\mu_i - \mu_{(i)}|}{\sigma}$. The simulation studied truncated normal distributions with bounds $a = 0$ and $b = 100$ and a common $\sigma = 10$ were used. The μ parameter was set to be 80. The same choices of Δ used to find the power curve for the $m = n$ in Figure 3.3 were repeated in this study.

For each value of Δ selected, four independent random samples were taken from a truncated normal (μ, σ, a, b) and a fifth independent sample was taken from a truncated normal $(\mu + \Delta\sigma, \sigma, a, b)$. The LRT was calculated for the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ and compared to a critcal value of χ^2 with 4 degrees of freedom. The study used an $\alpha = 0.05$ and the critical value from the $\chi_4^2 = 9.4877$. Finding the LRT at each Δ was done for $10,000$ simulations. Figure 3.6 shows the power curve from the simulation study. The power at $\Delta = 0$ is similar to the values found in the two population case. From the curve, 90% power is achieved at about $\Delta = 0.875$.
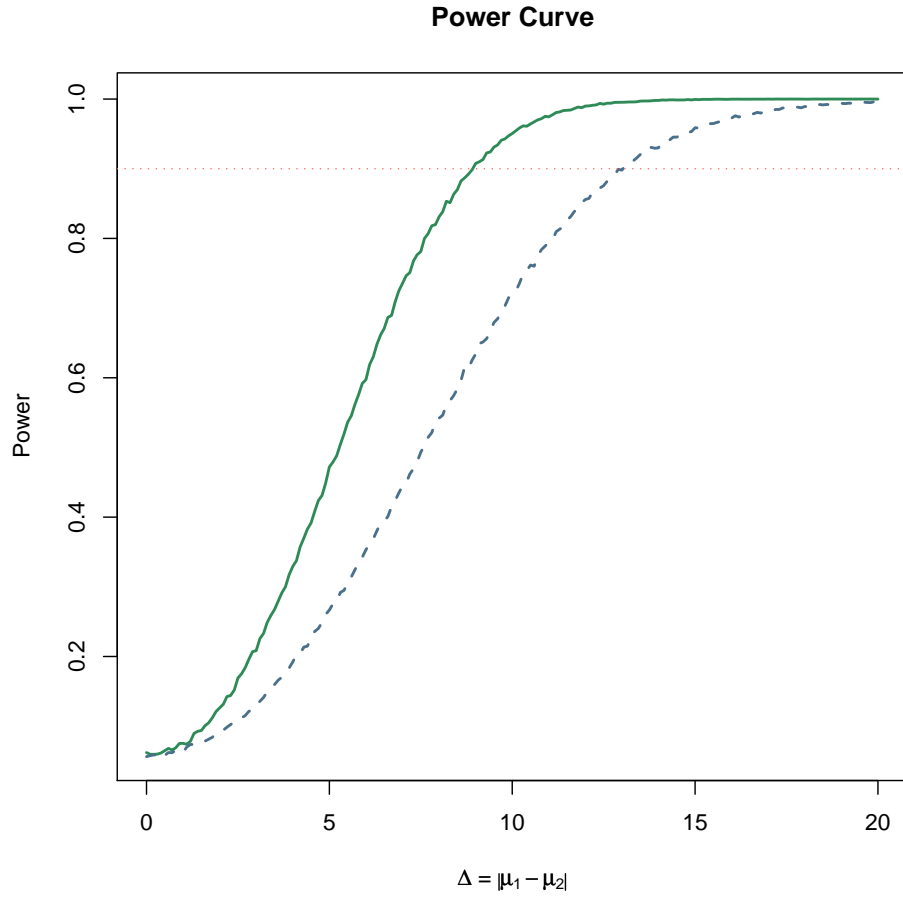
Power Curve

Figure 3.5: The power curve of the LRT testing the hypothesis of $H_0 : \mu_1 = \mu_2$, where the samples sizes are equal ($n_1 = n_2 = \{15, 30\}$). The Δ is measured in absolute difference of $\mu_1 - \mu_2$.
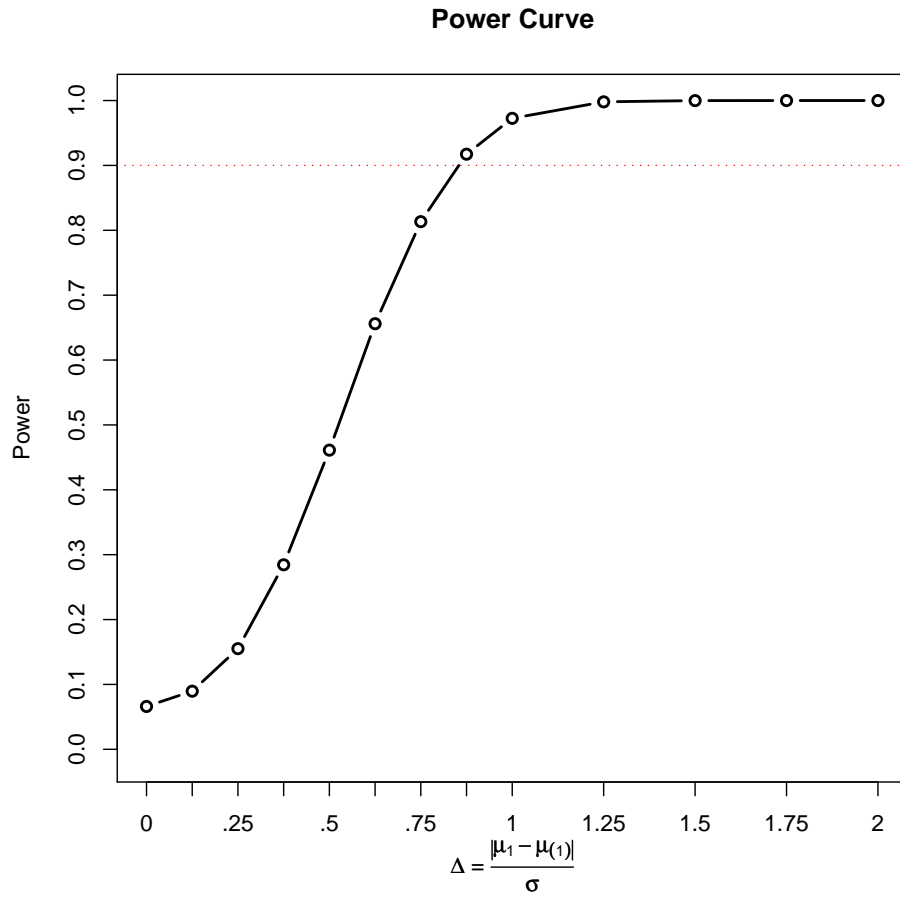
Power Curve

Figure 3.6: The power curve of the LRT testing the hypothesis of $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$, where the samples sizes are all 30. The detectable difference is measured in $\Delta = \frac{|\mu_i - \mu_{(i)}|}{\sigma}$.

_____

APPLICATION

BYU's Department of Statistics offers different introductory statistics courses. The largest of these courses is Stat 221, which is taught to about 4,500 students each year. The course is available through traditional in-class lecture, online, or independent study. Lectures are available both during the day and evening and are taught at the main campus in Provo, Utah, and through the Salt Lake Center Annex in Salt Lake City, Utah.

The learning outcomes for Stat 221 are to

1. Understand the importance of how data is collected and how data collection dictates the appropriate statistical method,

2. Understand and communicate using technical language about probability and variation, and

3. Interpret and communicate the outcomes of estimation and hypothesis tests in the context of a problem.

These outcomes are measured through the completion of homework assignments and test questions. Questions are designed to have students work through simple statistical problems and draw correct conclusions. The steps to solve a statistical problem and the correct way to report the findings are presented to the students in both the lecture as well as in the optional weekly lab.

The learning outcomes are measured by the final grade in the course. To remove potential confounding due to differences in homework assignments and tests, every section of the Stat 221 courses share the same homework assignments and tests. This common set

of homework and tests across the various sections allows for easier comparisons between the different course types and the different instructors.

There are two types of optional labs for the students to attend each week. The first type is a scheduled lab in which a teaching assistant (TA) will review the concepts taught that week in lecture and how to use computer software to do the statistical procedures learned. This type of lab is scheduled to have 40 people learning what to do for each homework assignment. The second type is an open lab in which one or two TAs are available in a room to help with particular questions that a student may have. This type of lab provides more one-on-one time between students and TAs, but the lab is designed to have TAs answer simple questions or problems before moving on to other students.

The final grade for the course is composed of the weighted average of homework, midterms, and final exam. There are additional in-class quizzes and other assignments that add extra credit to the final grade. A typical homework assignment is done online and is composed of multiple choice and free response questions covering the basic concepts taught. Each assignment is about 20 questions long and is due on Fridays every week. There are 13 assignments that equate to 15% of the final grade. There are three midterms given, averaging about 50 questions each and covering the material taught. Together these midterms account for 55% of the final grade. A comprehensive final is given and is about 100 questions in length. The final is the remaining 30% of the grade. The four tests are all given in BYU's Testing Center. Each test is available for five days, providing ample opportunity for students to take it as they have time.

Figure 4.1 shows a histogram of all of the final grades from the Fall 2009 semester of Stat 221. The vertical red line indicates the mean of all of the grades, which is pulled down into the tail. The blue lines show the distribution if it were from a truncated normal, with the blue vertical line showing the mode of the distribution of grades. Given how well the histogram follows the blue curve, it is reasonable to think that the grades can be modeled by a truncated normal distribution.

Not only does the range of grades provide truncation to the distribution, other factors might affect the truncation for the Stat 221 grades. Not all of the students at BYU are required to take Stat 221. Some students take AP Statistics or AP Calculus in high school, which allows them to not take Stat 221. The students who do not take Stat 221 because they already had Statistics in high school would be expected to perform well in the course and be in the upper tail of the distribution. Depending on their majors at BYU, they might end up in other statistics courses geared more towards their major, or more advanced courses in the Statistics Department.

4.1 Results

The final grades from the Fall 2009 semester were analyzed using the truncated normal distribution. The MLEs were calculated and the LRTs were performed for two comparisons of sections. The first comparison was the performance of student learning between the morning and afternoon lecture sections on the main BYU campus. The second comparison was between the $k = 5$ sections of Stat 221 taught at the Salt Lake Center.

Prior to comparing the sections, the data were filtered to remove incomplete observations. The final grades of students who took all four exams were the only ones considered. The filtered grades represents students that began Stat 221 but failed to complete the course. Initially there were 2,007 grades, but 133 were removed for not having scores for all four exams, leaving 1,874 students in the data.

Morning and Afternoon Stat 221 sections

Figure 4.2 depicts the MLEs for the Fall 2009 semester grades of Stat 221 from the morning and afternoon lecture sections. The morning lecture sections of Stat 221 had 1,017 students, and the afternoon lecture sections had 425 students. The MLEs for the two groups of sections were found, with the morning having $\hat{\mu}_1 = 92.5$ and the afternoon having $\hat{\mu}_2 = 91.45$ with $\hat{\sigma} = 13.3$. Based on the model, both of the populations had a significant amount of

truncation.The morning lecture grades had 28.6% truncation and the afternoon section had 26% truncation. The hypothesis test for $H_o : \mu_1 = \mu_2$ was done using the LRT. The test statistic was found to be 4.1221 and a p-value of 0.0423. From the test statistic and p-value, the conclusion would be that the morning lecture sections performed better than the afternoon lecture sections. Figure 4.3 shows the power curve for the unequal sample sizes with the dash blue line representing the difference $|\mu_1 - \mu_2| = 1.05$ seen from the data. The test rejects the null hypothesis that the two populations have the same μ, and the power curve indicates a low power of about 31%. The low power is significant because the weak test was able to detect a difference. The dotted red line indicates 90% power. Based on the power curve, in order to have 90% power the difference $|\mu_1 - \mu_2|$ would need to be near 2.4. The two-sample t-test is robust to the violation on the assumption of normality, but the t-test was done on the sample means, which are in the tails. Comparing the results to a two-sample t-test, the t statistic is 1.7996 and the p-value is 0.0363, which leads to the same conclusion as the LRT. The t-test seems reasonable to use because the sample sizes are large enough that an appeal to the Central Limit Theorem can be made. The mean of the morning sections was $\bar{x}_1 = 84.98$; the mean of the afternoon sections was $\bar{x}_2 = 83.70$, and the pooled standard deviation was $s_p^2 = 12.31$. Both means were adjusted down into the lower tail of the distribution based on the percent truncation. Reporting the mean of the samples would place the average grade a full letter grade lower than the MLE mean grade for the population. Like the means, the variance was smaller than the MLE due to the adjustment made for percent truncation.

Histogram of Stat 221 Grades
Fall 2009
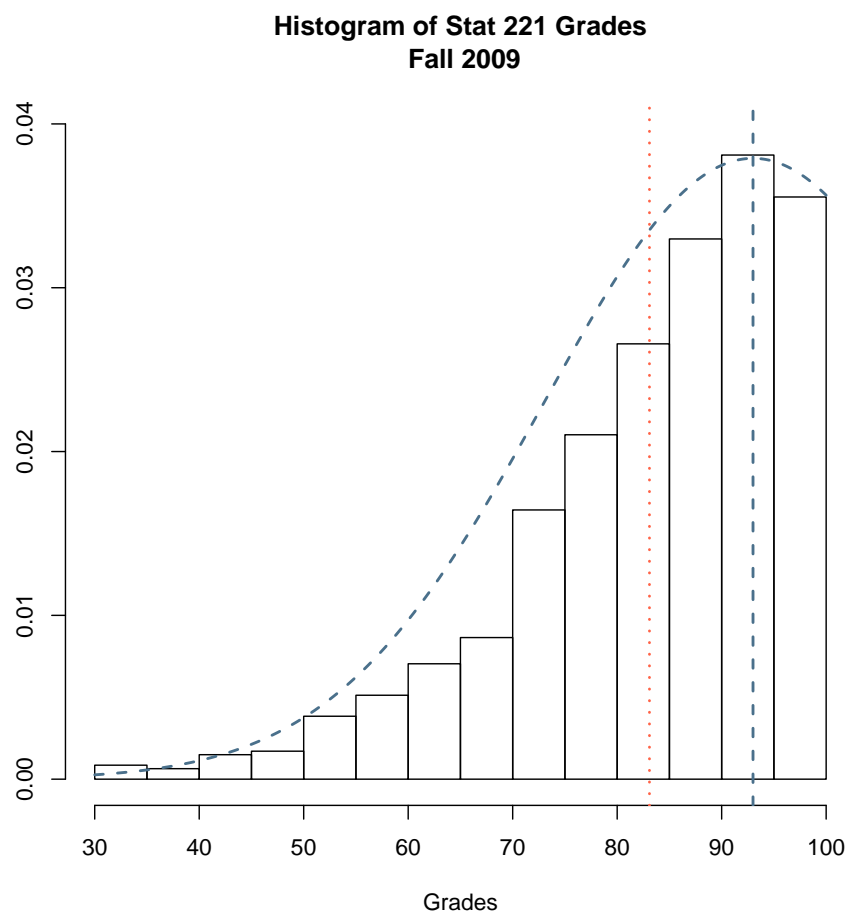
Figure 4.1: Histogram of all grades from Stat 221, Fall 2009. The red line denotes the mean of all of the grades, while the blue line indicates the mode of the grades. The curve represents the distribution if the grades came from a truncated normal. The MLE of the complete data ($\hat{\mu} = 92.5$ and $\hat{\sigma} = 16$) were used to draw the curve.

Figure 4.2: Graph of the densities of the morning (green) and afternoon (blue) lecture grades from the Fall 2009 semester of Stat 221. The dashed line represents the MLE $\hat{\mu}$ for each population. The MLE for the morning is $\hat{\mu}_1 = 92.5$ and for the afternoon is $\hat{\mu}_2 = 91.45$ with $\hat{\sigma} = 13.3$. The dotted curves are the *original* distribution as if there had been no truncation.

37
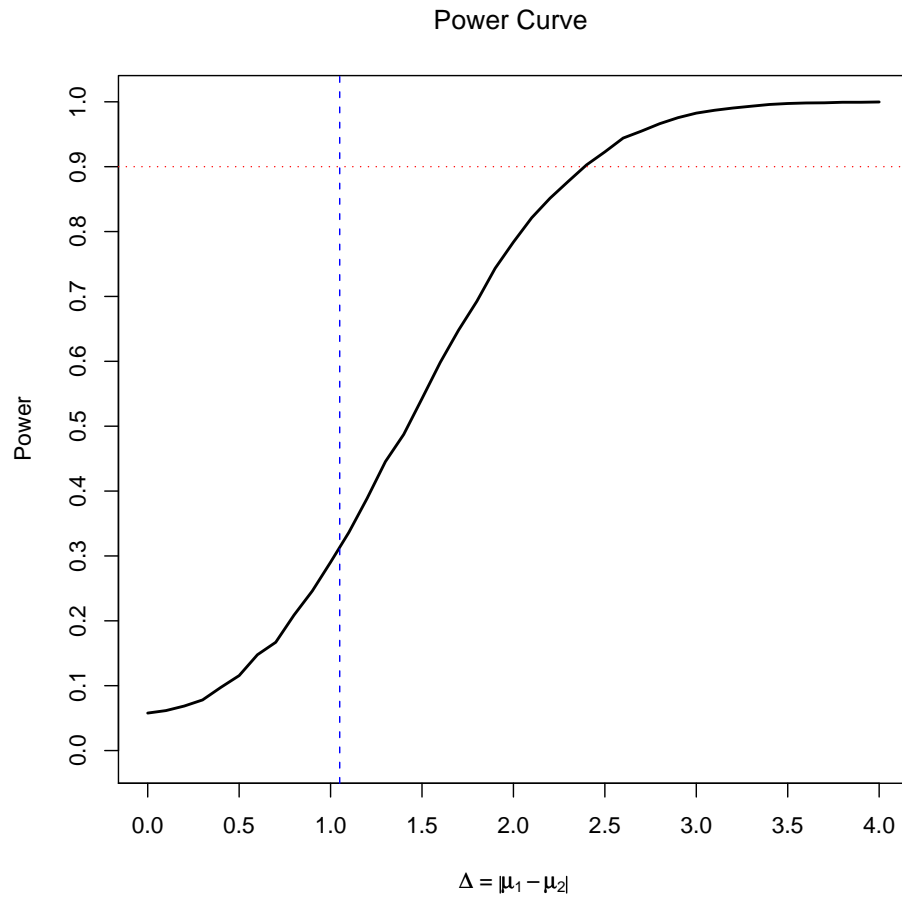
Power Curve

Figure 4.3: Power curve for the $m \neq n$ and large sample sizes testing the hypothesis $H_0 :$ $\mu_1 = \mu_2$. The dotted red line represents 90% power, which is achieved when $|\mu_1 - \mu_2| = 2.4$. The dashed blue line is the difference $|\mu_1 - \mu_2| = 1.05$ from the Fall 2009 semester grades of Stat 221 for the morning versus the afternoon lecture sections.

The five sections of Stat 221 at the Salt Lake Center were compared using the k population LRT. The sections averaged 20 students. The MLEs were found using the k population likelihood and seen in Table 4.1. The percent truncation for each of these populations was small. The two populations with $\hat{\mu}$ near 70 had about 1.7% truncation, the two populations with $\hat{\mu}$ near 76 had about 4.2% truncation, and the population with $\hat{\mu}$ near 80 had about 7% truncation.

Table 4.1: Table of the MLEs for the Salt Lake Center's Stat 221 sections for testing the hypothesis $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$.

Instructor	n_i	k Population $\hat{\mu}$	Same Population $\tilde{\mu}$	$\bar{x}_i$
A	13	70.26		70.15
B	11	80.11		79.11
C	49	76.49	75.82	76.39
D	9	72.57		72.59
E	25	76.85		76.73
	$\hat{\sigma} = 13.5$	$\tilde{\sigma} = 14.02$		

The hypothesis test for $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ was done using the LRT. The test statistic was 5.5017 with p-value of 0.2396 when compared to a χ^2_4 distribution. This test fails to reject the hypothesis, so the conclusion is that there is no difference between the five Salt Lake Center sections of Stat 221. Comparing these results to an one-way ANOVA, the ANOVA produces an F statistics of 1.0938 and p-value of 0.3637. This results in the same conclusion. The low percent truncation in each of the five populations contributes to the robustness of the ANOVA procedure. Overall the ANOVA procedure seems to be robust to violations, but like the t-test it relies on the sample moments. In this case there was only a small difference between the MLE and sample moments. Figure 4.5 shows the power curve with the dotted red line representing 90% power, which is achieved at $\Delta = 0.8$. The dashed blue line shows $\Delta = 0.5$, as in the data. The test had power of about 45%.
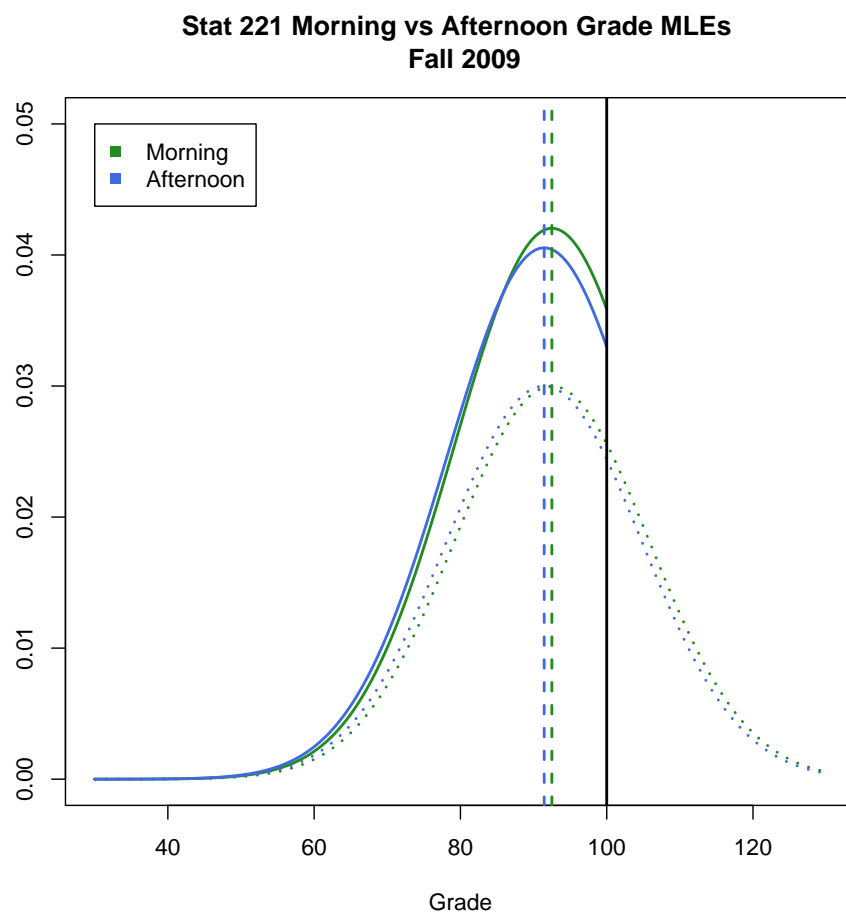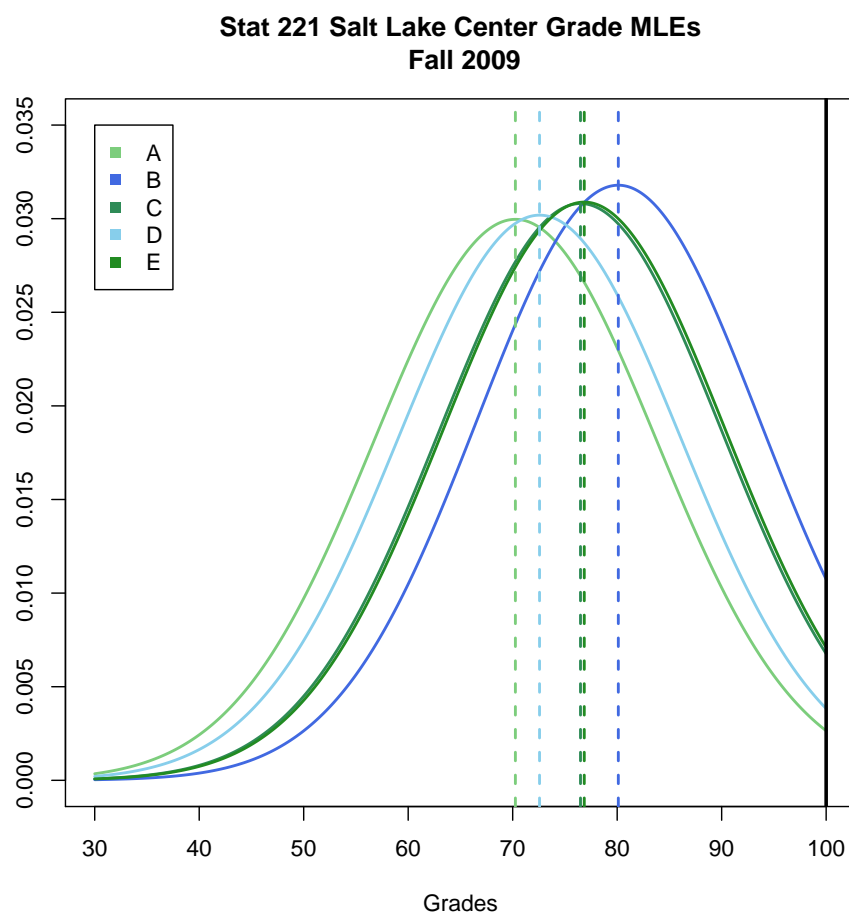
Figure 4.4: Graph of the densities of the five Salt Lake Center lecture grades from the Fall 2009 semester of Stat 221. The dashed line represents the MLE $\hat{\mu}$ for each population. The MLE for $\hat{\mu}$ and $\hat{\sigma}$ can be found in Table 4.1.
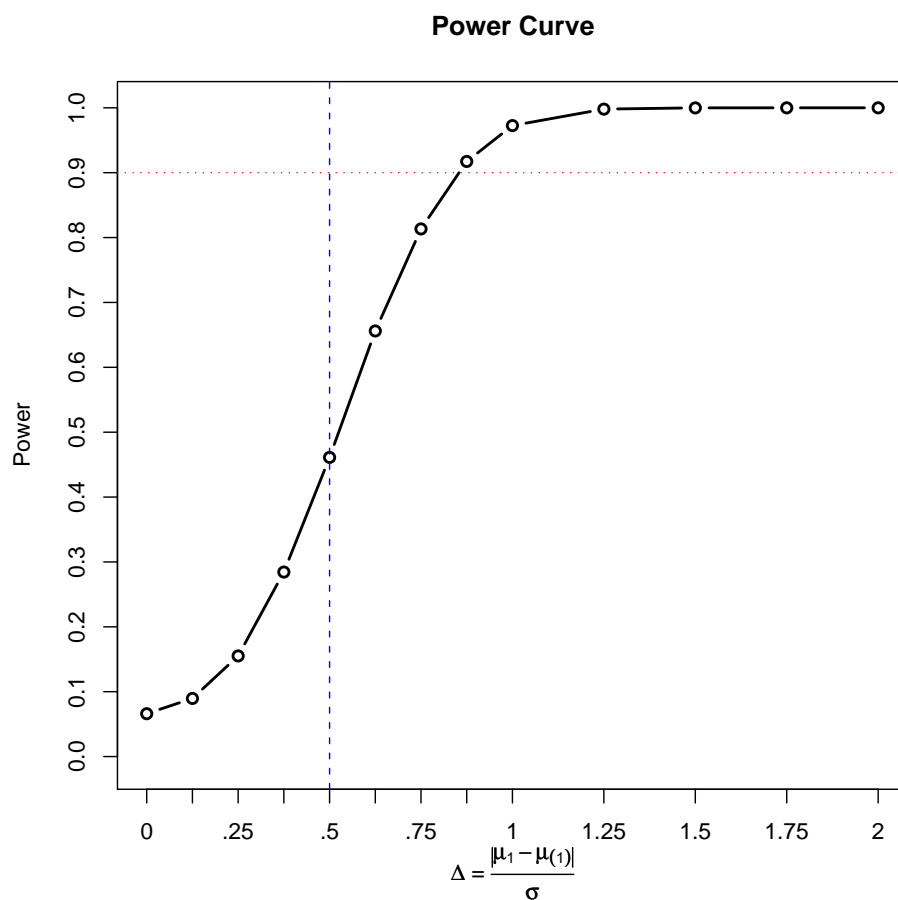
Power Curve

Figure 4.5: Power curve for the $n_1 = n_2 = n_3 = n_4 = n_5 = 30$ sample sizes testing the hypothesis $H_0 : \mu_1 = \cdots = \mu_k$. The dotted red line represents 90% power, which is achieved when $|\mu_i - \mu_{(i)}| = 0.875$. The dashed blue line is the difference $|\mu_i - \mu_{(i)}| = 0.5$ from the Fall 2009 semester grades of Stat 221 for the Salt Lake Center lecture sections.

This paper presented truncated distributions and how truncation can be applied to the normal distributions. The expected value and variance of the truncated normal distribution were presented in (1.3) and (1.4). Parameter estimation of the truncated normal distribution was presented, focusing on understanding the original population and not the truncated population actually sampled. The maximum likelihood estimators (MLE) and method of moment (MoM) estimators were derived in Sections 2.1 and 2.2. For the MLE, the careful examination of the log-likelihood and its derivatives was discussed. The gradient and the Hessian of the log-likelihood were derived and shown in (2.5) and (2.6). Newton-Raphson (NR) optimization was discussed to find the MLE with the code presented. A modified MoM estimator was presented that relied on standardizing the sample data before finding parameter estimates. The NR optimization was implemented for the MoM estimators. A simulation study evaluating the performance of the MLE and MoM estimators was performed and the results were discussed in Section 2.3.

Hypothesis testing using the Likelihood Ratio Test (LRT) was discussed for the truncated normal distributions in Chapter 3. The level α LRT for testing the hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ was derived in (3.5) and code for evaluating the LRT was presented in Appendix B. Simulation studies examining the power of the LRT to detect differences in a two sample setting were performed and discussed in Section 3.1. The first simulation study evaluated the power of the LRT to detect the absolute standardized differences between the μ_is for small sample sizes ($n_1 = n_2 = \{15, 30, 100\}$). The LRT was evaluated at various $\Delta = \frac{|\mu_1 - \mu_2|}{\sigma}$ to approximate the power curve. At each Δ, 10,000 simulations were performed. The approximate power curve is shown in Figure 3.3. The second

simulation study evaluated the power of the LRT to detect the absolute difference between the μ_is for large unequal sample sizes ($n_1 = 1000$ and $n_2 = 250$). The approximate power curve is shown in Figure 3.4. The final simulation study evaluated the power of the LRT for detecting the absolute difference for small equal sized samples ($n_1 = n_2 = \{15, 30\}$). The approximate power curve is shown in Figure 3.5. The level α LRT for testing the hypothesis $H_0 : \mu_1 = \cdots = \mu_k$ was derived in Section 3.2 and the code for evaluating this LRT was presented in Appendix B.1. A simulation study was performed evaluating the power of this LRT to detect absolute standardized differences. Various $\Delta = \frac{|\mu_i - \mu_{(i)}|}{\sigma}$ were selected to approximate the power curve. The approximate power curve is shown in Figure 3.6.

The Fall 2009 semester grades of Stat 221 were analyzed. The morning and afternoon lecture sections were compared. The hypothesis test $H_o : \mu_1 = \mu_2$ was done for these sections and a significant difference was found. When comparing the difference to the power curve, the results had about 45% power. This significance could simply be due to the large sample sizes of the morning and afternoon lecture sections. The significance of the test leads to the question of whether the difference is practically significant. The five different Salt Lake Center sections were compared using the k population LRT and the hypothesis $H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ was tested. The results indicated that there was no difference in the five sections. The failure to reject raises concern given the low power of the test.

There are other areas of interest that can be studied with the truncated normal population. This paper has introduced the two population and the k population LRT for hypothesis testing from truncated normal distribution with the same σ. An extension of this research would focus on the LRT when the populations have unequal variance. Another extension of the truncated normal is to evaluate the performance of the t test when the data are from a truncated normal distribution. It was seen that similar conclusions were reached for the Stat 221 data, but the robustness of the t test should be evaluated to see how well it does with large departures from normality.

Another extension of the truncated normal distribution is examining the application of the truncated normal distribution in the regression and linear models setting, or in other settings. The regression model utilizing the truncated normal distribution would be ideal for modeling a response variable bounded in such a way that it is impossible to measure responses beyond the boundary.

BIBLIOGRAPHY

Cohen, A. (1949), "On Estimating The Mean and Standard Deviation of Truncated Normal Distributions," *Journal of the American Statistical Association*, 45, 518–525.

—— (1950), "Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples," *The Annals of Mathematical Statistics*, 21, 557–569.

Fisher, R. (1930), "The Moments of the Distribution for Normal Samples of Measures of Departure from Normality," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 130, 16–28.

Halperin, M. (1952), "Estimation in the Truncated Normal Distribution," *Journal of the American Statistical Association*, 47, 457–465.

Harter, H. L., and Moore, A. H. (1966), "Iterative Maximum-likelihood Estimation of the Parameters of Normal Populations from Singly and Doubly Censored Samples," *Biometrika*, 53, 205–213.

Johnson, A., and Thomopoulos, N. (2002), "Characteristics and Tables of the Left-Truncated Normal Distribution," *Proceedings of Midwest Decision Sciences Institute*, 133–139.

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions, Vol. 1 (Wiley Series in Probability and Statistics)*, Wiley-Interscience.

Pearson, K., and Lee, A. (1908), "On the Generalised Probable Error in Multiple Normal Correlation," *Biometrika*, 6, 59–68.

Tiku, M. L. (1989), "Modified Maximum Likelihood Estimation," in *Encyclopedia of Statistical Sciences (9 vols. plus Supplement), Volume S*, eds. Banks, D. L., Read, C. B., and Kotz, S., Wiley-Interscience, pp. 98–103.

APPENDICES

PARAMETER ESTIMATION SIMULATION STUDY

A.1   MAXIMUM LIKELIHOOD CODE

```
library(msm) # used for a 'built in' rtnorm function.

#functions that will be integrated using integrate()

# Normal PDF. When intergrate it is the probability to be in
#the truncated region of the Normal Distribution
psi<-function(y,mu,sigma){
exp(-(y-mu)^2/(2*sigma^2))/(sigma*sqrt(2*pi))
}


# the derivate of psi wrt mu.  Used in the gradient.
psi.mu<-function(y,mu,sigma){
exp(-(y-mu)^2/(2*sigma^2)) * ((y-mu)/(sigma^3*sqrt(2*pi)))
}
# the derivate of psi wrt sigma.  Used in the gradient.
psi.sigma<-function(y,mu,sigma){
exp(-(y-mu)^2/(2*sigma^2)) *
        ( ((y-mu)^2)/(sigma^4*sqrt(2*pi)) - 1/(sigma^2*sqrt(2*pi)) )
}


# the second derivate of psi wrt mu.  Used in the Hessian.
psi2.mu<-function(y,mu,sigma){
exp(-(y - mu)^2/(2*sigma^2)) *
        ( ((y - mu)^2)/(sigma^5*sqrt(2*pi))-1/(sigma^3*sqrt(2*pi)))
}


# the second derivative of psi wrt sigma.  Used in the Hessian.
psi2.sigma<-function(y,mu,sigma){
exp(-(y-mu)^2/(2*sigma^2)) *
        ( (2)/(sigma^3*sqrt(2*pi)) - (5*(y-mu))/(sigma^5*sqrt(2*pi)) +
            ((y-mu)^4)/(sigma^7*sqrt(2*pi)))
}
# the second derivative of psi taking mu then sigma or sigma than mu.
# Used in the Hessian.
psi12.musig<-function(y,mu,sigma){
exp(-(y-mu)^2/(2*sigma^2)) *
        ( ((y-mu)^3)/(sigma^6*sqrt(2*pi)) - (3*(y-mu))/(sigma^4*sqrt(2*pi)))
}


# The Truncated Normal PDF.  p<- c(p[1],p[2]) where p[1] is the mu parameter
```

```
# and p[2] is the sigma parameter.  t<-c(t[1],t[2]) where t[1] is the lower
# truncation point and t[2] is the upper truncation point.  This is the first
# of the functions that uses the numerical integration function integrate().

# integrate() takes as parameters psi<-function to be integrated, then lower
# and upper bounds, and then the rest of the inputs for the function.  We are
# only interested in the value of the function.
tnorm.pdf<-function(p,t,data){
  exp(-(data-p[1])^2/(2*p[2]^2))/(sqrt(2*pi*p[2]^2) *
    (pnorm(max(t),p[1],p[2])-pnorm(min(t),p[1],p[2])))
}


# Gradient of the log-likelihood.  This is used in the function nlminb().
grad.tnorm<-function(p,t,data){
  n  <-length(data)
  g1 <- -n*(integrate(psi.mu,t[1],t[2],mu=p[1],sigma=p[2])$value) /
          (pnorm(max(t),p[1],p[2])-pnorm(min(t),p[1],p[2])) -
          (n*p[1]-sum(data))/p[2]^2
  g2 <- -n*(integrate(psi.sigma,t[1],t[2],mu=p[1],sigma=p[2])$value) /
          (pnorm(max(t),p[1],p[2])-pnorm(min(t),p[1],p[2])) -
          n/(2*p[2]^2) + (2*sum((data-p[1])^2))/(4*p[2]^4)
  out <- cbind(g1,g2)
  out
}


#Hessian of the log-likelihood.  This is used in the function nlmib().
hessian.tnorm<-function(p,t,data){
n<-length(data)
h1<- -n*(integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value *
                integrate(psi2.mu,t[1],t[2],mu=p[1],sigma=p[2])$value -
                integrate(psi.mu,t[1],t[2],mu=p[1],sigma=p[2])$value^2) /
                (integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value^2) -
             n/(p[2]^2)
h3<- -n*(integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value *
                integrate(psi12.musig,t[1],t[2],mu=p[1],sigma=p[2])$value -
                integrate(psi.mu,t[1],t[2],mu=p[1],sigma=p[2])$value *
                integrate(psi.sigma,t[1],t[2],mu=p[1],sigma=p[2])$value) /
                 (integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value^2) +
                (n*p[1]-sum(data))/(p[2]^4)
h2<- -n*(integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value *
                integrate(psi2.sigma,t[1],t[2],mu=p[1],sigma=p[2])$value -
                integrate(psi.sigma,t[1],t[2],mu=p[1],sigma=p[2])$value^2) /
                (integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value^2) +
                (2*n)/(4*p[2]^4)-(4*sum((data-p[1])^2))/(8*p[2]^6)
H<-matrix(0,nrow=2,ncol=2)
H[1,1]<-h1
H[2,2]<-h2
H[1,2]<-H[2,1]<-h3
H
}


# Negative Log Likelihood of the Truncated Normal PDF.
# Used as the objective function in nlminb().
```

```
ll.tnorm2<-function(p,t,data){
  t<-sort(t)
  n    <- length(data)
  out <- -n*log(pnorm(t[2],p[1],p[2])-pnorm(t[1],p[1],p[2])) -
    (n/2) * log(2*pi) -(n/2)*log(p[2]^2) -
      sum((data-p[1])^2)/(2*p[2]^2)
  -1*out
}


# Log-Likelihood.
ll.tnorm<-function(p,t,data){
  t<-sort(t)
  n    <- length(data)
  out <- -n*log(pnorm(max(t),p[1],p[2])-pnorm(min(t),p[1],p[2])) -
    (n/2) * log(2*pi) -(n/2)*log(p[2]^2) -
      sum((data-p[1])^2)/(2*p[2]^2)
  out
}


# This is repeated for each of the sample size and parameter combination
# For each iteration start off with a clean output.
out<-NULL
print("Simulation Study for mu=4, sigma=1");
for(i in 1:2500){
  # Determine the starting values of the Newton-Raphson optimization based on
  # the sample mean and std deviation. This is based on Cohen (1949,1950) and
  # Hap (1952) and what they suggested for starting values.
  p<-c(mean(samN1000m4s1[i,]),sd(samN1000m4s1[i,]))
  # Perform the NR optimization using nlminb() function
  nlminb(p, # parameters to be optimized (must be first input of objective)
         ll.tnorm2, # objective function (log-likelihood) to be maximized
         gradient=grad.tnorm,   # gradient of the objective function (if blank,
                                # defaults to Numerical Approximations.)
         hessian=hessian.tnorm, # hessian of the objective function (if blank,
                                # defaults to Numerical Approximations.)
         data=samN1000m4s1[i,], # rest of the the inputs for the objective.
         t=c(0,Inf),
         lower=c(-Inf,0), # Bounds on the parameters. Matches the order in p
         upper=c(Inf,Inf))->OUT
  if(i%%500==0) {print(i);}
  # Output for each dataset as a row of out,
  # The first two columns are parameters,
  # The next column is the number of iterations for that sample. Should be >1.
  # The last column is the sample number. (should go from 1 to 2500), but
  # represents samples that converged only.
  if(OUT$iter>1) out <- rbind(out,
                              cbind(t(OUT$par),     # MLE estimates
                                    OUT$iterations, # Iterations taken
                                    OUT$convergence, # convergence type
                                    i));            # sample number
}
# Save off to the appropriate object
outN1000m4s1<-out
```

A.2 Method of Moments Code

```
# The two functions that need to be solved.  Here the objective is the min of
# the absolute value of the two functions.  Given an optimizer, when this
# stops moving around, then there is convergence.
mme.tnorm<-function(param,data){
lamalph<- ( dnorm(-param[1]/sqrt(param[2]) ) /
                 (1 - pnorm(-param[1]/sqrt(param[2]))))
delalph<-lamalph*(lamalph-(-param[1])/sqrt(param[2]))
n<-length(data)

g1<-param[1] + sqrt(param[2])*lamalph-mean(data)
g2<-param[2]*(1- delalph) + param[1]^2 + 2*param[1]*param[2]*lamalph +
              param[2]^2*lamalph^2 - sum(data^2)/n
min(abs(g1),abs(g2))
}


#This is repeated for each of the sample sizes.
#For each iteration start off with a clean output.
out<-NULL
for(i in 1:2500){
  # Determine the starting values of the Newton-Raphson optimization based on
  # the sample mean and std deviation.  This is based on Cohen (1949,1950) and
  # Hap (1952) and what they suggested for starting values.
  p<-c(mean(samN15m4s1[i,]),sd(samN15m4s1[i,]))
  # Perform the NR optimization using nlminb() function
  nlminb(p,                       # Parameters to be optimized
                                  # (must be first input of objective)
        mme.tnorm,                # Objective function
        data=samN15m4s1[i,],   # Rest of the the inputs for the objective.
        lower=c(-Inf,0),          # Bounds on the parameters.
        upper=c(Inf,Inf))->OUT # Matches the order in p
  if(i%%500==0) {print(i);}
  #output for each dataset as a row of out,
  #  the first two columns are parameters,
  # the next column is the number of iterations for that sample.  Should be >1.
  # the last column is the sample number.  (should go from 1 to 2500), but
  # represents samples that converged only.
  out<-rbind(out,cbind(t(OUT$par),OUT$iterations,i));
}
#Save off to the appropriate object
out15<-out
```

The following are the tables of bias and mean square error (MSE) from the simulation study evaluating the performance of the MLE and MoM estimators. The tables record the results of the simulation study, and the figures for these tables are seen in Section 2.3.

Table A.1: Table of bias for μ for the different parameter combinations.

	Size	$\mu = 0$ $\sigma = 1$	$\mu = 1$ $\sigma = 1$	$\mu = 2$ $\sigma = 1$	$\mu = 2$ $\sigma = 2$	$\mu = 4$ $\sigma = 1$
MLE	15	-0.00913	-0.26895	-0.06684	-0.48727	-0.00563
	30	-0.00730	-0.08465	-0.01158	-0.17409	0.00255
	100	-0.00708	-0.01429	-0.00226	-0.05477	-0.00145
	250	-0.00665	-0.00929	0.00153	-0.00962	0.000142
	1000	-0.00494	-0.00296	-0.00141	-0.00141	-0.000090
MME	15	0.57759	0.13057	0.00008	0.51400	-0.00590
	30	0.54555	0.13069	0.00864	0.49227	0.00234
	100	0.53903	0.12819	0.01293	0.50528	-0.00159
	250	0.55339	0.12003	0.01259	0.50768	0.00005
	1000	0.56079	0.11178	0.00945	0.50302	-0.00071

Table A.2: Table of MSE for μ for the different parameter combinations.

	Size	$\mu = 0$ $\sigma = 1$	$\mu = 1$ $\sigma = 1$	$\mu = 2$ $\sigma = 1$	$\mu = 2$ $\sigma = 2$	$\mu = 4$ $\sigma = 1$
MLE	15	0.14475	9.72737	10.71898	14.61722	0.065329
	30	0.06857	0.24911	0.16298	1.29298	0.034171
	100	0.02272	0.03618	0.01184	0.16463	0.009449
	250	0.00731	0.01462	0.00469	0.05452	0.003967
	1000	0.00014	0.00317	0.00118	0.00118	0.000918
MME	15	0.37978	0.07537	0.07497	0.45775	0.06546
	30	0.32957	0.04463	0.03624	0.34366	0.03422
	100	0.30548	0.02324	0.01029	0.28532	0.00945
	250	0.31138	0.01716	0.00422	0.26939	0.00397
	1000	0.31497	0.01344	0.00111	0.25632	0.00099

Table A.3: Table of bias for σ for the different parameter combinations.

	Size	$\mu = 0$ $\sigma = 1$	$\mu = 1$ $\sigma = 1$	$\mu = 2$ $\sigma = 1$	$\mu = 2$ $\sigma = 2$	$\mu = 4$ $\sigma = 1$
	15	-0.00754	-0.02817	-0.02752	-0.01784	-0.03991
	30	-0.00318	-0.00899	-0.01607	-0.01347	-0.01622
MLE	100	-0.00534	-0.00695	-0.004356	0.00779	-0.00575
	250	-0.00561	0.00042	-0.00312	0.00264	-0.00276
	1000	-0.00242	0.00007	0.00012	-0.00135	-0.00096
	15	-0.54265	-0.26528	-0.07572	-0.44357	-0.02044
	30	-0.57193	-0.24456	-0.07417	-0.43702	-0.00718
MME	100	-0.57389	-0.22451	-0.06526	-0.42334	-0.00336
	250	-0.55825	-0.22434	-0.06030	-0.41967	-0.00190
	1000	-0.54718	-0.22473	-0.05828	-0.42116	-0.00082

Table A.4: Table of MSE for σ for the different parameter combinations.

	Size	$\mu = 0$ $\sigma = 1$	$\mu = 1$ $\sigma = 1$	$\mu = 2$ $\sigma = 1$	$\mu = 2$ $\sigma = 2$	$\mu = 4$ $\sigma = 1$
	15	0.00282	0.55071	0.08206	0.90511	0.040639
	30	0.00084	0.11957	0.03536	0.40114	0.01812
MLE	100	0.00014	0.01459	0.00723	0.06596	0.00524
	250	0.00001	0.00629	0.00304	0.02459	0.00208
	1000	< 0.00001	0.00147	0.00072	0.00634	0.00052
	15	0.31984	0.09253	0.03258	0.28557	0.03567
	30	0.34812	0.07172	0.01818	0.23284	0.01645
MME	100	0.34118	0.05293	0.00819	0.19180	0.00495
	250	0.31535	0.05144	0.00544	0.18097	0.00201
	1000	0.29971	0.05080	0.00382	0.17868	0.00052

LRT POWER STUDY

B.1 2 POPULATION LRT

Code

```
# Functions not present are used from MLE code.
# Log Likelihood of the 2 population Truncated Normal.
ll.tnorm.2pop<-function(p,t,d1,d2){
  m<-length(d1)
  n<-length(d2)
  out<- ( -m * log(pnorm(t[2],p[1],p[3]) - pnorm(t[1],p[1],p[3])) -
             (m/2)*log(2*pi) - m * log(p[3]) - sum((d1 - p[1])^2)/(2*p[3]) ) +
           ( -n * log(pnorm(t[2],p[2],p[3]) - pnorm(t[1],p[2],p[3])) -
             (n/2)*log(2*pi) - n * log(p[3]) - sum((d2 - p[2])^2)/(2*p[3]))
  -1*out
}


# Gradient for the LogLikelihood 2 pop
grad.tnorm.2pop<-function(p,t,d1,d2){
   m <- length(d1)
   n <- length(d2)
   g1<- p[1] - mean(d1) + integrate(psi.mu,t[1],t[2],mu=p[1],sigma=p[3])$value /
          integrate(psi,t[1],t[2],mu=p[1],sigma=p[3])$value
   g2<- p[2] - mean(d2) + integrate(psi.mu,t[1],t[2],mu=p[2],sigma=p[3])$value /
          integrate(psi,t[1],t[2],mu=p[2],sigma=p[3])$value
   g3<- p[3] - sum((d1-p[1])^2)/m -sum((d2-p[2])^2)/m +
          integrate(psi.sigma,t[1],t[2],mu=p[1],sigma=p[2])$value /
          integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value
   out<-cbind(g1,g2,g3)
   out
}


#truncation points for Simulation Study
trunc <- c(0,100)

# Simulation Study other things
del <- seq( ) # sequence of delta values for power study.
samMean <- 80 # mean of Application data.
samSD   <- 10 # standard deviation of Application data.

out<-NULL
for(i in 1:10000){
```

```
  # Create a random sample
  X<-rtnorm(n,samMean, samSD )
  Y<-rtnorm(n,samMean + del * samSD, samSD )

  # Starting values for the iterations

  # 2 Population Case
  p2<-c(mean(X),
        mean(Y),
         sd(c(X,Y)))

  # 1 Population Case
  p<-c(mean(c(X,Y)),
        sd(c(X,Y)))

  # Find the parameter combinations
  nlminb(p,
       lower=c(-Inf,0),upper=c(Inf,Inf),
       ll.tnorm2,
       gradient=grad.tnorm,
       hessian=hessian.tnorm,
       data=c(X,Y),
       t=trunc)->OUT1
  nlminb(p2,
       ll.tnorm.2pop,
       lower=c(-Inf,-Inf,0),upper=c(Inf,Inf,Inf),
       gradient=grad.tnorm.2pop,
       d1=X,
       d2=Y,
       t=trunc)->OUT2
  # Calculate the LRT.
  lrt.out <- -2 * (ll.tnorm(OUT1$par,trunc,c(X,Y) -
                    ll.tnorm(OUT2$par[c(1,3)],trunc,X) -
                    ll.tnorm(OUT2$par[c(2,3)],trunc,Y))
  out <- rbind( out , lrt.out )
}

lrt<-out

mean(lrt > qchisq(.95 , 1) )
```

Table and Figures of Simulation Results

The first power simulation study originally looked at $\Delta = \frac{\mu_1 - \mu_2}{\sigma}$ values that ranged from -2 to 2 standard deviations. Plotting the values from the simulation study (see Figure B.1) and examing the values (see Table B.1), it was determined that it was symmetrical, so the power curve could be done using $\Delta = \frac{|\mu_1 - \mu_2|}{\sigma}$ and values in the range 0 to 2. The Δ changed by $\frac{1}{8}$ for values form 0 to 1 and by $\frac{1}{4}$ for values from 1 to 2. The unequal spacing of Δ helps approximate the power curve better than using equal spacing for the full range, by reducing the number of unnecessary values. The final curve is seen in Figure 3.3.

Table B.1: Results from the first simulation study looking at the power of the two sample LRT for samples sizes $n_1 = n_2$. The results are plotted in Figure B.1. Noting the symmetric appearance, it was decided to change to the absolute standardized difference.

$\Delta = \frac{\mu_1 - \mu_2}{\sigma}$	-2	-1.75	-1.5	-1.25	-1	-0.875	-0.75	-0.625	-0.5	-0.375	-0.25	-0.125	0
15	0.99970	0.99510	0.97530	0.90950	0.74400	0.63420	0.50560	0.38220	0.25980	0.17420	0.10750	0.06980	0.06140
30	1	1	1	0.99750	0.96650	0.90770	0.81120	0.65330	0.47490	0.30320	0.16490	0.08600	0.06320
100	1	1	1	1	1	1	0.99960	0.99410	0.94250	0.76340	0.44690	0.18030	0.06410

Δ	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1	1.25	1.5	1.75	2
15	0.07000	0.11490	0.17570	0.25920	0.36800	0.48460	0.59210	0.69500	0.84720	0.93860	0.97780	0.99330
30	0.08440	0.16880	0.28940	0.45280	0.62430	0.77220	0.87050	0.93810	0.98910	0.99810	0.99990	1
100	0.16190	0.40810	0.71710	0.90910	0.98150	0.99840	1	1	1	1	1	1

Table B.2: Table of power for 13 different values of Δ for the two sample LRT testing $H_o : \mu_1 = \mu_2$ with small sample sizes ($n_1 = n_2$). Delta is done in terms of absolute diviance because Figure B.1 showed symmetry. Figure 3.3 shows the plot of the table.

$\Delta = \frac{|\mu_1 - \mu_2|}{\sigma}$	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1	1.25	1.5	1.75	2
15	0.06140	0.06990	0.11120	0.17495	0.25950	0.37510	0.49510	0.61315	0.71950	0.87835	0.95695	0.98645	0.99650
30	0.06320	0.08520	0.16685	0.29630	0.46385	0.63880	0.79170	0.88910	0.95230	0.99330	0.99905	0.99995	1
100	0.06410	0.17110	0.42750	0.74025	0.92580	0.98780	0.99900	1	1	1	1	1	1

Figure B.1 shows a symmetric pattern in power to detect a difference in standardized differences in the μ_i's. The symmetric curve leads to the conclusion that Δ can be done on the absolute standardized difference instead.
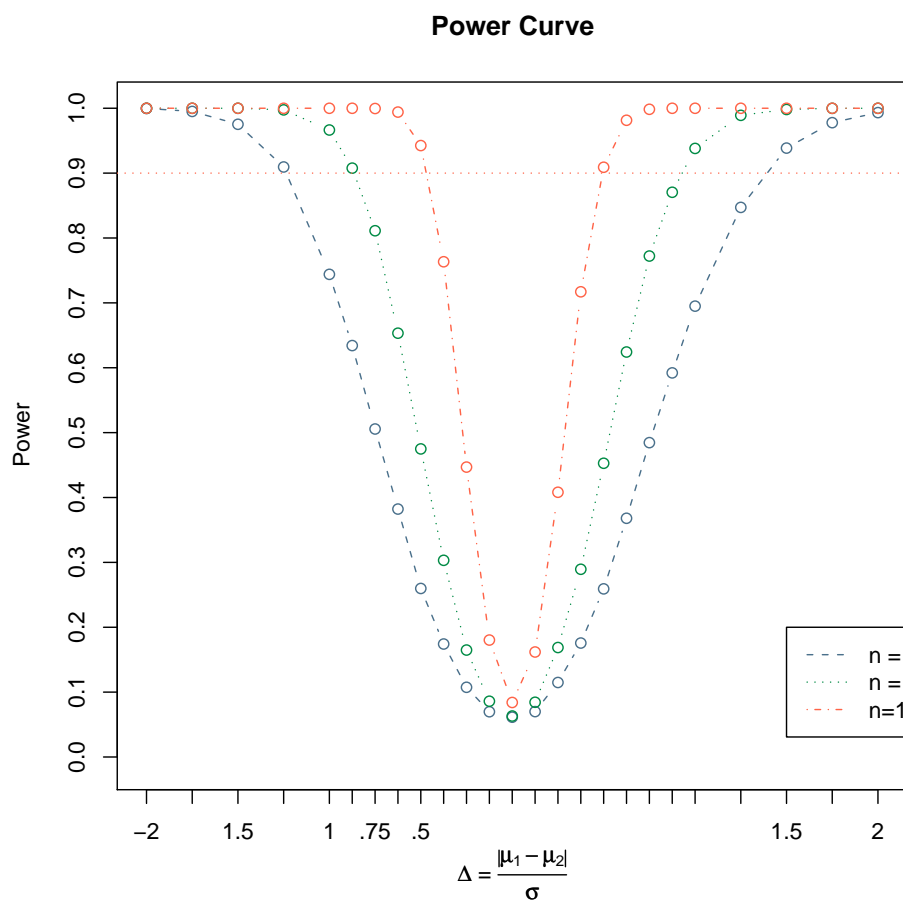

Figure B.1: The symmetric curves allow for Δ to be in absolute standardized deviance.

Figure B.2 shows a symmetric pattern in power to detect a difference in standardized in the μ_i's for large sample sizes. The symmetric curve leads to the conclusion that Δ can be done on the absolute difference instead.

Figure B.3 shows a symmetric pattern in power to detect differences in the μ_i's. The symmetric curve leads to the conclusion that Δ can be done on the absolute standardized difference instead.
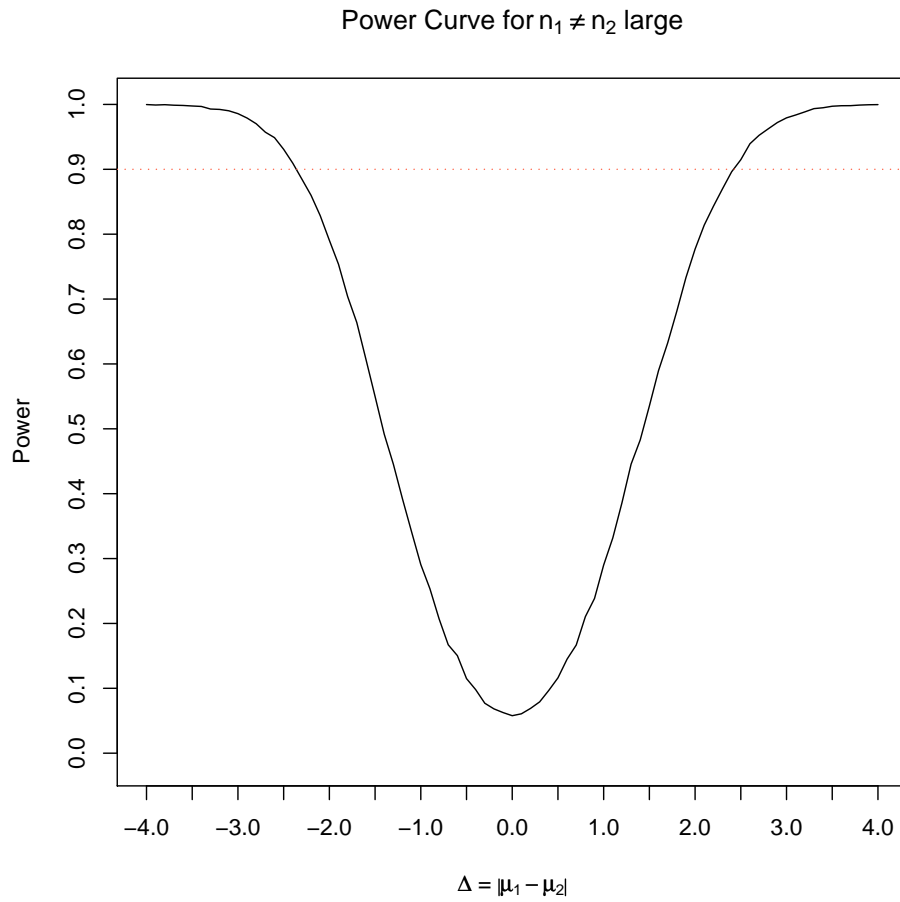
Figure B.2: The symmetry in the curve is an indication that we can instead focus Δ on the absolute deviance instead of the deviance.
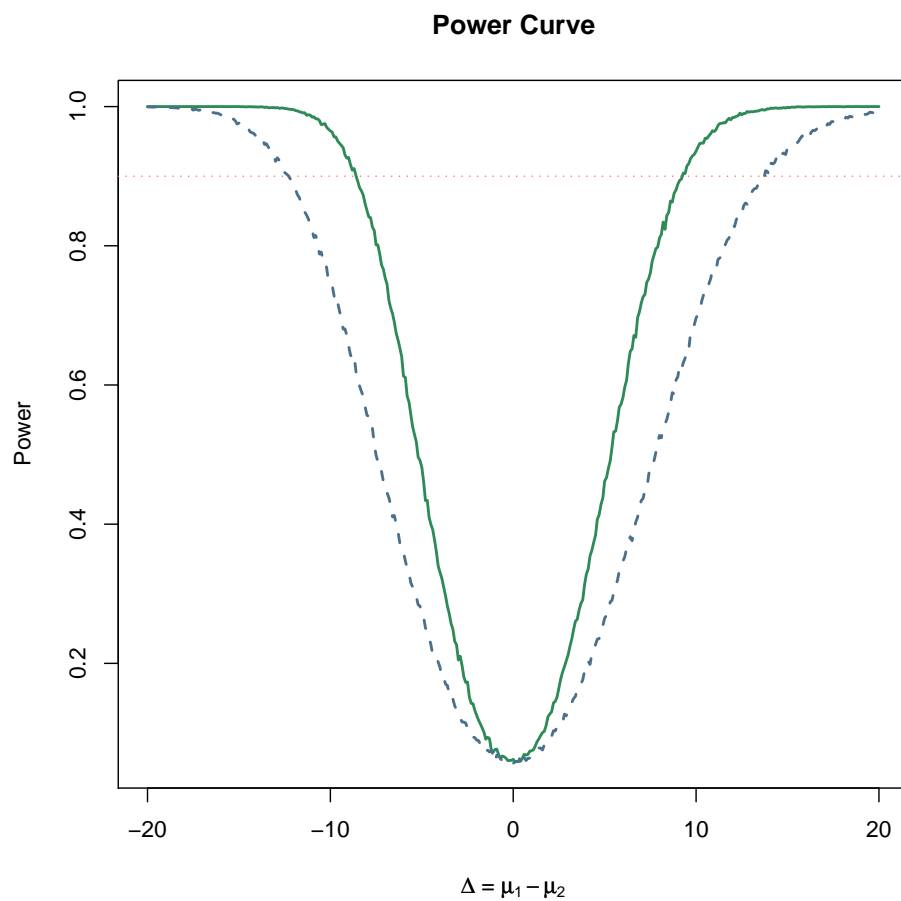
Power Curve

Figure B.3: The symmetric curves allow for Δ to be in absolute deviance.

B.2 k POPULATION LRT

Code

```
# Log Likelihood of the k Population
ll.tnorm.kpop<-function(p,t,d1,d2,d3,d4,d5){
m1 <- length(d1)
m2 <- length(d2)
m3 <- length(d3)
m4 <- length(d4)
n  <- length(d5)
out <- ll.tnorm2(c(p[1],p[6]),t,d1) +
               ll.tnorm2(c(p[2],p[6]),t,d2) +
               ll.tnorm2(c(p[3],p[6]),t,d3) +
               ll.tnorm2(c(p[4],p[6]),t,d4) +
               ll.tnorm2(c(p[5],p[6]),t,d5)
        out
}
# Used to optimize
ll.tnorm.kpop2<-function(p,t,d1,d2,d3,d4,d5){
m1 <- length(d1)
m2 <- length(d2)
m3 <- length(d3)
m4 <- length(d4)
n  <- length(d5)
out <- ll.tnorm2(c(p[1],p[6]),t,d1) +
               ll.tnorm2(c(p[2],p[6]),t,d2) +
               ll.tnorm2(c(p[3],p[6]),t,d3) +
               ll.tnorm2(c(p[4],p[6]),t,d4) +
               ll.tnorm2(c(p[5],p[6]),t,d5)
        -1*out
}


# Gradient of the k population Log Likelihood.
grad.tnorm.kpop<-function(p,t,d1,d2,d3,d4,d5){
   m1 <- length(d1)
   m2 <- length(d2)
   m3 <- length(d3)
   m4 <- length(d4)
   n  <- length(d5)
   g1 <- p[1] - mean(d1) + integrate(psi.mu,t[1],t[2],mu=p[1],sigma=p[6])$value /
           integrate(psi,t[1],t[2],mu=p[1],sigma=p[6])$value
   g2 <- p[2] - mean(d2) + integrate(psi.mu,t[1],t[2],mu=p[2],sigma=p[6])$value /
           integrate(psi,t[1],t[2],mu=p[2],sigma=p[6])$value
   g3 <- p[1] - mean(d3) + integrate(psi.mu,t[1],t[2],mu=p[3],sigma=p[6])$value /
           integrate(psi,t[1],t[2],mu=p[1],sigma=p[6])$value
   g4 <- p[2] - mean(d4) + integrate(psi.mu,t[1],t[2],mu=p[4],sigma=p[6])$value /
           integrate(psi,t[1],t[2],mu=p[2],sigma=p[6])$value
   g5 <- p[1] - mean(d5) + integrate(psi.mu,t[1],t[2],mu=p[5],sigma=p[6])$value /
           integrate(psi,t[1],t[2],mu=p[1],sigma=p[6])$value
```

```
    g6 <- p[6] - sum((d1-p[1])^2)/m1 - sum((d2-p[2])^2)/m2 - sum((d1-p[3])^2)/m3
             - sum((d2-p[4])^2)/m4 - sum((d1-p[5])^2)/n +
          integrate(psi.sigma,t[1],t[2],mu=p[1],sigma=p[2])$value /
          integrate(psi,t[1],t[2],mu=p[1],sigma=p[2])$value
    out<-cbind(g1,g2,g3,g4,g5,g6)
    out
}

del<-c(seq(-2,-1.25,by=.25),seq(-1,1,by=.125),seq(1.25,2,by=.25))
l.kpop<-rep(0,length(del))
for( j in 1:length(del)){
  lrt<-NULL
  print(del[j])
  for(i in 1:10000){
                                      #create a random sample
    X1<-rtnorm(n,samMean,samSD,lower=0,upper=100)
    X2<-rtnorm(n,samMean,samSD,lower=0,upper=100)
    X3<-rtnorm(n,samMean,samSD,lower=0,upper=100)
    X4<-rtnorm(n,samMean,samSD,lower=0,upper=100)
    Y<-rtnorm(n,samMean + del[j]*samSD,samSD,lower=0,upper=100)

                                      #starting values for the iterations
    p2<-c(mean(X1),mean(X2),mean(X3),mean(X4),
         mean(Y),
         sd(c(X1,X2,X3,X4,Y)))

    p<-c(mean(c(X1,X2,X3,X4,Y)),
         sd(c(X1,X2,X3,X4,Y)))

                                      #find the parameter combinations
    nlminb(p,
         lower=c(-Inf,0),upper=c(Inf,Inf),
         ll.tnorm,
         gradient=grad.tnorm,
         hessian=hessian.tnorm,
         data=c(X1,X2,X3,X4,Y),
         t=trunc)->OUT1
    nlminb(p2,
         ll.tnorm.kpop2,
         lower=c(-Inf,-Inf,-Inf,-Inf,-Inf,0),upper=c(Inf,Inf,Inf,Inf,Inf,Inf),
         gradient=grad.tnorm.kpop,
         d1=X1,
         d2=X2,
         d3=X3,
         d4=X4,
         d5=Y,
         t=trunc)->OUT2

                                      #find the Likelihood ratio test stat.
    bob<- -2*(ll.tnorm2(OUT1$par,trunc,c(X1,X2,X3,X4,Y)) -
            ( ll.tnorm2(OUT2$par[c(1,6)],trunc,X1) +
              ll.tnorm2(OUT2$par[c(2,6)],trunc,X2) +
              ll.tnorm2(OUT2$par[c(3,6)],trunc,X3) +
              ll.tnorm2(OUT2$par[c(4,6)],trunc,X4) +
```

```
                ll.tnorm2(OUT2$par[c(5,6)],trunc,Y)))
    lrt<-rbind(lrt,bob)
  }
  l.kpop[j]<-mean(lrt>qchisq(.95,4))
}
```

Results

Figure B.4 shows the results of the simulation study examining the power of the LRT for detecting differences in the standardized difference, $\Delta = \frac{\mu_i - \mu_{(i)}}{\sigma}$, when testing the hypothesis $H_o : \mu_1 = \cdots = \mu_k$ for $k = 5$. The symmetric nature of the curve lead to the conclusion Δ can be measured under the absolute standardized difference instead.
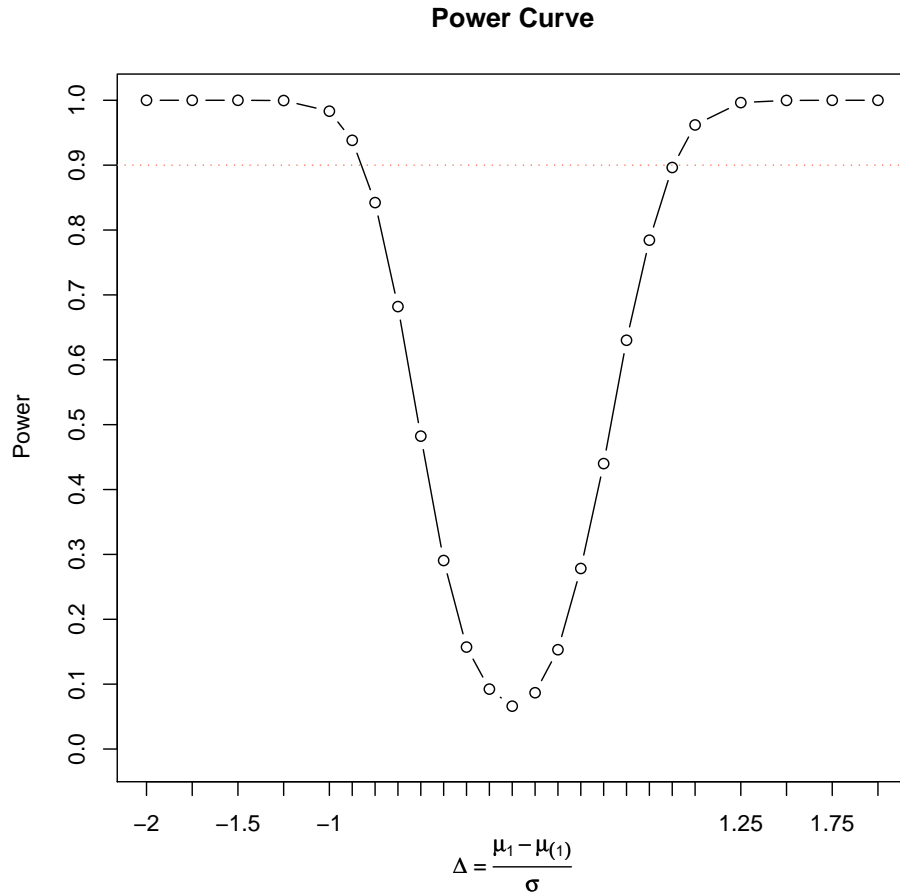

Figure B.4: The symmetric curves allow for Δ to be in absolute deviance.