A SIMULATION STUDY OF THE ROBUSTNESS OF HOTELLING'S T^2 TEST FOR THE MEAN OF A MULTIVARIATE DISTRIBUTION WHEN SAMPLING FROM A MULTIVARIATE SKEW-NORMAL DISTRIBUTION

by

YUN WU

B.S., Gui Zhou University of Finance and Economics, 1996

A REPORT

Submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Paul I, Nelson

Abstract

Hotelling's T^2 test is the standard tool for inference about the mean of a multivariate normal population. However, this test may perform poorly when used on samples from multivariate distributions with highly skewed marginal distributions. The goal of our study was to investigate the type I error rate and power properties of Hotelling's one sample T^2 test when sampling from a class of multivariate skew-normal (*SN*) distributions, which includes the multivariate normal distribution and, in addition to location and scale parameters, has a shape parameter to regulate skewness.

Simulation results of tests carried out at nominal type I error rate 0.05 obtained from various levels of shape parameters, sample sizes, number of variables and fixed correlation matrix showed that Hotelling's one sample T^2 test provides adequate control of type I error rates over the entire range of conditions studied. The test also produces suitable power levels for detecting departures from hypothesized values of a multivariate mean vector when data result from a random sample from a multivariate *SN*. The shape parameter of the *SN* family appears not to have much of an effect on the robustness of Hotelling's T^2 test. However, surprisingly, it does have a positive impact on power.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to take this occasion to express of my gratitude to:

Dr. Paul Nelson, my advisor. He provided invaluable guidance, suggestions and encouragement that helped me to produce this report. His attitude to the research and students inspired me to be a better researcher and person;

My other committee members Dr. Gary Gadbury and Dr. Weixing Song for their constructive suggestions and support;

Dr. Boyer and Dr. Neill for giving me the opportunity to study in the Department of Statistics at KSU;

All other professors, fellow students and staffs in the Department of Statistics at KSU for all their help;

My parents, brothers and sister for their spiritual support;

My beloved husband and daughter for their understanding and belief in me.

CHAPTER 1 - Introduction

Hotelling's (1931) well known T^2 is the standard statistic for testing hypotheses of the form $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}$ is the mean vector of a multivariate normal distribution with unknown covariance matrix, based on a random sample. In this setting, it is a uniformly most powerful, invariant test. Hotelling's statistic can also be used to construct exact confidence ellipsoids for $\boldsymbol{\mu}$**.**

However, in practice, not all data satisfy the multivariate normality assumption. The robutsness of Hotelling's one sample T^2 test when the assumption of multivariate normality is violated has been a topic of interest for researchers. See, for example, Chase, (1971), Mardia (1975), Everitt (1979), and Kariya (1981). These works showed that the size of Hotelling's one sample T^2 test is robust against slight departures from multivariate normality. But for multivariate distributions with highly skewed marginal distributions, such as the exponential and lognormal distributions (Everitt, 1979), Hotelling's T^2 test can be adversely affected by a departure from marginal symmetry.

This report studies the performance of Hotelling's one sample T^2 test in terms of size and power when sampling from a class of what are termed multivariate skew-normal (*SN*) distributions, developed in Azzalini (1985), Azzalini and Valle (1996), and Azzalini and Capitanio (1999). This family includes the multivariate normal distribution and, in addition to location and scale parameters, has a shape parameter which regulates skewness. The multivariate *SN* distributions are natural extensions of normal distributions and may prove to be more appropriate in practical situations in which marginal distributions maybe skewed. Examples given in Azzalini and Capitanio (1999) show that *SN* distributions can be useful in fitting real data.

In this report, we assess the effect of skewness on Hotelling's one sample T^2 test by conducting a simulation study. A literature review on the robustness of Hotelling's one sample T^2 test will be given in Chapter 2. The model and basic properties of skew-normal distributions will be described in Chapter 3. In Chapter 4, a sampling procedure and a simulation design will be presented. Chapter 5 summarizes the simulation results in terms of size and power when data are generated from a multivariate skew-normal distribution. Chapter 6 makes conclusions about this study and offers suggestions for further research.

CHAPTER 2 - Literature Review

Hotelling's T^2 test plays an important role in inference about the mean vector of a p-dimensional distribution. For $p > 1$, the variables being measured on each unit are often correlated and applying a set of separate, one sample t -tests p times, each carried out at level α, can make the overall type I error greater than α. However, Hotelling's T^2 allows us to test all p means simultaneously with overall type I error rate equal to α. The validity of this statement is derived under the assumption that the data are a random sample from a multivariate normal distribution.

Studies of the robustness of Hotelling's one sample T^2 test with respect to size and power when multivariate normality does not hold have been carried out by several researchers, including Arnold (1964), Mardia (1970, 1975), Chase and Bulgren (1971), Everitt (1979) and Kariya (1981). Arnold (1964) began the study of bivariate distributions with independent marginals. He showed that when sampling from the rectangular and the double exponential distributions, with a sample size of 8 and a nominal level of $\alpha = 0.05$, the empirical significance levels were close to 0.05. Chase and Bulgren (1971) studied 6 skewed and correlated bivariate distributions: bivariate normal (as a check on the procedure) with correlation coefficient $\rho = 0, 0.25, 0.50, 0.75$; bivariate uniform with $\rho = 0, 0.25$; bivariate exponential with $\rho = 0, 0.25, 0.50, 0.75$; bivariate gamma, bivariate lognormal and bivariate double exponential with $\rho = 0$. Most of the time, the difference between the empirical type I error rates and the nominal level $\alpha = 0.05$ was under 0.01 for the bivariate uniform distribution. The type I error rates for the gamma and exponential distributions were larger than the nominal level α = 0.05. For example, for ρ = 0 and n = 20, the estimated type I error rates were 0.10 for the exponential distribution and 0.068 for the gamma distribution. Again for tests carried out at nominal α = 0.05, the lognormal distribution yielded very large actual type I error rates even for sample sizes as large as 20. However, the observed type I error rates for the double exponential were slightly conservative. This study also showed that the effect of the correlation coefficient on Hotelling's T^2 test was not large, especially when the sample size was 20.

Mardia (1975) investigated the robustness of Hotelling's T^2 by summarizing earlier works with the help of the well-known Mardia (1970) measure of multivariate skewness. He showed that actual type I error rates of Hotelling's one sample T^2 test were sensitive to departures from multivariate normality in terms of skewness. Some values of skewness he examined were zero for the double exponential distribution and 8 for the exponential distribution when samples were drawn from bivariate distributions of independent random variables. A further study, which extended the number of variables p from 2 to 10, was conducted by Everitt (1979). He demonstrated that when sampling each variable independently from a uniform distribution, Hotelling's T^2 produced actual type I error rates close to nominal levels for α = 0.1, 0.05, 0.01. More specifically, for a sample size of 10, p = 6, samples from a distribution with independent uniform marginals had empirical type I error rates of 0.118, 0.059, and 0.013, respectively. The exponential and lognormal marginal distributions, both skewed, resulted in very high actual type I error rates, two to four times higher than the nominal for the exponential and three to sixteen times higher for the lognormal. This study also indicated that large sample sizes produced actual type I error rates similar to those attained from small sample sizes.

The studies described above were carried out using Monte Carlo simulation. Kariya (1981) derived the uniform most powerful invariance (UMPI) of Hotelling's T^2, a property that is robust with respect to the following departure from the assumption of multivariate normality. UMPI also holds for $n > p$ if the density of $\mathbf{y} = (y_1, \cdots, y_p)$, a random vector, has the form

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{C} \left| \boldsymbol{\Sigma} \right|^{-n/2} q((\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})), \tag{2.1}$$

where q is a non-increasing, convex function from $[0, \infty)$ into $[0, \infty)$. This family of distributions contains the multivariate *t*-distribution, the multivariate Cauchy distribution, the contaminated normal distribution, etc. But it does not include skew normal distributions.

In view of these previous studies, it seems evident that the one sample Hotelling's T^2 is fairly robust with respect to symmetric non-normal distributions, for example, the double exponential, rectangular and uniform distributions, Arnold(1964) and Chase and Bulgren (1971). Meanwhile, this test can result in inflated type I error rates for highly skewed marginal distributions, such as the exponential and lognormal distributions, as shown by Chase and Bulgren (1971) and Everitt (1979). The connection between the skewness and the robustness of Hotelling's T^2 motivated us to consider how the one sample Hotelling's T^2 test performs for skew normal distributions, which I will further describe in the next chapter.

CHAPTER 3 - The Skew Normal Distribution

The scalar skew-normal (*SN*) family of distributions, formally introduced by Azzalini (1985), attracted a great deal of attention in the literature because of their flexibility in modeling skewed data, "mathematical tractability" and inclusion of the normal distribution as a special case. Azzalini and Dalla Valle (1996) developed the multivariate version of these distributions. These two studies are considered to be the pioneering works in this area. Subsequently, there have been numerous further developments related to this *SN*. For example, the closed skew-normal (*CSN*) was presented in González-Farías *et al*. (2004) and the generalized skew-elliptical (*GSE*) in Genton and Loperfido (2002). Genton (2004) gives an extensive review of the research work in this area.

We chose to use the original *SN* distribution in this report because it is relatively easy to work with and it is the most thoroughly investigated of these distributions. In the following sections, I will introduce various properties of the Azzalini's *SN* distribution and its stochastic representation in both the univariate and mulitivariate cases.

3.1 Univariate Skew-Normal Distribution

In this section, some important properties and characterizations of the scalar *SN* distribution will be presented. This material can be found in Azzalini (1985) and Henze (1986).

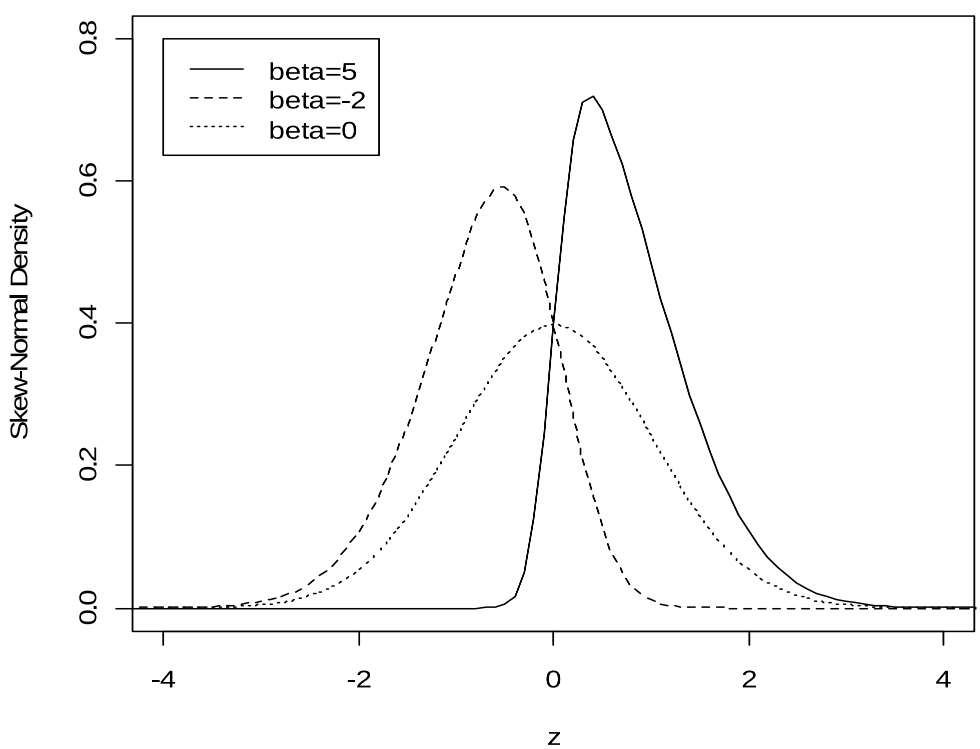
3.1.1 Definition and Basic Properties

Definition 3.1: A random variable Z is said to have a scalar $SN(\beta)$ distribution if the density function is of the form:

$$f(z, \beta) = 2\,\Phi(\beta z)\phi(z), \qquad z \in R, \text{ and } \beta \in R, \qquad (3.1)$$

where $\Phi(.)$ and $\phi(.)$ denote the standard normal cumulative distribution function and the standard normal probability density function, respectively. The parameter β controls the shape of the distribution. For instance, when $\beta = 0$, $f(z, \beta)$ corresponds to the standard normal distribution. Plots of the univariate density (3.1) for $\beta = 0, -2, 5$, given in Figure 3.1, illustrate the effects of changing β on the shape of the density.

Figure 3.1: Graphs of $SN_1(\beta)$ Densities (β = 0, -2, and 5)



We usually call the density (3.1) a "standard" *SN*. If we add location and scale components, ξ and σ, respectively, to this density, the variable

$$Y = \xi + \sigma Z, \qquad \xi \in R,\ \sigma \in R^{+}$$

written as $SN(\xi, \sigma, \beta)$, has the density function

$$g(y; \xi, \sigma, \beta) = \frac{2}{\sigma}\phi\left(\frac{y-\xi}{\sigma}\right)\Phi\left(\beta\frac{y-\xi}{\sigma}\right).$$

Some basic properties that follow from the definition 3.1 are:

Property 3.1.1: When $\beta = 0$, Z has a standard normal distribution, i.e., $SN(0) = N(0, 1)$.

Property 3.1.2: If $Z \sim SN(\beta)$, then $-Z \sim SN(-\beta)$.

Property 3.1.3: As $\beta \to +\infty$ ($\beta \to -\infty$), the density $SN(\beta)$ tends to positive (negative) half normal density.

Property 3.1.4: If $Z \sim SN(\beta)$ then $Z^2 \sim \chi_1^2$.

Properties 3.1.1 - 4 show that the scalar skew-normal density includes the normal distribution and shares similar properties to the normal density.

3.1.2 Stochastic Representation

The next three properties can be used to generate a random variable $Z \sim SN(\beta)$.

Property 3.1.5: If *Y* and *W* are independent $N(0, 1)$ variates, and *Z* is set equal to *Y* conditionally on $\beta Y > W$, for some real β, then $Z \sim SN(\beta)$.

Note: An efficient way to use the random variables generated by Property 3.1.5 is to set

$$Z = \begin{cases} Y & \text{if } \beta Y > W, \\ -Y & \text{if } \beta Y \le W. \end{cases}$$

Property 3.1.6: If (X, Y) is a bivariate normal random variable with standardized marginals and correlation δ, the conditional distribution of Y given $X > 0$ is $Z \sim SN$ ($\beta(\delta)$).

Note: $\beta(\delta)$ means β is related to δ through the following relationships

$$\beta(\delta) = \frac{\delta}{\sqrt{(1-\delta^2)}}, \qquad \delta(\beta) = \frac{\beta}{\sqrt{(1+\beta^2)}}. \tag{3.2}$$

Property 3.1.7: If Y_0 and Y_1 are independent N (0, 1) variables and $\delta \in$ (-1, 1), then

$$Z = \delta |Y_0| + (1-\delta^2)^{1/2}\, Y_1 \tag{3.3}$$

is SN ($\beta(\delta)$).

3.1.3 Cumulative Distribution Function and Moments

The Cumulative Distribution Function (cdf) of Z, as given in (3.3), denoted by G (z, β), is given by

$$G\,(z, \beta) = 2\int_{-\infty}^{z}\int_{-\infty}^{\beta t} \phi(t)\,\phi(u)\,du\;dt\,. \tag{3.4}$$

Property 3.1.8: $1 - G\,(-z; \beta\,) = G\,(z; -\beta\,)$.

Property 3.1.9: The cdf of SN (1) is equal to the square of the standard normal cdf, i.e.,

$$G\,(z;\,1) = (\Phi\,(z))^2\,.$$

The moment generating function (mgf) of Z, denoted by $M_z(t)$, is given by

$$M_z(t) = 2\,exp\,(\frac{t^2}{2})\,\Phi\,(\delta\, t)\,. \tag{3.5}$$

Taking derivatives and evaluating at $t = 0$ yields the following:

The mean of Z, $E\,(Z) = \sqrt{\frac{2}{\pi}}\ \delta$. (3.6)

The variance of Z, $Var\,(Z) = 1 - (\sqrt{\frac{2}{\pi}}\ \delta\,)^2$. (3.7)

The third standardized moment (a measure of skewness) of Z,

$$S\,(Z) = \frac{1}{2}(4-\pi)\,\mathrm{sign}\,(\beta)\,(\frac{\beta^2}{\frac{\pi}{2}+(\frac{\pi}{2}-1)\,\beta^2})^{3/2}, \tag{3.8}$$

varies from -0.9953 to 0.9953.

Note: The third standardized moment, is written as γ_1, and defined as

$$\gamma_1 = \frac{\mu_3}{\sigma^3},$$

where $\mu_3 = E[(X - E[X])^3]$ is the third central moment and σ is the standard deviation.

The fourth standardized moment (a measure of kurtosis) of Z,

$$K\,(Z) = 2\,(\pi-3)(\frac{\beta^2}{\frac{\pi}{2}+(\frac{\pi}{2}-1)\,\beta^2})^2, \tag{3.9}$$

varies from -0.869 to 0.869.

Note: The fourth standardized moment, is written as γ_2, and defined as

$$\gamma_2 = \frac{\mu_4}{\sigma^4},$$

where $\mu_4 = E[(X - E[X])^4]$ is the fourth central moment and σ is the standard deviation.

The computation of higher moments can be obtained by using properties 3.1.14 and 3.1.15.

Property 3.1.14: The even moments of Z are the same as the even moments of the standard normal distribution.

Property 3.1.15: The odd moments of Z are given by

$$E(z^{2k+1}) = \sqrt{\frac{2}{\pi}}\beta(1+\beta^2)^{-(k+1/2)}\, 2^{-k}\,(2k+1)!\sum_{t=0}^{k}\frac{t!\,(2\beta)^{2t}}{(2t+1)!(k-t)!}$$

3.2 Multivariate Skew-Normal Distribution

In this section, we will retain the same notation as given in the above section, except changing them into matrix form. Vectors are represented by lower case bold text and matrices are represented in upper case bold text. The material in this section can be found in Azzalini (1996, 1999, and 2005).

3.2.1 Definition and Basic Properties

Definition 3.2: A p-dimensional random vector $\mathbf{z}$ is said to has a multivariate skew-normal distribution, denoted by $\mathbf{z} \sim SN_p(\mathbf{\Omega}, \boldsymbol{\beta})$, if it is continuous with density function

$$f(\mathbf{z}) = 2\,\phi_p(\mathbf{z};\mathbf{\Omega})\,\Phi(\boldsymbol{\beta}^T\mathbf{z}) \qquad \mathbf{z} \in R^p, \tag{3.10}$$

where $\phi_p(\mathbf{z};\mathbf{\Omega})$ is the p-dimensional normal density with zero mean and correlation matrix $\mathbf{\Omega}$, $\Phi(.)$ is the standard, univariate $N(0,1)$ distribution function, and $\boldsymbol{\beta}$ is a vector of shape parameters.

Contour plots of the bivariate case of the density (3.10) when β_1 = -5, β_2 = 10, and $\omega = 0.75$ (where ω is the off-diagonal element of $\mathbf{\Omega}$) are given in Figure 3.2. It is clearly different from an ellipse and presents some skewness. If we set the shape parameters β_1 = 0, β_2 = 0, and keep $\omega = 0.75$, the density (3.10) reduce to the multivariate normal distribution, as shown in Figure 3.3. More plots for different values of $\boldsymbol{\beta}$ and $\mathbf{\Omega}$ are shown in appendix D.

Similarly as in the univariate case, if we add location parameter $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)^T$ and positive scale parameter $\mathbf{S}$ = diag $(\sigma_1, \ldots, \sigma_p)$ to the density (3.10), the random vector $\mathbf{y} = \boldsymbol{\xi} + \mathbf{S}\mathbf{z}$, denoted by $\mathbf{y} \sim SN_p(\boldsymbol{\xi}, \mathbf{S}\mathbf{\Omega}\mathbf{S}, \boldsymbol{\beta})$, has the density function

$$g(\mathbf{y}) = 2\phi_p(\mathbf{y}; \boldsymbol{\xi}, \mathbf{S}\mathbf{\Omega}\mathbf{S})\, \Phi(\boldsymbol{\beta}^{\mathbf{T}}\mathbf{S}^{-1}(\mathbf{y} - \boldsymbol{\xi})\,.$$

Figure 3.2: Contour Plot of the SN_2 Density for $\beta_1 = -5$, $\beta_2 = 10$, and $\omega = 0.75$

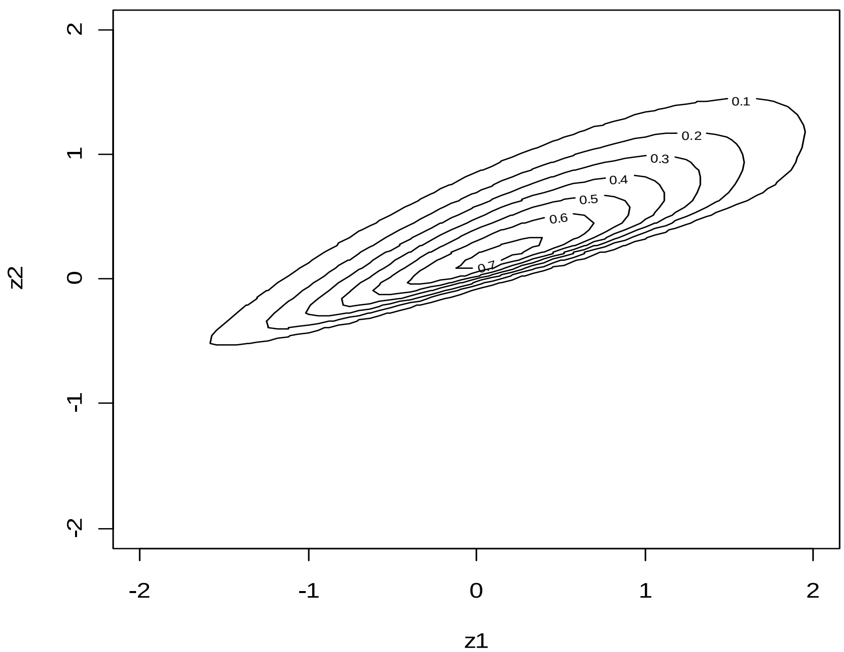
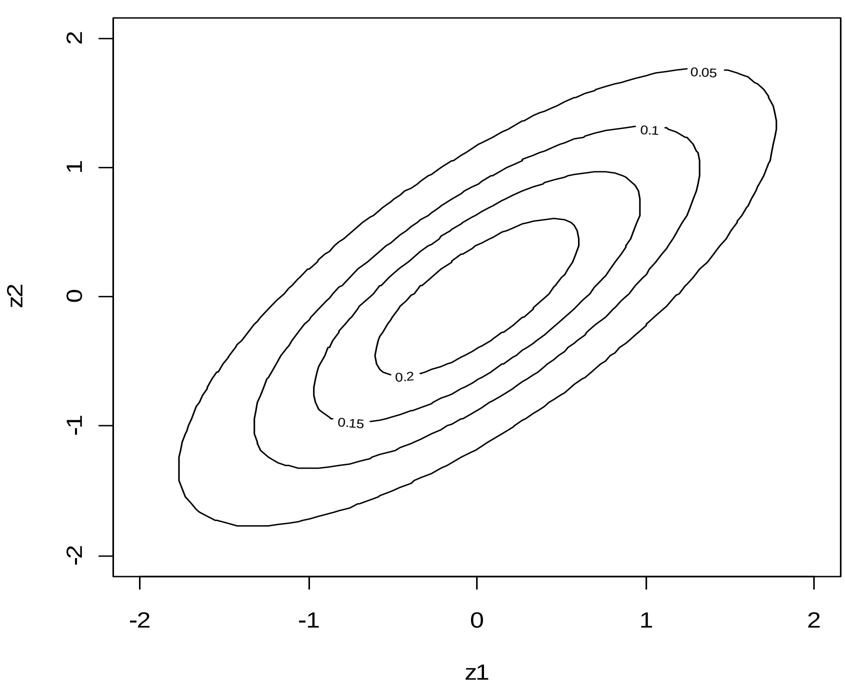


Figure 3.3: Contour Plot of the SN_2 Density for $\beta_1 = 0$, $\beta_2 = 0$, and $\omega = 0.75$



Some properties of $\mathrm{S}N_p\ (\boldsymbol{\Omega},\boldsymbol{\beta})$ and its quadratic form are given below.

Property 3.2.1: If $\mathbf{z} \sim \mathrm{S}N_p\ (\boldsymbol{\Omega},\boldsymbol{\beta})$, then

$$-\mathbf{z} \sim SN_p(\boldsymbol{\Omega}, -\boldsymbol{\beta})$$

Property 3.2.2: If $\mathbf{z} \sim \mathrm{S}N_p\ (\boldsymbol{\Omega},\boldsymbol{\beta})$, and $\mathbf{A}$ is a $p \times p$ non-singular matrix then

$$\mathbf{A}^T\mathbf{z} \sim \mathbf{SN}_p(\mathbf{A}^T\boldsymbol{\Omega}\mathbf{A},\ \mathbf{A}^{-1}\boldsymbol{\beta})$$

Property 3.2.3: If $\mathbf{z} \sim \mathrm{S}N_p\ (\boldsymbol{\Omega},\boldsymbol{\beta})$, then

$$\mathbf{z}^{\mathbf{T}}\boldsymbol{\Omega}^{-1}\mathbf{z} \sim \chi_p^2.$$

Property 3.2.4: If $\mathbf{z} \sim \mathrm{S}N_p\ (\boldsymbol{\Omega},\boldsymbol{\beta})$ and $\mathbf{B}$ is a symmetric positive semi-definite $p \times p$ matrix of rank k such that $\mathbf{B}\,\boldsymbol{\Omega}\,\mathbf{B} = \mathbf{B}$, then

$$\mathbf{z}^{\mathbf{T}}\mathbf{B}\,\mathbf{z} \sim \chi_k^2$$

Property 3.2.5: Let $\mathbf{x} \sim SN_p(\mathbf{0},\ \boldsymbol{\Omega},\ \boldsymbol{\beta})$, then the distribution of $\mathbf{x}\mathbf{x}^T$ is Wishart with scale parameter $\boldsymbol{\Omega}$ and 1 degree of freedom:

$$\mathbf{x}\mathbf{x}^T \sim W(\boldsymbol{\Omega}, 1)\,\mathbf{.}$$

3.2.2 Stochastic Representation

There are two ways to generate a random variable having the density given in (3.10).

[1] Transformation method

Consider a p-dimensional normal random vector $\mathbf{y}$ with standardized marginals and correlation matrix $\boldsymbol{\Psi}$, independent of $Y_0 \sim N\,(0,\,1)$, so that

$$\begin{pmatrix} Y_0 \\ \mathbf{y} \end{pmatrix} \sim N_{p+1}\left(\mathbf{0}, \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \Psi \end{pmatrix}\right) \quad . \tag{3.11}$$

With $\delta_{1,\dots,}\delta_p$ each in the interval (-1, 1), define

$$Z_j = \delta_j |Y_0| + (1-\delta_j^2)^{1/2} Y_j . \qquad (j = 1, \dots, p) \tag{3.12}$$

Then $Z_j \sim SN(\beta(\delta_j))$, and after some algebra, the pdf of $\mathbf{z} = (Z_{1,\dots,}Z_p)^T$ is given in (3.10), where

$$\boldsymbol{\beta}^{\mathbf{T}} = \frac{\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\Delta}^{-1}}{(\mathbf{1}+\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda})^{\mathbf{1/2}}}, \tag{3.13}$$

$$\boldsymbol{\Delta} = \operatorname{diag}((1-\delta_1^2)^{1/2}, \dots, (1-\delta_p^2)^{1/2},$$

$$\boldsymbol{\Omega} = \boldsymbol{\Delta}(\boldsymbol{\Psi} + \boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathbf{T}})\boldsymbol{\Delta}, \tag{3.14}$$

$$\boldsymbol{\lambda} = (\beta(\delta_1), \dots, \beta(\delta_p))^T .$$

[2] Conditioning method

Let $\mathbf{z}^* = (Z_0, Z_1 \dots Z_p)^T$ be a (p+1)-dimensional multivariate normal random vector such that $\mathbf{z}^* \sim N_{p+1}(\mathbf{0}, \boldsymbol{\Omega}^*)$, with standardized marginals and correlation matrix

$$\boldsymbol{\Omega}^* = \begin{pmatrix} 1 & \delta_1 & \dots & \delta_p \\ \delta_1 & & & \\ \vdots & & \boldsymbol{\Omega} & \\ \delta_p & & & \end{pmatrix}. \tag{3.15}$$

Using 3.1.6, the vector $(Z_1 \dots Z_p)^T$ conditionally on $Z_0 > 0$ is a multivariate SN random vector. Notice that for $\boldsymbol{\Omega}^*$ being positive definite matrix, we have some restrictions on the elements of $\boldsymbol{\Omega}$. In the bivariate case, it requires ω (where ω is the off-diagonal element of $\boldsymbol{\Omega}$) satisfying

$$\delta_1\delta_2 - \left\{(1-\delta_1^2)(1-\delta_2^2)\right\}^{1/2} < \omega < \delta_1\delta_2 + \left\{(1-\delta_1^2)(1-\delta_2^2)\right\}^{1/2}.$$

Combining the conditional method and property 3.2.1, we have the following property, which we will use in this report for generating random samples from $\mathbf{z} \sim$ $\mathrm{S}N_p\ (\mathbf{\Omega}, \boldsymbol{\beta})$.

Property 3.2.6: If X_0 is a scalar random variable and $\mathbf{x}$ is *p*-dimensional such that

$$\begin{pmatrix} X_0 \\ \mathbf{x} \end{pmatrix} \sim N_{p+1}\left(\mathbf{0}, \mathbf{\Omega}^*\right), \qquad \Omega^* = \begin{pmatrix} 1 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \mathbf{\Omega} \end{pmatrix}$$

and $\mathbf{z}$ is defined by

$$\mathbf{z} = \begin{cases} \mathbf{x} & \text{if } X_0 > 0, \\ -\mathbf{x} & \text{otherwise,} \end{cases}$$

then $\mathbf{z} \sim \mathrm{S}N_p\ (\mathbf{\Omega}, \boldsymbol{\beta})$, where

$$\boldsymbol{\beta} = \frac{\mathbf{\Omega}^{-1}\boldsymbol{\delta}}{(\mathbf{1} - \boldsymbol{\delta}^{\mathbf{T}}\mathbf{\Omega}^{-1}\boldsymbol{\delta})^{\mathbf{1/2}}} \ . \tag{3.16}$$

3.2.3 Cumulative Distribution Function and Moments

The Cumulative Distribution Function (cdf) of $\mathbf{z} \sim \mathrm{S}N_p\ (\mathbf{\Omega}, \boldsymbol{\beta})$ is

$$G\,(\mathbf{z}) = 2\int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_p} \phi_p\,(\mathbf{z};\mathbf{\Omega})\Phi(\boldsymbol{\beta}^T\mathbf{z})dz_1 \ldots dz_p\,. \qquad \mathbf{z} \in R^p$$

The moment generating function (mgf) of $\mathbf{z}$, denoted by $M\,(\mathbf{t})$, is given by

$$M\,(\mathbf{t}) = 2\int_{R^p} \exp(\mathbf{t}^{\mathbf{T}}\mathbf{z})\,\phi_p\,(\mathbf{z};\mathbf{\Omega})\,\Phi(\boldsymbol{\beta}^{\mathbf{T}}\mathbf{z})\,d\mathbf{z}$$

$$= 2\exp\left\{\frac{1}{2}\mathbf{t}^{\mathbf{T}}\,\mathbf{\Omega}\,\mathbf{t}\right\}\ \Phi\left\{\frac{\boldsymbol{\beta}^{\mathbf{T}}\mathbf{\Omega}\mathbf{t}}{(1+\boldsymbol{\beta}^{\mathbf{T}}\mathbf{\Omega}\boldsymbol{\beta})^{1/2}}\right\}.$$

Hence, the mean vector and the covariance matrix are

$$E\,(\mathbf{z}) = (2/\pi)^{1/2}\,\boldsymbol{\delta}\,, \tag{3.17}$$

$$Cov\,(\mathbf{z}) = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^T\,, \tag{3.18}$$

where

$$\boldsymbol{\delta} = \frac{\mathbf{1}}{(\mathbf{1} + \boldsymbol{\beta}^{\mathbf{T}}\boldsymbol{\Omega}\boldsymbol{\beta})^{\mathbf{1/2}}}\boldsymbol{\Omega}\boldsymbol{\beta}\,. \tag{3.19}$$

The multivariate indices of skewness and kurtosis are

$$\boldsymbol{S}\,(\mathbf{z}) = \left(\frac{4-\pi}{2}\right)^2 \left(\frac{\boldsymbol{\mu}_{\mathbf{z}}^{\mathbf{T}}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\mu}_{\mathbf{z}}}{1-\boldsymbol{\mu}_{\mathbf{z}}^{\mathbf{T}}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\mu}_{\mathbf{z}}}\right)^3, \tag{3.20}$$

$$\boldsymbol{K}\,(\mathbf{z}) = 2(\pi-3)\left(\frac{\boldsymbol{\mu}_{\mathbf{z}}^{\mathbf{T}}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\mu}_{\mathbf{z}}}{1-\boldsymbol{\mu}_{\mathbf{z}}^{\mathbf{T}}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\mu}_{\mathbf{z}}}\right)^2, \tag{3.21}$$

where **S** (**z**) and **K** (**z**) range from about 0 to 0.9905, and from 0 to 0.869, respectively.

The above framework will allow me to use simulation to investigate the performance of Hotelling's T^2 test in terms of size and power when sampling from a multivariate skew-normal distribution.

CHAPTER 4 - Simulation Experiment

4.1 General Remarks

Our simulation study was conducted to evaluate the performance of Hotelling's one sample T^2 test when sampling from a multivariate *SN* distribution in order to answer two main questions. Recall that a multivariate *SN* distribution is a multivariate normal distribution when the skewness parameter β is zero.

1. How well does Hotelling's one sample T^2 test perform in terms of size and power when β differs from zero?

2. Are there combinations of parameters of the *SN* distribution under which Hotelling's T^2 performs particularly poorly?

My simulations were carried out using the R language, version 2.7.2. In this chapter, I describe the procedures I used for generating skew normal data and the parameter settings I chose. The notations are the same as in Chapter 3. Vectors are represented by lower case bold text and matrices are represented in upper case bold text. Moreover, when we write $\boldsymbol{\mu}$ = **a**, it means μ_i = a, where a is a real number, $i = 1, \ldots, p$, and **a** = (a, a, . . . , a)$'$. Likewise for $\boldsymbol{\beta}$.

4.2 Steps of Simulation

The algorithm I used for generating multivariate *SN* random variables having specified mean vector $\boldsymbol{\mu}$ is based on property 3.2.6. First, specify a $p \times p$ positive definite correlation matrix $\mathbf{\Omega}$, a $p \times 1$ vector of constants $\boldsymbol{\beta}$ (shape parameter), and a $p \times 1$ vector of constants $\boldsymbol{\mu}$. Then, and define the $p \times 1$ vector

$$\boldsymbol{\delta} = (1 + \boldsymbol{\beta}^{\mathbf{T}} \mathbf{\Omega} \boldsymbol{\beta})^{-1/2} \mathbf{\Omega} \boldsymbol{\beta},$$

and the (p+1) $\times$(p+1) correlation matrix

$$\mathbf{\Omega}^* = \begin{pmatrix} 1 & \delta_1 \; \ldots \; \delta_p \\ \delta_1 & \\ \vdots & \mathbf{\Omega} \\ \delta_p & \end{pmatrix}.$$

[I] The Algorithm

(1) Generate $\mathbf{z}^* = (Z_0, Z_1, \ldots, Z_p)^T \sim N_{p+1}(\mathbf{0}, \mathbf{\Omega}^*)$.

(2) If $Z_0 > 0$, let $\mathbf{z} = (Z_1, \ldots, Z_p)^T$, otherwise let $\mathbf{z} = -(Z_1, \ldots, Z_p)^T$. Then, $\mathbf{z}$ is an observation from a p-dimensional skew-normal distribution with

$$E(\mathbf{z}) = (2/\pi)^{1/2} \boldsymbol{\delta}, \qquad Cov(\mathbf{z}) = \mathbf{\Omega} - \frac{2}{\pi} \boldsymbol{\delta}\boldsymbol{\delta}^T.$$

(3) Let $\mathbf{x} = \mathbf{z} + \boldsymbol{\mu} - (2/\pi)^{1/2} \boldsymbol{\delta}$.

As noted in Chapter 3, we then have that $\mathbf{x}$ is a multivariate skew-normal vector with

$$E(\mathbf{x}) = \boldsymbol{\mu}, \qquad Cov(\mathbf{x}) = \mathbf{\Omega} - \frac{2}{\pi} \boldsymbol{\delta}\boldsymbol{\delta}^T.$$

[II] Generating Multivariate Skew-Normal Data

(A) Set sample size n, dimension p, $\boldsymbol{\mu}$, $\boldsymbol{\beta}$ and $\mathbf{\Omega}$.

(B) Using [I] independently generate multivariate skew-normal observations $\{\mathbf{x}_i, i = 1, 2, \ldots, n\}$.

[III] Hotelling's T^2 Test

Test H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ *vs* H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ at nominal type I error rate 0.05 using Hotelling's T^2 based on data $\{\mathbf{x}_i, i = 1,2,\ldots,n\}$. Specifically, reject H_0 at nominal type I error rate α if

$$T^2 = n\,(\bar{\mathbf{x}} - \boldsymbol{\mu}_\mathbf{0})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_\mathbf{0}) > \frac{(n-1)p}{(n-p)} F_{p,\,n-p,\,1-a}$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the sample mean and covariance, respectively and $F_{1-\alpha,\nu_1,\nu_2}$ is the (1-α) quantile of an F distribution with degree of freedom (ν_1, ν_2). To carry out my simulation study, I independently repeated [II] and [III] N = 10,000 times and recorded the proportion of times (denoted by $\hat{\alpha}$) that H_0 was rejected. If $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, this is an estimate of type I error rate. Otherwise $\hat{\alpha}$ estimates power.

4.3 Simulation Settings

The parameters involved in our study are: sample size n, dimension p, mean vector $\boldsymbol{\mu}$, shape parameter $\boldsymbol{\beta}$ and correlation matrix $\boldsymbol{\Omega}$. They were varied for the purpose of assessing the robustness and power of the test across a range of settings. The specifications of the simulation design are described as follows. For simplicity, we took $\boldsymbol{\Omega}$ to be the identity matrix. The sample sizes n were set at 10, 20, and 50. We chose the number of variables p to be 1, 3 and 7. Without loss of generality, to study the attained type I error rate we set the mean vector $\boldsymbol{\mu} = \mathbf{0}$ and for the power part of the study, we selected values of the mean vector $\boldsymbol{\mu}$ to be the constant vectors **-1, -0.8, -0.5, -0.2, 0.2, 0.5, 0.8**, and **1** to represent increasing departures from H_0. Recall that the statement $\boldsymbol{\mu} = \mathbf{1}$

indicates that all of the components of the vector $\boldsymbol{\mu}$ are 1. The skewness parameter $\boldsymbol{\beta}$ is difficult to interpret in a multidimensional space. Therefore, before choosing the value of $\boldsymbol{\beta}$, we investigated how this parameter is related to skewness in the univariate case. Using equation (3.8) given in the last chapter, we calculated the skewness for various values of β. The results are shown in Tables 4.1, 4.2 and Figure 4.1. As can be seen, the positive (negative) values of β correspond to positive (negative) skewness. And the skewness increases as the value of β goes up. When β is equal to 10 (-10), the skewness of the univariate *SN* approaches its maximum 0.9953 (minimum -0.9953). Hence, in the multivariate case, the values of $\boldsymbol{\beta}$ were set equal to the constant vectors **-8**, **-4**, **-2**, **0**, **2**, **4**, and **8** to represent increasing amounts of skewness.

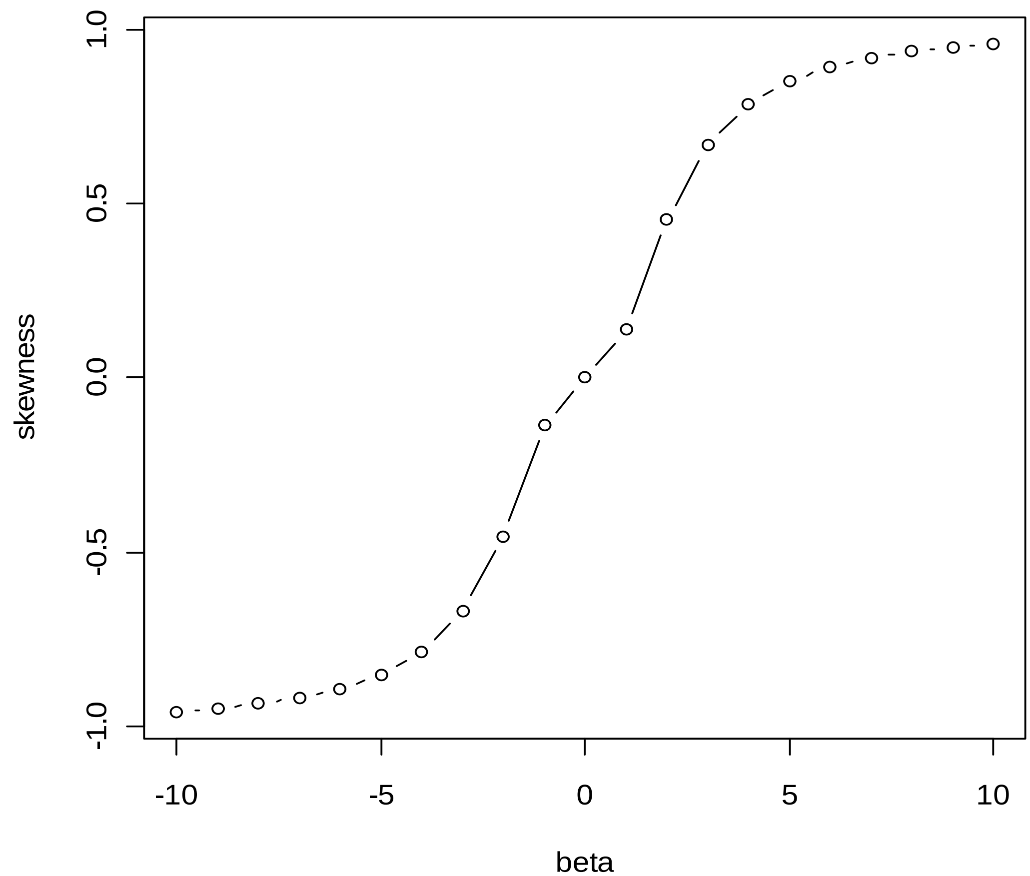
Table 4.1: Skewness for Positive Beta

	beta										
	0	1	2	3	4	5	6	7	8	9	10
skewness	0	0.14	0.45	0.67	0.78	0.85	0.89	0.92	0.93	0.95	0.96

Table 4.2: Skewness for Negative Beta

	beta										
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10
skewness	0	-0.14	-0.45	-0.67	-0.78	-0.85	-0.89	-0.92	-0.93	-0.95	-0.96

Figure 4.1: Skewness against Beta



In summary, the parameters of study are stated below:

1. Three levels of sample size *n*: 10, 20, 50;

2. Three levels of number of variables *p*: 1, 3, 7;

3. Nine levels of the constant mean vetor $\boldsymbol{\mu}$: **-1**, **-0.8**, **-0.5**, **-0.2**, **0**, **0.2**, **0.5**, **0.8**, **1**;

4. Seven levels of the constant vector of shape parameter $\boldsymbol{\beta}$: **-8**, **-4**, **-2**, **0**, **2**, **4**, **8**;

5. One level of correlation matrix $\boldsymbol{\Omega}$: **I**.

The total number of parameter settings was therefore equal to $3 \times 3 \times 9 \times 7 \times 1$ = 567. For each parameter setting N=10,000 independent data sets were simulated. The nominal type I error rate α was set equal to 0.05.

CHAPTER 5 - Results and Analysis

This chapter presents the results of my simulation study of estimated type I error rates and powers of Hotelling's one sample T^2 test when data are generated from multivariate *SN* distributions using the parameter combinations discussed in Chapter 4. The results are summarized in tables and plots. The first section of this chapter analyzes attained type I error rates. The second section deals with power and the third section summarizes my results. Recall that the mean $\boldsymbol{\mu}$ is zero under the null hypothesis.

5.1 Type I Error Rates

Table 5.1 gives estimated type I error rates, denoted $\hat{\alpha}$, organized by the dimension of the observations, sample size and skewness parameter $\boldsymbol{\beta}$. The values colored in green represent results for $\boldsymbol{\beta} = \mathbf{0}$, which corresponds to the multivariate normal distribution. Values marked with a '*' indicate results that we designate as being *non-robust* since they lie outside of the interval $\alpha \pm 2\sqrt{\alpha(1-\alpha)/N}$, where α = 0.05 is the nominal level and *N*= 10,000 is the simulation sample size. This interval ranges from 0.046 to 0.054 and represents the values of the estimate type I error rate that would lead to rejection of $H_0: \alpha = 0.05$ in favor of $H_a: \alpha \neq 0.05$ using a test whose type I error rate is approximately 0.05.

From the table 5.1, we see that except for $\boldsymbol{\beta} = \mathbf{0}$, all the rates are statistically significantly different from 0.05 and that none of the entries in the column corresponding to multivariate normal data is starred. Although most of the other entries are starred, type I error rates for the skew normal data settings I studied appear to have values close enough to 0.05 to be considered as satisfactory from a practical standpoint. For example, the largest value of $\hat{\alpha}$, 0.068, corresponding to p = 1, n = 10 and β = 8, is only 0.018 higher than the nominal type I error rate 0.05. Moreover, $\hat{\alpha}$ becomes stable and close to the nominal level 0.05 for the large sample size (n = 50) or large number of variables (p = 7) across all the levels of $\boldsymbol{\beta}$. Thus, overall, I judge that the Hotelling's one sample T^2 test satisfactorily holds its nominal type I error rate for most parameter combination conditions of the multivariate skew-normal distribution under investigation. The multivariate normal case ($\boldsymbol{\beta} = \mathbf{0}$) produces the best results and all the values from this situation are in our acceptance interval (0.046, 0.054).

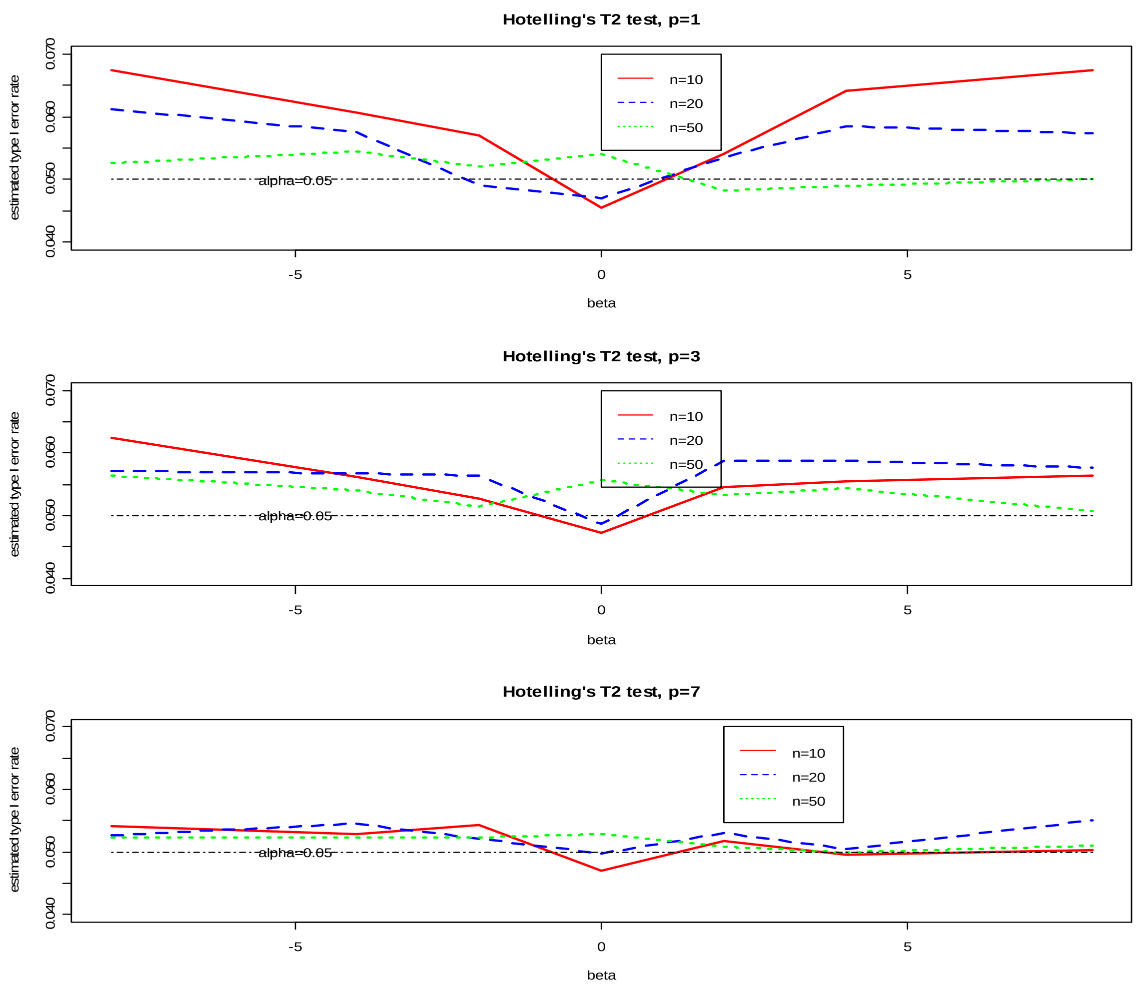
Figures 5.1 is graphical representation of the values in Table 5.1 and shows plots of estimated type I error rate versus shape parameter, $\boldsymbol{\beta}$, for p = 1, 3, 7 and n = 10, 20, 50. Four different profiles were drawn for each graph. The solid red color profile represents values for n = 10. The dashed with blue profile represents values for n = 20. The dotted green profile represents values for n = 50 and the black dot-dash straight line is the nominal value α = 0.05. A very gradual decrease of the type I error rate can be seen as the number of variables increases. Moreover, as the sample size increases from 10 to 50, the estimate type I error rate gets smaller for p = 1 or p = 3. The effect of sample size seems not to be important for p = 7. The plot also shows that the estimated type I error rates for both negative and positive $\boldsymbol{\beta}$ are almost symmetric.

Table 5.1: Type I Error Rates for Multivariate *SN*

		β						
p	*n*	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.067*	0.061*	0.057*	0.045	0.054	0.064*	0.068*
	20	0.067*	0.058*	0.049	0.047	0.054	0.058*	0.057*
	50	0.053	0.055*	0.052	0.054	0.048	0.049	0.050
3	10	0.063*	0.056*	0.053	0.047	0.055*	0.056*	0.056*
	20	0.057*	0.057*	0.056*	0.049	0.059*	0.059*	0.058*
	50	0.056*	0.054	0.051	0.056	0.053	0.054	0.051
7	10	0.054	0.053	0.054	0.047	0.052	0.050	0.050
	20	0.053	0.055*	0.052	0.050	0.053	0.050	0.055*
	50	0.052	0.052	0.052	0.053	0.051	0.050	0.051

* Indicates that the estimated type I error rate is more than $2\{(0.05)(0.95)/10{,}000\}^{1/2}$ from the nominal 0.05 level.

Figure 5.1: Type I Error Rates for Multivariate *SN*



5.2 Power Results

Results of the powers are analyzed in two ways. One way is to divide the results into three parts, corresponding to what I subjectively call a small effect size ($\boldsymbol{\mu} = \pm \mathbf{0.2}$), medium effect size ($\boldsymbol{\mu} = \pm \mathbf{0.5}$), and large effect sizes ($\boldsymbol{\mu} = \pm \mathbf{0.8}$ and $\boldsymbol{\mu} = \pm \mathbf{1}$). In each part, a tabular analysis and graphical display is performed on the data to determine the effect of the factors under study. Another way to accomplish this is through the development of a regression model.

5.2.1 Small Effect Size ($\boldsymbol{\mu} = \pm \mathbf{0.2}$)

Values of estimated powers in Tables 5.2 -5.9 are organized the same way as in Table 5.1 for the estimated type I error rates. The symbols used in the plots given in Figures 5.2-5.9 are consistent with the ones in Figure 5.1 has.

When the degree of departure from the null hypothesis is small ($\boldsymbol{\mu} = \pm \mathbf{0.2}$), as displayed in Tables 5.2-5.3 and Figures 5.2-5.3, Hotelling's T^2 does not appear to have good power to detect the fact that the null hypothesis is false for small and moderate sample sizes (n =10 and 20), across all the levels of $\boldsymbol{\beta}$. When the sample size increases to 50 and number of variables increases to 3, changes in the value of $\boldsymbol{\beta}$ have a sizeable effect on the power of Hotelling's T^2 test in my judgment. Specifically, the larger $\boldsymbol{\beta}$ values correspond to the higher powers. For example, for $\boldsymbol{\mu} = \mathbf{0.2}$, p = 3 and n = 50, the estimated power of Hotelling's T^2 is 0.479 for $\boldsymbol{\beta} = \mathbf{0}$ and 0.911 for $\boldsymbol{\beta} = \mathbf{2}$. However the power curves level off when absolute values of $\boldsymbol{\beta}$ is larger than **2**. We always have the lowest power when $\boldsymbol{\beta} = \mathbf{0}$ under various conditions. This may be do to the somewhat larger than nominal type I error rates indicated in Table 5.1 when $\boldsymbol{\beta} \neq \mathbf{0}$. The sign of $\boldsymbol{\beta}$ appears to have a slight effect on the power for $p = 1$ or $p = 3$. Specifically, the powers for negative $\boldsymbol{\beta}$ are slightly higher than those for positive $\boldsymbol{\beta}$ when the components of the mean vector are positive. The situation reverses for a negative mean vector. We have almost symmetric results for different signs of $\boldsymbol{\beta}$ for $p = 7$.

Table 5.2: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{0.2}$

		$\boldsymbol{\beta}$						
p	*n*	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.221	0.194	0.162	0.087	0.105	0.100	0.097
	20	0.323	0.305	0.254	0.136	0.205	0.230	0.234
	50	0.622	0.590	0.511	0.283	0.504	0.601	0.625
3	10	0.251	0.245	0.219	0.095	0.132	0.135	0.136
	20	0.488	0.485	0.445	0.184	0.382	0.420	0.426
	50	0.886	0.884	0.852	0.479	0.911	0.950	0.954
7	10	0.161	0.162	0.152	0.074	0.111	0.113	0.118
	20	0.622	0.590	0.511	0.219	0.550	0.584	0.588
	50	0.984	0.985	0.983	0.701	0.999	1.000	1.000

Table 5.3: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = -\mathbf{0.2}$

		$\boldsymbol{\beta}$						
p	*n*	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.098	0.098	0.102	0.085	0.155	0.194	0.218
	20	0.235	0.231	0.208	0.132	0.254	0.299	0.318
	50	0.627	0.587	0.504	0.283	0.512	0.591	0.612
3	10	0.140	0.138	0.132	0.096	0.212	0.246	0.258
	20	0.423	0.410	0.377	0.191	0.441	0.478	0.491
	50	0.955	0.947	0.911	0.479	0.851	0.880	0.883
7	10	0.120	0.116	0.114	0.073	0.155	0.162	0.166
	20	0.627	0.587	0.504	0.228	0.572	0.588	0.591
	50	1.000	1.000	0.998	0.709	0.980	0.985	0.988

Figure 5.2: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{0.2}$

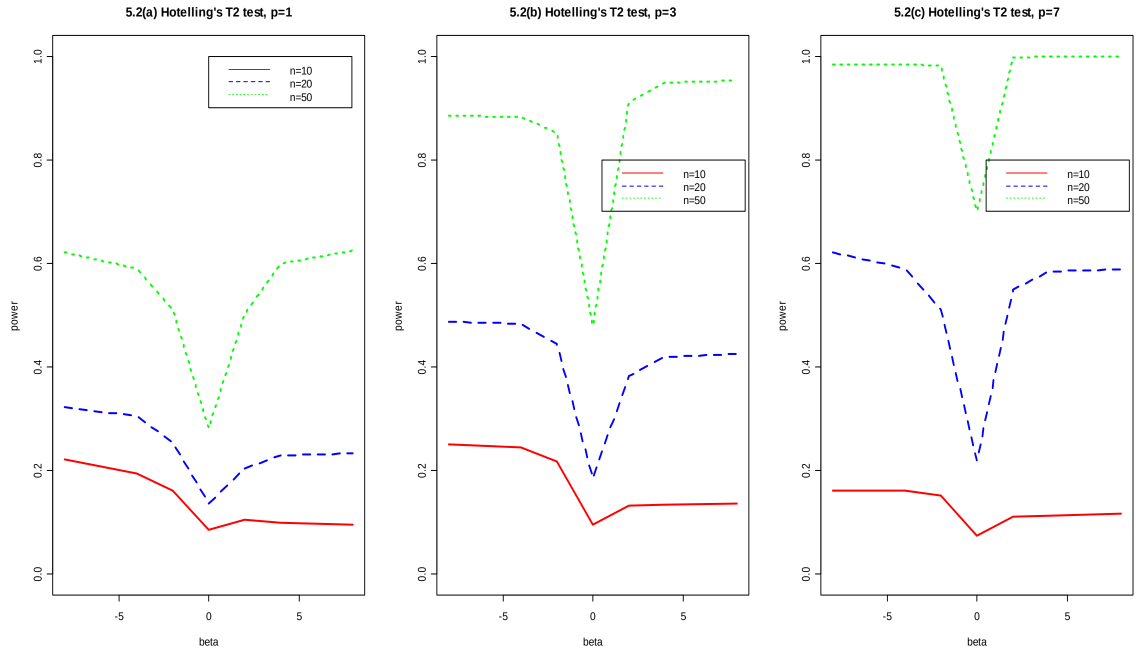
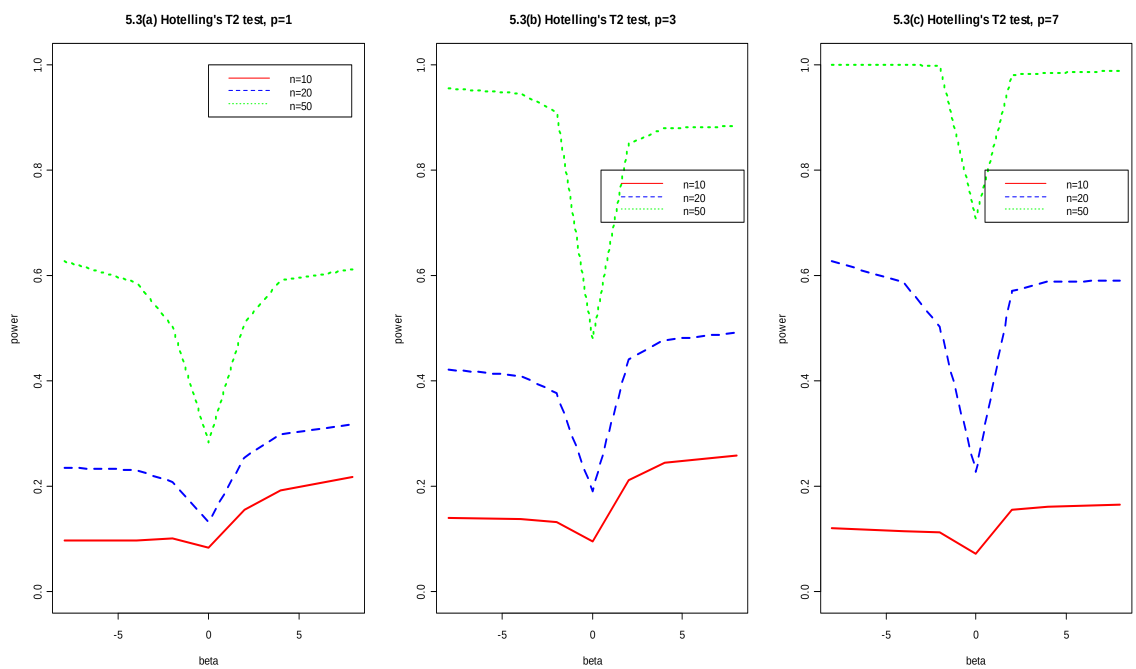


Figure 5.3: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{-0.2}$



5.2.2 *Medium Effect Size* ($\boldsymbol{\mu} = \pm\mathbf{0.5}$)

Tables 5.4-5.5 and Figures 5.4-5.5 show the results obtained when the effect size is medium ($\boldsymbol{\mu} = \pm\mathbf{0.5}$). When the sample size is small (n = 10), excellent power is observed only for p = 3 and $\boldsymbol{\beta}$ is larger than **2** or less than **-2**. As the sample size increases to 20, the increase of $\boldsymbol{\beta}$ has a sizeable effect on the power for p = 1, but not for p =3 or 7. When p is 7, the Hotelling's T^2 test attains the highest power, one, no matter how $\boldsymbol{\beta}$ changes. As the sample size increases to 50, the power curve is almost a straight line equal to one. Contrary to the results from the small effect size, the powers for negative $\boldsymbol{\beta}$ are slightly lower than those of positive $\boldsymbol{\beta}$ for a positive mean vector when p = 3. The situation reverses for negative mean vector. For p = 1 or 7, we have almost symmetric results for different signs of $\boldsymbol{\beta}$.

5.2.3 *Large Effect Size* ($\boldsymbol{\mu} = \pm\mathbf{0.8}$ **and** $\boldsymbol{\mu} = \pm\mathbf{1}$)

Tables 5.6-5.9 and Figures 5.6-5.9 show the results obtained when the effect sizes are what I call large ($\boldsymbol{\mu} = \pm\mathbf{0.8}$ and $\boldsymbol{\mu} = \pm\mathbf{1}$). By looking Figures 5.6-5.9, the powers vary a bit as $\boldsymbol{\beta}$ changes for small sample sizes. But, as the sample size increases to 20, we have very high powers across all the levels of $\boldsymbol{\beta}$ and p. The signs of $\boldsymbol{\beta}$ here only have effect when n = 10 and p = 1.In this case, the powers of negative $\boldsymbol{\beta}$ are slightly lower than those of positive $\boldsymbol{\beta}$ for a positive mean vector and situation reverses for negative mean vector.

Table 5.4: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{0.5}$

		β						
p	n	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.619	0.601	0.524	0.292	0.515	0.620	0.668
	20	0.897	0.884	0.840	0.561	0.880	0.958	0.982
	50	0.999	0.999	0.997	0.934	1.000	1.000	1.000
3	10	0.784	0.770	0.743	0.389	0.832	0.879	0.902
	20	0.991	0.991	0.989	0.834	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	10	0.565	0.554	0.532	0.234	0.524	0.547	0.551
	20	0.999	0.999	0.997	0.941	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.5: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{-0.5}$

		β						
p	n	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.673	0.622	0.529	0.288	0.531	0.531	0.594
	20	0.982	0.957	0.886	0.558	0.837	0.888	0.897
	50	1.000	1.000	0.999	0.939	0.997	0.999	0.999
3	10	0.898	0.882	0.830	0.391	0.751	0.778	0.788
	20	0.999	1.000	1.000	0.840	0.988	0.991	0.993
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	10	0.547	0.543	0.519	0.232	0.533	0.552	0.564
	20	1.000	1.000	0.999	0.944	0.999	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 5.4: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{0.5}$

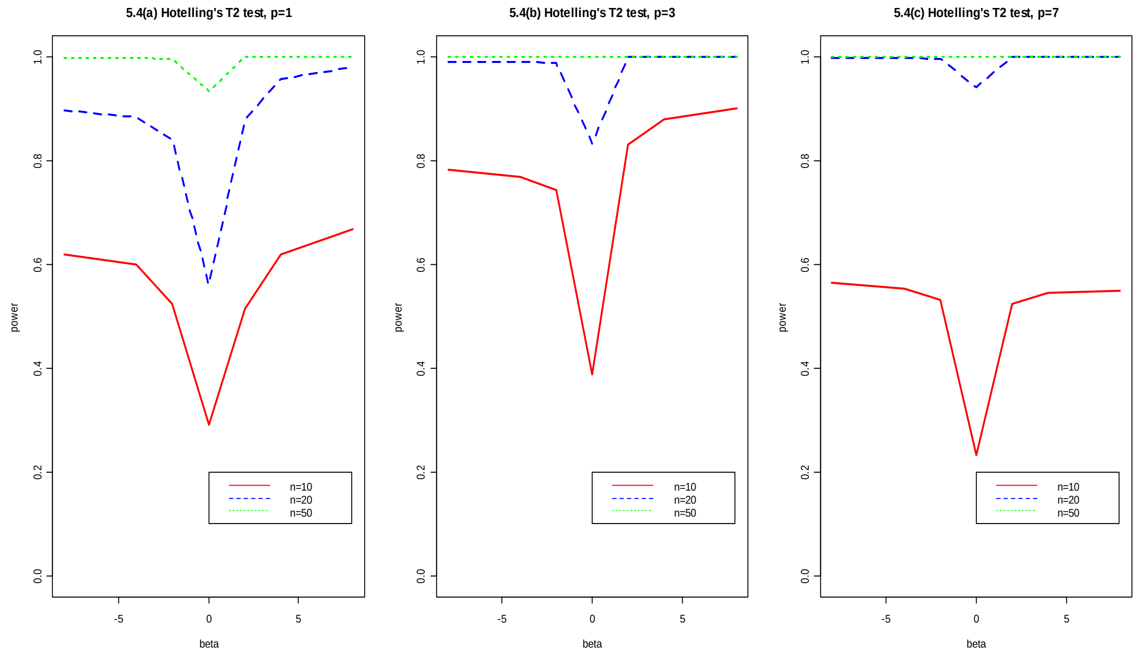


Figure 5.5: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{-0.5}$

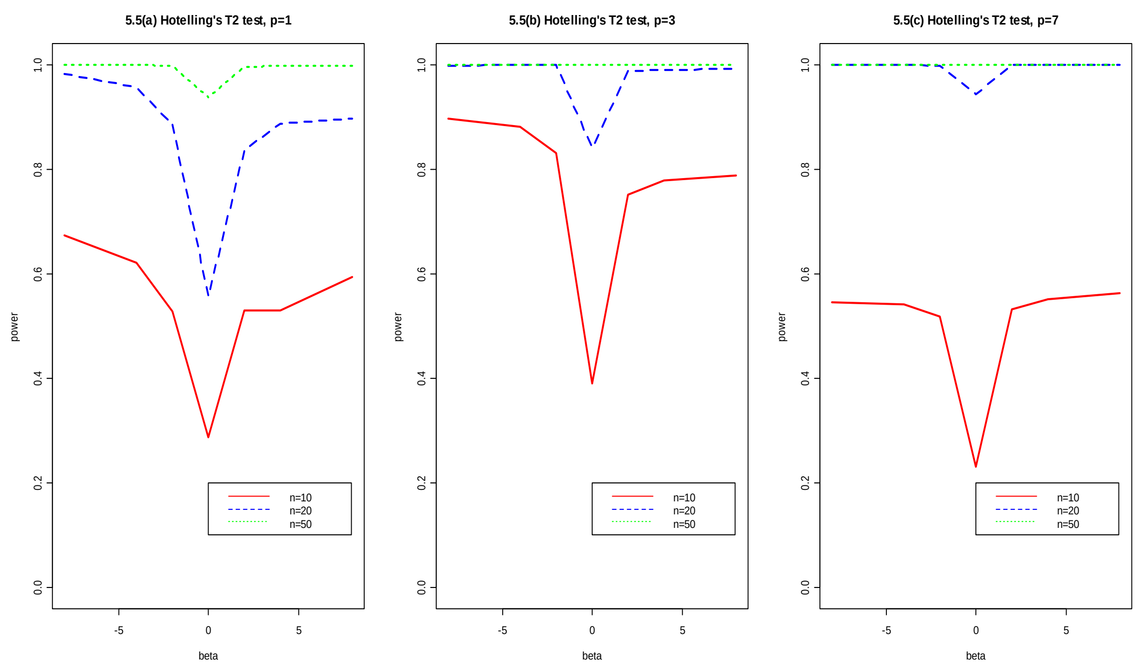


Table 5.6: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{0.8}$

		β						
p	*n*	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.912	0.900	0.912	0.621	0.930	0.986	0.997
	20	0.998	0.998	0.994	0.923	0.999	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	10	0.982	0.981	0.979	0.802	0.999	1.000	1.000
	20	1.000	1.000	1.000	0.998	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	10	0.862	0.865	0.857	0.512	0.889	0.905	0.906
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.7: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{-0.8}$

		β						
p	*n*	**-8**	**-4**	**-2**	**0**	**2**	**4**	**8**
1	10	0.997	0.987	0.997	0.614	0.868	0.902	0.909
	20	1.000	1.000	1.000	0.926	0.994	0.998	0.998
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	10	1.000	1.000	1.000	0.797	0.975	0.980	0.979
	20	1.000	1.000	1.000	0.899	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	10	0.908	0.910	0.891	0.511	0.858	0.862	0.866
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 5.6: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{0.8}$

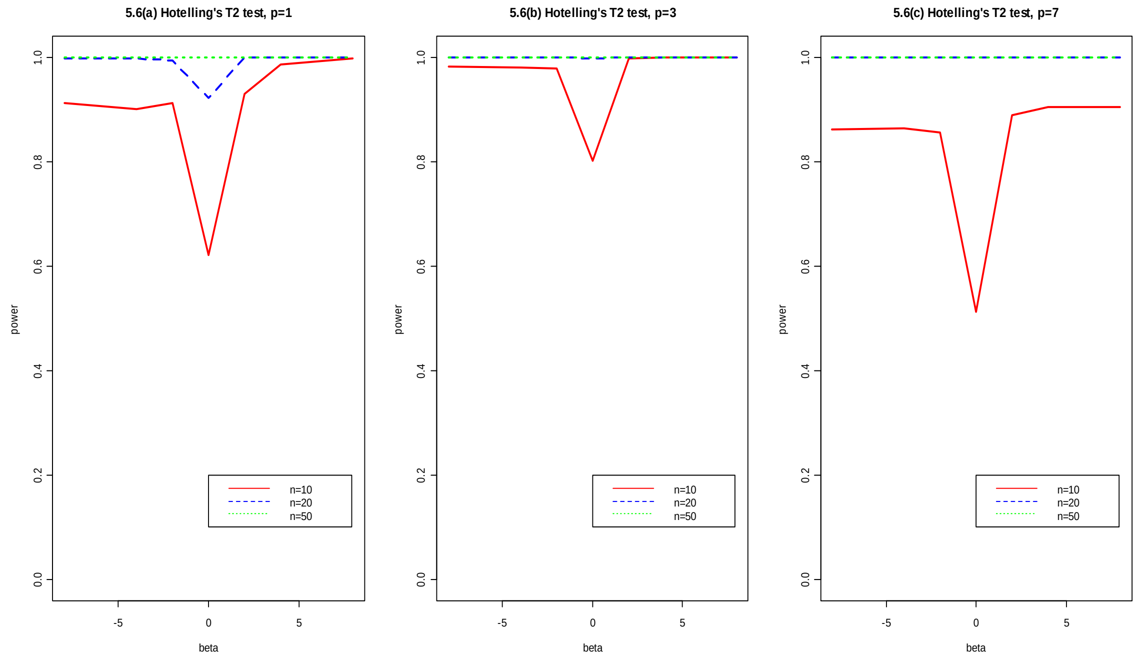


Figure 5.7: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = -\mathbf{0.8}$

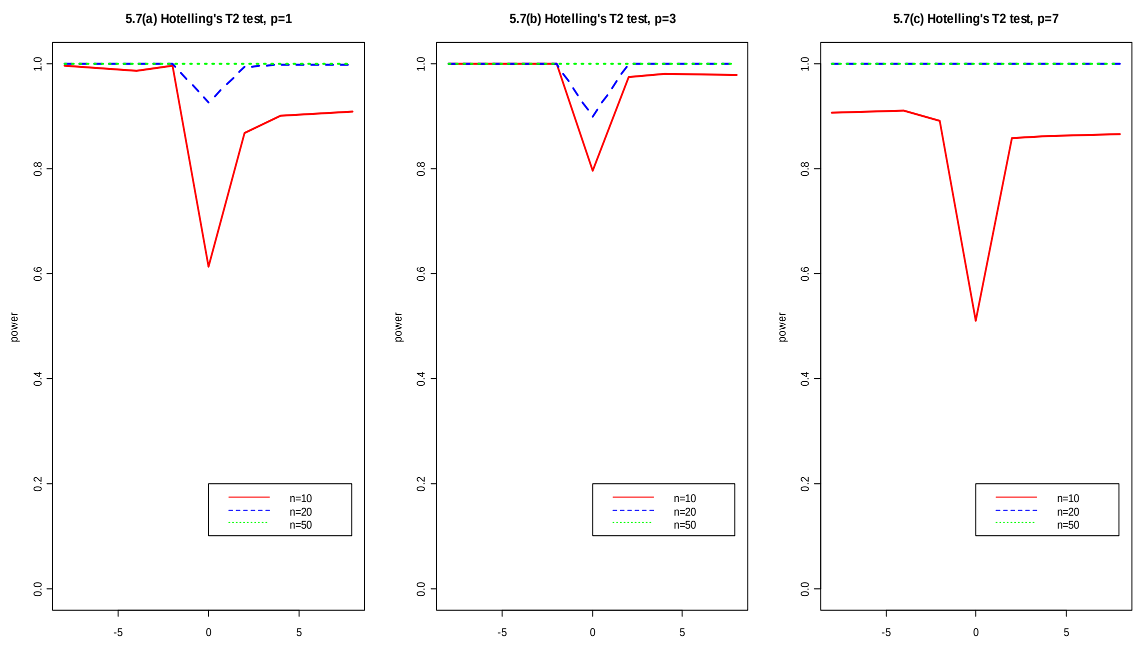


Table 5.8: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{1}$

p	n	β -8	-4	-2	0	2	4	8
1	10	0.977	0.972	0.960	0.804	0.992	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	10	0.997	0.998	0.998	0.942	1.000	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	10	0.956	0.958	0.949	0.688	0.972	0.979	0.978
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.9: Simulated Powers for Multivariate *SN* with $\boldsymbol{\mu} = -\mathbf{1}$

p	n	β -8	-4	-2	0	2	4	8
1	10	1.000	1.000	0.994	0.805	0.958	0.975	0.976
	20	1.000	1.000	1.000	0.991	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	10	1.000	1.000	1.000	0.940	0.999	0.999	0.998
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	10	0.980	0.978	0.971	0.688	0.950	0.952	0.957
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 5.8: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = \mathbf{1}$

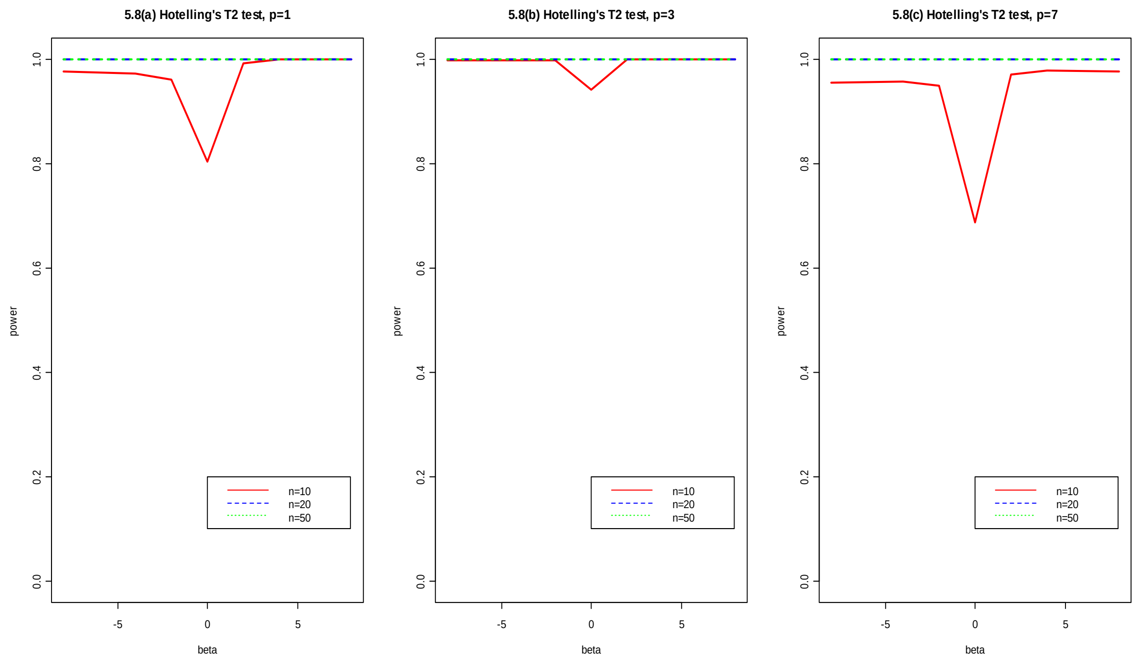
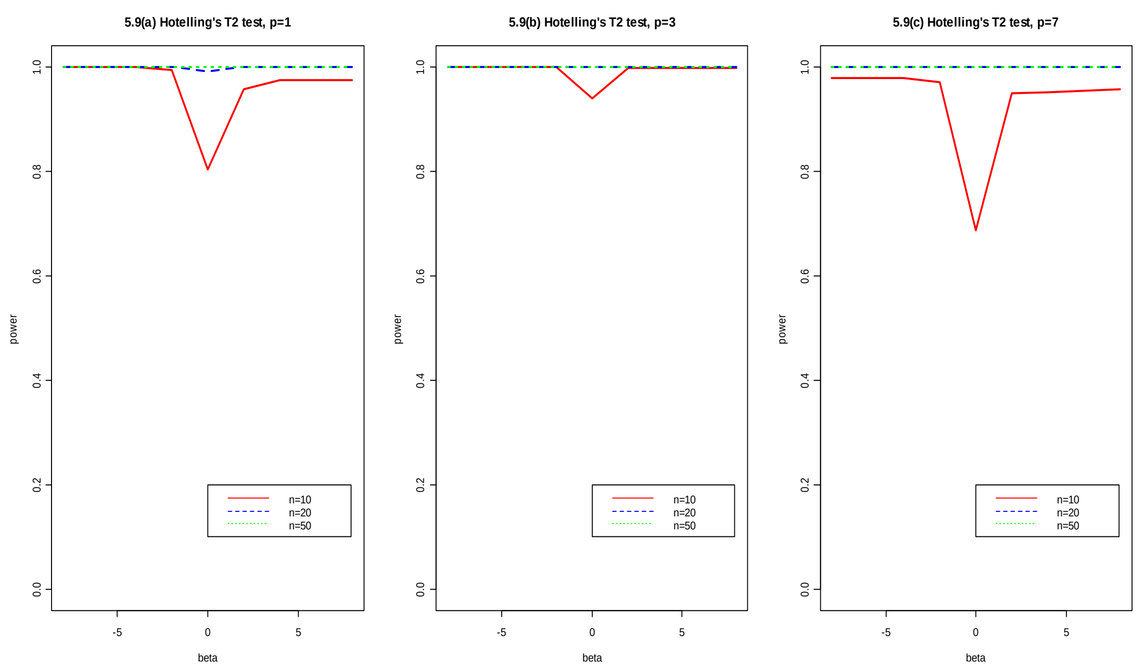


Figure 5.9: Power as Function of $\boldsymbol{\beta}$ for Multivariate *SN* with $\boldsymbol{\mu} = -\mathbf{1}$



5.2.4 Regression Analysis

Analyzing power results in tables and figures is not the only approach. A logistic regression analysis allows us to quantify the effects of the factors: x_1 = the shape parameter (*beta*), x_2 = the number of variables (*p*), x_3 = the sample size (*n*), and x_4 = the common element of the mean vector (*mu*) on power = $\pi(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3, x_4)$, by using estimated power $\hat{\pi}(\mathbf{x})$ as a response variable. For ease of interpretation, we only fit a model with main effects of the form:

$$Logit(\pi(\mathbf{x})) = a + b_1 beta + b_2 p + b_3 n + b_4 mu \tag{5.1}$$

where a and $\{b_i\}$ are constant coefficients, and all the explanatory variables are treated as being quantitative. The quantity $Logit(\pi(\mathbf{x}))$ is the logarithm of the odds $\pi(\mathbf{x})/(1-\pi(\mathbf{x}))$.

Recall that the estimated power curves in Figures 5.2-5.9 are close to being symmetric in beta especially for moderate and large sample sizes, which implies that the effect of the negative *beta* on power is similar to the effect of the positive *beta*. Moreover, the shape of the estimated power curve reverses from left to right if we change the sign of the *mu* from plus to minus, which indicates the signs of *mu* don't have a significant impact on power. The almost symmetric graphs, estimated power plotted against *mu*, shown in Figure 5.10 – 5.12, further verify our findings. The symmetric effects of the *beta* and *mu* on power suggest that we could fit the model (5.1) with data which only includes values of non-negative *beta* and non-negative *mu*.

Figure 5.10: Power as Function of *mu* for Multivariate *SN* with n = 10

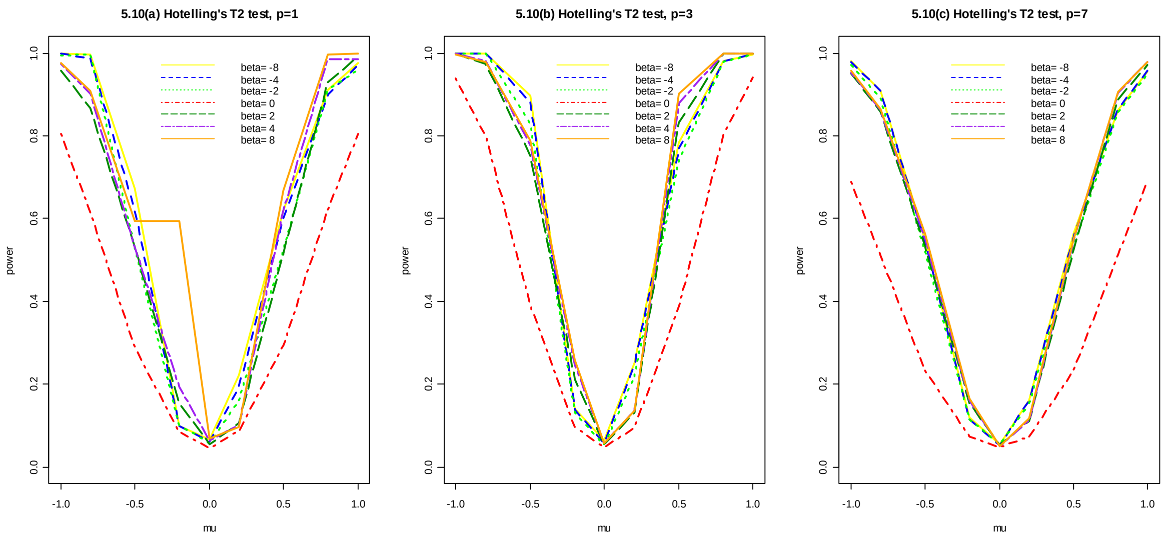


Figure 5.11: Power as Function of *mu* for Multivariate *SN* with n = 20

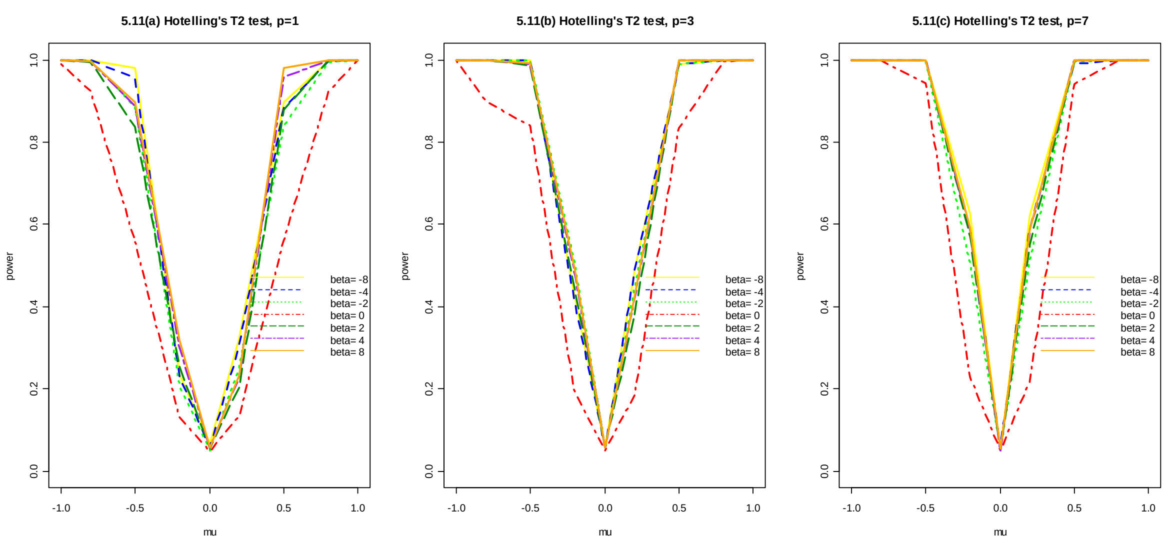
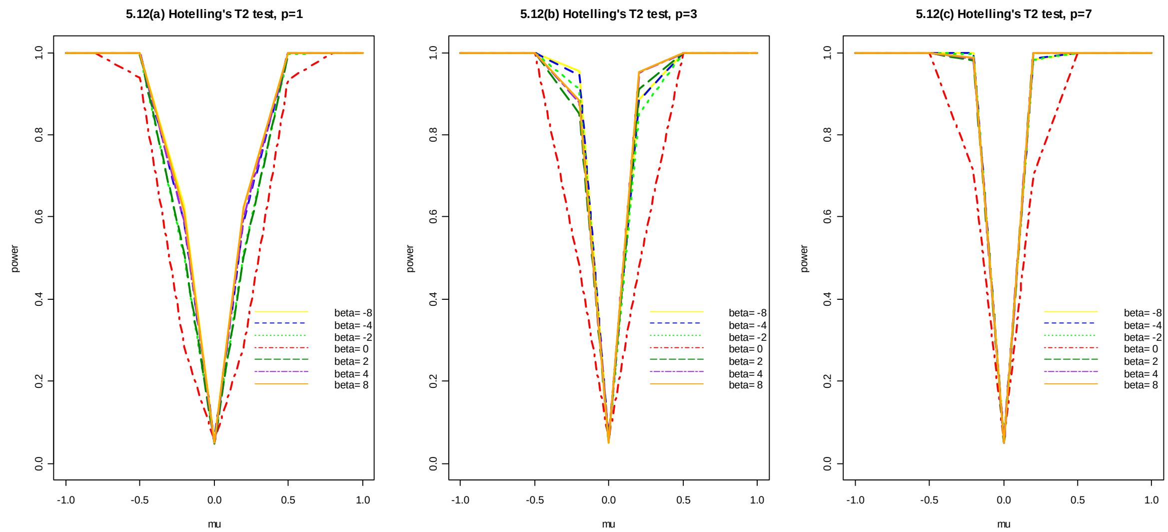


Figure 5.12: Power as Function of *mu* for Multivariate *SN* with n = 50



The *logistic* procedure in SAS was used to conduct the analysis. Eestimated powers ($\hat{\pi}(\mathbf{x})$) of 1 were changed to 0.9999 before doing analysis. The estimated coefficients, $\hat{a}$, $\left\{\hat{b}_i\right\}$, the quantities $\exp(\hat{b}_i)$ and the goodness-of-fit tests of the model are summarized in Table 5.10. Recall that we estimate that a unit increase in x_i while holding the other explanatory variables fixed, multiplies the power odds by $\exp(\hat{b}_i)$. Thus, for example, we see from Table 5.10 that we estimate for $\boldsymbol{\beta} \geq \mathbf{0}$ and $\boldsymbol{\mu} \geq \mathbf{0}$, increasing sample size by 10, the other factors remaining fixed, results in multiplying the power odds by 10.76. Although Hosmer and Lemeshow goodness-of-fit test for model gives large Chi-Square values, 29501.43 with 8 degrees of freedom, which suggests that the model doesn't fit decently. We still could obtain some information about the relationship between power and other factors by looking at the signs and values of the coefficients. The positive signs of the coefficients for p and n in the model provide evidence that power increases as the number of variables or the sample size increases.

These results are consistent with what is seen by looking at the graphs 5.2-5.9. There is also evidence that the effect size, *mu*, which has an estimated odds ratio larger than 999.99, has an important impact on the power. The positive coefficient of *beta* in model indicates that increasing magnitude of shape parameter would increase the power. Specifically, the estimated odds of power multiply by 1.156 for each unit increase in the shape parameter. Overall the regression analysis gives similar results to those discussed in Section 5.2.1-3.

Table 5.10: Results for the Model Fitting with $\boldsymbol{\beta} \geq \mathbf{0}$ and $\boldsymbol{\mu} \geq \mathbf{0}$

Variable	Estimate	Exp(Est)	Chi-Square	P_Value
intercept	-4.624		172218.758	<.0001
mu	7.492	>999.99	266591.71	<.0001
beta	0.145	1.156	20961.517	<.0001
p	0.045	1.046	1541.215	<.0001
n	0.074	1.076	184156.475	<.0001

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr-Chisq
29501.43	8	<.0001

5.3 Summary

For nominal type I error rate 0.05, the attained type I error rates under sampling from the multivariate skew distribution are slightly high, but satisfactory from a practical standpoint. In terms of power, the Hotelling's T^2 test was observed to have high power as the sample size and number of variables increase. Increases in the common component of $\boldsymbol{\beta}$ have a more positive impact on power for small and moderate effect size than those for large effect size. As a whole, the one sample Hotellng T^2 test performs best in terms of power when the common component of $\boldsymbol{\beta}$ increases and surprisingly, worst when $\boldsymbol{\beta} = \mathbf{0}$ under all conditions. Furthermore, the sign of the common component of $\boldsymbol{\beta}$ does not appear to play a big role in attained size or power.

CHAPTER 6 – Discussion

The purpose of this study was to investigate the type I error rate and power properties of Hotelling's one sample T^2 test when sampling from multivariate skew-normal distributions. Simulation results obtained from various levels of shape parameters, sample sizes, number of variables and fixed correlation matrix for nominal level 0.05 were used to determine the performance of the test.

The results from Chapter 5 indicate that Hotelling's one sample T^2 test provides adequate control of type I error rates over the entire range of conditions studied. The test also produces suitable power levels for detecting departures from hypothesized values of a multivariate mean vector when data results from a random sample from a multivariate *SN*. Shape parameter appears not affect the robustness of Hotelling's T^2 test. This may be due to the range of the skewness for beta is from -0.9953 to 0.9953 in the univariate *SN* distributions. Whereas, as the value of shape parameter increases, Hotelling's T^2 test generally shows increased power under all the situations invested in our study.

As with any simulation study, the choices of parameter settings may limit the generalization of the results. The choice of $\mathbf{\Omega}$ be identity matrix, although a mathematical convenience, does not cover all the possible situations. The nominal level 0.05 is also a limited choice. The analysis may have yield different values if we had conducted a study of 0.1 or .01 levels and changed the values of $\mathbf{\Omega}$. Further studies that examine all three alpha levels and other values of $\mathbf{\Omega}$ would give a much more complete view of Hotelling's one sample T^2 test performance.

References

Arellano-Valle, R. B. (2006). On the Unification of Families of Skew-Normal Distributions. *Scandinavian Journal of Statistics, Theory and Applications*, 33(3), 561.

Arnord, H.J. (1964). Permutation Support for Multivariate Techniques. *Biometrika*, 51(1-2), 65.

Azzalini, A (2005). The Skew-Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics, Theory and Applications*, 32(2), 159.

Azzalini, A. (1985). A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics, Theory and Applications*, 12(2), 171.

Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society. Series B, Statistical Methodology,* 61, 579.

Azzalini, A. and Dalla Valle, A. (1996). The Multivariate Skew-Normal Distribution. *Biometrika*, 83(4), 715.

Chase, G. R. and Bulgren, W. G. (1971). A Monte Carlo Investigation of the Robustness of T^2 . *Journal of the American Statistical Association*, 66(335), 499.

Everitt, B. S. (1979). A Monte Carlo Investigation of the Robustness of Hotelling's One- and Two-Sample T^2 Tests. *Journal of the American Statistical Association*, 74(365), 48.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall/CRC

Henze, N. (1986). A Probabilistic Representation of the Skew-Normal Distribution. *Scandinavian Journal of Statistics, Theory and Applications*, 13(4), 271.

Kariya, T. (1981). A Robustness Property of Hotelling's T^2 -test. *The Annals of Statistics*, 9(1), 211-214.

Mardia, K. V. (1970). Measure of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, 57(3), 519.

Mardia, K. V. (1975). Assessment of Multinormality and the Robustness of Hotelling's T^2 . *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 24(2), 163.

Appendix A - R code for Simulation Experiment

The following R code is for generating multivariate *SN* random samples and calculating type I error rates and powers for simulated data.

*## The following function '*rsnp*' generates multivariate SN, where* **n** *is the sample size;* **mux** *is the specified mean vector,* **beta** *is the shape parameter,* **Omega** *is the correlation matrix*.

```
rsnp = function(n=1, mux=rep(0,length(beta)), beta, Omega)
{
	p = length(beta)
	p1 = p + 1
	z0 = matrix(rep(0,n),n,1)
	z = matrix(rep(0,p*n),n,p)
	x = matrix(rep(0,p*n),n,p)
	Z = matrix(rep(0,n),n,1)
	mu = rep(0, p1)
	tmp = as.vector(sqrt(1 + t(as.matrix(beta))%*%Omega%*%beta))
	delta = as.vector(Omega %*%beta)/tmp
	delta1 = c(1,delta)
	om = cbind(delta,Omega,deparse.level = 0)
	omega = rbind(delta1,om,deparse.level = 0)
	if(det(omega)< 0) stop("omega must be positive definite matrix")
	for (i in (1:n))
	{
		Z = matrix(matrix(rnorm(p1),1,p1) %*% chol(omega),p1,1)
		z0[i,1] = Z[1,1]
		if (z0[i, 1]>0) z[i, 1:p]=Z[2:p1, 1] else z[i, 1:p] = -Z[2:p1, 1]
```

```
        x[i, 1:p] = z[i, 1:p] + mux - sqrt(2/pi)*delta
    }
    return (x)
}
```

The following function ‘HT2_test’ *calculates type I error rates as* **mux** = **0** *and powers as* **mux** *is away from* ***0****, where* **N** *is the simulation size; other parameters are the same as those in function* ‘rsnp’.

Note: Function ‘HotellingsT2’ *is in the package* ‘ICSNP’.

```
library(ICSNP)

HT2_test = function(N, n, mux, beta, Omega)
{
    p.value=numeric()
    p.value=0
    for (j in c(1:N))
    {
        x1 = rsnp(n,mux,beta,Omega)
        result = HotellingsT2(x1, mu=rep(0,length(beta)),test="f")
        p.value[j] = result[[2]]
    }
    test = length(p.value[p.value<0.05])/N
    return(test)
}
```

Appendix B - R code for Figures

All graphs showed in my report are produced by using R language. Figures 3.2 – 3.7 require downloading package ‘sn’.

Figure 3.1

```
sn_01 = function(z, beta=0)

{

    y = 2*dnorm(z)*pnorm(z*beta)
        return (y)

}

z = seq(-5, 5, 0.1)

y1 = sn_01(z, beta=5)

y2 = sn_01(z, beta=-2)

y3 = sn_01(z, beta=0)

plot(c(z,z), c(y1,y2), xlim=c(-4,4), ylim=c(0,0.8), xlab="z", ylab="Skew-Normal
Density",type="n")

lines(z,y1,lty=1)

lines(z,y2,lty=2)

lines(z,y3,lty=3)

legend(-4,0.8, c("beta=5","beta=-2","beta=0"), lty=c(1,2,3))
```

Figure 3.2

Figure 3.3 and Figures D.1-D.4 use the same code as Figure 3.2 except changing the values of beta(shape parameter) and Omega(correlation matrix)

```
# Download package 'sn'
library(sn)

Omega = matrix(c(1,0.75,0.75,1),2)

beta = c(-5,10)

tmp = as.vector(sqrt(1 + t(as.matrix(beta))%*%Omega%*%beta))

delta = as.vector(Omega %*%beta)/tmp

muZ = delta*sqrt(2/pi)

omega = Omega-outer(muZ,muZ)

# Function 'dsn2.plot' being used to produce contour plot
x = y = seq(-3, 3, length=100)

dsn2.plot(x, y, c(0,0), omega, beta, nlevels = 6, ylim = c(-2,2), xlim = c(-2,2), ylab =
"z2",xlab = "z1")
```

*Figure 4.1*

```
skewness.sn = function(n, beta)

{

        for (i in n)

        {

        gamma_1 = 0.5*(4-pi)*sign(beta)*sqrt((beta^2/(pi/2+(pi/2-1)*beta^2))^3)

        gamma_1[i] = gamma_1[i]

        }
```

```
	return (gamma_1)

}

n = 10

beta = c(-n:n)

skewness = skewness.sn(n, beta)

plot(beta, skewness, type = "b")
```

Figure 5.1

Note: t.10, t.20 and t.50 are the empirical type I error rates when beta changes from -8 to 8 and sample sizes are 10, 20 and 50, respectively.

```
par(mfrow = c(3,1))
# Type I error rate plots for multivariate SN under study: p = 1, n = 10, 20, 50
beta = c(-8,-4,-2,0,2,4,8)
t.10 = c(0.0674, 0.0607, 0.057, 0.0454, 0.0541, 0.0642, 0.0675)
t.20 = c(0.0613, 0.0576, 0.0492, 0.0469, 0.0535, 0.0584, 0.0573)
t.50 = c(0.0526, 0.0545, 0.0521, 0.054, 0.0483, 0.0489, 0.05)
p.level = rep(0.05, 7)
plot(c(beta, beta),c(t.10, t.20),ylim = c(0.04, 0.07), xlab = "beta", ylab = "estimated type I
error rate", type= "n", main = "Hotelling's T2 test, p=1")
lines(beta, t.10, lty = 1, col = "red", lwd = 2)
lines(beta, t.20, lty = 2, col = "blue", lwd = 2)
lines(beta, t.50, lty = 3, col = "green", lwd = 2)
lines(beta, p.level, lty = 4,)
legend(0, 0.07, c("n=10","n=20","n=50"), lty = c(1,2,3), col= c("red", "blue", "green"))
text(-5,0.05,"alpha=0.05")


# Type I error rate plots for multivariate SN under study: p = 3, n = 10, 20, 50
```

```
beta = c(-8,-4,-2,0,2,4,8)
t.10 = c(0.0625, 0.0563, 0.0528, 0.0473, 0.0545, 0.0555, 0.0564)
t.20 = c(0.0572, 0.0568, 0.0564, 0.0488, 0.0588, 0.0587, 0.0576)
t.50 = c(0.0564, 0.054, 0.0514, 0.0556, 0.0533, 0.0543, 0.0508)
p.level = rep(0.05, 7)
plot(c(beta, beta),c(t.10, t.20),ylim = c(0.04, 0.07), xlab = "beta", ylab = " estimated type
I error rate ", type= "n", main = "Hotelling's T2 test, p=3")
lines(beta, t.10, lty = 1, col = "red", lwd = 2)
lines(beta, t.20, lty = 2, col = "blue", lwd = 2)
lines(beta, t.50, lty = 3, col = "green", lwd = 2)
lines(beta, p.level, lty = 4,)
legend(0, 0.07, c("n=10","n=20","n=50"), lty = c(1,2,3), col= c("red", "blue", "green"))
text(-5,0.05,"alpha=0.05")

# Type I error rate plots for multivariate SN under study: p = 7, n = 10, 20, 50
beta = c(-8,-4,-2,0,2,4,8)
t.10 = c(0.0542, 0.0528, 0.0543, 0.0471, 0.0518, 0.0495, 0.0503)
t.20 = c(0.0526, 0.0545, 0.0521, 0.0498, 0.053, 0.0504, 0.0551)
t.50 = c(0.0523, 0.0523, 0.0524, 0.0528, 0.0509, 0.0499, 0.051)
p.level = rep(0.05, 7)
plot(c(beta, beta),c(t.10, t.20),ylim = c(0.04, 0.07), xlab = "beta", ylab = "estimated type
Ierror rate ", type= "n", main = "Hotelling's T2 test, p=7")
lines(beta, t.10, lty = 1, col = "red", lwd = 2)
lines(beta, t.20, lty = 2, col = "blue", lwd = 2)
lines(beta, t.50, lty = 3, col = "green", lwd = 2)
lines(beta, p.level, lty = 4,)
legend(2, 0.07, c("n=10","n=20","n=50"), lty = c(1,2,3), col= c("red", "blue", "green"))
text(-5,0.05,"alpha=0.05")

#### Figure 5.2 ####
```

Note: p.10, p.20 and p.50 are the empirical powers when beta changes from -8 to 8 and sample sizes are 10, 20 and 50, respectively. Figures 5.3 – 5.9 basically are produced by using the same code below except changing the values of p.10, p.20, p.50 and some arguments in functions 'plot' and 'legend'.

 # Power plots for multivariate SN under study: mu = 0.2, p = 1, n = 10,20,50

```
par(mfrow = c(1,3))

beta = c(-8,-4,-2,0,2,4,8)

p.10 = c(0.2213, 0.1938, 0.1619, 0.0866, 0.1048, 0.1002, 0.0965)

p.20 = c(0.3231, 0.3053, 0.2542, 0.1355, 0.2049, 0.2302, 0.2338)

p.50 = c(0.6223, 0.5901, 0.5114, 0.2826, 0.5041, 0.6009, 0.6252)

plot(c(beta, beta),c(p.10, p.20),ylim = c(0, 1), xlab = "beta", ylab = "power", type= "n",
main = "5.2(a) Hotelling's T2 test, p=1")

lines(beta, p.10, lty = 1, col = "red", lwd = 2)

lines(beta, p.20, lty = 2, col = "blue", lwd = 2)

lines(beta, p.50, lty = 3, col = "green", lwd = 2)

legend(0, 1, c("n=10","n=20","n=50"), lty = c(1,2,3), col= c("red", "blue", "green"))
```

 # Power plots for multivariate SN under study: mu=0.2, p=3, n=10,20,50

```
beta = c(-8,-4,-2,0,2,4,8)

p.10 = c(0.2505, 0.2453, 0.2188, 0.0947, 0.1317, 0.1346, 0.1362)

p.20 = c(0.4883, 0.4846, 0.4446, 0.184, 0.3824, 0.4202, 0.4262)

p.50 = c(0.8862, 0.8835, 0.8517, 0.4794, 0.911, 0.9496, 0.9543)

plot(c(beta, beta),c(p.10, p.20),ylim = c(0, 1), xlab = "beta", ylab = "power", type= "n",
main = "5.2(b) Hotelling's T2 test, p=3")

lines(beta, p.10, lty = 1, col = "red", lwd = 2)

lines(beta, p.20, lty = 2, col = "blue", lwd = 2)

lines(beta, p.50, lty = 3, col = "green", lwd = 2)
```

```
legend(0.5, 0.8, c("n=10","n=20","n=50"), lty = c(1,2,3), col= c("red", "blue", "green"))


# Power plots for multivariate SN under study:mu=0.2, p=7, n=10,20,50

beta = c(-8,-4,-2,0,2,4,8)

p.10 = c(0.1611, 0.162, 0.1522, 0.0744, 0.1105, 0.113, 0.1178)

p.20 = c(0.6223, 0.5901, 0.5114, 0.2192, 0.5498, 0.5839, 0.5877)

p.50 = c(0.9844, 0.9851, 0.9828, 0.7007, 0.9989, 0.9997, 1)

plot(c(beta, beta),c(p.10, p.20),ylim = c(0, 1), xlab = "beta", ylab = "power", type= "n",
main = "5.2(c) Hotelling's T2 test, p=7")

lines(beta, p.10, lty = 1, col = "red", lwd = 2)

lines(beta, p.20, lty = 2, col = "blue", lwd = 2)

lines(beta, p.50, lty = 3, col = "green", lwd = 2)

legend(0.5, 0.8, c("n=10","n=20","n=50"), lty = c(1,2,3), col= c("red", "blue", "green"))


#### Figure 5.10 ####

## Note: beta_8 to beta8 are the empirical powers when mu changes from -1 to 1 and
beta are -8 to 8, respectively. Figures 5.11 – 5.12 basically are produced by using the
same code below except changing the values of beta_8 to beta8 and some arguments in
functions 'plot' and 'legend'.


# Power plot for multivariate SN under study: n=10, p=1,

par(mfrow = c(1,3))

mu = c(-1,-0.8,-0.5,-0.2,0,0.2,0.5,0.8,1)

beta_8 = c(1.000,0.997,0.673,0.098,0.067,0.221,0.619,0.912,0.977)

beta_4 = c(1.000,0.987,0.622,0.098,0.061,0.194,0.601,0.900,0.972)

beta_2 = c(0.994,0.997,0.529,0.102,0.057,0.162,0.524,0.912,0.960)

beta0 = c(0.805,0.614,0.288,0.085,0.045,0.087,0.292,0.621,0.804)
```

```
beta2 = c(0.958,0.868,0.531,0.155,0.054,0.105,0.515,0.930,0.992)
beta4 = c(0.975,0.902,0.531,0.194,0.064,0.100,0.620,0.986,0.986)
beta8 = c(0.976,0.909,0.594,0.594,0.068,0.097,0.668,0.997,1.000)

plot(c(mu,mu),c(beta_8,beta8),ylim = c(0, 1), xlab = "mu", ylab = "power", type= "n",
main = "5.10(a) Hotelling's T2 test, p=1")
lines(mu,beta_8, lty = 1, col = "yellow", lwd = 2)
lines(mu,beta_4, lty = 2, col = "blue", lwd = 2)
lines(mu,beta_2, lty = 3, col = "green", lwd = 2)
lines(mu,beta0, lty = 4, col = "red", lwd = 2)
lines(mu,beta2, lty = 5, col = "green4", lwd = 2)
lines(mu,beta4, lty = 6, col = "purple", lwd = 2)
lines(mu,beta8, lty = 7, col = "orange", lwd = 2)
legend(-0.5, 1, c("beta= -8","beta= -4","beta= -2","beta= 0","beta= 2", "beta= 4","beta=
8"), lty = c(1,2,3,4,5,6,7,8), col= c("yellow", "blue", "green","red", "green4", "purple",
"orange"), bty="n")

# Power plot for multivariate SN under study: n=10, p=3,
mu = c(-1,-0.8,-0.5,-0.2,0,0.2,0.5,0.8,1)
beta_8 = c(1.000,1.000,0.898,0.140,0.063,0.251,0.784,0.982,0.997)
beta_4 = c(1.000,1.000,0.882,0.138,0.056,0.245,0.770,0.981,0.998)
beta_2 = c(1.000,1.000,0.830,0.132,0.053,0.219,0.743,0.979,0.998)
beta0 = c(0.940,0.797,0.391,0.096,0.047,0.095,0.389,0.802,0.942)
beta2 = c(0.999,0.975,0.751,0.212,0.055,0.132,0.832,0.999,1.000)
beta4 = c(0.999,0.980,0.778,0.246,0.056,0.135,0.879,1.000,1.000)
beta8 = c(0.998,0.979,0.788,0.258,0.056,0.136,0.902,1.000,1.000)

plot(c(mu,mu),c(beta_8,beta8),ylim = c(0, 1), xlab = "mu", ylab = "power", type= "n",
main = "5.10(b) Hotelling's T2 test, p=3")
lines(mu,beta_8, lty = 1, col = "yellow", lwd = 2)
lines(mu,beta_4, lty = 2, col = "blue", lwd = 2)
```

```
lines(mu,beta_2, lty = 3, col = "green", lwd = 2)
lines(mu,beta0, lty = 4, col = "red", lwd = 2)
lines(mu,beta2, lty = 5, col = "green4", lwd = 2)
lines(mu,beta4, lty = 6, col = "purple", lwd = 2)
lines(mu,beta8, lty = 7, col = "orange", lwd = 2)
legend(-0.5, 1, c("beta= -8","beta= -4","beta= -2","beta= 0","beta= 2", "beta= 4","beta=
8"), lty = c(1,2,3,4,5,6,7,8), col= c("yellow", "blue", "green", "red", "green4", "purple",
"orange"), bty="n")

# Power plot for multivariate SN under study: n=10, p=7,
mu = c(-1,-0.8,-0.5,-0.2,0,0.2,0.5,0.8,1)
beta_8 = c(0.980,0.908,0.547,0.120,0.054,0.161,0.565,0.862,0.956)
beta_4 = c(0.978,0.910,0.543,0.116,0.053,0.162,0.554,0.865,0.958)
beta_2 = c(0.971,0.891,0.519,0.114,0.054,0.152,0.532,0.857,0.949)
beta0 = c(0.688,0.511,0.232,0.073,0.047,0.074,0.234,0.512,0.688)
beta2 = c(0.950,0.858,0.533,0.155,0.052,0.111,0.524,0.889,0.972)
beta4 = c(0.952,0.862,0.552,0.162,0.050,0.113,0.547,0.905,0.979)
beta8 = c(0.957,0.866,0.564,0.166,0.050,0.118,0.551,0.906,0.978)

plot(c(mu,mu),c(beta_8,beta8),ylim = c(0, 1), xlab = "mu", ylab = "power", type= "n",
main = "5.10(c) Hotelling's T2 test, p=7")
lines(mu,beta_8, lty = 1, col = "yellow", lwd = 2)
lines(mu,beta_4, lty = 2, col = "blue", lwd = 2)
lines(mu,beta_2, lty = 3, col = "green", lwd = 2)
lines(mu,beta0, lty = 4, col = "red", lwd = 2)
lines(mu,beta2, lty = 5, col = "green4", lwd = 2)
lines(mu,beta4, lty = 6, col = "purple", lwd = 2)
lines(mu,beta8, lty = 7, col = "orange", lwd = 2)
```

```
legend(-0.5, 1, c("beta= -8","beta= -4","beta= -2","beta= 0","beta= 2", "beta= 4","beta= 8"), lty = c(1,2,3,4,5,6,7,8), col= c("yellow", "blue", "green", "red", "green4", "purple", "orange"), bty="n")
```

Appendix C - SAS code for Logistic Regression Analysis

The following SAS code is for power analysis in Section 5.2.4

```
* Import data from excel file

proc import out=analyze_data_01
datafile="G:\MS_file\simulation\analyze_data_01.xls"
DBMS=EXCEL REPLACE;
run;

* Changing power 1 to 0.9999

data power_data_01;
set analyze_data_01;
if accept=10000 then accept=9999;
run;

* Logistic Regression for beta ≥ 0 and mu ≥ 0

data pos_beta_mu;
set power_data_01;
where beta in (0, 2 ,4, 8) and mu in (0, 0.2, 0.5, 0.8, 1);
run;

proc logistic data = pos_beta_mu;
model accept/size = mu beta p n / lackfit;
run;
```

Appendix D - Contour Plots for Multivariate *SN*

Figure 5.13: Contour plot of the SN_2 for β_1 = 10, β_2 = -5 and ω = 0.5

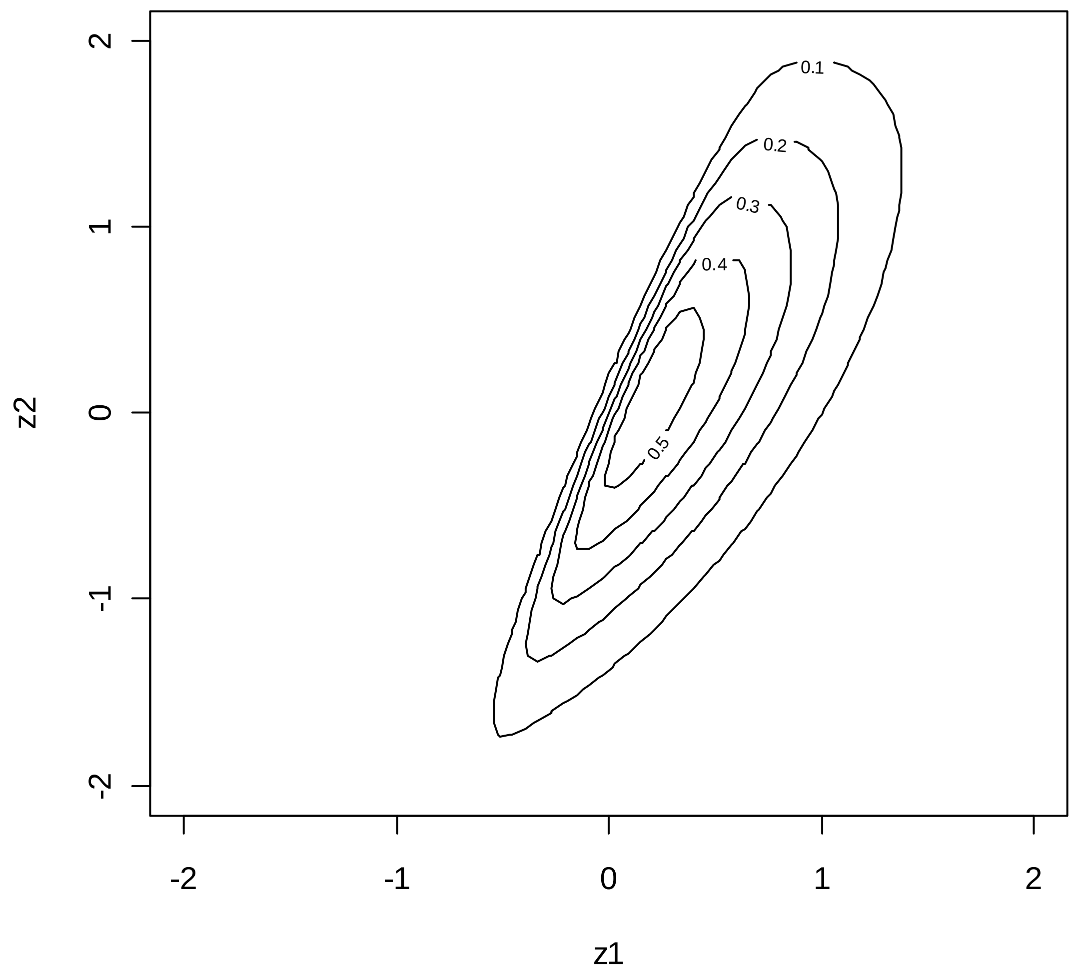


Figure 5.14: Contour plot of the SN_2 for β_1 = 0, β_2 = -5 and ω = 0

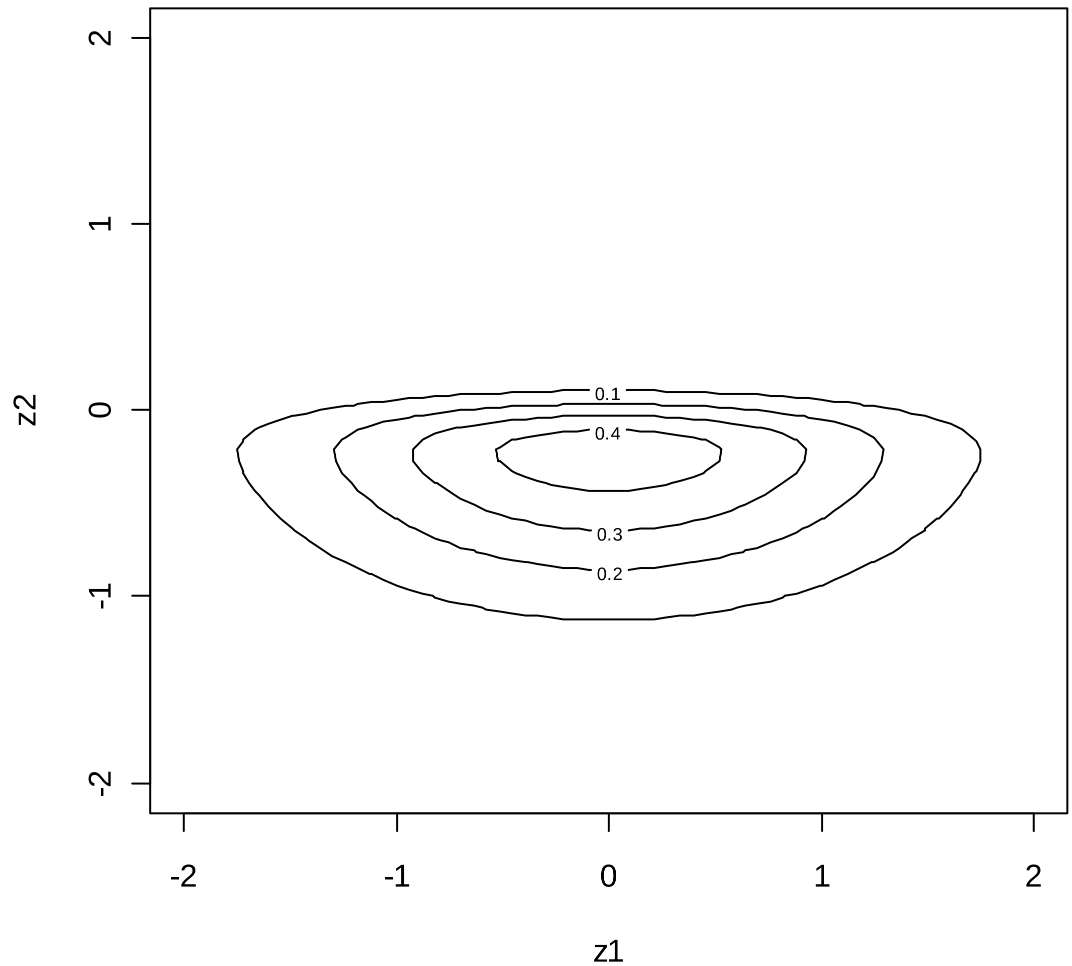


Figure 5.15: Contour plot of the SN_2 for $\beta_1 = 2$, $\beta_2 = 5$ and $\omega = 0.2$

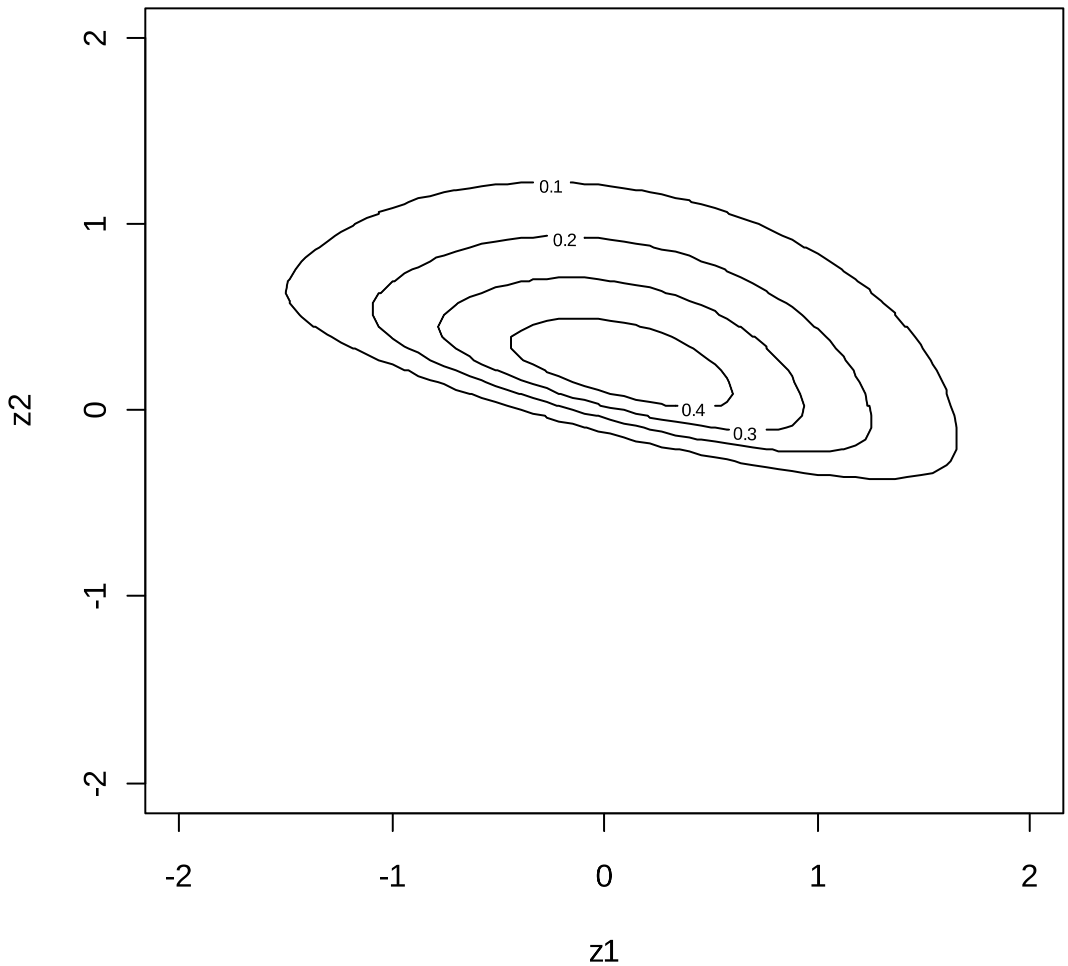


Figure 5.16: Contour plot of the SN_2 for $\beta_1 = -2$, $\beta_2 = -5$ and $\omega = 0.2$