

Using Pearson's System of Curves to Approximate the Distributions of the Difference between Two Correlated Estimates of Signal-to-Noise Ratios: The Cases of Bivariate Normal and Bivariate Lognormal Distributions

Mohamed M. Shoukri

Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada

Email: mmshoukr@uwo.ca, Shoukri.mohamed@gmail.com

How to cite this paper: Shoukri, M.M. (2024) Using Pearson's System of Curves to Approximate the Distributions of the Difference between Two Correlated Estimates of Signal-to-Noise Ratios: The Cases of Bivariate Normal and Bivariate Lognormal Distributions. *Open Journal of Statistics*, 14, 207-227.

<https://doi.org/10.4236/ojs.2024.143010>

Received: April 15, 2024

Accepted: May 27, 2024

Published: May 30, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Background: The signal-to-noise ratio (SNR) is recognized as an index of measurements reproducibility. We derive the maximum likelihood estimators of SNR and discuss confidence interval construction on the difference between two correlated SNRs when the readings are from bivariate normal and bivariate lognormal distribution. We use the Pearson's system of curves to approximate the difference between the two estimates and use the bootstrap methods to validate the approximate distributions of the statistic of interest.

Methods: The paper uses the delta method to find the first four central moments, and hence the skewness and kurtosis which are important in the determination of the parameters of the Pearson's distribution. **Results:** The approach is illustrated in two examples; one from veterinary microbiology and food safety data and the other on data from clinical medicine. We derived the four central moments of the target statistics, together with the bootstrap method to evaluate the parameters of Pearson's distribution. The fitted Pearson's curves of Types I and II were recommended based on the available data. The R-codes are also provided to be readily used by the readers.

Keywords

Signal-to-Noise Ratio, Bivariate Distributions, Bootstrap Methods, Delta Method, Pearson System of Curves

1. Introduction

Signal-to-noise is an important measure of quality measurements and it has applications in many fields including medicine, engineering and genomics. Generally speaking, a signal is what the investigator tries to measure, and the noise is the amount of uncertainty that surrounds the value of the signal making it hard to identify the actual value of the signal. For example, in a trial that uses weight reduction drugs, the Body Mass Index (BMI) may be a signal of interest while the noise may be attributed to failure to account for participants' baseline measurements.

Radiologists use the signal-to-noise ratio (SNR), defined as the ratio of the mean signal to its standard deviation, as a measure of image quality (Cunningham and Shaw [1]). For example, in the case of Magnetic Resonance Images (MRI), the noise may be distributed uniformly throughout the image, and one way of measuring the effect of noise is to calculate the SNR. For MRI data, radiologists evaluate the SNR by computing the mean signal intensity over a certain region of interest (ROI) and dividing this by the standard deviation of the signal from outside the image. In cancer diagnosis, Jung *et al.* [2] applied SNR to investigate the accuracy and inter-observer variability of a standardized evaluation system for endorectal three-dimensional MR spectroscopic imaging of the prostate. They noted that most of current image processing applied to MRI image data can be formulated as a parameter estimation problem, and in the case of noise filtering, the SNR is the target parameter. Sim and Kamel [3] proposed an autoregressive (AR) model for SNR estimation. In more general settings, Taguchi, and Wu [4] used the SNR as an indicator of "closeness to target" and [5] provided references for indices of measurement quality and reliability.

Applications of SNR in Genomics

One of the interesting applications of SNR is in genomics. The unique conditions in gene expression analysis [6] are: 1) the level of a transcript depends roughly on the concentration of the related factors that control the rate of production of the transcript; 2) the random variations for any particular transcript are normally distributed. In several applications, it is assumed that the variation of any transcript is constant relative to most of the other transcripts in the genome which means the coefficient of variation σ/μ is considered constant across the genome. Alternatively, we may take μ/σ (signal-to-noise ratio) which we use in this paper as a measure of relative variation.

For example, suppose that we have a micro assay with the R genes with red and green expression values, R_1, \dots, R_k , and G_1, G_2, \dots, G_k . Let μ_R and σ_R denote the mean and standard deviation of R_j . It is of interest to test the null hypothesis.

$$H_0 : \mu_R / \sigma_R = \mu_G / \sigma_G$$

Our paper has three-fold objectives: The first is developing nonparametric methods for testing the equality of two correlated SNR parameters. The second

is to use Pearson's system of curves to construct a probability distribution for the difference between two estimates of SNR in the case of bivariate normal and the bivariate lognormal distributions. This is quite important because providing the best approximate distribution allows us to study the characteristics of the distribution of the statistic of interest beyond the point and interval estimation of the corresponding population parameter. The last objective is to illustrate the methodologies on real life data.

2. Testing the Equality of Two Correlated SNRs

Let (X_1, X_2, \dots, X_n) be a random sample drawn from $N(\mu, \sigma^2)$. We define the signal-to-noise ration as:

$$\theta = \mu/\sigma$$

The maximum likelihood estimator (MLE) of θ is

$$\hat{\theta} = \bar{X} / S^*$$

where,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i,$$

and

$$S^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

and respectively the maximum likelihood estimators of μ , and σ^2 . Instead of S^2 , we shall use

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

which is unbiased estimator of σ^2 instead of S^{*2} .

Since one of the characteristics of the normal distribution is the stochastic independence of \bar{X} from S^2 , we have:

$$E(\hat{\theta}) = E(\bar{X}) \cdot E[1/S], \text{ with } E(\bar{X}) = \mu$$

Now, since $(n-1)S^2 = \sigma^2 X_{(n-1)}^2$, where $X_{(n-1)}^2$ denotes a chi-square random variable with $(n-1)$ degrees of freedom, we can evaluate the exact moments of $\hat{\theta}$. The r th non-central moment of $\hat{\theta} = \bar{X} / S^*$ are:

$$\mu'_r(\hat{\theta}) = E(\bar{X})^r \cdot E[S^{-r}] \quad (1)$$

$$\mu'_1(\hat{\theta}) = \frac{\mu}{\sigma} \sqrt{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \quad (2)$$

$$\mu'_2(\hat{\theta}) = \left[\mu^2 + \frac{\sigma^2}{n} \right] \left[\frac{n-1}{\sigma^2(n-3)} \right] \quad (3)$$

Hence, from (1) and (2) we have

$$\text{var}(\hat{\theta}) = \mu_2(\hat{\theta}) = \frac{n-1}{n(n-3)} + \frac{\mu^2}{\sigma^2} \left[\frac{n-1}{n-3} - \frac{n-1}{2} \frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} \right] \tag{4}$$

This estimator is biased and has an expected value:

$$E(\hat{\theta} - \theta) = \theta \left[\sqrt{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - 1 \right] \tag{5}$$

From (5) the relative bias of $\hat{\theta}$ is given by:

$$\text{RB} = E\left[\frac{\hat{\theta} - \theta}{\theta}\right] = \sqrt{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} - 1 \tag{6}$$

For selected values of the sample size n , the corresponding values RB given in (6) are provided in **Table 1**.

From **Table 1**, we can see that the RB of the MLE of $\hat{\theta}$ is almost negligible when the sample size n gets larger. Now, if two independent samples are available, the difference between estimated two independent SNRs defined as D :

$$D = \hat{\theta}_1 - \hat{\theta}_2.$$

We can show that under the assumption of independence that:

$$\text{var}(D) = \text{var}(\hat{\theta}_1 - \hat{\theta}_2) = \frac{2(n-1)}{n(n-3)} + (\theta_1^2 + \theta_2^2)H(n) \tag{7}$$

where

$$H(n) = \frac{n-1}{n-3} - \left(\frac{n-1}{2}\right) \frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)}$$

We may therefore construct a $(1 - \alpha)100\%$ CI on $(\hat{\theta}_1 - \hat{\theta}_2)$ as:

$$(\hat{\theta}_1 - \hat{\theta}_2) \pm z_{1-\alpha/2} \sqrt{\text{var}(\hat{\theta}_1 - \hat{\theta}_2)} \tag{8}$$

When neither the assumptions of independence nor the normality are satisfied, we propose several approaches to test the hypothesis on the equality of two correlated SNRs. The first approach is nonparametric and is suitable when the distributional assumptions of the parents populations cannot be verified.

Table 1. This relative bias is negligible even for small sample size.

n	10	15	20	30	40	50
RB	0.094	0.058	0.042	0.027	0.02	0.016

3. Nonparametric Test of the Null Hypothesis H_0 : SNR1 = SNR2

Let (x_{1j}, x_{2j}) denote a random sample ($j = 1, 2, \dots, n$) available from any bivariate distribution such that, $i = 1, 2, \dots, n$ denote a random sample of size n , with parameter vector given by $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_{12})$. The samples summary statistics of the data are:

$\bar{X}_1 = \text{mean}(X_1)$, $\bar{X}_2 = \text{mean}(X_2)$, $S_1^2 = \text{variance}(X_1)$, $S_2^2 = \text{variance}(X_2)$, and ρ_{12} is the correlation between X_1 and X_2 .

We define the signal-to-noise ratio (SNR) of the i th sample by:

$\theta_i = \mu_i / \sigma_i$, ($i = 1, 2$) whose moment estimators are: $\hat{\theta}_i = \bar{x}_i / s_i$

It is evident that only if $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$, then $\theta_1 = \theta_2$. Therefore, if the available data support the null hypothesis:

$$H_0 : \mu_1 = \mu_2 \cap \sigma_1^2 = \sigma_2^2 \quad (9)$$

then we conclude that $\theta_1 = \theta_2$.

Testing the hypothesis (9) is equivalent to using the test developed in [7]. Earlier, [7] suggested tests of equality of correlated means and correlated variances using the statistic (10):

$$\text{WILK} = \frac{S_1^2 S_2^2 (1 - \rho_{12}^2)}{S^2 (1 + \rho_{12}) [S^2 (1 - \rho_{12}) + C]} \quad (10)$$

where $S^2 = \frac{1}{2} [S_1^2 + S_2^2]$, $C = (\bar{X}_1 - \bar{X}_2) / 2$, and ρ_{12} is the correlation between the two sets of observations.

Then $Q = -2 \log(\text{WILK}) \sim X_{(2)}^2$.

Example 1:

The Petrifilm test, a quantitative microbiological test for *Escherichia coli* O157:H7, was evaluated for its performance as a beef-carcass monitoring test compared with an *E. coli* O157:H7 detection method using a hydrophobic grid membrane filter (HGMF) [8]. The Petrifilm test showed excellent agreement with the HGMF method when test samples were obtained from pure cultures and experimentally contaminated meat. Here we shall compare the SNR of the two methods using the nonparametric indirect approach suggested in (10), using a sample of 23 pairs of observations.

We provide the R code in which we use the bootstrap methodology to establish the large sample properties of the proposed statistic in **Appendix 2**.

RESULTS

SNR_Pertifilm = 0.979

SNR_hgmf = 1.05

Correlation = 0.996

quantile (-2 * log(WILK), probs = c (0.05, 0.95))

5% 95%

0.006 0.863

quantile (xx, probs = c (0.05, 0.95))

5% 95%
0.101 5.983

As can be seen the (5%, 95%) bootstrap quantiles of the test on the equality of the two SNRs (0.006, 0.863) are completely contained within the theoretical quantiles of a chi-square distribution χ^2 , with 2 degrees of freedom. Thus, we established the equality of the two SNRs. The approximate distribution of the statistic (10) is depicted graphically in **Figure 1**, **Figure 2**.

Figure 1 shows that the histogram is skewed to the right, while **Figure 2** shows that the distribution of $-2 * \log(WILK)$ is in close agreement with χ^2 (chi square random variable with $n - 2$ degrees of freedom).

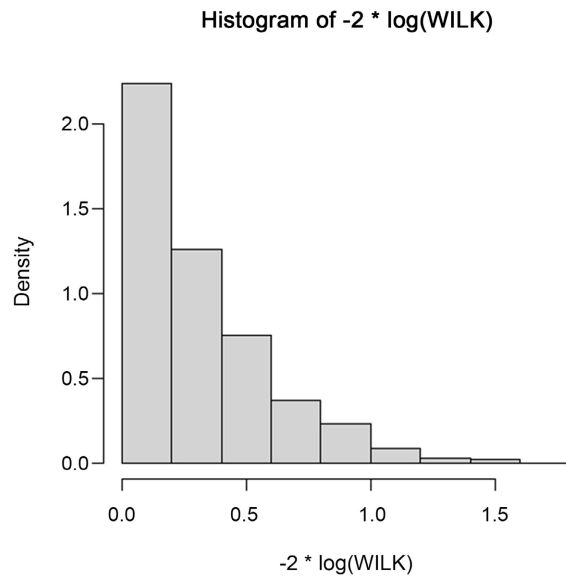


Figure 1. Histogram of 1000 bootstrap samples of the statistic (10).

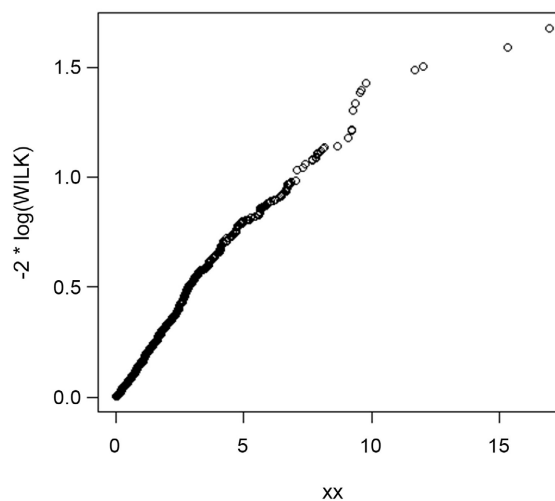


Figure 2. The q-q plot of the statistic (10) plotted against 1000 samples xx , from the chi-square distribution with 2 degrees of freedom.

4. The Pearson's System of Curves

Several years ago, Karl Pearson introduced a system of distributions that have been found useful in many applications. Pearson's family of distributions generates different densities (f) that are solutions of the differential equation:

$$\frac{df}{dx} = \frac{(x-a)f}{b_0 + b_1x + b_2x^2} \quad (11)$$

The system of curves was discussed in detail in [9], and in this section we present a summary of their results.

The moments of a PSC (Pearson's System of Curves) are determined by the values of the constants in (1) in addition to the constant of integration; conversely if we have the four moments of a PC we can solve for the constants in Equation (11). Pearson classified the solutions into types numbered 1 to 6. In constructing confidence intervals on the population parameter, the quartiles of x are more important than the form of the probability density " f ".

To clarify this point, we assume that x has a Pearson curve density, with mean μ and central moments μ_2, μ_3 and μ_4 . The percentage points of the standardized x given as $x = (x - \mu) / \sqrt{\mu_2}$ can be obtained as functions of skewness $= \beta_1 = \mu_3^2 / \mu_2^3$ and kurtosis $= \beta_2 = \mu_4 / \mu_2^2$, and are tabulated in double entry tables against $\sqrt{\text{skewness}}$ and kurtosis. The percentage points of x are then easily obtained from those of x .

In summary, the steps of fitting a Pearson curve to theoretical distribution as summarized in [9] are:

- 1) We calculate skewness $= s = \mu_3^2 / \mu_2^3$ and kurtosis $= k = \mu_4 / \mu_2^2$
- 2) Based on s and k we obtain X_α , the percentage point at upper or lower level α

- 3) Calculate $x'_\alpha = x_\alpha \sqrt{\mu_2} + \mu$ if μ_3 is positive or

$$x'_{1-\alpha} = \mu - X_\alpha \sqrt{\mu_2}$$

if μ_3 is negative

There are six types of Pearson's distributions:

Type I = These distributions are (location scale transformation of) Beta distributions. The probability density function with parameters, a, b , scale $= s$ and location $= m$ is given by:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)|ss|} \left(\frac{x-m}{ss} \right)^{a-1} \left(1 - \frac{x-m}{ss} \right)^{b-1} \quad (12)$$

For $a > 0, b > 0, ss \neq 0, 0 < \frac{x-m}{s} < 1$

Type II = Pearson Type II is the symmetric Beta is a special case of Type I and is obtained when $a = b$. Therefore, the probability density function is given by:

$$f(x) = \frac{\Gamma(2a)}{\Gamma^2(a)|ss|} \left[\left(\frac{x-m}{ss} \right) \left(1 - \frac{x-m}{ss} \right) \right]^{a-1} \quad (13)$$

Type III = This is the Gamma distribution. The probability density function with shape = a , scale = ss , and location = m is given by:

$$f(x) = \frac{1}{\Gamma(a)|ss|^a} (x-m)^{a-1} \exp\left[-\frac{x-m}{ss}\right] \tag{14}$$

For $ss \neq 0$, $a > 0$ and $\frac{x-m}{ss} \geq 0$.

Type IV = The Pearson Type IV with location parameter = m , scale parameter = ss and shape parameters δ_1, δ_2 has pdf given by:

$$f(x) = \frac{k}{ss} \left[1 + \left(\frac{x-m}{ss}\right)^2\right]^{-\delta_1} \exp\left[-\delta_1 \tan^{-1}\left(\frac{x-m}{ss}\right)\right] \tag{15}$$

where

$$k = \left[\frac{\Gamma\left(\delta_1 + i\frac{\delta_2}{2}\right)}{\Gamma(\delta_1)} \right]^2 / B\left(\delta_1 - \frac{1}{2}, \frac{1}{2}\right)$$

where $ss > 0$, $\delta_1 > \frac{1}{2}$ ($\delta_2 = 0$ corresponds to the Pearson Type VII distribution family).

Type V = This is the Inverse Gamma distributions. The pdf is given by:

$$f(x) = \frac{|ss|^\alpha}{\Gamma(\alpha)} (x-m)^{-\alpha-1} \exp\left[-\frac{x}{x-m}\right] \tag{16}$$

where ss = scale parameter, α = shape parameter, and m = location parameter. The scale parameter is permitted to have negative values to allow for left skewness.

Type VI = Known as the Beta Prime distribution and in fact are scaled F-distributions. The pdf is given by:

$$f(x) = \frac{\Gamma(a+b)}{|ss|\Gamma(\alpha)\Gamma(b)} \left(\frac{x-m}{ss}\right)^{a-1} \left(1 + \frac{x-m}{ss}\right)^{-a-b} \tag{17}$$

ss = scale parameter

m = location parameter

$$a > 0, b > 0, s \neq 0, \frac{x-m}{ss} > 0$$

Clearly, to determine the approximate Pearson curve for any random variable, we should have the numerical values of the skewness and kurtosis, that is we have to determine its first four central moments. To find the first 4 central moments of $\tilde{\Delta}$, we shall use the delta method by employing the Taylor's expansion on $\tilde{\Delta}$. For ready calculations of the moments, we shall use some of the results from [10] together with the expectations of products of correlated chi-square variables and the higher order moments of bivariate normal distribution. These expectations are given in **Appendix 1**.

5. The Case of the Bivariate Normal Distribution

Let $x_{ij} \sim BVN(\mu_i, \sigma_i^2)$, $i=1,2; j=1,2,\dots,n$ be a random sample from bivariate normal distribution.

Our main objective are:

- 1) Construct a large sample confidence interval on $\theta_1 - \theta_2 = (\mu_1/\sigma_1) - (\mu_2/\sigma_2)$
- 2) Use the first four central moments of $\tilde{\Delta} = \hat{\theta}_1 - \hat{\theta}_2 = (\bar{x}_1/s_1) - (\bar{x}_2/s_2)$ to fit a Pearson family and find the appropriate quantiles.

We shall use the delta method by employing the Taylor's expansion on the statistic $\tilde{\Delta}$.

Now we derive the first 4 central moments of the targeted statistic $\tilde{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$.

On using the Taylor's expansion, we define the variance of $\tilde{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$ as:

$$\mu_2(\tilde{\Delta}) = E\left[(\bar{x}_1 - \mu_1)\alpha_1 + (\bar{x}_2 - \mu_2)\alpha_2 + (s_1^2 - \sigma_1^2)\gamma_1 + (s_2^2 - \sigma_2^2)\gamma_2\right]^2$$

where

$$\alpha_j = \frac{\partial \tilde{\Delta}}{\partial x_j}, \quad j=1,2$$

$$\gamma_j = \frac{\partial \tilde{\Delta}}{\partial s_j^2}, \quad j=1,2$$

$$\alpha_1 = 1/\sigma_1, \quad \alpha_2 = -1/\sigma_2$$

$$\gamma_1 = -\frac{\mu_1}{2\sigma_1^3}, \quad \gamma_2 = \frac{\mu_2}{2\sigma_2^3}$$

Substituting these values in $\mu_2(\tilde{\Delta})$, using the appropriate expectations in the **Appendix 1**, and after simplifications, to the first order of approximation we get:

$$\mu_2(\tilde{\Delta}) = \frac{2(1-\rho)}{n-1} + \frac{1}{2(n-1)}[\theta_1^2 + \theta_2^2 - 2\theta_1\theta_2\rho^2] \quad (18)$$

The third central moment is defined as:

$$\mu_3(\tilde{\Delta}) = E\left[(\bar{x}_1 - \mu_1)\alpha_1 + (\bar{x}_2 - \mu_2)\alpha_2 + (s_1^2 - \sigma_1^2)\gamma_1 + (s_2^2 - \sigma_2^2)\gamma_2\right]^3$$

Again, using the expectations in **Appendix 1**, and after simplifications we get:

$$\mu_3(\tilde{\Delta}) = \frac{\theta_1^3 + \theta_2^3}{(n-1)^2} - \frac{3n\rho^2\theta_1\theta_2}{2(n-1)}(\theta_1 + \theta_2) \quad (19)$$

The fourth central moment is defined as:

$$\mu_4(\tilde{\Delta}) = E\left[(\bar{x}_1 - \mu_1)\alpha_1 + (\bar{x}_2 - \mu_2)\alpha_2 + (s_1^2 - \sigma_1^2)\gamma_1 + (s_2^2 - \sigma_2^2)\gamma_2\right]^4$$

Again, using the expectations in **Appendix 1** and after simplifications we get:

$$\begin{aligned} \mu_4(\tilde{\Delta}) = & \frac{12(1-\rho)^2}{n} + \frac{6\rho^2}{n(n-1)}\theta_1\theta_2 - \frac{3(n+3)}{(n-1)^3}\rho^2\theta_1\theta_2(\theta_1^2 + \\ & + \frac{3(n+1)}{4(n-1)^3}(\theta_1^4 + \theta_2^4) \\ & + \frac{3}{8}\theta_1^2\theta_2^2\left[\frac{n+1}{(n-1)^2} + \frac{8(n+1)}{(n-1)^3}\rho^4 + \frac{4}{(n-1)^3}\rho^2\right] \end{aligned} \quad (20)$$

Example 2: Vitamin B12 and Folic Acid as predictors of Osteoporosis

Osteoporosis and bone health remain a major health problem often associated with significant morbidity, mortality, and healthcare costs [11]. An epidemiological analysis reported that 34% of healthy Saudi women and 30.7% of healthy Saudi men aged between 50 and 79 years are osteoporotic [12]. Furthermore, approximately 40% - 50% of women and 25% - 33% of men sustain osteoporotic fractures in their lifetimes. Hence, identifying the risk factors for osteoporosis is crucial in reducing the incidence of fractures. One study investigated the relationship between vitamin B12 and folate levels and BMD in Saudi Arabia [13] [14]. A recent study established the association between B12 and folate levels and BMD [15]. We used part of the data to investigate which factor has a higher SNR so it can be used reliably in the prediction of osteoporotic fractures.

Results of data analysis are shown in **Table 2**.

On using the normal Q-Q plot, we noted that the log transformed B12, and Folate are approximately normally distributed with correlation = 0.290 (p-value < 0.001).

Figure 3 and **Figure 4** show respectively that the log-transformed variables B12 and Folate are approximately normally distributed. Preliminary data analysis gave:

Table 2. Summary statistics of the B12 and Folate data.

	N	Mean	Std. Deviation
B12 level	218	326.14	174.485
Folate level	218	2133.47	667.676
logB12 (X1)	218	5.6546	.51782
logFolate (X2)	218	7.6180	.30944

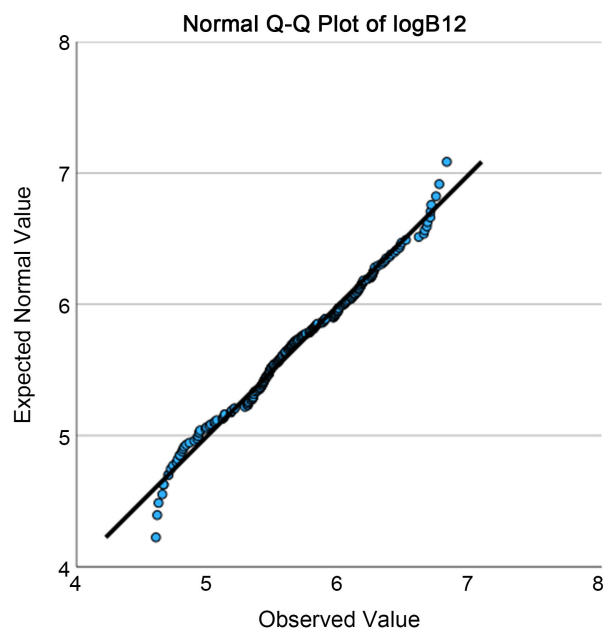


Figure 3. Q-Q normal plot of the log-transformed B12.

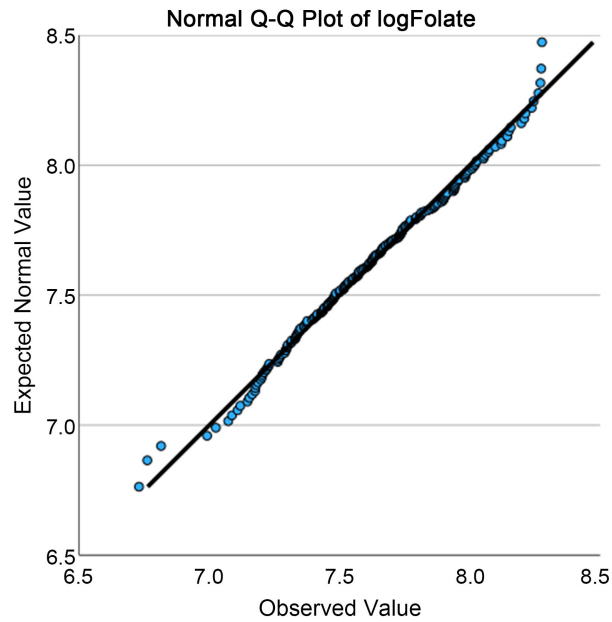


Figure 4. Q-Q normal plot of the log-transformed Folate.

$$\text{SNR} (X1) = 5.6546/.51782 = 10.92$$

$$\text{SNR} (X2) = 7.618/.30944 = 24.62$$

$$\text{Difference} = \text{SNR} (X2) - \text{SNR} (X1) = 13.7$$

$$\text{SE} (\text{Difference}) = 0.41$$

$$\text{Lower 95\% confidence limit} = \text{Difference} - 1.96 * \text{SE} (\text{Difference}) = 12.9 \quad (21)$$

$$\text{Upper 95\% confidence limit} = \text{Difference} + 1.96 * \text{SE} (\text{Difference}) = 14.5 \quad (22)$$

Apparently, logFolate has a significantly higher SNR than the logB12.

Again, we use the bootstrap of the data to find the best fitting Pearson curve for the distribution of: $\tilde{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$. The R code is given in **Appendix 3**.

Results

Variance = 1.106637, mean = 13.78

quantile (PIVOT, probs = c (0.05, 0.95))

5%	95%
12.155	15.613

Since the 95% empirical confidence limits (12.155, 15.613) do not include zero, we conclude that the two SNRs are significantly different. Note that the empirical limits are almost identical to the 95% confidence limits (21) and (22) (**Figure 5** and **Figure 6**).

PEARSON’S CURVE:

Parameters estimates of the Pearson curve

type	a	b	location	scale
II	15.75	30.53	8.58	15.26

Pearson moments

mean	variance	skewness	kurtosis
13.779	1.107	0.192	2.932

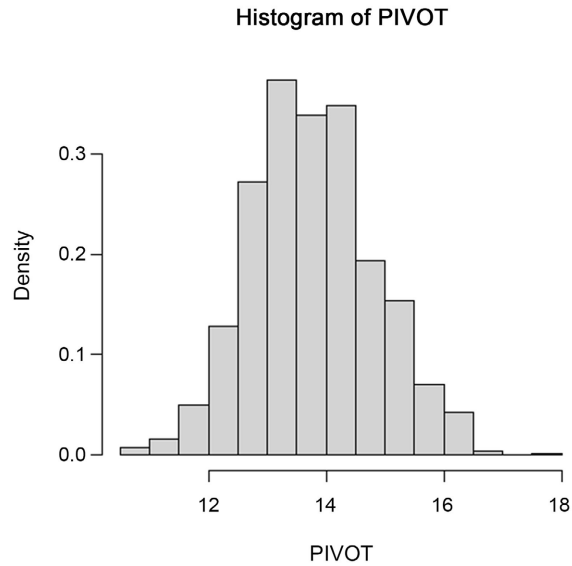


Figure 5. The histogram of 1000 bootstrap values of the statistic $PIVOT = \tilde{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$.

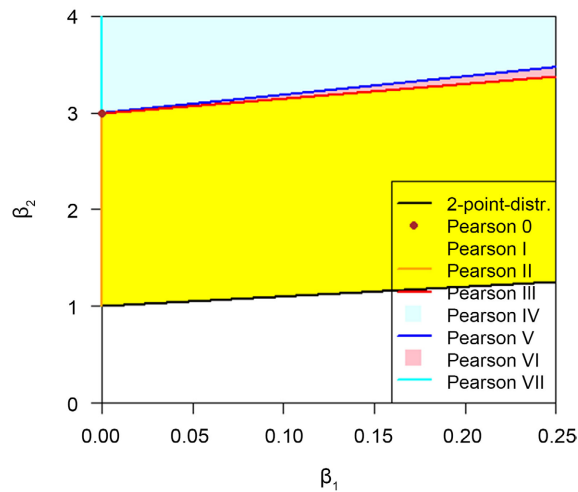


Figure 6. Region of definition (yellow) of the Pearson’s curve showing that the Type II is the best curve to fit the distribution of the difference between the estimated SNRs.

6. The Bivariate Lognormal Case

Let V_1 and V_2 be two log normally distributed random variables defined on the intervals $0 < V_1 < \infty$, $0 < V_2 < \infty$, and X_1, X_2 denote the transformed normal variables; that is

$$X_1 = \log V_1, \quad X_2 = \log V_2$$

where the joint distribution of (X_1, X_2) is a bivariate normal. The joint pdf of (V_1, V_2) is given by:

$$f(V_1, V_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \frac{1}{V_1V_2} \exp\left(\frac{-q}{2(1-\rho^2)}\right) \tag{23}$$

where $q = A^2 - 2\rho AB + B^2$

$$A = (\log V_1 - \mu_1)/\sigma_1, \quad B = (\log V_2 - \mu_2)/\sigma_2$$

μ_i are the means of X_i , σ_i are their standard deviations and ρ is the correlation between (X_1, X_2) .

The Joint probability density function given in (23) is that of a bivariate log-normal distribution which was studied extensively in [16]. The means and the variance of the lognormal distributions are given respectively as follows

$$\text{Mean}(V_i) = \exp(\mu_i + \sigma_i^2/2)$$

$$\text{Var}(V_i) = \exp(2\mu_i + 2\sigma_i^2) - \exp(\mu_i + \sigma_i^2)^2$$

The SNR is thus given by $\text{Mean}(V_i)/\sqrt{\text{Var}(V_i)} = (e^{\sigma_i^2} - 1)^{-1/2}$.

We are interested in deriving the distribution of the maximum likelihood estimator of

$$\Delta = (e^{\sigma_1^2} - 1)^{-1/2} - (e^{\sigma_2^2} - 1)^{-1/2} \quad (24)$$

This estimator is given by:

$$\tilde{\Delta} = g(S_1^2, S_2^2) = (e^{S_1^2} - 1)^{-1/2} - (e^{S_2^2} - 1)^{-1/2} \quad (25)$$

Note that the case of two independent lognormal populations was investigated in [17].

To find the first 4 central moments and similar to the case of bivariate normal distribution discussed in the previous section, we employ the Taylor's series expansion on $g(S_1^2, S_2^2)$, as shown in (26).

In general for the pair of estimators $(\hat{\theta}_1, \hat{\theta}_2)$ assumed unbiased for (θ_1, θ_2) we have:

$$g(\hat{\theta}_1, \hat{\theta}_2) \approx g(\theta_1, \theta_2) + (\hat{\theta}_1 - \theta_1)\delta_1 + (\hat{\theta}_2 - \theta_2)\delta_2 \quad (26)$$

where $\delta_j = \frac{\partial g}{\partial \theta_j}$, g means the partial derivative is evaluated at the true value of the parameter. This partial derivative is given in (27).

$$\delta_j = \frac{-e^{-\sigma_j^2}}{2[e^{\sigma_j^2} - 1]^{3/2}} \quad (27)$$

Let $\mu_r(\tilde{\Delta})$ denote the r th central moment of $\tilde{\Delta}$, that is:

$$\mu_r(\tilde{\Delta}) = E[g(S_1^2, S_2^2) - g(\theta_1, \theta_2)]^r$$

$$r = 2, 3, 4$$

Therefore, using (27) we get:

$$\mu_2(\tilde{\Delta}) = \frac{\sigma_1^4}{2(n-1)}\delta_1^2 + \rho^2 \frac{\sigma_1^2\sigma_2^2}{n-1}\delta_1\delta_2 + \frac{\sigma_2^4}{2(n-1)}\delta_2^2 \quad (28)$$

$$\mu_3(\tilde{\Delta}) = -\frac{\sigma_1^6}{(n-1)^2} \delta_1^3 + \frac{\sigma_2^6}{(n-1)^2} \delta_2^3 - \frac{3n}{2(n-1)} \rho^2 \sigma_1^2 \sigma_2^2 [\sigma_1^2 \delta_1^2 \delta_2 + \sigma_2^2 \delta_1 \delta_2^2] \tag{29}$$

$$\begin{aligned} \mu_4(\tilde{\Delta}) = & -\frac{12(n+3)}{(n-1)^3} [\sigma_1^8 \delta_1^4 + \sigma_2^8 \delta_2^4] + \frac{48(n+3)}{(n-1)^3} \sigma_1^6 \sigma_2^2 \rho^2 \delta_1^3 \delta_2 \\ & + \frac{48(n+3)}{(n-1)^3} \sigma_1^2 \sigma_2^6 \rho^2 \delta_1 \delta_2^3 \\ & + 6\sigma_1^4 \sigma_2^4 \delta_1^2 \delta_2^2 \left[\frac{n+1}{(n-1)^2} + \frac{4(n+1)}{(n-1)^3} \rho^4 + \frac{4\rho^2}{(n-1)^3} \right] \end{aligned} \tag{30}$$

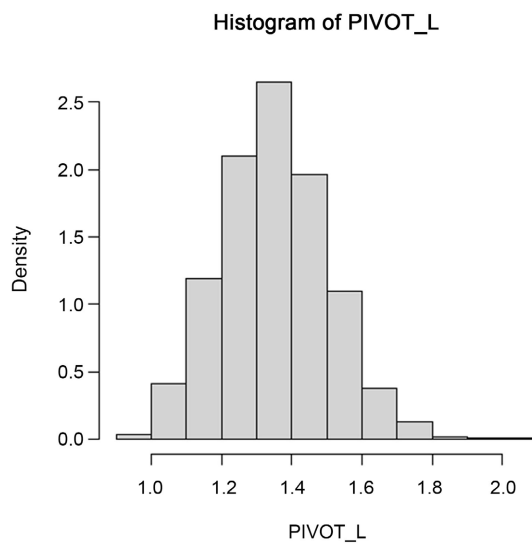


Figure 7. The histogram of the 1000 bootstrap samples for the distribution of $\tilde{\Delta}$.

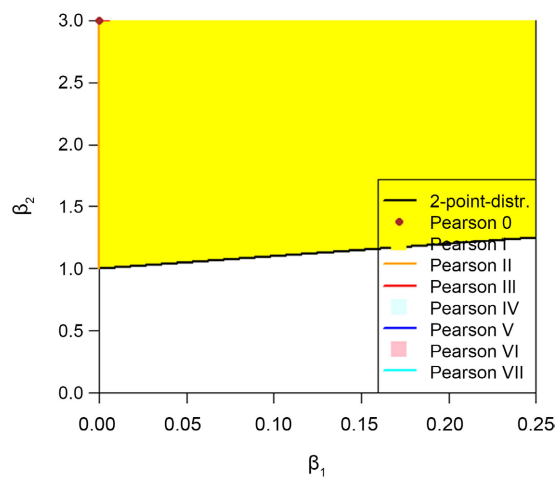


Figure 8. Pearson’s diagram shows that in the case of the correlated lognormal distribution the Type I curve is the best distribution of $\tilde{\Delta}$ for the bivariate log-normally distributed.

The moments of $\tilde{\Delta}$ given in (28), (29), and (30) are used to derive the skewness and kurtosis that are needed to calculate the parameters of the Pearson's curve.

Example 3:

We have shown that the B12 and Folate data in the previous example, after applying the log transformation, the transformed variables were quite close to the normal distribution. Therefore, we may conclude that the untransformed variables may have lognormal distributions. Similar Bootstrap approach may then be used to find the empirical Pearson's distribution, when the difference between the estimated SNRs is $\tilde{\Delta}$. The R code in this case is the same as in the case of bivariate normal distribution. The results are summarized below.

Variance ($\tilde{\Delta}$) = 0.023, Mean ($\tilde{\Delta}$) = 1.363, Skewness = 0.070, and Kurtosis = 2.916.

The empirical 5%, 95% quantiles are (1.118, 1.612). Since the quantiles do not include zero, we conclude that with 95% confidence, that there is a significant difference between the two SNRs under the bivariate log-normal set up. The parameters of the Pearson's curve are:

type	a	b	location	scale
1	27.14	36.028	0.32	2.428

Note that, the notation PIVOT_L is used to denote the random variable $\tilde{\Delta}$.

As can be seen from the histogram in **Figure 7**, the distribution of $\tilde{\Delta}$ is slightly skewed to the right.

From the diagram in **Figure 8**, it seems that the Pearson's Type 1 is the best approximating distribution.

7. Discussion and Conclusions

In this paper, we considered two approaches to estimate and test the difference between two SNR ratios. Through bootstrap sampling, the nonparametric approach confirmed the large sample distribution of the test statistic on the hypothesis of simultaneously testing the equality of means and variances, and hence equality of two correlated SNRs.

When the bivariate distribution is either normal or lognormal we derived the first four central moments of the target statistics using delta the method. Thereafter, we used the bootstrap to find the Pearson's family that best fits the distribution of the estimated difference between SNRs. Pearson's Type I&II gave the best fit based on the available data. Several R [18] packages were used in the computational steps to achieve the objectives of the paper.

Acknowledgements

The author acknowledges the anonymous reviewer's comments that have greatly improved the presentation of the paper.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Cunningham, I. and Shaw, R. (1999) Signal-to-Noise Optimization of Medical Imaging System. *Journal of the Optical Society of America A*, **16**, 621-632. <https://doi.org/10.1364/JOSAA.16.000621>
- [2] Jung, J.A., Coakley, F.V., Vigneron, D.B., Swanson, M.G., Qayyum, A., Weinberg, V., Jones, K.D., Carroll, P.R. and Kurhanewicz, J. (2004) Prostate Depiction at Endorectal MR Spectroscopic Imaging: Investigation of a Standardized Evaluation System. *Radiology*, **233**, 701-708. <https://doi.org/10.1148/radiol.2333030672>
- [3] Sim, K.S. and Kamel, N.S. (2004) Image Signal-to-Noise Ratio Estimation Using the Autoregressive Model. *Scanning*, **26**, 135-139. <https://doi.org/10.1002/sca.4950260306>
- [4] Taguchi, G. and Wu, Y. (1985) Introduction to Off-Line Quality Control. Central Japan Quality Control Association, Tokyo.
- [5] Shoukri, M.M. (2010) Measures of Inter-Observer Agreement and Reliability. 2nd Edition, Chapman & Hall/CRC Press, Boca Raton, FL.
- [6] Chen, Y., Dougherty, E.R., Bittner, M.L., Meltzer, P. and Trent, J. (2002) Chapter 1. Microarray Image Analysis and Gene Expression Ratio Statistics. In: Zhang, W. and Shmulevich, I., Eds., *Computational and Statistical Approaches to Genomics*, Springer, Boston, MA, 1-21. https://doi.org/10.1007/0-306-47825-0_1
- [7] Gulliksen, H. and Wilks, S.S. (1950) Regression Tests for Several Samples. *Psychometrika*, **15**, 91-114. <https://doi.org/10.1007/BF02289195>
- [8] Power, C.A., McEwen, S.A., Johnson, R., Shoukri, M., Rahn, K., Griffith, M. and De Grandis, S. (1998) Repeatability of the Petrifilm™ HEC Test and Agreement with a Hydrophobic Grid Membrane Filtration Method for the Enumeration of *Escherichia coli* O157:H7 on Beef Carcasses. *Journal of Food Protection*, **61**, 402-408. <https://doi.org/10.4315/0362-028X-61.4.402>
- [9] Stuart, M. and Ord, K. (2009) Advanced Theory of Statistics. 6th Edition, Wiley, Hoboken.
- [10] Joarder, A.H. (2009) Moments of the Products and Ratio of Two Correlated Chi-Square Variables. *Statistical Papers*, **50**, 581-592. <https://doi.org/10.1007/s00362-007-0105-0>
- [11] Jawad, A.S. (2016) Osteoporosis in Saudi Arabia. *Saudi Medical Journal*, **37**, 468. <https://doi.org/10.15537/smj.2016.4.14776>
- [12] Alwahhabi BK. (2015) Osteoporosis in Saudi Arabia: Are We Doing Enough? *Saudi Medical Journal*, **36**, 1149-1150. <https://doi.org/10.15537/smj.2015.10.11939>
- [13] Alharbi, A., Awlia, J., Ardawi, M. and Hussein, Y.M. (2012) Relation of Vitamin B12, Folate, and Methylenetetrahydrofolate Reductase Polymorphism to Bone Mass Density in Healthy Saudi Men. *Journal of American Science*, **8**, 110-116.
- [14] Tucker, K.L., Hannan, M.T., Qiao, N., Jacques, P.F., Selhub, J., Cupples, L.A., *et al.* (2005) Low Plasma Vitamin B12 Is Associated with Lower BMD: The Framingham Osteoporosis Study. *Journal of Bone and Mineral Research*, **20**, 152-158. <https://doi.org/10.1359/JBMR.041018>
- [15] Hussain, S.I., AlKhenizan, A., Mahmoud, A. and Qashlaq, H. (2023) The Correlation between Vitamin B12 and Folate Levels and Bone Mineral Density among the Saudi Population in a Primary Care Setting. *Journal of Family Medicine and Primary Care*, **12**, 1063-1068. https://doi.org/10.4103/jfmpc.jfmpc_1209_22
- [16] Mostafa, M.D. and Mahmoud, M.W. (1964) On the Problem of Estimation for the Bivariate Lognormal Distribution. *Biometrika*, **51**, 522-527.

<https://doi.org/10.1093/biomet/51.3-4.522>

- [17] Thangjai, W. and Niwitong, S.-A. (2019) Confidence Intervals for the Signal-to-Noise Ratio and difference of Signal-to-Noise Ratios of Log-Normal Distributions. *Stats*, **2**, 164-173. <https://doi.org/10.3390/stats2010012>
- [18] <http://www.R-project.org>

Appendix 1. Cross Product Expectations of Higher Order Functions of Bivariate Normal and Bivariate Chi-Square Random Variables

$$E\left[(\bar{x}_i - \mu_i)^4\right] = 3\sigma_i^4/n, \quad i=1,2$$

$$E\left[(s_i^2 - \sigma_i^2)^2\right] = 2\sigma_i^4/(n-1)$$

$$E\left[(s_i^2 - \sigma_i^2)^3\right] = 8\sigma_i^4/(n-1)^2$$

$$E\left[(s_i^2 - \sigma_i^2)^4\right] = \frac{12(n+3)}{(n-1)^3}\sigma_i^8$$

$$E\left[(s_1^2 - \sigma_1^2)(s_2^2 - \sigma_2^2)\right] = 2\rho^2\sigma_1^2\sigma_2^2/(n-1)$$

$$E\left[(s_1^2 - \sigma_1^2)^2(s_2^2 - \sigma_2^2)^2\right] = \sigma_1^4\sigma_2^4M(n, \rho)$$

where

$$M(n, \rho) = \frac{n+1}{(n-1)^2} + \frac{8(n+1)}{(n-1)^3}\rho^4 + \frac{4}{(n-1)^3}\rho^2$$

$$E\left[(\bar{x}_1 - \mu_1)^2(\bar{x}_2 - \mu_2)\right] = 0$$

$$E\left[(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)^2\right] = 0$$

$$E\left[(s_1^2 - \sigma_1^2)^2(s_2^2 - \sigma_2^2)\right] = \frac{4n}{n-1}\rho^2\sigma_1^4\sigma_2^2$$

$$E\left[(s_1^2 - \sigma_1^2)(s_2^2 - \sigma_2^2)^2\right] = \frac{4n}{n-1}\rho^2\sigma_1^2\sigma_2^4$$

$$E\left[(\bar{x}_1 - \mu_1)^2(\bar{x}_2 - \mu_2)^2\right] = \frac{\sigma_1^2\sigma_2^2}{n}[1 + 2\rho^2]$$

$$E\left[(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)\right] = \frac{\rho\sigma_1\sigma_2}{n}$$

$$E\left[(\bar{x}_1 - \mu_1)^3(\bar{x}_2 - \mu_2)\right] = 3\rho\sigma_1^3\sigma_2/n$$

By symmetry

$$E\left[(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)^3\right] = 3\rho\sigma_1\sigma_2^3/n$$

$$E\left[(\bar{x}_1 - \mu_1)^2(\bar{x}_2 - \mu_2)^2\right] = \frac{\sigma_1^2\sigma_2^2}{n}[1 + 2\rho^2]$$

$$E\left[(s_1^2 - \sigma_1^2)^3(s_2^2 - \sigma_2^2)\right] = \frac{12(n+3)}{(n-1)^3}\sigma_1^6\sigma_2^2\rho^2$$

Appendix 2. R-CODE for Example 3.1

```
library(boot)
B=1000
n=nrow(data)
n
WILK=numeric(B)
b<-data$petrifilm
f<-data$hgmf
mb<-mean(b)
sb<-sd(b)
mf<-mean(f)
sf<-sd(f)
for (bb in 1:B){
i<-sample(1:n,size=n,replace=TRUE)
b<-data$petrifilm[i]
f<-data$hgmf[i]
mb=mean(b)
mf=mean(f)
sb=sd(b)
sf=sd(f)
rr=cor(b,f)
up=sb^2*sf^2*(1-rr^2)
down1=((sb^2+sf^2)/2)*(1+rr)
down2=(((sb^2+sf^2)/2)*(1-rr))+((mb-mf)^2/2)
WILK[bb]=up/(down1*down2)
}
print(vv<-var(-2*log(WILK)))
print(mm<-mean(-2*log(WILK)))
hist(-2*log(WILK),prob=TRUE)
xx<-rchisq(1000,2)
qqplot(xx,-2*log(WILK))
quantile(-2*log(WILK), probs = c(.05,.95))
quantile(xx,probs=c(.05,.95))
```

Appendix 3. R Code for Fitting Pearson's Curve for the B12 and Folate Data of Example 3

R-Code

Bootstrapping the distribution of the PIVOT $\tilde{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$

```

library(ggplot2)
library(boot)
B=1000
n=nrow(data)
n
PIVOT=numeric(B)
b<-data$logB12
f<-data$logFolate
mb<-mean(b)
sb<-sd(b)
s1<-var(b)
mf<-mean(f)
sf<-sd(f)
s2<-var(f)
for (bb in 1:B){
i<-sample(1:n,size=n,replace=TRUE)
b<-data$logB12[i]
f<-data$logFolate[i]
mb=mean(b)
mf=mean(f)
sb=sd(b)
sf=sd(f)
r=cor(b,f)
r
t1=mb/sb
t2=mf/sf
MU2<-(2*(1-r)/(n-1))+((t1^2+t2^2-2*t1*t2*r^2)/(2*(n-1)))
MU2
MU31<-(t1^3+t2^3)/(n-1)^2
MU32<-(3*n*r^2*t1*t2*(t1+t2))/(2*(n-1))
MU3<- MU31-MU32
A1<-(12*(1-r)^2/n)
A2<-(6*r^2*t1*t2/(n*(n-1)))
A3<-((3*(n+3)*r^2*t1*t2*(t1^2+t2^2))/(n-1)^3)
A4<-((3*(n+1)*(t1^4+t2^4))/(4*(n-1)^3))
A51<-(3*t1^2*t2^2/8)
A52<-(n+1)/(n-1)^2
A53<-((8*(n+1)*r^4)/(n-1)^3)+(4*r^4/(n-1)^3)

```

```
MU4<- A1+A2-A3+A4+A51*(A52+A53)
PIVOT[bb]=(t2-t1)
}
print(vv<-var(PIVOT))
print(mm<-mean(PIVOT))
hist(PIVOT,prob=TRUE)
xx<-rnorm(1000)
qqplot(xx,PIVOT)
quantile(PIVOT, probs = c(.05,.95))
den <- density(PIVOT)
plot(den, frame = FALSE, col = "blue",main = "Density plot")
#####MOMENTS AND PEARSON CURVE
library(moments)
all.moments(PIVOT,central=TRUE,order.max=4)
s=skewness(PIVOT)
k=kurtosis(PIVOT)
m=mean(PIVOT)
v=var(PIVOT)
library(PearsonDS)
##### PEARSON CODE#####
library(PearsonDS)
moments<-c(mean=m,variance=v,skewness=s,kurtosis=k)
ppar<-pearsonFitM(moments=moments)
print(unlist(ppar))
pearsonMoments(params=ppar)
pearsonDia-
gram(max.skewness=sqrt(.25),max.kurtosis=4,squared.skewness=TRUE,lwd=2,l
egend=TRUE,n=999)
pearsonMoments(moments=moments)
```