# Similarities of Hotelling`s $T^2$ and Students statistics

**MILLARD R. MAMHOT**
**Lecturer**
Mathematics Department
College of Arts Sciences
Silliman University
Dumaguete City,Philippines
mllrdmmht@yahoo.com

**MAHERA R.QASEM**
**Lecturer**
Mathematics Department
College of Education
Tikrit University
mahera_rabee@yahoo.com

**Nadwa S. Yonis**
**Assistant Lecturer**
Mathematics
Department
College of Education
Mousl University

**ABSTRACT**

In this paper, the probability coverage of $P\left(T^2 \le \frac{p(n-1)}{n-p} F_{n,n-p}(\alpha)\right)$ is investigated to samples of different sizes that come from non−normal bivariate distributions, particularly, from skewed distributions. When samples come from univariate skewed distributions, the objective of this paper to develop an expression for a given underlying density of a random sample in terms of skewness coefficient, it is found that the relationship between skewness and confidence intervals is evident when the variance of the population is not given also that populations with big coefficients of skewness required bigger n for the probability coverage for $\mu$ to be exactly equal to $(1-\alpha)\times 100\%$.

**KEYWORDS:**Statistics, Bivariate distribiution, skewed distribution.

## 1- Introduction

There is a general tendency in the statistical literature towards more flexible methods, to represent features of the data as adequately as possible and reduce unrealistic assumptions. For the treatment of continuous multivariate observations within a parametric approach, one aspect which has been little affected by the above process is the overwhelming role played by the assumption of normality which underlies most methods for multivariate analysis. A major reason for this state of affairs is certainly the unrivaled mathematical tractability of the multivariate normal distribution, in particular its simplicity when dealing with fundamental operations like linear combinations, marginalization and conditioning, and indeed its closure under these operations. From a practical viewpoint, the most commonly adopted approach is transformation of the variables to achieve multivariate normality, and in a number of cases this works satisfactorily. There are however also problems: (i) the transformations

are usually on each component separately, and achievement of joint normality is only hoped for; (ii) the transformed variables are more difficult to deal with as for interpretation, especially when each variable is transformed using a different function; (iii) when multivariate homoscedasticity is required, this often requires a different transformation from the one for normality. Alternatively, there exist several other parametric classes of multivariate distributions to choose from, although the choice is not as wide as in univariate case; many of them are reviewed by Johnson & Kotz (1972). A special mention is due to the hyperbolic distribution and its generalized version, which form a very flexible and mathematically fairly tractable parametric class; see Barndorff-Nielsen & Blæsild (1983) for a summary account, and Blæsild (1981) for a detailed treatment of the bivariate case and a numerical example. As for extensions of distribution theory of classical statistical methods, the direction which seems to have been explored more systematically in this context is the extension of distribution theory of traditional sample statistics to the case of elliptical distribution of the underlying population; elliptical distributions represent a natural extension of the concept of symmetry to the multivariate setting.

The Hotelling's T2 statistic $n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)$ has many properties similar to that of the t statistic, $\frac{(\bar{x}-u)\sqrt{n}}{s}$ .One of them is that the probability coverage, $P\left(\bar{X}-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}<\mu<\bar{X}+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\right)$, is exactly equal to (1–α)×100% if the random variable X comes from a normal distribution and s, the sample standard deviation. This property is very much alike to the probability coverage of the following

$$P\left(n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)\le\frac{p(n-1)}{n-p}F_{n,(n-p)}(\alpha)\right)$$

where S is the sample covariance of a multivariate random variable X coming from a multivariate normal distribution.

Another property in which these two statistics share is on the effect of skewness of the underlying distribution of X and the sample size n. Boos and Oliver (2000) investigated the

effects of these on the probability coverage on the univariate population mean μ and found out that different sample sizes yield different probability coverage and likewise with corresponding different coefficients of skewness. Boos et. al.'s findings were supported by the one–term expansion of the Edgeworth series (Hall, 1987), which states that

$$P(tn \leq t) \approx P(Z \leq t) + \frac{\sqrt{\beta_1}(X)}{\sqrt{n}} \frac{(2t^2+1)}{6} \phi(t),$$

where $\sqrt{\beta_1}(X)$ is the skewness coefficient of the distribution of X and n the sample size. Sen et. al. (1992) offered a discussion of the minimum sample size needed to ensure the validity of classical intervals for means with platykurtic distributions. Chen (1995) mentioned how skewness may affect the accuracy of tests of hypothesis about means of normal populations using classical t–tests. This paper investigates the same problem for bivariate populations.

**Generalization of Hall's Findings to Multivariate Case**

In a multivariate case, if X1, X2, …Xn are random sample from a multivariate normal distribution with mean μ and covariance Σ, then the statistic, $(X - \mu)\Sigma{-}1(X - \mu)$ has a chi–square distribution with the corresponding probability coverage:

$$P\left(n(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu) \leq \chi_p^2(\alpha)\right) = 1-\alpha$$

if Σ is known. If Σ is unknown, then

$$P\left(n(\overline{X}-\mu)'S^{-1}(\overline{X}-\mu) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right) = 1-\alpha$$

where S is the estimated covariance of p×n matrix X, Fp, n–p(α) is the upper (100α)th percentile of the Fp, n–p distribution. This is true as long as X comes from a multivariate normal distribution. Thus, a natural question to ask is, what happens when X comes form a non–normal distribution, specifically form a skewed distribution?.

Let KX(β) be the cumulant generating function for the random variable X, β, an n×1 vector. To estimate the density fX(x) using saddlepoint approximation, fX(x) is embedded into an exponential family and a density in the exponential family is chosen to be approximated. An approximation of the chosen density results in an approximation of fX(x) since members of the exponential family differ only be a factor of exp(xβ – KX(β)).

The approximation of fX(x) is finally accomplished upon expanding the chosen member using the Edgeworth series. The Edgeworth series approximation of a density of a multivariate variable X is given as

$$fX(x) = \phi(x,\Sigma) \sum_{j=0}^{\infty} \sum_{s\in S(j)} \frac{1}{j!} \mu^{*s_1 \ldots s_j} (-1)^j (h_{s_1 \ldots s_j}(x,\Sigma))$$

where $\phi(x,\Sigma)$ is a multivariate normal distribution, s, a 1×p vector of integers, μ*s are pseudo–moments, hs(x,Σ) are generalized Hermite polynomials.

Suppose fX(x) is the underlying density of random sample X1, X2, …, Xn. Then embedding fX(x) into an exponential family, the following expression for fX(x) is obtained:

$$fX(x) = f_X(x,\hat{\beta}) \exp[K_X(\hat{\beta}) - \hat{\beta}^T x]$$

where $\hat{\beta}$ Is the solution of K'(β) = x and KX is the cumulant generating function of $fX(x,\hat{\beta})$, the chosen member from the exponential family with mean x.

Approximating $fX(x,\hat{\beta})$ with a normal density with mean0,we have the following

$$fX(x) = \exp\left(n[K_X(\hat{\beta}) - \hat{\beta}^T x]\right)\left(\frac{n}{2\pi}\right)^{p/2} \det[K_X''(\hat{\beta})]^{-1/2}\left[1 + \frac{b(\hat{\beta})}{2n} + O(n^{-2})\right]$$

where $b(\hat{\beta})$ is the tilt measure for $fX(x,\hat{\beta})$ and n, the sample size. Then by Edgeworth expansion,

$$fX(x) = \sum_{j=0}^{\infty} \sum_{s\in S(j)} \frac{1}{j!} \mu^{*s_1 \ldots s_j} (-1)^j \frac{d^j}{dx^{s_1} \cdots dx^{s_j}}$$

$$= \sum_{j=0}^{\infty} \sum_{s\in S(j)} \frac{1}{j!} \mu^{*s_1 \ldots s_j} (-1)^j (h_{s_1 \ldots s_j}(0;K''(\hat{\beta}))(\phi(0,K''(\hat{\beta})))$$

$$= \phi(x,K''(\hat{\beta}))\left(\sum_{j=0}^{\infty} \sum_{s\in S(j)} \frac{1}{j!} \mu^{*s_1 \ldots s_j} (-1)^j (h_{s_1 \ldots s_j}(0;K''(\hat{\beta})))\right)$$

$$= \phi(x,K''(\hat{\beta}))\left(1 + \sum_{j=1}^{\infty} \sum_{s\in S(j)} \frac{1}{j!} \mu^{*s_1 \ldots s_j} (-1)^j (h_{s_1 \ldots s_j}(0;K''(\hat{\beta})))\right)$$

Now, using cumulants instead of pseudo–moments, we get

$$fX(x) = \phi(x,K"(\hat{\beta}))\left(\exp\left(\sum_{j=3}^{\infty}\sum_{s\in S(j)}\frac{1}{j!}\mu^{*s_1\ldots s_j}(-1)^j(h_{s_1\ldots s_j}(0;K"(\hat{\beta})))\right)\right)$$

$$= \phi(x,K"(\hat{\beta}))\left(\left(1+\left(\sum_{j=3}^{\infty}\sum_{s\in S(j)}\frac{1}{j!}\mu^{*s_1\ldots s_j}(-1)^j(h_{s_1\ldots s_j}(0;K"(\hat{\beta})))\right)\right)\right.$$

$$+\frac{1}{2}\left(\sum_{j=3}^{\infty}\sum_{s\in S(j)}\frac{1}{j!}\mu^{*s_1\ldots s_j}(-1)^j(h_{s_1\ldots s_j}(0;K"(\hat{\beta})))\right)^2$$

$$\left.+\frac{1}{3}\left(\sum_{j=3}^{\infty}\sum_{s\in S(j)}\frac{1}{j!}\mu^{*s_1\ldots s_j}(-1)^j(h_{s_1\ldots s_j}(0;K"(\hat{\beta})))\right)^3+\ldots\right)$$

$$= \phi(x,K"(\hat{\beta}))[1+(\frac{1}{3!}\hat{\kappa}^{ijk}(-1)h_{ijk}(0;K"(\hat{\beta}))+\frac{1}{4!}\hat{\kappa}^{ijkl}h_{ijkl}(0;K"(\hat{\beta}))+\cdots)$$

$$+\frac{1}{2}(\frac{1}{3!}\hat{\kappa}^{ijk}(-1)h_{ijk}(0;K"(\hat{\beta}))+\frac{1}{4!}\hat{\kappa}^{ijkl}hijkl(0;K"(\hat{\beta}))+\cdots)^2+\cdots]$$

Since $hijk(0,K"(\hat{\beta})) = 0$, $\frac{1}{3!}\hat{\kappa}^{ijk}h_{ijk}(0;K"(\hat{\beta}))$. The terms of order O(n–1) are:

$$\frac{1}{4!}\hat{\kappa}^{ijkl}hijkl(0;K"(\hat{\beta}))+\cdots)+\frac{1}{2\times 3!3!}\hat{\kappa}^{ijk}\hat{\kappa}^{lmo}h_{ijklmo}(0;K"(\hat{\beta}))$$

Equating this with $2b(\hat{\beta})$, we get

$$2b(\hat{\beta}) = \frac{1}{4!}\hat{\kappa}^{ijkl}hijkl(0;K"(\hat{\beta}))+\cdots)+\frac{1}{2\times 3!3!}\hat{\kappa}^{ijk}\hat{\kappa}^{lmo}h_{ijklmo}(0;K"(\hat{\beta}))$$

Since $hijkl(0;\Sigma) = \kappa ij\kappa kl^{[3]} = \kappa ij\kappa kl + \kappa ik\kappa jl + \kappa il\kappa jk$

and from McCullagh (1987), $hijklmn(0;\Sigma) = \kappa ij\kappa kl\kappa mn^{[15]}$,

we have $2b(\hat{\beta}) = \frac{1}{4!}\hat{\kappa}^{ijkl}(\hat{\kappa}_{ij}\hat{\kappa}_{kl}[3])-\frac{10}{6!}\hat{\kappa}^{ijk}\hat{\kappa}^{lmo}(\hat{\kappa}_{ij}\hat{\kappa}_{kl}\hat{\kappa}_{mo}[15])$.

Thus,

$$b(\hat{\beta}) = \frac{1}{4}\hat{\kappa}^{ijkl}(\hat{\kappa}_{ij}\hat{\kappa}_{kl}) - \frac{25}{12}\hat{\kappa}^{ijk}\hat{\kappa}^{lmo}(\hat{\kappa}_{ij}\hat{\kappa}_{kl}\hat{\kappa}_{mo})$$

Since b1,p = $\hat{\kappa}^{ijk}\hat{\kappa}^{lmo}(\hat{\kappa}_{ij}\hat{\kappa}_{kl}\hat{\kappa}_{mo})$ and b2, p = $\hat{\kappa}^{ijkl}(\hat{\kappa}_{ij}\hat{\kappa}_{kl})$ (Mardia, 1970), we have

$$b(\hat{\beta}) = \frac{1}{4}b_{2,p} - \frac{25}{12}b_{1,p}$$

Substituting this to (7), we get

$$f_X(x) = \exp\left(n[K_X(\hat{\beta}) - \hat{\beta}^T x]\right)\left(\frac{n}{2\pi}\right)^{p/2} \det[K_X''(\hat{\beta})]^{-1/2}\left[1 + \frac{\frac{1}{4}b_{2,p} - \frac{25}{12}b_{1,p}}{2n} + O(n^{-2})\right]$$

$$= \exp\left(n[K_X(\hat{\beta}) - \hat{\beta}^T x]\right)\left(\frac{n}{2\pi}\right)^{p/2} \det[K_X''(\hat{\beta})]^{-1/2}\left[1 + \frac{b_{2,p}}{8n} - \frac{25b_{1,p}}{24n} + O(n^{-2})\right]$$

Hence,

$$f_X(x) = MVNp(\mu,\Sigma)\left[1 + \frac{b_{2,p}}{8n} - \frac{25b_{1,p}}{24n} + O(n^{-2})\right]$$

This result has two important implications, namely:

1. If f is the underlying distribution of a random sample X1, X2, …, Xn with KX(β) as its cumulant generating function, then $f_X(x) \to MVNp(\mu,\Sigma)$ as $n \to \infty$.

This means that even if the underlying distribution of the given random sample, we can still apply the following probability coverage

$$P\left(n(\overline{X} - \mu)'S^{-1}(\overline{X} - \mu) \le \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)\right) = 1 - \alpha$$

where S is the estimated covariance of p×n matrix X, $F_{p,n-p}(\alpha)$ is the upper $(100\alpha)$th percentile of the Fp, n–p distribution for large n since as $n \to \infty$, $f_X(x) \approx MVNp(\mu,\Sigma)$.

How large should n be so that $f_X(x) \approx MVNp(\mu,\Sigma)$? The following simulation results give us some estimate on the size of n:

**Table1. Percentages of times the population mean μ is within the confidence interval when X comes from distributions with different measures of skewness b1, p, unknown variance, different sample sizes, and with $\alpha = 0.05$.**

| Sample Size N | Sample from Normal(0, 1) | Sample from Skewed Dist. With b1,p ≈3.50 | Sample from Skewed Dist. With b1,p ≈7.60 |
|---|---|---|---|
| 20 | 94.68 | 92.48 | 89.19 |
| 30 | 95.04 | 92.98 | 91.13 |
| 50 | 94.68 | 93.6 | 92.21 |
| 100 | 94.80 | 93.98 | 93.05 |
| 300 | 952.2 | 94.17 | 94.32 |
| 500 | 94.70 | 94.21 | 93.93 |

Thus, at about n = 500, $P\left(n(\overline{X}-\mu)'S^{-1}(\overline{X}-\mu) \le \frac{(n-1)2}{(n-2)}F_{p,n-2}(0.05)\right)$ = 94.21 when b1, p = 3.5 and $P\left(n(\overline{X}-\mu)'S^{-1}(\overline{X}-\mu) \le \frac{(n-1)2}{(n-2)}F_{p,n-2}(0.05)\right)$ = 93.93 when b1, p = 7.60.

These findings are very much similar to that of the probability coverage for the univariate distribution using the t–test. The following table show the simulation of the probability coverage of $P\left(\overline{X}-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}} < \mu < \overline{X}+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\right)$, where s, the sample standard deviation

.

**Table 2. . Percentages of times the population mean μ is within the confidence interval $\left(\overline{X}-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}, \overline{X}+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\right)$ when X comes from a distribution with different measures of skewness $\sqrt{\beta_1}(X)$, unknown**

**variance, different sample sizes, and with $\alpha = 0.05$.**

| Sample Size n | $\sqrt{\beta_1}(X)$ | Percentage of times μ is within the confidence interval |
|---|---|---|
| 30 | 3.52 | 89.61 |
| 50 | 3.52 | 91.49 |
| 100 | 3.52 | 92.29 |
| 200 | 3.52 | 93.96 |
| 300 | 3.52 | 94.77 |
| 500 | 3.52 | 93.76 |
| 30 | 6.57 | 83.98 |
| 50 | 6.57 | 87.38 |
| 100 | 6.57 | 90.47 |
| 200 | 6.57 | 91.80 |
| 300 | 6.57 | 92.20 |
| 500 | 6.57 | 94.04 |

From the table, we see that when n is about 500, and the skewness of the distribution, $\sqrt{\beta_1}(X) = 3.52$, the probability coverage is about 93.76% and when $\sqrt{\beta_1}(X) = 6.57$, for the same sample size, the probability coverage is about 94.04%

2. The other important result is that when the underlying distribution of a random sample X1, X2, …, Xn

with KX(β) as the cumulant generating function of f. Then $fX(x) \approx MVNp(\mu,\Sigma)$ as $b1,p \to 0$.

**Conclusion and Recommendation**

The relationship between skewness and confidence intervals is evident when the variance of the population is not given and this relationship seemed to be a direct one since, as the simulation suggests, populations with big coefficients of skewness required bigger n for the probability coverage for μ to be exactly equal to $(1-\alpha)\times 100\%$. The univariate case was supported by the one–term expansion of the Edwort

h series which

$$P(tn \le t) \approx P(Z \le t) + \frac{\sqrt{\beta_1}(X)}{\sqrt{n}}\frac{(2t^2+1)}{6}\phi(t)$$

As for the multivariate case the relationship is

$$f_X(x) = MVN_p(\mu,\Sigma)\left[1+\frac{b_{2,p}}{8n}-\frac{25b_{1,p}}{24n}+O(n^{-2})\right]$$

The simulation was not really able to find the exact value of the sample size n so that

$$P\left(\overline{X}-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}} < \mu < \overline{X}+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\right) = 1-\alpha$$

and

$$P\left(n(\overline{x}-\mu)'S^{-1}(\overline{x}-\mu) \le \frac{p(n-1)}{n-p}F_{n,(n-p)}(\alpha)\right) = 1-\alpha .$$

It is recommended therefore that further simulation is performed to identify n for a given skewness;

Once a number of simulation is performed for several n and several skewness coefficients, it is recommended that an equation be formulated expressing explicitly the relationship of n and skewness, such as the following: n = f(b1, 2) or n = f( $\sqrt{\beta_1}x$ ) where b1, 2 and $\sqrt{\beta_1}x$ are coefficients of skewness for bivariate and univariate distributions, respectively.

**References**

Barndorff-Nielsen, O. & Blæsild, P. (1983). Hyperbolic distributions. In: Encyclopedia of Statistical Sciences (ed. N.L.Johnson, S.Kotz & C.B.Read), vol. 3, 700.707. Wiley, New York.

Blæsild, P. (1981). The two-dimensional hyperbolic distribution and related distributions,

with an application to Johansen's bean data. Biometrika, 68, 251.63.

Boos, D.D.,and Oliver, J.H. (2000), "How Large Does n Have to be for Z and t Intervals?" Journal of American Statistical Association, 54, 121–128.

Hall, P. (1992), "On the Removal of Skewness by Transformation" Journal of Royal Statistical Society, Series B, 221–228.

Johnson, N. L. & Kotz, S. (1972). Distributions in statistics: continuous multivariate distributions. Wiley, New York

Johnson, N.J. (1978), "Modified t Tests and Confidence Intervals for Asymmetric Populations" Journal of American Association, 73, 536–547.

Johnson, R.A. and Wichern, D.W. (1998). Applied Multivariate Statistical Analysis, Prentice–Hall International, Inc.

Kolassa, J.E. (1997), Series Approximation Methods in Statistics. 2nd Edition, Springer–Verlag.

Mardia, K.V. (1970), "Measures of Multivariate Skewness and Kurtosis with Applications" Biometrika, 57, 519–530.

Zhou, X.H.,and Gao, S. (2000), "One–sided Confidence Intervals for Means and Positively Skewed Distributions" Journal of American Statistical Association, 54, 100–104.

# التشابه بين معاملات هوتلنغ T2 و الدوال الاحصائية للطلبة

**ماهرة ربيع قاسم**
**قسم الرياضيات**
**كلية التربية**
**جامعة تكريت**

**ندوى سالم يونس**
**قسم الرياضيات**
**كلية التربية**
**جامعة الموصل**

**ميلارد ماهوت**
**كلية العلوم – جامعة سيليمان**
**مدينة ديماكوت-الفلبين**

**الخلاصة**

في هذا البحث تم دراسة احتمالية تغطية المعامل P باستخدام عينات بحجوم مختلفة و التي تم استحصالها من التوزيعات غير الطبيعية لبايفاريت و خاصة من توزيعات سكويد. و قد تبين من خلال البحث ان العلاقة بين تقسيمات الانحراف و الموثوقية واضحة جداً عندما لا تعطى تباينات المجتمع و تبدو هذه العلاقة مباشرة جدا حيث ان كان المجتمع بمعاملات انحراف عالية– كما تفترض المحاكاة- تحتاج معامل n أكبر لتغطية الاحتمالية حيث أن العلاقة تكون مساوية تماماً لـ :

$$P(tn \le t) \approx P(Z \le t) + \frac{\sqrt{\beta_1}(X)}{\sqrt{n}} \frac{(2t^2+1)}{6} \phi(t)$$

ان الحالات التي لا ينطبق عليها الاختبار متعدد المتغيرات فقد تم اسنادها بتوسع الفقرة الواحدة باستخدام سلسلة إدوارد و كما هو موضح من العلاقة:

$$f_X(x) = MVN_p(\mu,\Sigma)\left[1+\frac{b_{2,p}}{8n}-\frac{25b_{1,p}}{24n}+O(n^{-2})\right]$$

و بالنسبة لعلاقة متعدد المتغيرات فقد كانت العلاقة المستحصلة هي :

$$P\left(\overline{X}-\frac{t_{\alpha/2,n-1}s}{\sqrt{n}} < \mu < \overline{X}+\frac{t_{\alpha/2,n-1}s}{\sqrt{n}}\right) = 1-\alpha$$

و لم تستطع المحاكاة ان تصل الى قيمة حقيقية لمعامل n .