

TABLES OF PERCENTAGE POINTS OF THE k -VARIATE
NORMAL DISTRIBUTION FOR LARGE VALUES OF k

William C. Horrace

Department of Economics
University of Arizona
Tucson, Arizona 85721-0108

Key words: multivariate normal distribution; multiple comparisons; simultaneous confidence intervals

ABSTRACT

This paper gives tabulations of the upper α percentage points of the maximum absolute value of the k -variate normal distribution with common correlation ρ for values of k as high as 500. The tables are useful for performing multiple comparison procedures in experiments with large numbers of treatments.

1. INTRODUCTION

This paper gives tabulations of the upper α percentage points of the maximum absolute value of the k -variate normal distribution with common correlation ρ for large values of k . A random vector $Z = (Z_1, \dots, Z_k)$ has a k -variate normal distribution with mean vector μ and covariance matrix Σ , if its density is given by:

$$f(z) = (2\pi)^{-k/2} (\det \Sigma)^{-1/2} \exp\{-\frac{1}{2} (z-\mu)' \Sigma^{-1} (z-\mu)\}.$$

If the Z_i 's are standardized so that $E(Z_i) = 0$ and $var(Z_i) = 1$, then Σ is a correlation matrix with off-diagonal elements $\rho_{ij} = corr(Z_i, Z_j)$ for $i \neq j$. In this paper we tabulate percentage points for $\max |Z_i|$ for the special case $\rho_{ij} = \rho$ for $i \neq j$, which we denote $|z_{k,\rho,\alpha}|$. Previous tabulations for $|z_{k,\rho,\alpha}|$ are due to Odeh (1982) and only go as high as k

= 50. Tables for $z_{k,\rho,\alpha}$, one-sided percentage points, are due to Milton (1963), and tables for $(|z_{k,\rho,\alpha}|)^2$ are due to Krishnaiah and Armitage (1965).

It should be noted that the distribution of Z_i 's is the limiting distribution of a k -variate Student t distribution with common correlation ρ and degrees of freedom ν , where $\nu \rightarrow \infty$. Similar tabulations of the k -variate Student t were originally due to Dunnett and Sobel (1954, 1955), Cornish (1954) and most recently Bechhofer and Dunnett (1986). In the interim there have been several other tabulations; see Bechhofer and Dunnett (1986) for a complete and detailed survey. These tabulations contain $|z_{k,\rho,\alpha}|$ as a special case, when $\nu = \infty$, but only go as high as $k = 20$.

These types of multivariate tabulations (normal or Student's t) have many well known applications, but are quite frequently used for confidence interval construction and inference in multiple comparison procedures. See Hochberg and Tamhane (1987) and Hsu (1996). As the scope and applicability of multiple comparison procedures expands and as data set sizes continue to grow, there has arisen a need for tables for large values of k . See Horrace and Schmidt (1996, 1995) for some applications with k large. Tong (1970) gives some conservative approximations for large values of k based on small k percentage points. However, this is no substitute for carefully constructed tables. The tables herein attempt to fill this void by going as high as $k = 500$. The tables are only for the k -variate normal. However, since large k typically implies large degrees of freedom, these tables can, for all practical purposes, be thought of as also being for the k -variate Student t .

2. DESCRIPTION OF THE TABLES.

Tables I - III provide tabulations of the factor $|z_{k,\rho,\alpha}| = z(k, \rho, \alpha)$ as the solution in z of the probability integral:

$$pr\{\max|Z_i| \leq z\} = \int_{-\infty}^{\infty} \left[\Phi \left\{ \frac{u\rho^{.5} + z}{(1-\rho)^{.5}} \right\} - \Phi \left\{ \frac{u\rho^{.5} - z}{(1-\rho)^{.5}} \right\} \right]^k d\Phi(u) = 1 - \alpha, \quad (1)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The critical value $|z_{k,\rho,\alpha}|$ is tabulated for all combinations $k = 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 250, 300, 350, 400$ and 500 ; $\rho = 0.1, 0.3, 0.5, 0.7$ and 0.9 ; and $\alpha = 0.10$ (Table I), 0.05 (Table II) and 0.01 (Table III). Interpolation rules are recommended as follows.

1. Interpolation with respect to k is done linearly in $\log k$.
2. Interpolation with respect to ρ is done linearly in $(1-\rho)^{-1}$.
3. Interpolation with respect to α is done linearly in $\log\alpha$.

TABLE I

Upper $\alpha \times 100$ percentage points of the distribution of the largest absolute value of k normal variates with common correlation ρ
 $\alpha = 0.10$

k	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
30	2.91	2.86	2.76	2.57	2.23
40	3.00	2.94	2.83	2.63	2.27
50	3.07	3.01	2.88	2.67	2.29
60	3.12	3.06	2.93	2.71	2.32
70	3.17	3.10	2.96	2.74	2.33
80	3.20	3.13	3.00	2.77	2.35
90	3.24	3.17	3.02	2.79	2.36
100	3.27	3.19	3.04	2.81	2.37
120	3.32	3.24	3.09	2.84	2.39
140	3.36	3.28	3.13	2.87	2.41
160	3.40	3.32	3.15	2.89	2.42
180	3.43	3.34	3.18	2.91	2.44
200	3.46	3.37	3.20	2.93	2.45
250	3.52	3.43	3.25	2.97	2.47
300	3.56	3.47	3.30	3.00	2.49
350	3.60	3.51	3.32	3.02	2.50
400	3.64	3.54	3.35	3.05	2.51
500	3.69	3.59	3.39	3.08	2.53

TABLE II
*Upper $\alpha \times 100$ percentage points of the distribution of the largest absolute
value of k normal variates with common correlation ρ*
 $\alpha = 0.05$

k	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
30	3.13	3.10	3.01	2.85	2.53
40	3.22	3.18	3.08	2.91	2.57
50	3.28	3.24	3.14	2.95	2.60
60	3.33	3.28	3.18	2.99	2.62
70	3.37	3.33	3.21	3.02	2.64
80	3.41	3.36	3.25	3.04	2.65
90	3.44	3.39	3.27	3.06	2.66
100	3.47	3.41	3.30	3.08	2.67
120	3.52	3.46	3.34	3.11	2.69
140	3.56	3.50	3.37	3.14	2.71
160	3.59	3.53	3.40	3.17	2.72
180	3.62	3.56	3.43	3.18	2.74
200	3.65	3.59	3.45	3.21	2.75
250	3.71	3.64	3.50	3.24	2.77
300	3.75	3.69	3.53	3.27	2.79
350	3.79	3.72	3.57	3.30	2.80
400	3.82	3.75	3.59	3.32	2.81
500	3.88	3.80	3.64	3.35	2.83

Due to some of the large values of k involved, numerical integration of equation (1) was impractical, therefore the $|z_{k,\rho,\alpha}|$ were simulated using the Gauss programming language. The simulation algorithm follows.

1. Draw $k+1$ independent standard normal random numbers, $z_{1n}, \dots, z_{(k+1)n}$.
2. Generate $z_{in}^* = z_{in}(1 - \rho)^.5 + z_{(k+1)n}\rho^>.5$, $i = 1, \dots, k$.
3. Find $y_n = \max |z_{in}^*|$, $i = 1, \dots, k$.
4. Perform steps 1, 2, and 3 for $n = 1, \dots, N$.
5. Calculate $(1-\alpha) \times 100$ percentiles from y_n , $n = 1, \dots, N$.

TABLE III
*Upper $\alpha \times 100$ percentage points of the distribution of the largest absolute
value of k normal variates with common correlation ρ
 $\alpha = 0.01$*

k	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
30	3.59	3.57	3.52	3.40	3.12
40	3.66	3.64	3.59	3.46	3.16
50	3.72	3.70	3.64	3.49	3.18
60	3.76	3.74	3.68	3.53	3.21
70	3.80	3.78	3.71	3.56	3.23
80	3.83	3.81	3.74	3.59	3.24
90	3.86	3.84	3.76	3.60	3.25
100	3.89	3.87	3.80	3.63	3.27
120	3.93	3.90	3.83	3.66	3.28
140	3.97	3.94	3.86	3.69	3.30
160	4.00	3.97	3.89	3.70	3.32
180	4.03	4.00	3.91	3.73	3.33
200	4.05	4.02	3.94	3.74	3.34
250	4.10	4.08	3.98	3.78	3.36
300	4.15	4.12	4.01	3.81	3.38
350	4.18	4.15	4.05	3.83	3.39
400	4.21	4.18	4.07	3.86	3.40
500	4.26	4.23	4.11	3.89	3.42

Here N , the simulation sample size, was set to 10^6 . A similar simulation procedure was used by Freeman and Kuzmak (1972) to estimate the percentage points of the Student t distribution, however the accuracy of their tables is suspect since their sample size was set to a mere 25,000.

3. ACCURACY OF THE TABLES

With a sample size N drawn from distribution $F(\)$, under certain regularity conditions satisfied in this context, $F\{(1-\alpha)\times 100\%ile\}$ is asymptotically normal with mean $(1-\alpha)$ and variance $(1-\alpha)\alpha/N$; see Wilks (1962, p. 271). This being the case, the standard deviation of the estimate of each percentile is known and confidence intervals for the estimates can be constructed. This gives a rough idea of the accuracy of the

tabulated values. Thus, 95% confidence intervals for each value of α are shown in Table IV. These are fairly tight confidence intervals.

TABLE IV

Confidence Intervals for Percentiles

Nominal Percentile	95% Confidence Interval
0.90	[0.8994, 0.9006]
0.95	[0.9496, 0.9504]
0.99	[0.9898, 0.9902]

Since N was very large we could accurately draw percentage points of the upper and lower bound for each confidence interval to compare them to the nominal percentage points, except for $\alpha = 0.10$ in which case we only had the upper bound. In almost all cases the bounds matched the nominal value to three significant digits with the values for $\alpha = 0.10$ being the most accurate and those for $\alpha = 0.01$ being least accurate. We found typical critical value tolerances of ± 0.002 , ± 0.003 and ± 0.005 for $\alpha = 0.10$, $\alpha = 0.05$ and $\alpha = 0.01$, respectively. As presented, the tables are probably more accurate than necessary for most applied problems and do give a sense the effects of changes in the parameters k , α and p .

As an additional check, the simulation was performed for small values of k and compared to the results of the Bechhofer and Dunnett (1986) tabulations, which are the most accurate tabulations available. In all cases the results exhibited exact correspondence and accuracy to at least three significant digits.

4. AN APPLICATION

A number of applications for small k are given in Hahn and Hedrickson (1971), Hochberg and Tamhane (1987) and Hsu (1996) to name a few. These applications are typically for tabulations of the multivariate t -distribution. However with large enough

k the tables presented here can be used instead as long as the degrees of freedom are concomitantly large. What follows is an application of the tables called *unconstrained multiple comparisons with the best*.

4.1 *Unconstrained Multiple Comparisons with the Best.* Unconstrained Multiple comparisons with the best were originally due to Edwards and Hsu (1983). Suppose that for the i^{th} treatment a random sample Y_{i1}, \dots, Y_{iT} of size T is taken. Further assume that observations between treatments $i = 1, \dots, k$ are independent. Then under a normality and equality of variances assumption:

$$Y_{it} = \mu_i + \varepsilon_{it}, \quad i = 1, \dots, k, \quad t = 1, \dots, T$$

where $\varepsilon_{11}, \dots, \varepsilon_{kT}$ are identically and independently distributed normal errors with mean 0 and variance σ^2 . Define the usual sample means and pooled sample variance

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}, \quad \hat{\sigma}^2 = \frac{1}{k(T-1)} \sum_{i=1}^k \sum_{t=1}^T (Y_{it} - \hat{\mu}_i)^2.$$

Then for large k or T (i.e. k or T large enough that $v \rightarrow \infty$) a set of $(1-\alpha) \times 100\%$ simultaneous confidence intervals for all distances from the best treatment mean, $\max_j \mu_j - \mu_i$, for $i = 1, \dots, k$, is given by $[L_i, U_i]$ where:

$$L_i = \max\{\min_{j \in \zeta} (\hat{\mu}_j - \hat{\mu}_i - d), 0\}$$

$$U_i = \max\{\max_{j \neq i} (\hat{\mu}_j - \hat{\mu}_i + d), 0\}$$

$$\zeta = \{j: \hat{\mu}_j \geq \max_i \hat{\mu}_i - d\}$$

and $d = z_{k,0.5,\alpha} \hat{\sigma} (2/T)^{.5}$.

The $|z_{k,0.5,\alpha}|$ can be drawn from the tables and confidence intervals constructed.

ACKNOWLEDGEMENTS

Portions of this research were supported by the National Science Foundation and the Ford Foundation.

BIBLIOGRAPHY

- Bechhofer, R.E. & C.W. Dunnett (1986), Tables of the percentage points of multivariate Student t distributions, R.E. Odeh and J.M. Davenport, eds. *Selected Tables in Mathematical Statistics* (Providence: American Mathematical Society).
- Cornish, E.A. (1954), The multivariate small t-distribution associated with a set of normal sample deviates, *Australian J. of Physics.* **7**, 531-42.
- Dunnett, C.W. & M. Sobel (1954), A bivariate generalization of Student's t-distribution with tables for certain special cases, *Biometrika* **41**, 153-69.
- Dunnett, C.W. & M. Sobel (1955), Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution, *Biometrika*, **42**, 258-60.
- Edwards, D.G. & J.C. Hsu (1983), Multiple comparisons with the best treatment, *J. Amer. Stat. Assoc.* **78**, 965-71. Corrigenda (1984), *J. Amer. Stat. Assoc.* **79**, 965.
- Freeman, H & A.M. Kuzmack (1972), Tables of multivariate t in six and more dimensions, *Biometrika*, **59**, 217-9.
- Hahn, G.J. & R.W. Hendrickson (1971), A table of percentage points of the distribution of the largest absolute value of k Student t variates and its applications, *Biometrika* **58**, 23-32.
- Hochberg, Y. & A.C. Tamhane (1987), *Multiple Comparison Procedures* (Wiley, New York).
- Horrace, W.C. & P. Schmidt (1996), Confidence statements for efficiency estimates from stochastic frontier models *J. of Productivity Analysis* **7**, 257-82.
- Horrace, W.C. & P. Schmidt (1995), Multiple comparisons with the best, with applications to the efficiency measurement problem, Unpublished Manuscript, Michigan State University Department of Economics.
- Hsu, J.C. (1996), *Multiple Comparisons, Theory and Methods* (Chapman and Hall, London).

- Krishnaiah, P.R. & J.V. Armitage (1965), Tables for the distribution of the maximum of correlated chi-square variates with one degree of freedom, *Trabajos Estadist.* **16**, 91-96
- Milton, R.C. (1963), Tables of the equally correlated multivariate normal probability integral, Technical Report No 27, Department of Statistics, University of Minnesota, Minneapolis.
- Odeh, R.E (1982), Tables of the percentage points of the distribution of the maximum absolute value of equally correlated normal random variates, *Comm. Stat., Series B* **11**, 65-87.
- Tong, Y.L. (1970), Some Probability inequalities of multivariate normal and multivariate t , *J. Amer. Stat. Assoc.* **65**, No 331, 1243-7.
- Wilks, S.S. (1962), *Mathematical Statistics* (Wiley, New York).