

Two tests using more assumptions but lower power

Sang Kyu Lee^a, Hyoung-Moon Kim^{1,b}

^aDepartment of Statistics and Probability, Michigan State University, USA;

^bDepartment of Applied Statistics, Konkuk University, Korea

Abstract

Intuitively, a test with more assumptions has greater power than a test with fewer assumptions. This kind of examples are abundant in the nonparametric tests vs corresponding parametric ones. In general, the nonparametric tests are less efficient in terms of asymptotic relative efficiency (ARE) compared to corresponding parametric tests (Daniel, 1990). However, this is not always true. To test equal means under independent normal samples, the usual test involves using the t -distribution with the pooled estimator of the common variance. Adding the assumption of equal sample size, we may derive another test. In this case, two tests using more assumptions were performed for univariate (multivariate) cases. For these examples, it was found that the power function of a test with more assumptions is less than or equal to that of a test with fewer assumptions. This finding can be used as an expository example in master's mathematical statistics courses.

Keywords: assumption, power function, test

1. Introduction

It is well known that a test with more (or strong) assumptions has greater power than one with fewer assumptions. This kind of examples are abundant in the nonparametric tests vs corresponding parametric ones. In general, the nonparametric tests are less efficient in terms of asymptotic relative efficiency (ARE) compared to corresponding parametric tests (Daniel, 1990). Some examples are as follows: One-sample sign test, Cox-Stuart test for trend, median test, Ansari-Bradley test, Moses test, two related samples sign test, Wilcoxon matched-pairs signed-ranks test, Kruskal-Wallis one-way analysis of variance by ranks, Friedman two-way analysis of variance by ranks, and Kendall's tau among others.

For a illustration, we first provide an example using ARE. Another example is illustrated using power functions.

The first example is as follows. ARE is used to compare two different tests (Chapter 3 of Lehmann, 1999). ARE is defined as

$$e_{2,1} \doteq \frac{n_1}{n_2},$$

The research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07045603), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1A4A5032622).

¹ Corresponding author: Department of Applied Statistics, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Korea. E-mail: hmkim@konkuk.ac.kr

where n_1 and n_2 are the sample sizes required by the two tests to achieve the same power against the same alternatives at the same level α . If ARE is 1/2, then approximately $n_2 = 2n_1$; test 2 is half as efficient as test 1 because it requires double the sample size to achieve the same power.

Let X_1, \dots, X_n be a random sample from a distribution $F(x - \theta)$, pdf $f(x)$, and median θ . To test $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$, consider the test statistic of sign test (S)

$$S = \sum_{i=1}^n I(X_i > 0).$$

Adding the additional assumption that the pdf $f(x)$ is symmetric, we have the test statistic of Wilcoxon test (W)

$$W = \sum_{i=1}^n \text{sgn}(X_i) R|X_i|,$$

where $\text{sgn}(X_i)$ is the sign function and $R|X_i|$ denotes the rank of $|X_i|$, $i = 1, \dots, n$, from low to high. Further adding F is the standard normal distribution, we have the test statistic of usual t -test (t)

$$t = \frac{\bar{X}}{S_X / \sqrt{n}},$$

where S_X is the sample standard deviation of X_i , $i = 1, \dots, n$. AREs in the case where F is the standard normal distribution, it is well known that:

$$e_{S,t} = \frac{2}{\pi} \quad \text{and} \quad e_{W,t} = \frac{3}{\pi}.$$

See Examples 3.4.1 and 3.4.2 of Lehmann (1999) for more information.

A bivariate extension is possible. Let X_1, \dots, X_n and Y_1, \dots, Y_m be independent random samples from distributions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y) = F(y - \theta)$, respectively. To test $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$, we may use the t -test given in (2.2) and the Wilcoxon test. Similar to the one-sample case, the ARE of the Wilcoxon test with respect to the t -test is $3/\pi$ when F is normal. See Example 3.4.3 of Lehmann (1999) for more information.

Based on one-sample and two-sample examples, we find that the test with more (or strong) assumptions (t -test) has greater power than the tests with fewer assumptions (the sign test and the Wilcoxon test for the one-sample test case; the Wilcoxon test for the two-sample case) in an asymptotic sense.

The second example is as follows. Let X_1, \dots, X_n and Y_1, \dots, Y_m be independent random samples taken from two populations with (μ_i, σ^2) , where $\sigma^2 < \infty$, $i = 1, 2$. It is well known (Chapter 3 of Lehmann, 1999) that

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{1/n + 1/m}} \xrightarrow{d} N(0, 1), \quad (1.1)$$

where $S_p^2 = ((n-1)S_X^2 + (m-1)S_Y^2)/(n+m-2)$ is the pooled estimator of common variance, $S_X^2 = 1/(n-1) \sum_{i=1}^n (X_i - \bar{X})^2$, and $S_Y^2 = 1/(m-1) \sum_{i=1}^m (Y_i - \bar{Y})^2$. Based on (1.1), the power function for testing $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$ can be derived as follows:

$$\beta_{\phi_0} \left(\frac{\delta}{\sigma} \right) \doteq 1 - \Phi \left(z_\alpha - \frac{\delta}{\sigma \sqrt{1/n + 1/m}} \right),$$

where $\delta = \mu_1 - \mu_2$.

Adding the assumption of normality, that is, let X_1, \dots, X_n and Y_1, \dots, Y_m be independent random samples obtained from $N(\mu_i, \sigma^2)$, where $\sigma^2 < \infty$, $i = 1, 2$. The power function is derived and given in (2.4). We easily find that $\beta_{\phi_0}(\delta/\sigma) \geq \beta_{\phi_1}(\delta/\sigma)$ if $\delta/\sigma > 0$ and that the inequality is reversed if $\delta/\sigma < 0$. The two power functions are identical to α if $\delta = 0$.

It is also intuitive that a test with more assumptions has greater power than one with fewer assumptions. However, this study found that a test with more assumptions has less or equal power. To test equal means under independent normal samples, the usual test involves using the t -distribution with the pooled estimator of the common variance. Under the assumption of equal sample size, the distribution of the differences is an independent normal distribution. We derive a test and the corresponding confidence interval in this case. It is noteworthy that the power function with equal sample size and distribution of the differences following an independent normal distribution is less than or equal to that of the usual t -test using the pooled estimator of the common variance. In the multivariate case, there is a similar situation. This finding can be used as a notable example in master's mathematical statistics courses.

2. Univariate test

Let independent random samples be obtained from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$. We denote these as X_1, \dots, X_n and Y_1, \dots, Y_m . It is well known (Chapter 8 of Hogg *et al.*, 2019) that

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{1/n + 1/m}} \sim t(n + m - 2), \quad (2.1)$$

where S_p^2 and the related statistics are defined in (1.1). Based on (2.1), we obtained a $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ as $\bar{X} - \bar{Y} \pm t_{\alpha/2}(n + m - 2)S_p \sqrt{1/n + 1/m}$. Furthermore, we reject $H_0 : \mu_1 - \mu_2 \leq \delta_0$ and accept the one-sided alternative $H_1 : \mu_1 - \mu_2 > \delta_0$ if

$$t_0 = \frac{\bar{X} - \bar{Y} - \delta_0}{S_p \sqrt{1/n + 1/m}} > t_{\alpha}(n + m - 2). \quad (2.2)$$

WLOG, we assume that $\delta_0 = 0$.

Here, we derive a test and the corresponding confidence interval based on the differences, if we further assume that $n = m$. Let $D_i = X_i - Y_i$, $i = 1, \dots, n$, $D_i \sim N(\mu_1 - \mu_2, 2\sigma^2)$ independently. Hence,

- (1) $\bar{D} = \bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, 2\sigma^2/n)$,
- (2) $((n - 1)S_D^2)/2\sigma^2 \sim \chi^2(n - 1)$, where $S_D^2 = (\sum_{i=1}^n (D_i - \bar{D})^2)/n - 1$,
- (3) \bar{D} and S_D^2 are independent.

Then, by using the definition of the t -distribution,

$$t = \frac{\bar{D} - (\mu_1 - \mu_2)}{S_D / \sqrt{n}} \sim t(n - 1).$$

Therefore, we obtained a $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ as $\bar{D} \pm t_{\alpha/2}(n - 1)S_D / \sqrt{n}$. Similarly, we reject $H_0 : \mu_1 - \mu_2 \leq \delta_0$ and accept the one-sided alternative $H_1 : \mu_1 - \mu_2 > \delta_0$ if

$$t_0 = \frac{\bar{D} - \delta_0}{S_D / \sqrt{n}} > t_{\alpha}(n - 1). \quad (2.3)$$

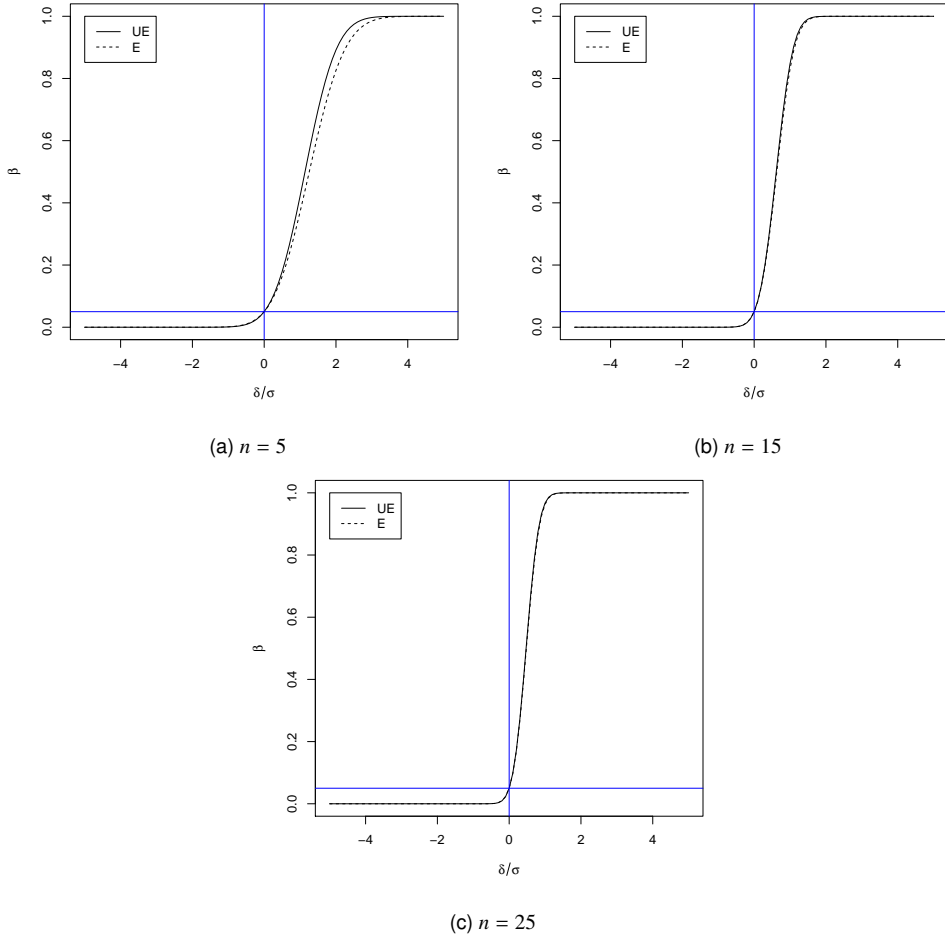


Figure 1: Power functions (2.4) denoted as UE and (2.5) denoted as E with $n = m$.

As before, WLOG, we can assume that $\delta_0 = 0$. The developed test and corresponding confidence interval are similar to the paired t -test. However, in the paired t -test, X_i and Y_i are not independent. Furthermore, we might think that the statistic \bar{D} is valid if one of the variables is re-numbered. However, this is not true because $\bar{D} = \bar{X} - \bar{Y}$ and we use only the sample means of X_i and Y_i , $i = 1, \dots, n$.

In this situation, we do not know which test is better in terms of power function. The power functions for (2.2) and (2.3) are:

$$\beta_{\phi_1}\left(\frac{\delta}{\sigma}\right) = 1 - T_{n+m-2}\left(t_\alpha(n+m-2), \sqrt{\frac{nm}{n+m}} \frac{\delta}{\sigma}\right) \quad \text{and} \quad (2.4)$$

$$\beta_{\phi_2}\left(\frac{\delta}{\sigma}\right) = 1 - T_{n-1}\left(t_\alpha(n-1), \sqrt{\frac{n}{2}} \frac{\delta}{\sigma}\right) \quad \text{respectively,} \quad (2.5)$$

where $\delta = \mu_1 - \mu_2$, $T_\nu(\cdot, ncp)$ denotes the cumulative distribution function (CDF) of the noncentral t -distribution with degrees of freedom ν , and the noncentrality parameter ncp . The derivation of (2.4) is

given in Appendix for completeness. We derive (2.5). Remark that $\sqrt{n}\bar{D}/\sqrt{2\sigma^2} \sim N(\sqrt{n/2}\delta/\sigma, 1)$. By the definition of the noncentral t -distribution,

$$\frac{\sqrt{n}\bar{D}/\sqrt{2\sigma^2}}{\sqrt{((n-1)S_D^2)/2\sigma^2}/(n-1)} = \frac{\sqrt{n}\bar{D}}{S_D} \sim t_{n-1}\left(ncp = \sqrt{\frac{n}{2}}\frac{\delta}{\sigma}\right).$$

We note the following facts:

- (i) $t_\alpha(v_1) < t_\alpha(v_2)$ if $v_1 > v_2$ and $\alpha < 0.5$.
- (ii) For $v_1 > v_2$, $T_{v_1}(t_\alpha(v_1), ncp) < T_{v_2}(t_\alpha(v_2), ncp)$ if $ncp > 0$, and the inequality is reversed if $ncp < 0$ since the right tail of the noncentral t -distribution will be heavier than the left when $ncp > 0$.
- (iii) $\beta_{\phi_1}(\delta/\sigma) \geq \beta_{\phi_2}(\delta/\sigma)$ if $ncp > 0$, and the inequality is reversed if $ncp < 0$. The two power functions are the same as α when $ncp = 0$.

Three examples are shown in Figure 1 depending on the sample size. From Figure 1, we see that fact (iii) holds. Thus, we find that the power function is smaller or equal, even though we use more assumptions. Furthermore, note that both power functions strictly increase δ/σ . However, the differences between the values of the two power functions became negligible as the sample size increased. We obtained similar power functions for the left-tail test in Figure 3 given in Appendix, which are the mirror images in Figure 1. The power function of the two-tailed test is bowl-shaped, similar to the case of the known σ .

3. Multivariate test

In this section, we extend the results of the previous section to include a multivariate case. If $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ is a random sample of size n from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$ is an independent random sample of size m from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, then

$$T^2 = [\bar{\mathbf{X}} - \bar{\mathbf{Y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^\top \left[\left(\frac{1}{n} + \frac{1}{m} \right) \mathbf{S}_p \right]^{-1} [\bar{\mathbf{X}} - \bar{\mathbf{Y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \quad (3.1)$$

is Hotelling's T^2 -distribution (Chapter 5 of Johnson and Wichern, 2007), and is distributed as

$$\frac{(n+m-2)p}{n+m-p-1} F_{p, n+m-p-1},$$

where $\mathbf{S}_p = ((n-1)\mathbf{S}_X + (m-1)\mathbf{S}_Y)/(n+m-2)$, $\mathbf{S}_X = 1/(n-1) \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$ and $\mathbf{S}_Y = 1/(m-1) \sum_{i=1}^m (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^\top$. This can be proven by the sampling distribution of the multivariate normal distribution, additivity of the Wishart distribution, and Theorem 1 in Appendix. Further details are provided in Appendix.

Hence, we reject $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ if

$$T^2 > \frac{(n+m-2)p}{n+m-p-1} F_{p, n+m-p-1}(\alpha),$$

where T^2 is given in (3.1) under the null hypothesis and $F_{\nu_1, \nu_2}(\alpha)$ is the upper α quantile of the F -distribution with dfs ν_1 and ν_2 . Consequently, the confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is obtained as

$$T^2 \leq \frac{(n+m-2)p}{n+m-p-1} F_{p, n+m-p-1}(\alpha),$$

where T^2 is expressed by (3.1). The power function for the two-tailed test is given by:

$$\beta_{\phi_3}(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}) = 1 - F_{p, n+m-p-1}(F_{p, n+m-p-1}(\alpha), \tau^2), \quad (3.2)$$

where $F_{\nu_1, \nu_2}(\cdot, \tau^2)$ is the CDF of the noncentral F -distribution with degrees of freedom, ν_1 and ν_2 , and the noncentrality parameter, τ^2 , $\tau^2 = ((1/n) + (1/m))^{-1} \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}$. Here, $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. This can be obtained using an approach similar to that used to derive the distribution in (3.1).

When $n = m$, let $\mathbf{D}_i = \mathbf{X}_i - \mathbf{Y}_i$, $i = 1, \dots, n$ then, $\mathbf{D}_i \sim N_p(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, 2\boldsymbol{\Sigma})$. Let $\mathbf{S}_D = 1/(n-1) \sum_{i=1}^n (\mathbf{D}_i - \bar{\mathbf{D}})(\mathbf{D}_i - \bar{\mathbf{D}})^\top$. We note that

- (1) $n^{1/2} \bar{\mathbf{D}} = n^{1/2}(\bar{\mathbf{X}} - \bar{\mathbf{Y}}) \sim N_p(n^{1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), 2\boldsymbol{\Sigma})$,
- (2) $(n-1)\mathbf{S}_D \sim W_p(2\boldsymbol{\Sigma}, n-1)$,
- (3) $\bar{\mathbf{D}}$ and \mathbf{S}_D are independent.

Hence, by Theorem 1 in Appendix of Hotelling's T^2 -distribution, we reject $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ if

$$T_D^2 = (\bar{\mathbf{D}} - \boldsymbol{\delta})^\top \left[\frac{1}{n} \mathbf{S}_D \right]^{-1} (\bar{\mathbf{D}} - \boldsymbol{\delta}) > \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha) \quad (3.3)$$

under H_0 . Consequently, the confidence region for $\boldsymbol{\delta}$ is given by:

$$T_D^2 \leq \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha),$$

where T_D^2 is given by (3.3). The developed test and corresponding confidence region are similar to those of paired Hotelling's T^2 test. However, note that in paired Hotelling's T^2 -test, \mathbf{X}_i and \mathbf{Y}_i are not independent. Therefore, the power function is given by:

$$\beta_{\phi_4}(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}) = 1 - F_{p, n-p}(F_{p, n-p}(\alpha), \tau_*^2), \quad (3.4)$$

where $\tau_*^2 = (n/2) \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}$. When $n = m$, $\tau_*^2 = \tau^2$; that is, we have the same noncentrality parameter.

We note the following facts:

- (i) $F_\alpha(\nu, \nu_1) < F_\alpha(\nu, \nu_2)$ if $\nu_1 > \nu_2$.
- (ii) For $\nu_1 > \nu_2$, $F_{\nu, \nu_1}(F_\alpha(\nu, \nu_1), ncp) < F_{\nu, \nu_2}(F_\alpha(\nu, \nu_2), ncp)$ if $ncp > 0$.
- (iii) $\beta_{\phi_3}(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}) \geq \beta_{\phi_4}(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta})$ if $ncp > 0$. The two power functions are identical to α when $ncp = 0$.

Similar to the univariate case, we present examples with various values of n and p in Figure 2. We find that both power functions increase strictly in $\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}$. Furthermore, when we increased n and decreased p , the differences between the values of the two power functions were negligible. Note that, similar to the univariate case, we have lower or equal power, even though more assumptions are used.

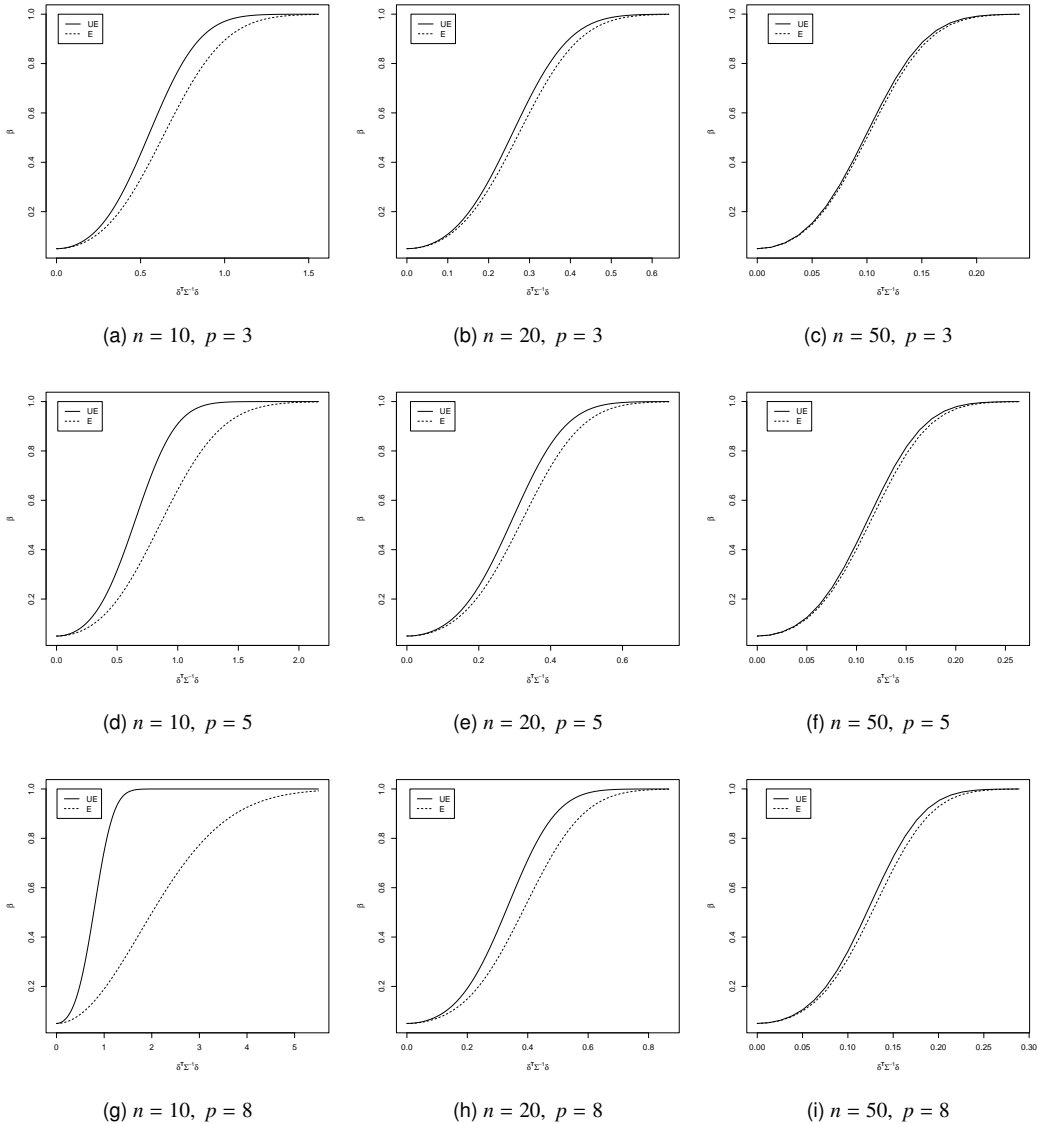


Figure 2: Power functions (3.2) denoted as UE and (3.4) denoted as E with $n = m$.

4. Discussion and conclusions

In this study, two tests using additional assumptions for equal means were conducted. Power functions were obtained from a random sample of independent normal distributions. The power of using more assumptions is smaller than or equal to that of fewer assumptions, which contradicts intuition; a test with more assumptions has greater power than that with fewer assumptions. We have a similar situation for the multivariate normal distribution. This note can be used as an illustrative example in master's mathematical statistics courses.

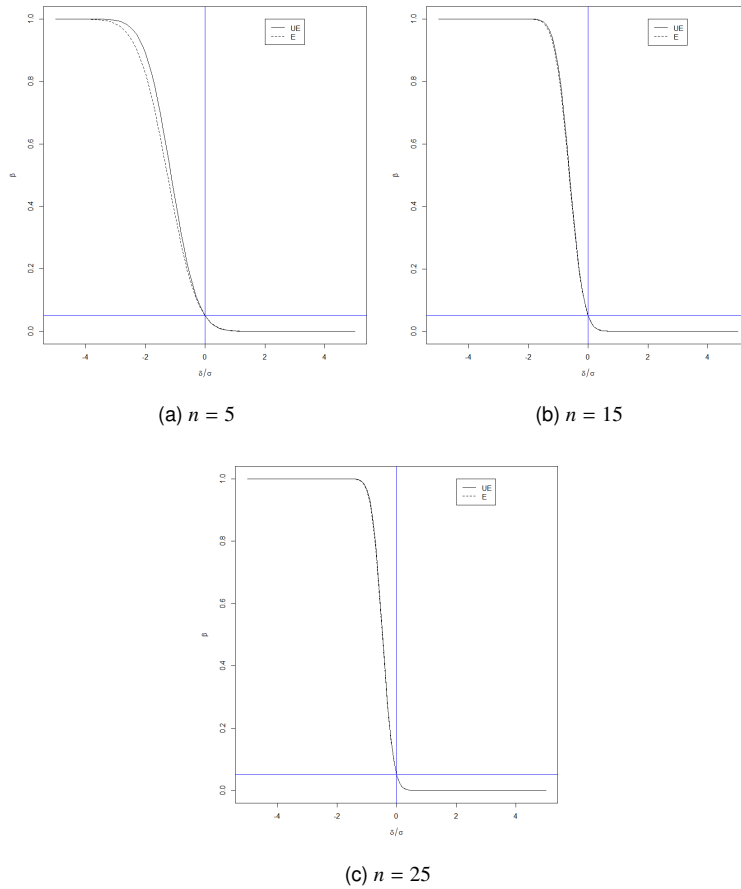


Figure 3: Power functions for the left-tail test denoted as UE with $n \neq m$ and denoted as E with $n = m$.

Acknowledgments

The corresponding author's research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1B07045603), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1A4A5032622).

Appendix

First, we derive the power function (2.4). Note that

- (1) $(\bar{X} - \bar{Y})/\sigma \sqrt{1/n + 1/m} \sim N((\mu_1 - \mu_2)/\sigma \sqrt{1/n + 1/m}, 1)$,
- (2) $((n + m - 2)S_p^2)/2\sigma^2 \sim \chi^2(n + m - 2)$, where S_p^2 is the pooled estimator of common variance,
- (3) $\bar{X} - \bar{Y}$ and S_p^2 are independent.

By the definition of noncentral t -distribution,

$$\frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n + 1/m}} \sim t_{n+m-2} \left(ncp = \frac{1}{\sqrt{1/n + 1/m}} \frac{\delta}{\sigma} \right).$$

Second, we obtain similar power functions for the left-tail test, which are the mirror images in Figure 1.

Some well-known facts are also summarized in the Appendix. First, the sampling distribution of the multivariate normal distribution was provided. Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be random samples of size n from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then,

- (1) $\bar{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$.
- (2) $(n-1)\mathbf{S} \sim W_p(\boldsymbol{\Sigma}, n-1)$, where $\mathbf{S} = 1/(n-1) \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^\top$.
- (3) $\bar{\mathbf{X}}$ and \mathbf{S} are independent.

For the properties of the Wishart distribution, we have the following additivity. That is, If $\mathbf{A}_1 \sim W_p(\boldsymbol{\Sigma}, m_1)$ is independent of \mathbf{A}_2 , which follows $W_p(\boldsymbol{\Sigma}, m_2)$, $\mathbf{A}_1 + \mathbf{A}_2 \sim W_p(\boldsymbol{\Sigma}, m_1 + m_2)$. Finally, the theorem for Hotelling's T^2 -distribution is given as follows:

Theorem 1. *If \mathbf{X} and \mathbf{M} are independently distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $W_p(\boldsymbol{\Sigma}, n)$, respectively, then: $\alpha^* = n\mathbf{X}^\top \mathbf{M}^{-1} \mathbf{X} \sim \frac{np}{n-p+1} F_{p, n-p+1}(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})$, where $F_{p,q}(\lambda)$ denotes a noncentral F -distribution with p and q degrees of freedom and a noncentrality parameter λ .*

References

- Daniel WW (1990). *Applied Nonparametric Statistics (2nd ed)*, PWS-KENT Publishing Company, Boston.
- Hogg RV, McKean JW, and Craig AT (2019). *Introduction to Mathematical Statistics (8th ed)*, Pearson, London.
- Johnson RA and Wichern DW (2007). *Applied Multivariate Statistical Analysis (6th ed)*, Pearson, London.
- Lehmann EL (1999). *Elements of Large-Sample Theory*, Springer-Verlag, New York.

Received July 15, 2022; Revised September 22, 2022; Accepted October 17, 2022