5.2 The Plausibility of μ_0 as a Value for a Normal Population Mean

- Remember, when we run a test for a specific predictor variable to a response variable, we often refer this to a hypothesis testing, $H_0: \mu = \mu_0$ and $H_1: \mu \neq \mu_0$. When we draw the conclusion by assessing the t-statistic,

$$t = \frac{(\bar{X} - \mu_0)}{S/\sqrt{n}}, \text{ where } \bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \text{ and } S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2,$$ which follows t-distribution with n-1 degree of freedom. We reject the H_0 if the observed $|t|$ exceeds a specified p-value, and having $|t|$ value also means t^2, which is the square of the distance from the sample mean $\bar{X}$ to the test value μ_0.

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{S^2/n} = n(\bar{X} - \mu_0)(S^2)^{-1}(\bar{X} - \mu_0)$$

$$\bar{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}$$

=> This is a random interval because the endpoints depend upon the random variables $\bar{X}$ and S.

- When the test is multivariate, which X is a $n \times p$ matrix and $\bar{X}$ is a $p \times 1$ matrix,

$$T^2 = (\bar{X} - \mu_0)'\left(\frac{1}{n}S\right)^{-1}(\bar{X} - \mu_0) = n(\bar{X} - \mu_0)'S^{-1}(\bar{X} - \mu_0), \text{ where}$$

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \quad S = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})',$$ and this T^2 is called Hotelling's T^2.

- T^2 is distributed as $\frac{(n-1)p}{(n-p)} F_{p,n-p}$, where $F_{p,n-p}$ denotes a random variable with an F-distribution p and $n-p$ degrees of freedom.

$$\Rightarrow \quad \alpha = P\left[T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] = P\left[n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right]$$

- Remember that the Wishart distribution generalizes the χ^2 distribution. As for T^2, we can write,

$$T^2_{p,n-p} = \sqrt{n}(\bar{X} - \mu_0)'\left(\frac{\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'}{n-1}\right)^{-1}\sqrt{n}(\bar{X} - \mu_0)$$

$$= \begin{pmatrix}\text{multivariate normal} \\ \text{random vector}\end{pmatrix}'\begin{pmatrix}\text{Wishart random matrix} \\ \text{degrees of freedom}\end{pmatrix}^{-1}\begin{pmatrix}\text{multivariate normal} \\ \text{random vector}\end{pmatrix}$$

$$= N_p(0, \Sigma)'\left[\frac{1}{n-1} W_{p,n-1}(\Sigma)\right]^{-1} N_p(0, \Sigma),$$ and this is analogous to

$t^2_{n-1} = n(\bar{X}-\mu_0)(S^2)^{-1}(\bar{X}-\mu_0)$ for the univariate case.

1. Compute $\bar{X}$, S^{-1}
2. Compute T^2
3. Calculate $\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$, and Compare it with T^2
4. H_0 will be rejected if one or more of the component means, or some combination of means, differs too much from the hypothesized values.

* One feature of the T^2-statistic is that it is invariant under changes in the units of measurements for X of the form,

$Y = CX + d$, C: nonsingular. Premultiplication of the centered and scaled quantities by any nonsingular matrix will be invariant.

proof

Given observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and the transformation in (5-9), it immediately follows from Result 3.6 that

$$\bar{\mathbf{y}} = \mathbf{C}\bar{\mathbf{x}} + \mathbf{d} \quad \text{and} \quad \mathbf{S}_y = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})' = \mathbf{CSC}'$$

Moreover, by (2-24) and (2-45),

$$\boldsymbol{\mu}_Y = E(\mathbf{Y}) = E(\mathbf{CX}+\mathbf{d}) = E(\mathbf{CX}) + E(\mathbf{d}) = \mathbf{C}\boldsymbol{\mu} + \mathbf{d}$$

Therefore, T^2 computed with the $\mathbf{y}$'s and a hypothesized value $\boldsymbol{\mu}_{Y,0} = \mathbf{C}\boldsymbol{\mu}_0 + \mathbf{d}$ is

$$\begin{aligned} T^2 &= n(\bar{\mathbf{y}} - \boldsymbol{\mu}_{Y,0})'\mathbf{S}_y^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}_{Y,0}) \\ &= n(\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0))'(\mathbf{CSC}')^{-1}(\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)) \\ &= n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{C}'(\mathbf{CSC}')^{-1}\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \\ &= n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{C}'(\mathbf{C}')^{-1}\mathbf{S}^{-1}\mathbf{C}^{-1}\mathbf{C}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \end{aligned}$$

The last expression is recognized as the value of T^2 computed with the $\mathbf{x}$'s.

5.3 Hotelling's T^2 and Likelihood Ratio Tests

- Likelihood ratio tests have several optimal properties for reasonably large samples, and they are particularly convenient for hypotheses formulated in terms of multivariate normal parameters.

- Likelihood Ratio $= \Lambda = \dfrac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu,\Sigma} L(\mu, \Sigma)} = \left(\dfrac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{\frac{n}{2}}$ or

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2} = \left(\frac{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right|}{\left|\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)'\right|}\right)^{n/2} < c_\alpha$$

where c_α is the lower (100α)th percentile of the distribution of Λ.

$$\Rightarrow \Lambda^{\frac{2}{n}} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$$, and this is called Wilks' lambda.

- H_0 is rejected if the likelihood ratio is too small.

* note that the likelihood ratio test statistic is a power of the ratio of generalized variances.

Theorem : * Result 5.1 pdf 239

- Let $X_1, X_2, \ldots, X_n$ be a random sample from an $N_p(\mu, \Sigma)$ population. Then the test based on T^2 is equivalent to the likelihood ratio test of $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ because,

$$\Lambda^{\frac{2}{n}} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$$

$$\approx> T^2 = \frac{(n-1)|\hat{\Sigma}_0|}{|\hat{\Sigma}|} - (n-1)$$

$$= \frac{(n-1)\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right|}{\left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|} - (n-1)$$

General Likelihood Ratio Method

와 여긴 진짜 못알아 듣겠는데...?

5.4 Confidence Regions and Simultaneous Comparisons of Component Means

- Confidence Regions = multivariate version of confidence interval
  $R(X)$

$\Rightarrow$ $P[R(X) \text{ covers the true } \theta] = 1-\alpha$

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

- $100(1-\alpha)\%$ confidence region for the mean of a p-dimensional normal distribution is the ellipsoid determined by all μ such that,

$$n(\bar{x}-\mu)S^{-1}(\bar{x}-\mu) \le \frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha) = C^2$$

- To determine whether any given μ_0 lies within the confidence region, we need to compute the generalized squared distance $n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)$ and compare it with $[p(n-1)/(n-p)]F_{p,n-p}(\alpha)$

$\Rightarrow$ $\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha) = C^2$

$\Rightarrow$ $\sqrt{\lambda_i}\, C/\sqrt{n} = \sqrt{\lambda_i}\sqrt{p(n-1)F_{p,n-p}(\alpha)/n(n-p)}$ half-lengths of the major & minor axes

$\bar{x} \pm \sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\cdot e_i$ the axes of confidence ellipsoids

※ the ratio of the λ_i's will help identify relative amounts of elongation along pairs of axes

Simultaneous Confidence Statements :

- For a fixed and σ_z^2 unknown, a $100(1-\alpha)\%$ confidence interval for $\mu_z = a'\mu$ is based on student's t-ratio,
- While the confidence region $n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \le C^2$ assesses the joint knowledge concerning plausible values of μ, any summary of conclusions ordinarily includes confidence statements about the individual component means.

  ✱ we adopt the attitude that all of the separate confidence statements should hold simultaneously with a specified high probability

- Let X have an $N_p(\mu, \Sigma)$ distribution and form the linear combination,

$$Z = a_1X_1 + a_2X_2 + \cdots + a_pX_p = a'X$$

$$\hookrightarrow \mu_Z = E(Z) = a'\mu, \quad \sigma_Z^2 = Var(Z) = a'\Sigma a$$

- Simultaneous confidence intervals can be developed from a consideration of confidence intervals for $a'\mu$ for various choices of a.

$$t = \frac{\bar{z}-\mu_z}{S_z/\sqrt{n}} = \frac{\sqrt{n}(a'\bar{x}-a'\mu)}{\sqrt{a'Sa}}$$

$$\Rightarrow \bar{z} - t_{n-1}(\alpha/2)\frac{S_z}{\sqrt{n}} \le \mu_z \le \bar{z} + t_{n-1}(\alpha/2)\frac{S_z}{\sqrt{n}}$$

$$\approx a'\bar{x} - t_{n-1}(\alpha/2)\frac{\sqrt{a'Sa}}{\sqrt{n}} \le a'\mu \le a'\bar{x} + t_{n-1}(\alpha/2)\frac{\sqrt{a'Sa}}{\sqrt{n}}$$

Inequality (5-21) can be interpreted as a statement about the components of the mean vector $\boldsymbol{\mu}$. For example, with $\mathbf{a}' = [1, 0, \ldots, 0]$, $\mathbf{a}'\boldsymbol{\mu} = \mu_1$, and (5-21) becomes the usual confidence interval for a normal population mean. (Note, in this case, that $\mathbf{a}'\mathbf{S}\mathbf{a} = s_{11}$.) Clearly, we could make several confidence statements about the components of $\boldsymbol{\mu}$, each with associated confidence coefficient $1 - \alpha$, by choosing different coefficient vectors $\mathbf{a}$. However, the confidence associated with all of the statements taken together is *not* $1 - \alpha$.

$$\left[a'\bar{x} - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\cdot a'Sa}\ ,\ a'\bar{x} + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\cdot a'Sa} \right]$$

T^2 intervals

It is convenient to refer to the simultaneous intervals of Result 5.3 as *T^2-intervals*, since the coverage probability is determined by the distribution of T^2. The successive choices $\mathbf{a}' = [1, 0, \ldots, 0]$, $\mathbf{a}' = [0, 1, \ldots, 0]$, and so on through $\mathbf{a}' = [0, 0, \ldots, 1]$ for the T^2-intervals allow us to conclude that

$$\begin{aligned}
\bar{x}_1 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} &\le \mu_1 \le \bar{x}_1 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} &\le \mu_2 \le \bar{x}_2 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \\
&\ \ \vdots \\
\bar{x}_p - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}} &\le \mu_p \le \bar{x}_p + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}}
\end{aligned} \tag{5-24}$$

- Note that, without modifying the coefficient $1-\alpha$, we can make statements about the differences $\mu_i - \mu_k$ corresponding to $a' = [0, 0, \ldots a_i, 0, \ldots, a_k, 0, \ldots]$, where $a_i = 1$ and $a_k = -1$.

$a_k = -1$. In this case $\mathbf{a}'\mathbf{S}\mathbf{a} = s_{ii} - 2s_{ik} + s_{kk}$, and we have the statement

$$\bar{x}_i - \bar{x}_k - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}} \le \mu_i - \mu_k$$
$$\le \bar{x}_i - \bar{x}_k + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}} \qquad (5\text{-}25)$$

The simultaneous T^2 confidence intervals are ideal for "data snooping." The confidence coefficient $1 - \alpha$ remains unchanged for any choice of $\mathbf{a}$, so linear combinations of the components μ_i that merit inspection *based upon an examination of the data* can be estimated.

In addition, according to the results in Supplement 5A, we can include the statements about (μ_i, μ_k) belonging to the sample mean-centered ellipses

$$n[\bar{x}_i - \mu_i, \quad \bar{x}_k - \mu_k]\begin{bmatrix} s_{ii} & s_{ik} \\ s_{ik} & s_{kk} \end{bmatrix}^{-1}\begin{bmatrix} \bar{x}_i - \mu_i \\ \bar{x}_k - \mu_k \end{bmatrix} \le \frac{p(n-1)}{n-p}F_{p,n-p}(\alpha) \qquad (5\text{-}26)$$

and still maintain the confidence coefficient $(1 - \alpha)$ for the whole set of statements.

The simultaneous T^2 confidence intervals for the individual components of a mean vector are just the shadows, or projections, of the confidence ellipsoid on the component axes. This connection between the shadows of the ellipsoid and the simultaneous confidence intervals given by (5-24) is illustrated in the next example.

- The simultaneous T^2-intervals above are wider than univariate intervals because all three must hold with 95% confidence. They may also be wider than necessary, because we can make statements about differences with the same confidence.

A Comparison of Simultaneous Confidence Intervals with One-at-a-Time Intervals

- The probability of all intervals containing their respective μ_i's is not $1-\alpha$

EX:

- To guarantee a probability of $1-\alpha$ that all of the statements about the component means hold simultaneously, the individual intervals must be wider than the separate t-intervals. ✱ just how much wider depends on both p and n, as well as on $1-\alpha$

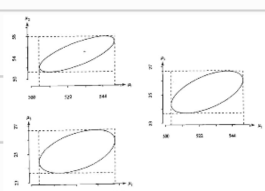

Figure 5.3 95% confidence ellipses for pairs of means and the simultaneous T^2-intervals—college test data.

✱

The Bonferroni Method of Multiple Comparisons

✱ Alternative method for multiple comparisons

- If the number of specified component means μ_i or linear combinations is small, simultaneous confidence intervals can be developed that are shorter than the simultaneous T^2-intervals

Suppose that, prior to the collection of data, confidence statements about m linear combinations $\mathbf{a}_1'\boldsymbol{\mu}, \mathbf{a}_2'\boldsymbol{\mu}, \ldots, \mathbf{a}_m'\boldsymbol{\mu}$ are required. Let C_i denote a confidence statement about the value of $\mathbf{a}_i'\boldsymbol{\mu}$ with $P[C_i \text{ true}] = 1 - \alpha_i$, $i = 1, 2, \ldots, m$. Now (see Exercise 5.6),

$$P[\text{all } C_i \text{ true}] = 1 - P[\text{at least one } C_i \text{ false}]$$
$$\geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m}(1 - P(C_i \text{ true}))$$
$$= 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m) \quad (5\text{-}28)$$

- The above special case of Bonferroni Inequality allows an investigator to control the overall error rate $\alpha_1 + \alpha_2 + \cdots \alpha_m$ regardless of the correlation structure behind the confidence statements.

$$P[\text{all } C_i \text{ true}] = 1 - P[\text{at least one } C_i \text{ false}]$$
$$\geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m}(1 - P(C_i \text{ true}))$$
$$= 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m) \quad (5\text{-}28)$$

Inequality (5-28), a special case of the Bonferroni inequality, allows an investigator to control the overall error rate $\alpha_1 + \alpha_2 + \cdots + \alpha_m$, regardless of the correlation structure behind the confidence statements. There is also the flexibility of controlling the error rate for a group of important statements and balancing it by another choice for the less important statements.

Let us develop simultaneous interval estimates for the restricted set consisting of the components μ_i of $\boldsymbol{\mu}$. Lacking information on the relative importance of these components, we consider the individual t-intervals

applied

$$\bar{x}_i \pm t_{n-1}\left(\frac{\alpha_i}{2}\right)\sqrt{\frac{s_{ii}}{n}} \qquad i = 1, 2, \ldots, m$$

with $\alpha_i = \alpha/m$. Since $P[\bar{X}_i \pm t_{n-1}(\alpha/2m)\sqrt{s_{ii}/n}$ contains $\mu_i] = 1 - \alpha/m$, $i = 1, 2, \ldots, m$, we have, from (5-28),

$$P\left[\bar{X}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \text{ contains } \mu_i, \text{ all } i\right] \geq 1 - \underbrace{\left(\frac{\alpha}{m} + \frac{\alpha}{m} + \cdots + \frac{\alpha}{m}\right)}_{m \text{ terms}}$$
$$= 1 - \alpha$$

Therefore, with an overall confidence level greater than or equal to $1 - \alpha$, we can make the following $m = p$ statements:

$$\bar{x}_1 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \bar{x}_1 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}}$$
$$\bar{x}_2 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} \leq \mu_2 \leq \bar{x}_2 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} \quad (5\text{-}29)$$
$$\vdots$$
$$\bar{x}_p - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \bar{x}_p + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}}$$

The Bonferroni intervals for linear combinations $\mathbf{a}'\boldsymbol{\mu}$ and the analogous T^2-intervals (recall Result 5.3) have the same general form:

$$\mathbf{a}'\bar{\mathbf{X}} \pm (\text{critical value})\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

Consequently, in every instance where $\alpha_i = \alpha/m$,

$$\frac{\text{Length of Bonferroni interval}}{\text{Length of } T^2\text{-interval}} = \frac{t_{n-1}(\alpha/2m)}{\sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}} \quad (5\text{-}30)$$

which does not depend on the random quantities $\bar{\mathbf{X}}$ and $\mathbf{S}$. As we have pointed out, for a small number m of specified parametric functions $\mathbf{a}'\boldsymbol{\mu}$, the Bonferroni intervals will always be shorter. How much shorter is indicated in Table 5.4 for selected n and p.

If we are interested only in the component means, the Bonferroni intervals provide more precise estimates than the T^2-intervals. On the other hand, the 95% confidence region for μ gives the plausible values for the pairs (μ_1, μ_2) when the correlation between the measured variables is taken into account.

- for a small number m of specified parametric functions $a'\mu$, the Bonferroni intervals will always be shorter.

$$\frac{\text{Length of Bonferroni Interval}}{\text{Length of } T^2\text{-interval}} = \frac{t_{n-1}\left(\frac{\alpha}{2m}\right)}{\sqrt{\frac{P(n-1)}{n-p}F_{p,n-p}(\alpha)}}$$

5.5 Large Sample Inference about a Population Mean Vector

- When the sample size is large, tests of hypotheses and confidence regions for μ can be constructed without the assumption of a normal population.
- The advantages associated with large samples may be partially offset by a loss in sample information caused by using only the summary statistics $\bar{x}$ and S. On the other hand, since they are sufficient summary for normal populations, the closer the underlying population is to multivariate normal, the more efficiently the sample information will be utilized in making inferences.

※ All large-sample inferences about μ are based on a χ^2-distribution

$$P[n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \le \chi^2_p(\alpha)] = 1-\alpha$$

Result 5.4. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. When $n - p$ is large, the hypothesis $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected in favor of $H_1: \boldsymbol{\mu} \ne \boldsymbol{\mu}_0$, at a level of significance approximately α, if the observed

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) > \chi^2_p(\alpha)$$

Here $\chi^2_p(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p d.f. ■

※ $\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$ and $\chi^2_p(\alpha)$ are approximately equal for n large relative to p.

- **Result 5.5.** Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive definite covariance $\boldsymbol{\Sigma}$. If $n - p$ is large,

$$\mathbf{a}'\bar{\mathbf{X}} \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

will contain $\mathbf{a}'\boldsymbol{\mu}$, for every $\mathbf{a}$, with probability approximately $1-\alpha$. Consequently, we can make the $100(1-\alpha)\%$ simultaneous confidence statements

$$\bar{x}_1 \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{s_{11}}{n}} \quad \text{contains } \mu_1$$

$$\bar{x}_2 \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{s_{22}}{n}} \quad \text{contains } \mu_2$$

$$\vdots \qquad\qquad \vdots$$

$$\bar{x}_p \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{s_{pp}}{n}} \quad \text{contains } \mu_p$$

and, in addition, for all pairs (μ_i, μ_k), $i, k = 1, 2, \ldots, p$, the sample mean-centered ellipses

$$n[\bar{x}_i - \mu_i, \ \bar{x}_k - \mu_k]\begin{bmatrix} s_{ii} & s_{ik} \\ s_{ik} & s_{kk}\end{bmatrix}^{-1}\begin{bmatrix}\bar{x}_i - \mu_i \\ \bar{x}_k - \mu_k\end{bmatrix} \le \chi^2_p(\alpha) \quad \text{contain} \quad (\mu_i, \mu_k)$$

χ는 ellipsoid 와 같이 특정 $\bar{X}_i$ 의 confidence interval 이던, $\bar{X}_i - \bar{X}_j$ 의 interval 이던 모든 조합의 confidence region 에 관한 것 같다.

어찌됐건 다름!

- When the sample size is large, the one-at-a-time confidence intervals for individual means are

$$\bar{x}_i - z\left(\frac{\alpha}{2}\right)\sqrt{\frac{s_{ii}}{n}} \le \mu_i \le \bar{x}_i + z\left(\frac{\alpha}{2}\right)\sqrt{\frac{s_{ii}}{n}} \qquad i = 1, 2, \ldots, p$$

where $z(\alpha/2)$ is the upper $100(\alpha/2)$th percentile of the standard normal distribution. The Bonferroni simultaneous confidence intervals for the $m = p$ statements about the individual means take the same form, but use the modified percentile $z(\alpha/2p)$ to give

$$\bar{x}_i - z\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{ii}}{n}} \le \mu_i \le \bar{x}_i + z\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{ii}}{n}} \qquad i = 1, 2, \ldots, p$$

Although the sample size may be large, some statisticians prefer to retain the F- and t-based percentiles rather than use the chi-square or standard normal-based percentiles. The latter constants are the infinite sample size limits of the former constants. The F and t percentiles produce larger intervals and, hence, are more conservative. Table 5.7 gives the individual, Bonferroni, and F-based, or shadow of the

- One-at-a-time
- Bonferroni
- Shadow of Ellipsoid

개념을 확실히 공부해서 차이점을 정확하게 알아야 할 듯

Table 5.6 gives the individual, Bonferroni, and chi-square-based (or shadow of the confidence ellipsoid) intervals for the musical aptitude data in Example 5.7.

Table 5.6 The Large Sample 95% Individual, Bonferroni, and T^2-Intervals for the Musical Aptitude Data

The one-at-a-time confidence intervals use $z(.025) = 1.96$.
The simultaneous Bonferroni intervals use $z(.025/7) = 2.69$.
The simultaneous T^2, or shadows of the ellipsoid, use $\chi^2_7(.05) = 14.07$.

Variable	One-at-a-time Lower	One-at-a-time Upper	Bonferroni Intervals Lower	Bonferroni Intervals Upper	Shadow of Ellipsoid Lower	Shadow of Ellipsoid Upper
X_1 = melody	26.95	29.25	26.52	29.68	25.90	30.30
X_2 = harmony	25.43	27.77	24.99	28.21	24.36	28.84
X_3 = tempo	34.64	36.16	34.35	36.45	33.94	36.86
X_4 = meter	33.18	35.22	32.79	35.61	32.24	36.16
X_5 = phrasing	22.85	24.35	22.57	24.63	22.16	25.04
X_6 = balance	21.21	22.79	20.92	23.08	20.50	23.50
X_7 = style	21.89	23.51	21.59	23.81	21.16	24.24

Although the sample size may be large, some statisticians prefer to retain the F- and t-based percentiles rather than use the chi-square or standard normal-based percentiles. The latter constants are the infinite sample size limits of the former constants. The F and t percentiles produce larger intervals and, hence, are more conservative. Table 5.7 gives the individual, Bonferroni, and F-based, or shadow of the confidence ellipsoid, intervals for the musical aptitude data. Comparing Table 5.7 with Table 5.6, we see that all of the intervals in Table 5.7 are larger. However, with the relatively large sample size $n = 96$, the differences are typically in the third, or tenths, digit.

Table 5.7 The 95% Individual, Bonferroni, and T^2-Intervals for the Musical Aptitude Data

The one-at-a-time confidence intervals use $t_{95}(.025) = 1.99$.
The simultaneous Bonferroni intervals use $t_{95}(.025/7) = 2.75$.
The simultaneous T^2, or shadows of the ellipsoid, use $F_{7,89}(.05) = 2.11$.

Variable	One-at-a-time Lower	One-at-a-time Upper	Bonferroni Intervals Lower	Bonferroni Intervals Upper	Shadow of Ellipsoid Lower	Shadow of Ellipsoid Upper
X_1 = melody	26.93	29.27	26.48	29.72	25.76	30.44
X_2 = harmony	25.41	27.79	24.96	28.24	24.23	28.97
X_3 = tempo	34.63	36.17	34.33	36.47	33.85	36.95
X_4 = meter	33.16	35.24	32.76	35.64	32.12	36.28
X_5 = phrasing	22.84	24.36	22.54	24.66	22.07	25.13
X_6 = balance	21.20	22.80	20.90	23.10	20.41	23.59
X_7 = style	21.88	23.52	21.57	23.83	21.07	24.33

☆ χ^2-based estimates are narrower than F and t based estimates of confidence intervals

one at a time : 표준정규분포를 이용하여 구간을 구하는 것

Bonferroni : T^2 보다 더 좋은 구간을 구함으로써 더 높은 정확도를 가짐

shadows : simultaneous T^2's