Ch 5. Inferences about a Mean Vector

- Univariate analysis

  - Determine whether a specific value μ_0 is a plausible value for the population mean μ.

  - Test H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$.

  - For a random sample X_1, X_2,… X_n from a normal distribution, the test statistic is

$$t = \frac{\left(\overline{X} - \mu_0\right)}{s/\sqrt{n}} \sim t_{n-1} \quad \text{under } H_0, \text{ where } \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \text{ and } s^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2.$$

  - Reject H_0 if the observed $|\, t \,| > t_{n\text{-}1}(\alpha/2)$.

  - This is equivalent to reject H_0 if

$$t^2 = \frac{\left(\overline{X} - \mu_0\right)^2}{s^2/n} = n\left(\overline{X} - \mu_0\right)\left(s^2\right)^{-1}\left(\overline{X} - \mu_0\right) > t_{n-1}^2(\alpha/2).$$

  - The value t^2 is the square of the distance from the sample mean $\overline{X}$ to the test value μ_0. The units of distance are expressed in terms of $s/\sqrt{n}$, the estimated standard deviations of $\overline{X}$.

Ch 5. Inferences about a Mean Vector

- Univariate analysis (continued)

  - If H_0 is not rejected, conclude that μ_0 is a plausible value for the normal population mean.

  - There is always a *set* of plausible values for a normal population mean.

  - From the correspondence between acceptance regions for tests of H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ and confidence intervals for μ,

    {Do not reject H_0: $\mu = \mu_0$ at level α} or $\left|\dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}\right| \le t_{n-1}(\alpha/2)$

    is equivalent to

    {μ_0 lies in the 100(1 - α)% confidence interval $\bar{x} \pm t_{n-1}(\alpha/2) s/\sqrt{n}$ }

    or

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

  - The confidence interval consists of all those values μ_0 that would not be rejected by the level α test of H_0: $\mu = \mu_0$.

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

- Multivariate analysis

  - Determine whether a given $p\times 1$ vector μ_0 is a plausible value for the mean of a multivariate normal distribution.

  - A natural generalization of the squared distance is its multivariate analog

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

where $\underset{(p\times 1)}{\overline{X}} = \frac{1}{n}\sum_{j=1}^{n} X_j$, $\underset{(p\times p)}{S} = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j-\overline{X}\right)\left(X_j-\overline{X}\right)'$, and $\underset{(p\times 1)}{\mu_0} = \begin{bmatrix}\mu_{10}\\ \mu_{20}\\ \vdots\\ \mu_{p0}\end{bmatrix}$.

  - The statistic T^2 is called **Hotelling's T^2**.

  - Note that $\frac{1}{n}S$ is the estimated covariance of $\overline{X}$.

  - T^2 is distributed as $\frac{(n-1)p}{(n-p)}F_{p,n-p}$.

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

- Let $X_1, X_2, \ldots X_n$ be a random sample from an $N_p(\mu, \Sigma)$ population.

With $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j,\ S = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)',$

$$\alpha = P\left[T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right]$$

$$= P\left[n\left(\bar{X} - \mu_0\right)' S^{-1}\left(\bar{X} - \mu_0\right) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right],$$

whatever the true μ and Σ. Here, $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of the $F_{p,n-p}$ distribution.

- At the α level of significance, reject H_0: $\mu = \mu_0$ in favor of H_1: $\mu \neq \mu_0$ if the observed

$$T^2 = n\left(\bar{x} - \mu_0\right)' S^{-1}\left(\bar{x} - \mu_0\right) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha).$$

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

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

  - T^2 combines a normal, $N_p(0, \Sigma)$, random vector and a Wishart, $W_{p,n-1}(\Sigma)$, random matrix in the form

$$T^2_{p,n-1} = \begin{pmatrix} \text{multivariate normal} \\ \text{random vector} \end{pmatrix}' \left(\frac{\text{Wishart random matrix}}{d.f.} \right)^{-1} \begin{pmatrix} \text{multivariate normal} \\ \text{random vector} \end{pmatrix}$$

$$= N_p(0, \Sigma)' \left[\frac{1}{n-1} W_{p,n-1}(\Sigma) \right]^{-1} N_p(0, \Sigma).$$

  - This is analogous to

$$t^2 = \sqrt{n}\left(\bar{X} - \mu_0\right)\left(s^2\right)^{-1} \sqrt{n}\left(\bar{X} - \mu_0\right)$$

or

$$t^2_{n-1} = \begin{pmatrix} \text{normal} \\ \text{random variable} \end{pmatrix} \left(\frac{\text{(scaled) chi-square random variable}}{d.f.} \right)^{-1} \begin{pmatrix} \text{normal} \\ \text{random variable} \end{pmatrix}$$

for the univariate case.

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

- Example 5.1 Evaluating T^2

  - Let the data matrix for a random sample of size n = 3 from a bivariate normal population be $X = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$. Evaluate the observed T^2 for $\mu_0' = [9, 5]$.

  - $\bar{x} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \quad S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$

  - $T^2 = 3[8-9, \quad 6-5]\begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}^{-1}\begin{bmatrix} 8-9 \\ 6-5 \end{bmatrix} = \frac{7}{9}.$

  - The sampling distribution of T^2 is $\frac{(3-1)2}{(3-2)}F_{2,3-2} = 4F_{2,1}.$

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

- T^2-statistic is **invariant** (unchanged) under changes in the units of measurements for X of the form

$$\underset{(p\times 1)}{Y} = \underset{(p\times p)}{C}\ \underset{(p\times 1)}{X} + \underset{(p\times 1)}{d}, \quad C \text{ nonsingular.}$$

- Given observations $x_1, x_2, \ldots x_n$ and the transformation $Y = CX + d$, it follows that

$$\bar{y} = C\bar{x} + d \ \text{ and } \ S_y = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})(y_j - \bar{y})' = CSC'.$$

- $\mu_Y = E(Y) = E(CX + d) = E(CX) + E(d) = C\mu + d$
- T^2 computed with the y's and a hypothesized value $\mu_{Y,0} = C\mu_0 + d$ is

$$T^2 = n(\bar{y} - \mu_{Y,0})' S_y^{-1}(\bar{y} - \mu_{Y,0}) = n(C(\bar{x} - \mu_0))'(CSC')^{-1}(C(\bar{x} - \mu_0))$$

$$= n(\bar{x} - \mu_0)' C'(CSC')^{-1} C(\bar{x} - \mu_0) = n(\bar{x} - \mu_0)' C'(C')^{-1} S^{-1} C^{-1} C(\bar{x} - \mu_0)$$

$$= n(\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0).$$

5.3. Hotelling's T^2 and Likelihood Ratio Tests

- The maximum of the multivariate normal likelihood as μ and Σ are varied over their possible values is given by

$$\max_{\mu,\Sigma} L(\mu,\Sigma) = \frac{1}{(2\pi)^{np/2}\left|\hat{\Sigma}\right|^{n/2}} e^{-\frac{1}{2}tr\left[\hat{\Sigma}^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' + n(\bar{x}-\hat{\mu})(\bar{x}-\hat{\mu})'\right)\right]}$$

$$= \frac{1}{(2\pi)^{np/2}\left|\hat{\Sigma}\right|^{n/2}} e^{-\frac{1}{2}tr\left[\hat{\Sigma}^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right)\right]} \quad \text{since } \hat{\mu} = \bar{x}$$

$$= \frac{1}{(2\pi)^{np/2}\left|\hat{\Sigma}\right|^{n/2}} e^{-np/2} \quad \text{since } \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})',$$

where $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'$ and $\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ are the maximum likelihood estimates.

5.3. Hotelling's T^2 and Likelihood Ratio Tests

- Under the hypothesis H_0: $\mu = \mu_0$, the normal likelihood specializes to

$$L(\mu_0,\Sigma)=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu_0)'\Sigma^{-1}(x_j-\mu_0)\right).$$

- The mean μ_0 is now fixed, but Σ can be varied to find the value that is "most likely" to have lead, with μ_0 fixed, to the observed sample. This value is obtained by maximizing $L(\mu_0, \Sigma)$ with respect to Σ.

- The normal likelihood

$$L(\mu_0,\Sigma)=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}e^{-\frac{1}{2}\sum_{j=1}^{n}tr\left[(x_j-\mu_0)'\Sigma^{-1}(x_j-\mu_0)\right]}=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}e^{-\frac{1}{2}\sum_{j=1}^{n}tr\left[\Sigma^{-1}(x_j-\mu_0)(x_j-\mu_0)'\right]}$$

$$=\frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}}e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right)\right]}$$

- It can be shown that with $\hat{\Sigma}_0=\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'$,

$$\max_{\Sigma}L(\mu_0,\Sigma)=\frac{1}{(2\pi)^{np/2}|\hat{\Sigma}_0|^{n/2}}e^{-\frac{np}{2}}.$$

5.3. Hotelling's T^2 and Likelihood Ratio Tests

- To determine whether μ_0 is a plausible value of μ, the maximum of $L(\mu_0, \Sigma)$ is compared with the unrestricted maximum of $L(\mu, \Sigma)$. The resulting ratio is called the **likelihood ratio statistic**:

$$\text{Likelihood ratio} = \Lambda = \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu,\Sigma} L(\mu, \Sigma)} = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}.$$

- The equivalent statistic $\Lambda^{\frac{2}{n}} = \dfrac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$ is called **Wilks' lambda**.

- If the observed value of this likelihood ratio is too small, H_0: $\mu = \mu_0$ is unlikely to be true and is, therefore, rejected.

- The likelihood ratio test of H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$, rejects H_0 if

$$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2} = \left(\frac{\left| \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' \right|}{\left| \sum_{j=1}^{n} (x_j - \mu_0)(x_j - \mu_0)' \right|} \right) < c_\alpha,$$

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

5.3. Hotelling's T^2 and Likelihood Ratio Tests

- Result 5.1. 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: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ because

$$\Lambda^{\frac{2}{n}} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1}.$$

Proof.

Let the $(p+1)\times(p+1)$ matrix

$$A = \begin{bmatrix} \sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})' & \sqrt{n}(\bar{x} - \mu_0) \\ \sqrt{n}(\bar{x} - \mu_0)' & -1 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Since $|A| = |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}| = |A_{11}|\,|A_{22} - A_{21}A_{11}^{-1}A_{12}|$,

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

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

5.3. Hotelling's T^2 and Likelihood Ratio Tests

- Result 5.1. *Proof.* (continued)

Since $\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)' = \sum_{j=1}^{n}(x_j-\bar{x}+\bar{x}-\mu_0)(x_j-\bar{x}+\bar{x}-\mu_0)'$

$$= \sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' + n(\bar{x}-\mu_0)(\bar{x}-\mu_0)',$$

$$(-1)\left|\sum_{j=1}^{n}(x_j-\mu_0)(x_j-\mu_0)'\right| = \left|\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right|(-1)\left(1+\frac{T^2}{(n-1)}\right)$$

or

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

Thus,

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

Here H_0 is rejected for small values of $\Lambda^{2/n}$ or, equivalently, large values of T^2. The critical values of T^2 are determined by $\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$.

5.3. Hotelling's T^2 and Likelihood Ratio Tests

- Since $\Lambda^{\frac{2}{n}} = \dfrac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} = \left(1 + \dfrac{T^2}{(n-1)}\right)^{-1}$, T^2 may be calculated from two determinants, thus avoiding the computation of S^{-1}.

  - Solving for T^2, we have

$$T^2 = \frac{(n-1)|\hat{\Sigma}_0|}{|\hat{\Sigma}|} - (n-1) = \frac{(n-1)\left|\sum_{j=1}^{n}(x_j - \mu_0)(x_j - \mu_0)'\right|}{\left|\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'\right|} - (n-1).$$

Generalized Likelihood Ratio Method

- θ: a vector consisting of all the *unknown* population parameters

  $L(\theta)$: the likelihood function obtained by evaluating the joint density of $X_1, X_2, \ldots X_n$ at their observed values $x_1, x_2, \ldots x_n$.

  The parameter vector θ takes its value in the parameter set Θ.

- Example. p-dimensional multivariate normal

  - $\theta' = [\mu_1, \ldots, \mu_p, \sigma_{11}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{2p}, \ldots, \sigma_{p-1,p}, \sigma_{pp}]$
  - Θ consists of the p-dimensional space, where $-\infty < \mu_1 < \infty, \ldots, -\infty < \mu_p < \infty$ combined with the $[p(p+1)/2]$-dimensional space of variances and covariances such that Σ is positive definite. Therefore, Θ has dimension $\nu = p + p(p+1)/2$.
  - Under the null hypothesis H_0: $\theta = \theta_0$, θ is restricted to lie in a subset Θ_0 of Θ.
  - For the multivariate normal situation with $\mu = \mu_0$ and Σ unspecified, $\Theta_0 = [\mu_1 = \mu_{10}, \mu_2 = \mu_{20}, \ldots, \mu_p = \mu_{p0}; \sigma_{11}, \ldots, \sigma_{1p}, \sigma_{22}, \ldots, \sigma_{2p}, \ldots, \sigma_{p-1,p}, \sigma_{pp}$ with positive definite $\Sigma]$, so Θ_0 has dimension $\nu_0 = 0 + p(p+1)/2 = p(p+1)/2$.

Generalized Likelihood Ratio Method

- A likelihood ratio test of H_0: $\theta \in \Theta_0$ rejects H_0 in favor of H_1: $\theta \notin \Theta_0$ if

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} < c,$$

where c is a suitably chosen constant.

- Reject H_0 if the maximum of the likelihood obtained by allowing θ to vary over the set Θ_0 is much smaller than the maximum of the likelihood obtained by varying θ over all values in Θ.

- When the maximum in the numerator is much smaller than the maximum in the denominator, Θ_0 does not contain plausible values for θ.

- In each application of the likelihood ratio method, must obtain the sampling distribution of the likelihood-ratio test statistic Λ.

- c can be selected to produce a test with a specified significance level α.

- When the sample size is large and certain regularity conditions are satisfied, the sampling distribution of $-2\ln\Lambda$ is well approximated by a chi-square distribution.

Generalized Likelihood Ratio Method

- Result 5.2. When the sample size n is large, under the null hypothesis H_0,

$$-2\ln\Lambda = -2\ln\left(\frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\theta)}\right)$$

is, approximately, a $\chi^2_{\nu - \nu_0}$ random variable. Here the degrees of freedom are $\nu - \nu_0$ = (dimension of Θ) - (dimension of Θ_0).

5.4. Confidence Regions and Simultaneous Comparisons of Component Means

- θ : a vector of unknown population parameters

  Θ : the set of all possible values of θ

  $X = [X_1, X_2, \ldots X_n]'$: data matrix

  \- A confidence region is a region of likely θ values.

  \- 100(1 - α)% confidence region $R(X)$ for θ is selected as

  $$P[R(X) \text{ will cover the true } \theta] = 1 - \alpha.$$

- A 100(1 – α)% **confidence region** for the mean of a p-dimensional normal distribution is the ellipsoid determined by all μ such that

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

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 $x_1, x_2, \ldots x_n$ are the sample observations.

5.4. Confidence Regions and Simultaneous Comparisons of Component Means

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

- If the squared distance is larger than $[p(n - 1)/(n - p)]F_{p,n-p}(\alpha)$, μ_0 is not in the confidence region.

- This is also consistent with a test of H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$.

5.4. Confidence Regions and Simultaneous Comparisons of Component Means

- For $p \geq 4$, cannot graph the joint confidence region for μ, but can calculate the axes of the confidence ellipsoid and their relative lengths.

- These are determined from the eigenvalues λ_i and eigenvectors e_i of S.

- The directions and lengths of the axes of

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

are determined by going

$$\sqrt{\lambda_i}\, c / \sqrt{n} = \sqrt{\lambda_i} \sqrt{p(n-1) F_{p,n-p}(\alpha) / n(n-p)}$$

units along the eigenvectors e_i.

- Beginning at the center $\bar{x}$, the axes of the confidence ellipsoid are

$$\pm \sqrt{\lambda_i} \sqrt{p(n-1) F_{p,n-p}(\alpha) / n(n-p)}\, e_i,$$

where $Se_i = \lambda_i e_i$, $i = 1,2,\ldots, p$.

- The ratios of the λ_i's identify relative amounts of elongation along pairs of axes.

5.4. Confidence Regions and Simultaneous Comparisons of Component Means

- Example 5.3 Constructing a confidence ellipse for μ

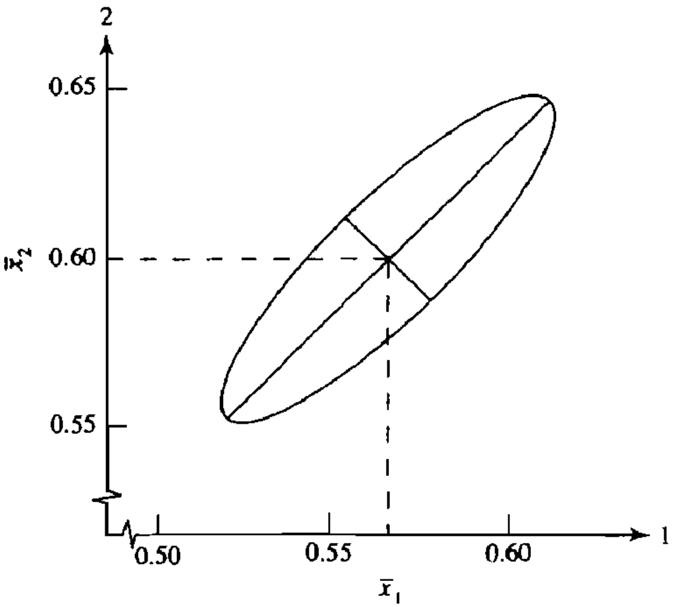

Figure 5.1 A 95% confidence ellipse for μ based on microwave-radiation data.

Simultaneous Confidence Statements

- While the confidence region $n(\bar{x}-\mu)' S^{-1}(\bar{x}-\mu) \le c^2$, for a constant c, correctly assesses the joint knowledge concerning plausible values of μ, any summary of conclusions ordinarily includes confidence statements about the individual component means.

 - All of the separate confidence statements should hold **simultaneously** with a specified high probability.

- Let $X \sim N_p(\mu, \Sigma)$ and consider the linear combination

$$Z = a_1X_1 + a_2X_2 + \ldots + a_pX_p = a'X.$$

 - Since $\mu_z = E(Z) = a'\mu$ and $\sigma^2_z = Var(Z) = a'\Sigma a$, $Z \sim N(a'\mu, a'\Sigma a)$.

 - With a random sample $X_1, X_2, \ldots X_n$ from $N_p(\mu, \Sigma)$ population, a corresponding sample of Z's can be created as

$$Z_j = a_1X_{j1} + a_2X_{j2} + \ldots + a_pX_{jp} = a'X_j\,, \quad \text{for } j = 1,2,\ldots,n.$$

 - The sample mean and variance of the observed values $z_1, z_2, \ldots, z_n$ are

$$\bar{z} = a'\bar{x} \quad \text{and} \quad s_z^2 = a'Sa,$$

where $\bar{x}$ and S are the sample mean vector and covariance matrix of the x_j's, respectively.

Simultaneous Confidence Statements

- Simultaneous confidence intervals can be developed from a consideration of confidence intervals for $a'\mu$ for various choices of a.
- For a *fixed* and unknown σ^2_z, a 100(1 - α)% confidence interval

  for $\mu_z = a'\mu$ is based on student's t-ratio

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

and leads to the statement

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

or

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

where $t_{n-1}(\alpha/2)$ is the upper 100(α/2)th percentile of a t-distribution with n - 1 d.f.

Simultaneous Confidence Statements

- The confidence interval

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

can be interpreted as a statement about the components of the mean vector μ.

- Ex) With $a' = [1,0,\ldots,0]$, $a'\mu = \mu_1$. This is the usual confidence interval for a normal population mean.
- Can make several confidence statements about the components of μ, each with associated confidence coefficient 1 - α, by choosing different coefficient vectors a. However, the confidence associated with all of the statements taken together is **not** 1 - α.

- It would be desirable to associate a "collective" confidence coefficient of 1 - α with the confidence intervals that can be generated by all choices of a.
- A price must be paid for the convenience of a large simultaneous confidence coefficient: intervals that are wider (less precise) than the interval for a specific choice of a.

Simultaneous Confidence Statements

- Given a data set $x_1, x_2, \ldots, x_n$ and a particular a, the confidence interval

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

is the set of $a'\mu$ values for which

$$|t| = \left|\frac{\sqrt{n}(a'\bar{x} - a'\mu)}{\sqrt{a'Sa}}\right| \le t_{n-1}(\alpha/2)$$

or, equivalently,

$$t^2 = \frac{n(a'\bar{x} - a'\mu)^2}{a'Sa} = \frac{n(a'(\bar{x} - \mu))^2}{a'Sa} \le t^2_{n-1}(\alpha/2).$$

- A simultaneous confidence region is given by the set of $a'\mu$ values such that t^2 is relatively small for *all* choices of a.

- Can expect that the constant $t^2_{n\text{-}1}(\alpha/2)$ will be replaced by a larger value c^2 when statements are developed for many choices of a.

- Considering the values of a for which $t^2 \le c^2$, lead to the determination of

$$\max_a t^2 = \max_a \frac{n(a'(\bar{x} - \mu))^2}{a'Sa} = n\left[\max_a \frac{(a'(\bar{x} - \mu))^2}{a'Sa}\right] = n(\bar{x} - \mu)' S^{-1}(\bar{x} - \mu) = T^2$$

with the maximum occurring for a proportional to $S^{-1}(\bar{x} - \mu)$.

Simultaneous Confidence Statements

- Result 5.3. Let $X_1, X_2, \ldots X_n$ be a random sample from an $N_p(\mu, \Sigma)$ population with positive definite Σ. Then, simultaneously for all a, the interval

$$\left(a'\overline{X} - \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha) a'Sa}, \quad a'\overline{X} + \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha) a'Sa} \right)$$

will contain $a'\mu$ with probability 1 - α.

- 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.

Simultaneous Confidence Statements

- The successive choices $a' = [1,0,\ldots 0]$, $a' = [0,1,\ldots 0]$, and so on through $a' = [0,0,\ldots,1]$ for the T^2-intervals allow us to conclude that

$$\bar{x}_1 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \le \mu_1 \le \bar{x}_1 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}}$$

$$\bar{x}_2 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \le \mu_2 \le \bar{x}_2 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}}$$

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

$$\bar{x}_p - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}} \le \mu_p \le \bar{x}_p + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}}$$

all hold simultaneously with confidence coefficient 1 - α.

Simultaneous Confidence Statements

- Without modifying the coefficient 1 - α, can make statements about the differences $\mu_i - \mu_k$ corresponding to $a' = [0,\ldots,0, a_i ,0,\ldots,0, a_k ,0,\ldots,0]$, where $a_i = 1$ and $a_k = -1$:

  With $a'Sa = s_{ii} - 2s_{ik} + s_{kk}$,

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

$$\le \bar{x}_i - \bar{x}_k + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}}.$$

  - The confidence coefficient 1 - α remains unchanged for any choice of a, so linear combinations of the components μ_i that merit inspection *based upon an examination of the data* can be estimated.

- The statements about (μ_i , μ_k) belonging to the sample mean-centered ellipses

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

  maintain the confidence coefficient (1 - α) for the whole set of statements.

Simultaneous Confidence 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.

- Example 5.4 Simultaneous confidence intervals as shadows of the confidence ellipsoid

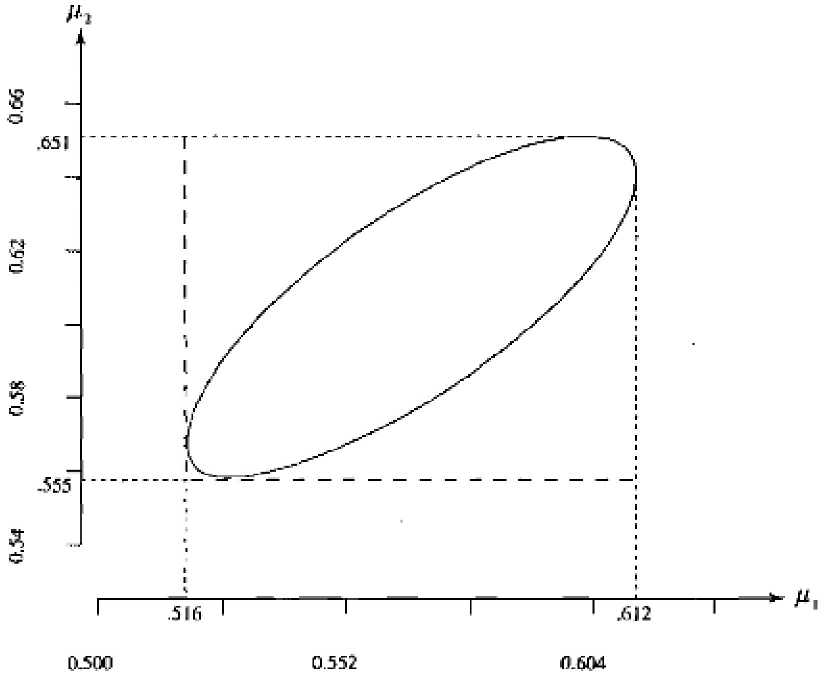

Figure 5.2 Simultaneous T^2-intervals for the component means as shadows of the confidence ellipse on the axes—microwave radiation data.

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

- An alternative approach to the construction of confidence intervals is to consider the components μ_i one at a time with $a' = [0,\ldots,0, a_i ,0,\ldots,0]$ where $a_i = 1$.

  - This approach ignores the covariance structure of the p variables and leads to the intervals

$$\bar{x}_1 - t_{n-1}(\alpha/2)\sqrt{\frac{s_{11}}{n}} \le \mu_1 \le \bar{x}_1 + t_{n-1}(\alpha/2)\sqrt{\frac{s_{11}}{n}},$$

$$\bar{x}_2 - t_{n-1}(\alpha/2)\sqrt{\frac{s_{22}}{n}} \le \mu_2 \le \bar{x}_2 + t_{n-1}(\alpha/2)\sqrt{\frac{s_{22}}{n}},$$

$$\vdots$$

$$\bar{x}_p - t_{n-1}(\alpha/2)\sqrt{\frac{s_{pp}}{n}} \le \mu_p \le \bar{x}_p + t_{n-1}(\alpha/2)\sqrt{\frac{s_{pp}}{n}}.$$

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

- Although prior to sampling, the ith interval has probability $1 - \alpha$ of covering μ_i, we do not know what to assert, in general, about the probability of **all** intervals containing their respective μ_i's.

  - This probability is not $1 - \alpha$.

  - Consider the special case where the observations have a joint normal distribution and

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}.$$

  - Since the observations on the first variable are independent of those on the second variable, and so on, the product rule for independent events can be applied:

$$P[\text{all } t\text{-intervals contain the } \mu_i\text{'s}] = (1 - \alpha)\,(1 - \alpha)\ldots\,(1 - \alpha) = (1 - \alpha)^p$$

  - Ex) If $1 - \alpha = .95$ and $p = 6$, this probability is $(.95)^6 = .74$.

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

- To guarantee a probability of 1 - α 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 - α.

  - Ex) 1 - α =.95, n = 15, and p = 4

    The multipliers of $\sqrt{\frac{s_{ii}}{n}}$ are $\sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)} = \sqrt{\frac{4\times 14}{11}\times 3.36} = 4.14$

    and $t_{n-1}(\alpha/2) = 2.145$, respectively. In this case, the simultaneous intervals are 100(4.14 – 2.145)/2.145 = 93% wider than those derived from the one-at-a-time t-method.

Table 5.3 Critical Distance Multipliers for One-at-a-Time t- Intervals and T^2-Intervals for Selected n and p $(1 - \alpha = .95)$

n	$t_{n-1}(.025)$	$\sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(.05)}$ $p = 4$	$\sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(.05)}$ $p = 10$
15	2.145	4.14	11.52
25	2.064	3.60	6.39
50	2.010	3.31	5.05
100	1.970	3.19	4.61
∞	1.960	3.08	4.28

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

- Note that the confidence level associated with any collection of T^2-intervals, for fixed n and p, is .95, and the overall confidence associated with a collection of individual t-intervals, for the same n, can be much less than .95.
- The one-at-a-time t-intervals are too short to maintain an overall confidence level for separate statements about all p means.
- The T^2-intervals are too wide if they are applied only to the p component means.

- In Figure 5.2, if μ_1 lies in its T^2-interval and μ_2 lies in its T^2-interval, then (μ_1,μ_2) lies in the rectangle formed by these two intervals. This rectangle contains the confidence ellipse and more. The confidence ellipse is smaller but has probability .95 of covering the mean vector μ with its components means μ_1 and μ_2. Consequently, the probability of covering the two individual means μ_1 and μ_2 will be larger than .95 for the rectangle formed by the T^2-intervals.

The Bonferroni Method of Multiple Comparisons

- Often, attention is restricted to a small number of individual confidence statements. In these situations, it is possible to do better than the simultaneous intervals of Result 5.3.

- **Bonferroni Method**

  - Suppose that confidence statements about m linear combinations $a_1'\mu, a_2'\mu, \ldots, a_m'\mu$ are required.

  - C_i : a confidence statement about the value of $a_i'\mu$ with $P[C_i \text{ true}] = 1 - \alpha_i$, $i = 1,2,\ldots, m$.

  - $P[\text{all } C_i \text{ true}] = 1 - P[\text{at least one } C_i \text{ false}]$

$$\geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m} \left(1 - P(C_i \text{ true})\right)$$

$$= 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m).$$

  - Can control the overall error rate $\alpha_1 + \alpha_2 + \cdots + \alpha_m$, regardless of the correlation structure behind the confidence statements.

The Bonferroni Method of Multiple Comparisons

- Example. Simultaneous interval estimates for the restricted set consisting of the components μ_i of μ

  - Consider the individual t-intervals

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

Since $P\left[\bar{X}_i \pm t_{n-1}(\alpha/2)\sqrt{s_{ii}/n} \text{ contains } \mu_i\right] = 1 - \dfrac{\alpha}{m},\ i = 1,2,\ldots,m,$

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

With an overall confidence level greater than or equal to 1 - α, can make the following $m = p$ statements:

$$\bar{x}_1 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \le \mu_1 \le \bar{x}_1 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}},$$

$$\bar{x}_2 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} \le \mu_2 \le \bar{x}_2 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}},$$

$$\vdots$$

$$\bar{x}_p - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}} \le \mu_p \le \bar{x}_p + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}}.$$

The Bonferroni Method of Multiple Comparisons

- Example 5.6 Constructing Bonferroni simultaneous confidence intervals and comparing them with T^2-intervals

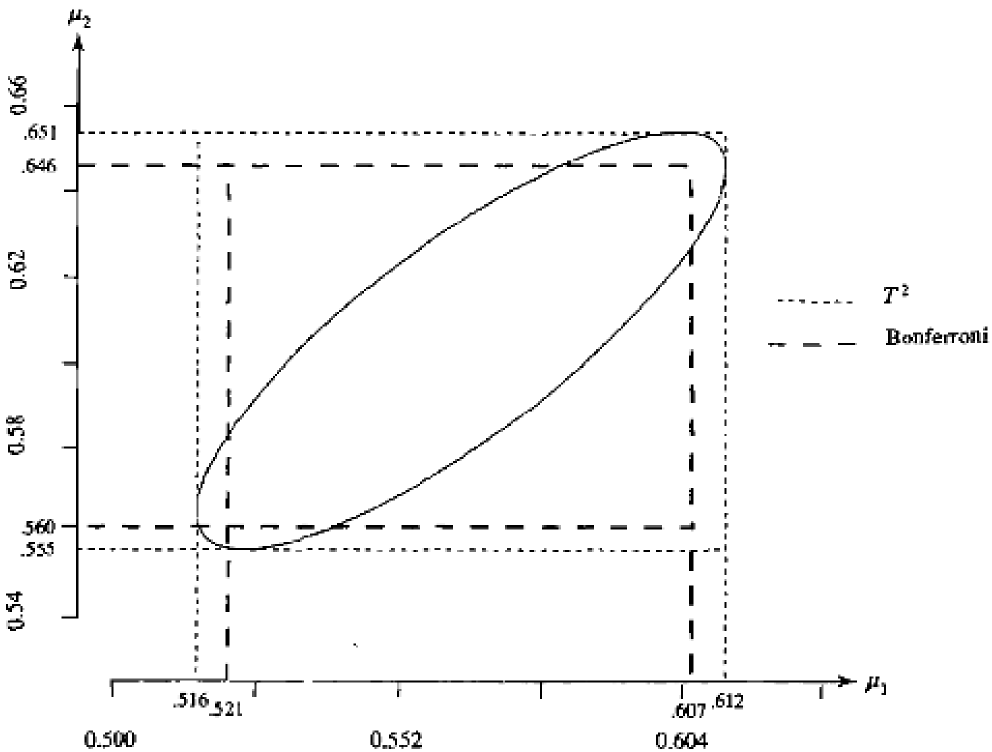

Figure 5.4 The 95% T^2 and 95% Bonferroni simultaneous confidence intervals for the component means—microwave radiation data.

The Bonferroni Method of Multiple Comparisons

- The Bonferroni intervals for linear combinations $a'\mu$ and the analogous T^2-intervals have the same general form:

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

- In every instance where $\alpha_i = \alpha/m$,

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

which does not depend on the random quantities $\bar{X}$ and S.

- For the ratio of the length of Bonferroni interval and the length of T^2-interval, refer Table 5.4:

Table 5.4 (Length of Bonferroni Interval)/(Length of T^2-Interval) for $1 - \alpha = .95$ and $\alpha_i = .05/m$

	$m = p$		
n	2	4	10
15	.88	.69	.29
25	.90	.75	.48
50	.91	.78	.58
100	.91	.80	.62
∞	.91	.81	.66

5.5. Large Sample Inferences 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.
  - Since ($\bar{x}$, S) is a 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.
  - Since $(\bar{X}-\mu)'(n^{-1}S)^{-1}(\bar{X}-\mu)=n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)$ is approximately χ^2 with p d.f.,

$$P\left[n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)\le \chi_p^2(\alpha)\right]\cong 1-\alpha,$$

where $\chi^2_p(\alpha)$ is the upper (100α)th percentile of the χ^2_p-distribution.

5.5. Large Sample Inferences about a Population Mean Vector

- Result 5.4. Let $X_1, X_2, \ldots X_n$ be a random sample from a population with mean μ and positive definite covariance matrix Σ. When $n-p$ is large, the hypothesis H_0: $\mu = \mu_0$ is rejected in favor of H_1: $\mu \neq \mu_0$, at a level of significance approximately α, if the observed
$$n\left(\overline{X}-\mu\right)' S^{-1}\left(\overline{X}-\mu\right) > \chi_p^2(\alpha).$$
Here $\chi^2_p(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p d.f.

- Comparing the test in Result 5.4 with the corresponding *normal theory test* using $T^2 = n(\bar{x}-\mu_0)S^{-1}(\bar{x}-\mu_0) > \dfrac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$, the test statistics have the same structure, but the critical values are different.

  - Both tests yield essentially the same result in situations where the χ^2-test of Result 5.4 is appropriate, since $\dfrac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)$ and $\chi^2_p(\alpha)$ are approximately equal for n large relative to p.

5.5. Large Sample Inferences about a Population Mean Vector

- Result 5.5. Let $X_1, X_2, \ldots X_n$ be a random sample from a population with mean μ and positive definite covariance Σ. If $n - p$ is large,

$$a'\overline{X} \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{a'Sa}{n}}$$

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

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

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

$$\vdots$$

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

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

$$n\left[\bar{x}_i - \mu_i, \bar{x}_k - \mu_k\right]\begin{bmatrix} s_{ii} & s_{ik} \\ s_{ik} & s_{kk} \end{bmatrix}^{-1}\begin{bmatrix} \bar{x}_i - \mu_i \\ \bar{x}_k - \mu_k \end{bmatrix} \leq \chi_p^2(\alpha) \quad \text{contains } (\mu_i, \mu_k).$$

5.5. Large Sample Inferences about a Population Mean Vector

- Example 5.7 Constructing large sample simultaneous confidence intervals

- 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

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