Statistics 135

Chapter 5

Inference about Mean Vector

Chris Drake
Department of Statistics
University of California, Davis

Hypothesis testing

1 *Univariate case:* $H_0: \ \mu = \mu_0$ vs $H_1: \ \mu \neq \mu_0$ tested using

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1} \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

2 $(1-\alpha)$ confidence interval given by

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

3 Relationship between $(1-\alpha)$ confidence interval and size α test: if a hypothesized value μ_0 is in the confidence interval, then the p-value corresponding to the hypothesis in (1) will be larger than α. Similarly, if μ_0 is not in the confidence interval $p - value < \alpha$.

Therefore, we can construct a confidence interval as a test of H_0.

4 *Multivariate case:* same hypothesis as above, except μ and μ_0 are vectors. Statistic

$$T^2 = (\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\left(\frac{1}{n}\mathbf{S}\right)^{-1}(\bar{\mathbf{X}} - \mu_{\mathbf{0}}) = (\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})$$

note: T^2 is the multivariate analogue of $t^2 = (\frac{\bar{X}-\mu_0}{s/\sqrt{n}})^2$ and T^2 is called $Hotelling's\ T^2$; its distribution is

$$T^2 \sim \frac{(n-1)p}{n-p}F_{p,n-p} \qquad where \qquad p = dimension(\mu)$$

and n is the sample size. We can calculate

$$\begin{aligned} \alpha &= P\left[T^2 > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)\right] \\ &= P\left[(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu_{\mathbf{0}}) > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)\right] \end{aligned}$$

5 T^2 is invariant under linear transformations, multiplying the data vector $\mathbf{X}$ by a matrix $\mathbf{C}$ and or adding a vector $\mathbf{d}$ to get $\mathbf{Y} = \mathbf{CX} + \mathbf{d}$; such transformations occur when data are re-scaled or shifted. Therefore $T^2_{\mathbf{X}} = T^2_{\mathbf{CX+d}} = T^2_{\mathbf{Y}}$ and we do not need to recalculate the statistic.

6 To calculate $Hotelling's\ T^2$ and carry out the test you proceed as follows:

a calculate $\bar{\mathbf{X}}$ from the data vector $\mathbf{X}_{n \times p}$

b calculate $\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{X_j} - \bar{\mathbf{X}})(\mathbf{X_j} - \bar{\mathbf{X}})'$

c calculate $(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})$ for the hypothesized value

c obtain the α^{th} percentile $F_\alpha = F_{p,n-p}(\alpha)$ from the F distribution with p numerator and $n-p$ denominator degrees of freedom and multiply F_α by $[(n-1)p/(n-p)]$; if T^2 exceeds that value, the null hypothesis is rejected at level α.

Likelihood Ratio Test (LRT)

6 General LRT: $\boldsymbol{\Theta}$ parameter space with $\boldsymbol{\Theta}_0$ the subspace in which the parameter vector θ lies if H_0 is true, then the LRT is given by

$$\Lambda = \frac{\max\limits_{\theta \in \boldsymbol{\Theta}_0} \mathcal{L}(\theta)}{\max\limits_{\theta \in \boldsymbol{\Theta}} \mathcal{L}(\theta)} < c$$

The test rejects $H_0: \ \theta \in \boldsymbol{\Theta}_0$ vs $H_1: \ \theta \in \boldsymbol{\Theta}$ if Λ is small. If the numerator is small then the restricted value that maximizes the likelihood is not very plausible. To find a p-value or critical value, we need to know the distribution of Λ. That is often difficult or intractable.

7 Fact: if the sample size n is large and some other conditions hold, then $-2ln(\Lambda) \sim \chi^2_{\nu-\nu_0}$ where $dim(\boldsymbol{\Theta}) = \nu$ and $dim(\boldsymbol{\Theta}_0) = \nu_0$

8 Fact: $Hotelling's\ T^2$ is the likelihood ratio test statistic for testing $H_0: \ \mu = \mu_0$ vs $H_1: \ \mu \neq \mu_0$

9 The MLE's by definition maximize the likelihood. Therefore

$$\max_{\theta \in \boldsymbol{\Theta}} \mathcal{L}(\theta) = (2\pi)^{(-np/2)} |\hat{\boldsymbol{\Sigma}}|^{(-n/2)} e^{-np/2}$$

with $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x_j} - \bar{\mathbf{x}})(\mathbf{x_j} - \bar{\mathbf{x}})'$ and $\hat{\mu} = \bar{\mathbf{x}}$.

10 Similarly

$$\max_{\theta \in \boldsymbol{\Theta}_0} \mathcal{L}(\theta) = (2\pi)^{(-np/2)} |\hat{\boldsymbol{\Sigma}}_{\mathbf{0}}|^{(-n/2)} e^{-np/2}$$

with $\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x_j} - \mu_{\mathbf{0}})(\mathbf{x_j} - \mu_{\mathbf{0}})'$ and $\hat{\mu} = \bar{\mathbf{x}}$.

11 Combining (9) and (10) yields $\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_{\mathbf{0}}|} \right)^{(n/2)}$ and Wilks' lambda

$$\Lambda^{(2/n)} = \left(1 + \frac{T^2}{n-1} \right)^{(-1)} \implies T^2 = \frac{(n-1)|\hat{\boldsymbol{\Sigma}}_{\mathbf{0}}|}{|\hat{\boldsymbol{\Sigma}}|} - (n-1)$$

Confidence Regions

1 $\boldsymbol{\Theta}$ the parameter space or set of all possible values of θ. The confidence region $R(\mathbf{X})$ is a set of points such that

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

2 A p-dimensional $(1-\alpha)$ confidence region for the mean vector is given by the set of values such that

$$P\Big[n(\bar{\mathbf{X}}-\mu)'\mathbf{S}^{-1}(\bar{\mathbf{X}}-\mu) \le \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)\Big]$$

For this region $\bar{\mathbf{X}}$ will be within $[\big((n-1)pF_{p,n-p}(\alpha)\big)/(n-p)]^{1/2}$ of the true mean μ for whatever the true values of μ and $\boldsymbol{\Sigma}$ are. For any sample value $\bar{\mathbf{x}}$ the region $n(\bar{\mathbf{x}}-\mu)'\mathbf{s}^{-1}(\bar{\mathbf{x}}-\mu) \le \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$ defines an ellipsoid centered at μ. This region can also be used to test the null hypothesis $H_0 : \ \mu = \mu_0$ vs $H_1 : \ \mu \neq \mu_0$

Simultaneous Confidence Statements

The confidence region of the previous page provides a statement about the vector μ, however, interest is often in the individual components. If we calculate individual confidence intervals, all at level $(1-\alpha)$ then the true confidence level for all components of μ simultaneously is typically less than $(1-\alpha)$. If we use the t-distribution for individual confidence intervals to get

$$\bar{x}_i - t_{n-1,(\alpha_i/2)}\sqrt{\frac{\sigma_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + t_{n-1,(\alpha_i/2)}\sqrt{\frac{\sigma_{ii}}{n}}$$

where by above $P(C_i \text{ contains } \mu_i) = 1 - \alpha_i$, then

$$P(\text{all } C_i \text{ contain } \mu_i) = \prod_{i=1}^{p}(1-\alpha_i) = (1-\alpha)^p < 1 - \alpha$$

if $\alpha_i = \alpha$ for $i = 1, .., p$

Bonferroni Method

The Bonferroni method adjusts the confidence level of the individual intervals to make the overall coverage level $1 - \alpha$; this requires that the number of individual confidence intervals to be calculated is determined before analysis.

$$\begin{aligned} P(\text{all } C_i \text{ contain } \mu_i) &= 1 - P(\text{at least one } C_i \text{ does not}) \\ &\geq 1 - \sum_{i=1}^{p} P(C_i \text{ does not}) \\ &= 1 - \sum_{i=1}^{p} (1 - P(C_i \text{ contains } \mu_i)) \\ &= 1 - \sum_{i=1}^{p} \alpha_i \end{aligned}$$

thus, we need $\alpha = \sum_{i=1}^{p} \alpha_i$, making the individual intervals wider.

To achieve a simultaneous confidence level for all components the intervals need to be wider. We want a procedure that provides coverage for all intervals simultaneously that results in the narrowest intervals possible and does not require the number of individual intervals to be pre-specified. If $\mathbf{X}$ is p-variate normal with mean normal with mean μ and covariance matrix **Sigma** then for any p-vector $\mathbf{a}$ the linear combination $\mathbf{a}'\mathbf{X}$ is univariate normal with mean $\mathbf{a}'\mu$ and variance $\mathbf{a}'\mathbf{\Sigma}\mathbf{a}$. Then a CI for $\mathbf{a}'\mu$ can be based on $\mathbf{a}'\bar{\mathbf{x}}$ for a sample $\mathbf{x}_1, .., \mathbf{x}_n$. We would like to find linear combination s.t.

$$\max_{\mathbf{a}} t^2 = \max_{\mathbf{a}} \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \mu))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}}$$

If we find this $\mathbf{a}_{max}$ then any $(1-\alpha)$ CI for $(\mathbf{a}'\mu)$ will also have confidence coefficient $(1-\alpha)$ for any other vector $\mathbf{a}$; in particular it will hold for $\mathbf{a}' = (0, .., 1, 0..0)$ with 1 in the i^{th} position, which leads to a CI for μ_i.

Fact

$$\mathbf{a}_{max} \propto S^{-1}(\mathbf{X} - \mu)$$

leading to

$$\max_{\mathbf{a}} t^2 = n(\bar{\mathbf{X}} - \mu)' S^{-1}(\bar{\mathbf{X}} - \mu) = T^2$$

from this we get for any other linear combination $\mathbf{a}'\mu$, the $(1-\alpha)$ CI is

$$\mathbf{a}'\bar{\mathbf{x}} - \sqrt{\frac{p(n-1)}{n(n-p} F_{p,n-p,\alpha} \mathbf{a}' S \mathbf{a}} \leq \mathbf{a}'\mu \leq \mathbf{a}'\bar{\mathbf{x}} + \sqrt{\frac{p(n-1)}{n(n-p} F_{p,n-p,\alpha} \mathbf{a}' S \mathbf{a}}$$