Chapter 5. Inferences About The Mean Vector

This chapter addresses the most basic and most standard statistical problem about the mean of a population, given a standard "one sample": some independent identically distributed (iid) observations from the population. This is the so-called one-sample problem that we have dealt with as the very first non-trivial statistical problem. The characteristic of the present problem is that each observation is multivariate, i.e., of high dimension. In other words, we are concerned with not a parameter of 1 dimension or a number, but a parameter of high dimension or a vector.

The estimation of the vector parameter of population mean is not problematic. We simply use the sample mean. It is the inference, the accuracy of the estimation, of the vector parameter, or several 1 dimensional parameter putting together, that needs a special treatment. As we know, the accuracy justification is typically formulated in terms of test of hypothesis or confidence intervals/regions.

5.1 Test of hypothesis "$\mu = \mu_0$".

(i). The univariate case — a review.

Suppose $X_1, ..., X_n$ are iid $\sim N(\mu, \sigma^2)$ with both μ and σ^2 unknown. The inference with μ is based on the fact that

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}.$$

Consider hypothesis

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_a: & \mu \neq \mu_0. \end{cases}$$

A test at significance level α is

$$\text{reject } H_0 \quad \text{when} \quad \frac{\bar{X} - \mu_0}{s/\sqrt{n}} > t_{n-1}(\alpha/2).$$

Correspondingly, a confidence interval for μ at confidence level $1 - \alpha$ is

$$\bar{X} \pm t_{n-1}(\alpha/2)s/\sqrt{n}.$$

(ii). The multivariate case.

Let $X_1, ..., X_n$ be iid $\sim MN(\mu, \Sigma)$ be random vectors of p dimension, where

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{pp} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}.$$

The sample mean and the sample variance are

$$\bar{X}_{p\times 1} = \frac{1}{n}\sum_{i=1}^n X_i \qquad \mathbf{S}_{p\times p} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'.$$

Then,

$$n(\bar{X} - \mu)'\mathbf{S}^{-1}(\bar{X} - \mu) \sim \frac{(n-1)p}{n-p} F_{p,n-p}.$$

This is key fact that enables inference with the population mean vector μ.

Consider hypotheses:

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_a: & \mu \neq \mu_0, \end{cases} \quad \text{where } \mu_0 = \begin{pmatrix} \mu_{01} \\ \vdots \\ \mu_{0p} \end{pmatrix} \text{ is a known } p\text{-vector.}$$

Set
$$T^2 = n(\bar{X} - \mu_0)'\mathbf{S}^{-1}(\bar{X} - \mu_0).$$
Then, a significance level α test is
$$\text{reject } H_0 \quad \text{when} \quad T^2 > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha).$$
Moreover, noticing that the larger the T^2, the more evidence against H_0, the p-value of the test is
$$P\Big(F_{p,n-p} > \frac{n-p}{(n-1)p} \times (\text{the observed value of } T^2 \text{ .})$$
As always, we can state that a significance level α test is to reject the null hypothesis when the p-value is smaller than α.

Example 1.1 (continued). Ten American companies. In this example, $p = 3$ and $n = 10$.
$$\bar{X} = \begin{pmatrix} 62.31 \\ 2.92 \\ 81.24 \end{pmatrix} \qquad \text{and} \qquad \mathbf{S} = \begin{pmatrix} 1000.5 & 25.6 & 1511.8 \\ 25.6 & 1.43 & 45.7 \\ 1511.8 & 45.7 & 2980.5 \end{pmatrix}.$$
Consider hypotheses:
$$\begin{cases} H_0: & \mu = \mu_0 \\ H_a: & \mu \neq \mu_0 \ , \end{cases} \qquad \text{where } \mu_0 = \begin{pmatrix} 50 \\ 2 \\ 60 \end{pmatrix}.$$
The T^2 statistic is
$$T^2 = n(\bar{X} - \mu_0)'\mathbf{S}^{-1}(\bar{X} - \mu_0) = 6.48.$$
By checking tables, we find $F_{3,7}(0.05) = 4.35$ and $[(n-1)p/(n-p)]F_{p,n-p}(0.05) = 16.76$. Since $T^2 = 6.48 < 16.76$, we conclude that, at significance level $\alpha = 0.05$, we accept H_0.

The p-value of the test is
$$P(F_{3,7} > \frac{10-3}{(10-1)\times 3} \times 6.48) = P(F_{3,7} > 1.68) = 0.257.$$
As the p-value is larger than 0.05, we accept H_0 at significance level 0.05.

5.2. Simultaneous confidence statements.

(i). Confidence regions.

Observe that
$$P\Big(n(\bar{X} - \mu)'\mathbf{S}^{-1}(\bar{X} - \mu) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)\Big) = 1 - \alpha.$$
As a result, a confidence region for μ at confidence level $1-\alpha$ is
$$\Big\{\mu \in R^p : n(\bar{X} - \mu)'\mathbf{S}^{-1}(\bar{X} - \mu) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)\Big\}$$
which is an ellipse in R^p centered at $\bar{X}$.

(ii). Simultaneous confidence intervals.

The confidence region is, in practice, often difficult to interpret, especially when covariates are all of totally different nature. In reality, it may be preferred to give a confidence interval for each component of the multivariate parameter, e.g., a confidence interval for each μ_i, and yet with an overall confidence level. If we call an interval not covering a targeted parameter a failure, we need to ensure a probability (confidence level) that no such failures for any parameters occur. Then, these confidence intervals can be called simultaneous confidence intervals, or confidence intervals with one overall confidence level. For the one sample problem, we have two approaches. One evolves from the T^2 method. The other, typically called Bonferroni's method, evolves from t-confidence interval for univariate parameter .

(a). Simultaneous T^2 confidence intervals.

The following lemma may be understood as a generalization of simple fact: the unit ball in R^p is same as $\{b \in R^p : (a'b)^2/\|a\|^2 \le 1, \text{ for all } a \in R^p\}$.

LEMMA. For any positive definite $p \times p$ matrix B,

$$\{b \in R^p : b'B^{-1}b \le c^2\} = \{b \in R^p : \frac{(a'b)^2}{a'Ba} \le c^2 \text{ for all } a \in R^p\}.$$

Proof. Let $\tilde{b} = B^{-1/2}b$ and $\tilde{a} = B^{1/2}a$. Then,

$$\begin{aligned} b'B^{-1}b \le c^2 \quad &\iff \quad \tilde{b}'\tilde{b} \le c^2 \iff \|\tilde{b}\|^2 \le c^2 \\ &\iff \quad \frac{(\tilde{a}'\tilde{b})^2}{\|\tilde{a}\|^2} \le c^2 \qquad \text{for all } \tilde{a} \in R^p \\ &\iff \quad \frac{(a'B^{1/2}B^{-1/2}b)^2}{a'Ba} \le c^2 \qquad \text{for all } a \in R^p \\ &\iff \quad \frac{(a'b)^2}{a'Ba} \le c^2 \qquad \text{for all } a \in R^p. \end{aligned}$$

□

It then follows form this lemma that

$$\begin{aligned} &\Big\{\mu \in R^p : n(\bar{X} - \mu)'\mathbf{S}^{-1}(\bar{X} - \mu) \le \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)\Big\} \\ = \quad &\Big\{\mu \in R^p : \frac{n[a'(\bar{X} - \mu)]^2}{a'\mathbf{S}a} \le \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha), \text{ for all } a \in R^p\Big\} \end{aligned}$$

Since it is with probability $1 - \alpha$, for any p-vectors $a_1, ..., a_k$,

$$P\Big(\{|a_j'(\bar{X} - \mu)|^2 \le a_j'\mathbf{S}a_j \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha) \quad \text{for } j = 1, .., k.\}\Big) \ge 1 - \alpha.$$

Consequently,

$$a_j'\bar{X} \pm \sqrt{a_j'\mathbf{S}a_j \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)}, \qquad j = 1, ..., k$$

are simultaneous confidence intervals for $a_j'\mu$, $j = 1, ..., k$ at an overall confidence level $1 - \alpha$. In particular, by choosing $a_1 = (1, 0, ...0)'$, $a_2 = (0, 1, 0, ...0)'$, ..., $a_p = (0, 0, ..., 0, 1)'$,

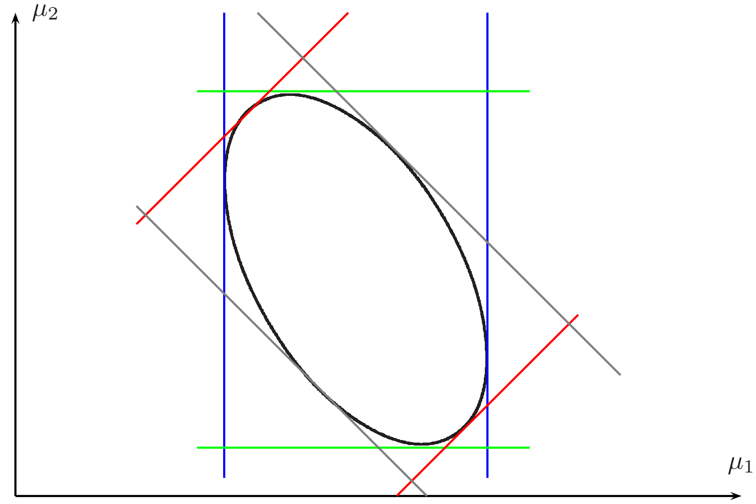
$$\bar{X}_j \pm c\sqrt{s_{jj}/n}, \qquad j = 1, ..., p$$

are confidence intervals for μ_i, $i = 1, ..., p$ at confidence level $1 - \alpha$, where

$$\bar{X}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}, \qquad s_{lj} = \frac{1}{n-1}\sum_{i=1}^n (x_{il} - \bar{X}_l)(x_{ij} - \bar{X}_j), \qquad c = \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)}$$

are, respectively, sample mean for the j-th variable/component, sample covariance between the l-th and the j-th variable/component and a constant.

The following diagram illustrates the geometry of the T^2 confidence intervals. Suppose $p = 2$, we wish to construct the T^2 simultaneous confidence intervals for μ_1, μ_2, $\mu_1 + \mu_2$ and $\mu_2 - \mu_1$.



The green lines are two endpoints for μ_2, the blue for μ_1, the red for $\mu_2-\mu_1$ and the gray for $\mu_1+\mu_2$. Each of the four pairs of parallel lines forms a trip. The intersection of the four strips is the region of the values of the vector parameter formed by the T^2 simulateous confidence intervals, which clearly cover the ellipse. As the ellipse has precisely confidence level $1-\alpha$. The T^2 simulateous confidence intervals with nominal level $1-\alpha$ have actua level larger than $1-\alpha$.

Remark. In general, suppose there exists a confidence region for a vector parameter $\theta \in R^p$ at confidence level $1-\alpha$:

$$\left\{\theta \in R^P : (b-\theta')A^{-1}(b-\theta) \le c_\alpha\right\}$$

which is an ellipse centered at $b \in R^p$, where c_α is a positive constant depending on α and A is a positive definite matrix. Then, simultaneous confidence intervals at confidence level $(1-\alpha)$ for parameters: $a_j'\theta, j = 1, ..., k$, where $a_j \in R^p$ are

$$\text{for } a_j'\theta : \qquad a_j'b \pm \sqrt{a_j'Aa_jc_\alpha}; \qquad j = 1, ..., k.$$

Note that k, the number of parameters in need of simultaneous inference, does not appear in the confidence intervals. This is a stark contrast to the following Bonferroni's method.

(b). Bonferroni's simultaneous confidence intervals for μ_i, $i = 1, ..., p$.

Recall that, for the j-th variable/component,

$$x_{1j}, x_{2j}, ..., x_{nj} \quad \text{are iid} \quad \sim N(\mu_j, \sigma_{jj}).$$

Hence,

$$\frac{\bar{X}_j - \mu_j}{\sqrt{s_{jj}/n}} \sim t_{n-1},$$

and, *for every fixed* j, $\bar{X}_j \pm t_{n-1}(\alpha^*/2)\sqrt{s_{jj}/n}$ is a confidence interval for μ_j at confidence level $1-\alpha^*$. It means that this interval fails to cover μ_j with probability α^*. Suppose we consider all these p confidence intervals all together. The overall confidence level cannot be $1-\alpha^*$ in general, as the chance of any of the p intervals failing to cover the target parameter should exceed α^*, but should be smaller than $p\alpha^*$. Set $\alpha = p\alpha^*$. Then $\alpha^* = \alpha/p$. The above confidence intervals is the so-called Bonferroni's simultaneous confidence at overall confidence level $1-\alpha$:

$$\bar{X}_j \pm t_{n-1}(\alpha/(2p))\sqrt{s_{jj}/n}, \text{for } \mu_j,\ j = 1, ..., p.$$

Suppose, in a general setting, we wish to construct Bonferroni's simultaneous confidence intervals at confidence level $1-\alpha$ for parameters: $a_1'\mu, ..., a_k'\mu$. They are (please DIY)

$$a_j'\bar{X}_j \pm t_{n-1}(\frac{\alpha}{2k})\sqrt{a_j'\mathbf{S}a_j/n}, \text{for } a_j'\mu_j,\ j = 1, ..., k.$$

Remark. The above (overall) confidence levels for simultaneous confidence intervals are only nominal levels, not the actual level. The nominal level is the required level that the confidence statements must be correct with chance *at least* as much. The actual level is the exact chance the confidence statements are correct.

Remark. In general, Bonferroni's method simply says that, if there is a confidence interval $[a_i(\alpha), b_i(\alpha)]$ for θ_i, *each* at confidence level $1-\alpha$, for every $i = 1, ..., K$. Then,

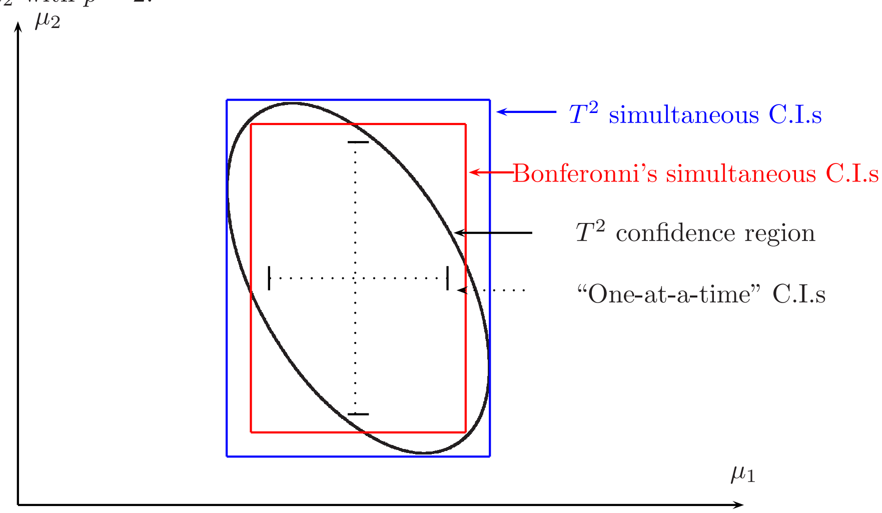
$$[a_i(\alpha/K), b_i(\alpha/K)], \qquad i = 1, ..., K,$$

are simultaneous confidence intervals for parameters $\theta_i, i = 1, ..., K$ at (overall) confidence level $1-\alpha$. The proof is as follows:

$$\begin{aligned}
& P\Big(\theta_i \in a_i(\alpha/K), b_i(\alpha/K)], \text{ for all } i = 1, ..., K\Big) \\
= \;& P\Big(\cap_{i=1}^K \{\theta_i \in a_i(\alpha/K), b_i(\alpha/K)]\}\Big) \\
= \;& 1 - P\Big(\cup_{i=1}^K \{\theta_i \notin a_i(\alpha/K), b_i(\alpha/K)]\}\Big) \\
\geq \;& 1 - \sum_{i=1}^K P(\theta_i \notin a_i(\alpha/K), b_i(\alpha/K)]) \\
= \;& 1 - \sum_{i=1}^K \alpha/K \\
= \;& 1 - \alpha.
\end{aligned}$$

It is important to observe that the lengths of the T^2 confidence intervals do not depend on the number of parameters, which we denote as k above, while those of the Bonferroni's confidence intervals do. We may conclude that, if the number of the parameters, for which we wish to construct simultaneous confidence intervals, is large, the T^2 method is preferred over the Bonferroni method. And vise versus otherwise.

The following is diagram illustrating a typical scenario of confidence statements at same confidence levels for μ_1 and μ_2 with $p = 2$.



Note that the Bonferroni's simultaneous C.I.s are not always smaller than T^2 simultaneous C.I.s shown is the above diagram.

Example 1.1 (continued) Ten American Companies. We wish to construct confidence region for μ and simultaneous confidence intervals for μ_i, $i = 1, 2, 3$, all at overall confidence level 95%.

Then, $\alpha = 0.05$. Recall that $n = 10$, $p = 3$ and

$$\bar{X} = \begin{pmatrix} 62.31 \\ 2.92 \\ 81.24 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 1000.5 & 25.6 & 1511.8 \\ 25.6 & 1.43 & 45.7 \\ 1511.8 & 45.7 & 2980.5 \end{pmatrix} \text{ and } \quad \mathbf{S}^{-1} = \begin{pmatrix} .0044 & -.0146 & -.0020 \\ & 1.4171 & -.0143 \\ & & .0016 \end{pmatrix}.$$

(a). The T^2 confidence region for μ at confidence level 95% is

$$\Big\{\mu \in R^3 : (\bar{X} - \mu)'\mathbf{S}^{-1}(\bar{X} - \mu) \le \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha) = \frac{9 \times 3}{10 \times 7} F_{3,7}(.05) = 1.676\Big\}.$$

(b). The T^2 simultaneous confidence intervals for μ_1, μ_2 and μ_3 at (overall) confidence level 95% are, with $c = \sqrt{[(n-1)p/(n-p)]F_{p,n-p}(\alpha)} = 4.09$,

$$\begin{aligned} &\text{for } \mu_1: && \bar{X}_1 \pm c\sqrt{s_{11}/n} = 62.31 \pm 4.09\sqrt{1000.5/10} = 62.31 \pm 40.95; \\ &\text{for } \mu_2: && \bar{X}_2 \pm c\sqrt{s_{22}/n} = 2.92 \pm 4.09\sqrt{1.43/10} = 2.92 \pm 1.58; \\ &\text{for } \mu_3: && \bar{X}_3 \pm c\sqrt{s_{33}/n} = 81.24 \pm 4.09\sqrt{2980.5/10} = 81.24 \pm 70.67. \end{aligned}$$

(c). Non-simultaneous "one-at-a-time" confidence for μ_i, each at confidence level 95%, $i = 1, 2, 3$: $\bar{X}_i + t_{n-1}(\alpha/2)\sqrt{s_{ii}/n}$. Specifically, with $t_{n-1}(\alpha/2) = t_9(.025) = 2.262$,

$$\begin{aligned} &\text{95\% C.I. for } \mu_1: && \bar{X}_1 \pm t_{n-1}(\alpha/2)\sqrt{s_{11}/n} = 62.31 \pm 2.262\sqrt{1000.5/10} = 62.31 \pm 22.63; \\ &\text{95\% C.I. for } \mu_2: && \bar{X}_2 \pm t_{n-1}(\alpha/2)\sqrt{s_{22}/n} = 2.92 \pm 2.262\sqrt{1.43/10} = 2.92 \pm 0.855; \\ &\text{95\% C.I. for } \mu_3: && \bar{X}_3 \pm t_{n-1}(\alpha/2)\sqrt{s_{33}/n} = 81.24 \pm 2.262\sqrt{2980.5/10} = 81.24 \pm 39.05. \end{aligned}$$

(d) Bonferroni's simultaneous confidence intervals for μ_1, μ_2 and μ_3 at (overall) confidence level 95%: $\bar{X}_i + t_{n-1}(\alpha/(2p))\sqrt{s_{ii}/n}$. Specifically, with $t_{n-1}(\alpha/(2p)) = t_9(.05/6) = 2.933$,

$$\begin{aligned} &\text{95\% C.I. for } \mu_1: && \bar{X}_1 \pm t_{n-1}(\alpha/(2p))\sqrt{s_{11}/n} = 62.31 \pm 2.933\sqrt{1000.5/10} = 62.31 \pm 29.34; \\ &\text{95\% C.I. for } \mu_2: && \bar{X}_2 \pm t_{n-1}(\alpha/(2p))\sqrt{s_{22}/n} = 2.92 \pm 2.933\sqrt{1.43/10} = 2.92 \pm 1.11; \\ &\text{95\% C.I. for } \mu_3: && \bar{X}_3 \pm t_{n-1}(\alpha/(2p))\sqrt{s_{33}/n} = 81.24 \pm 2.933\sqrt{2980.5/10} = 81.24 \pm 50.64. \end{aligned}$$

In this example, at the same nominal confidence level 95%, Bonferroni's simultaneous confidence intervals have smaller width/length than those of the T^2 simultaneous confidence intervals. Note that the "one-at-a-time" confidence intervals, despite their smallest width, is not comparable with the simultaneous confidence intervals.