ST 437/537: Applied Multivariate and Longitudinal Data Analysis

Summary of inference in the one sample case

Arnab Maity
NCSU Department of Statistics
SAS Hall 5240 919-515-1937 amaity[at]ncsu.edu

Univariate case

Let $X_1, X_2, \ldots, X_n$ be a sample from a normal distribution with mean μ and unknown variance σ^2. An estimator of μ is the sample mean $\bar{X}$.

> ***Confidence Interval*** *Let $t_{n-1}(\alpha/2)$ be the upper-tail probability corresponding to the t_{n-1} distribution. Then the $100(1-\alpha)\%$ confidence interval (CI) for μ is*
>
> $$\left(\bar{X} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}},\ \bar{X} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right).$$
>
> *In absence of normality, we can construct **a large sample interval ($n \geq 40$)** using the same formula above with replacing $t_{n-1}(\alpha/2)$ by $z(\alpha/2)$.*

> ***One sample t-test:*** *We reject $H_0 : \mu = \mu_0$ in favor of $H_a : \mu \neq \mu_0$, if*
>
> $$\left|\frac{\bar{X} - \mu_0}{s/\sqrt{n}}\right| > t_{n-1}(\alpha/2);$$
>
> *we fail to reject H_0 otherwise.*

R function: `t.test()` ; it does both estimation and testing.

Multivariate Inference

Suppose we have random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, each of them is a $p \times 1$ vector (in our example, $p = 4$), generated from a p-variate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ and unknown covariance matrix $\boldsymbol{\Sigma}$. We want to form confidence intervals for the mean parameters $\mu_1, \ldots, \mu_p$.

<u>Simultaneous confidence intervals</u>: *The simultaneous* $100(1-\alpha)\%$ *confidence intervals for* $\mu_1, \ldots, \mu_p$ *are*

$$\text{For } \mu_k\text{: } \left(\bar{X}_k - \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) \frac{S_{kk}}{n}},\ \ \bar{X}_k + \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) \frac{S_{kk}}{n}}\right),$$

where S_{kk} *is the* k*th element of the sample covariance* $\boldsymbol{S}$*.*

<u>Large sample simultaneous intervals</u>: *When* n *is large, the approximate simultaneous* $100(1-\alpha)\%$ *confidence intervals for* $\mu_1, \ldots, \mu_p$ *are*

$$\text{For } \mu_k\text{: } \left(\bar{X}_k - \sqrt{\chi^2_p(\alpha) \frac{S_{kk}}{n}},\ \ \bar{X}_k + \sqrt{\chi^2_p(\alpha) \frac{S_{kk}}{n}}\right),$$

where S_{kk} *is the* k*th element of the sample covariance* $\boldsymbol{S}$*.*

<u>The Bonferroni method for multiple correction</u>: *The Bonferroni* $100(1-\alpha)\%$ *confidence intervals for* μ_k*,* $k = 1, \ldots, p$ *are*

$$\bar{X}_k \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{S_{kk}/n}, \quad k = 1, \ldots, p$$

where S_{kk} *is the* k*th element of the diagonal of the sample covariance* $\boldsymbol{S}$*.*

Hotelling's T^2 ***test:*** *We reject* $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ *at level* α *if*

$$\frac{n(n-p)}{(n-1)p}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{s}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) > F_{p,n-p}(\alpha).$$

When we have a large sample size, n*, we can again relax the normality assumption and conduct an approximate test: reject* H_0 *at level* α *if*

$$n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{s}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \chi^2_p(\alpha).$$

R function: `HotellingsT2()` in the library `ICSNP` for Hotelling's T^2 testing.

Main page: **ST 437/537: Applied Multivariate and Longitudinal Data Analysis (https://maityst537.wordpress.ncsu.edu/)**