L14 One sample T^2-test

1. One-sample T^2-test

 (1) α-level test
 For $N(\mu, \Sigma)$ and given $\mu_0 \in R^p$, the following is an α-level test.

$H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ Test Statistic: $T^2 = (\overline{x} - \mu_0)' \left(\frac{S_u}{n}\right)^{-1} (\overline{x} - \mu_0)$ Reject H_0 if $T^2 > T^2_\alpha(p, n-1)$

 $$\textbf{Proof}\quad \begin{aligned} P(\text{Type I error}) &= P(\text{Rejecting } H_0 | H_0 \text{ is true}) = P\left(T^2 > T^2_\alpha(p, n-1) | \mu = \mu_0\right) \\ &= P(T^2(p, n-1) > T^2_\alpha(p, n-1)) = \alpha. \end{aligned}$$

 (2) Observed significance level
 For $N(\mu, \Sigma)$ and given $\mu_0 \in R^p$, p-value defined below is the observed significance level.

$H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ Test Statistic: $T^2 = (\overline{x} - \mu_0)' \left(\frac{S_u}{n}\right)^{-1} (\overline{x} - \mu_0)$ p-value: $P(T^2(p, n-1) > T^2_{ob})$.

 $$\textbf{Proof}\quad \begin{aligned} & \alpha\text{-level test rejects } H_0 \Longleftrightarrow T^2_{ob} > T^2_\alpha(p, n-1) \\ \Longleftrightarrow\ & P(T^2(p, n-1) > T^2_{ob}) < P(T^2(p, n-1) > T^2_\alpha(p, n-1)) = \alpha \\ \Longleftrightarrow\ & P(T^2(p, n-1) > T^2_{ob}) < \text{Significance level} \\ \Longleftrightarrow\ & P(T^2(p, n-1) > T^2_{ob}) \text{ is the observed significance level.} \end{aligned}$$

 (3) Remarks
 In (1) $T^2_\alpha(p, n-1) = \frac{(n-1)p}{n-p} F_\alpha(p, n-p)$.
 In (2) $P(T^2(p, n-1) > T^2_{ob}) = P\left(\frac{(n-1)p}{n-p} F(p, n-p) > T^2_{ob}\right) = P\left(F(p, n-p) > \frac{n-p}{(n-1)p} T^2_{ob}\right)$.
 In both (1) and (2) $T^2 = (\overline{x} - \mu_0) \left(\frac{S}{n-1}\right)^{-1} (\overline{x} - \mu_0)$.

 Ex1: With $\mu_0 = \begin{pmatrix} 9 \\ 6 \end{pmatrix}$ and $n = 3$, for $\alpha = 0.05$, $T^2_\alpha(p, n-1) = \frac{2\times 2}{1} F_{0.05}(2, 1) = 798$. With $T^2_{ob} = 0.7778$, $P(T^2(2, 2) > 0.7778) = P(4F(2, 1) > 0.7778) = 0.8485$. So

$H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ where $\mu_0 = \begin{pmatrix} 9 \\ 6 \end{pmatrix}$ Test Statistic: $T^2 = (\overline{X} - \mu_0)' \left(\frac{S_u}{3}\right)^{-1} (\overline{X} - \mu_0)$ Reject H_0 if $T^2 > 798$ for $\alpha = 0.05$ $T^2_{ob} = 0.7778$ Fail to reject H_0.

$H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$ where $\mu_0 = \begin{pmatrix} 9 \\ 6 \end{pmatrix}$ Test Statistic $T^2 = (\overline{X} - \mu_0)' \left(\frac{S_u}{n}\right)^{-1} (\overline{X} - \mu_0)$ p-value: $P(T^2(2, 2) > T^2_{ob})$ $T^2 = \left(\frac{1}{0.72} - 1\right) \times 2 = 0.7778$; p-value: $P(T^2(2,2) > 0.7778) = P(F(2,1) > 0.19) = 0.8485$ Fail to reject H_0

2. Likelihood ratio test

   (1) Likelihood ratio test (LRT)
   For testing H_0 on θ, LR $= \frac{\max[L(\theta):\,\theta]}{\max[L(\theta):\,\theta \in H_0]}$ is the likelihood ratio.
   By intuition we reject H_0 when LR $> c$. A such test is a LRT.
   If LR is a monotone function of T, then T can be used as a LRT test statistic.
   T should be selected such that the distribution of T under H_0 is known so that
   P(Type I error) $= P(T \in$ Rejection region$|H_0$ is true) can be calculated.

   (2) Test in (1) of 1
   For test in (1) of 1, LR is a decreasing function of Wilk's Lambda $\Lambda = -\frac{n}{2}\ln(\text{LR})$.
   But $\Lambda = \left(1 + \frac{T^2}{n-1}\right)^{-1}$ is a decreasing function of $T^2 = (\overline{x} - \mu_0)\left(\frac{S_u}{n}\right)^{-1}(\overline{x} - \mu_0)$.
   Also, $T^2|H_0 \sim T^2(p,\, n-1)$. Thus test in (1) of 1 is an α-level LRT.

3. Implementation

   (1) SAS output and T^2
   For the test in Ex1, SAS produces

Statistics	Value	F-value	Num DF	Den DF	Pr>F
Wilks' Lambda	0.7200	0.19	2	1	0.8485
Pillai's Trace	0.2800	0.19	2	1	0.8485
Hotellig-Lawley Trace	0.3889	0.19	2	1	0.8485
Roy's Greatest Root	0.3889	0.19	2	1	0.8485

   If using rejection region, one still need to find $T^2_\alpha(p,\, n-1) = \frac{(n-1)p}{n-p}F_\alpha(p,\, n-p)$.
   But with $\Lambda_{ob} = 0.72$, $T^2_{ob} = \left(\frac{1}{\Lambda_{ob}-1}\right)^{-1}(n-1) = 0.7778$.
   T^2_{ob} can also be calculated from the other three statistics:
   Pillai's trace$= 1 - \Lambda = 1 - \left(1 + \frac{T^2}{n-1}\right)^2$. So $T^2 = \left(\frac{1}{1-(\text{ptrace})} - 1\right)(n-1)$.
   Hotelling-Lawley Trace=Roy's Greatest Root$= \frac{T^2}{n-1}$. So $T^2 = (Hotelling - Lawleytrace)(n-1)$.

   (2) Observed significance level
   If using p-value, then the computation result is directly displayed.
   p-value: $P(T^2(p, n-1) > T^2_{ob}) = P(F(p, n-p) > F_{ob}) = P(F(2,1) > 0.19) = 0.8485$.

   (3) SAS code

```
data a;
   infile "D://example.txt";
   input x1 x2 @@;
   y1=x1-6; y2=x2-9;

proc reg;
   model y1 y2=/noprint;
   mtest intercept;
   run;
```

L15: Multivariate multiple linear regression

1. Multivariate multiple linear regression model

   (1) Multivariate multiple linear regression model
   $$y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} \beta_{11} \\ \vdots \\ \beta_{1p} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} \beta_{q1} \\ \vdots \\ \beta_{qp} \end{pmatrix} x_q + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_p \end{pmatrix} = \beta_1 x_1 + \cdots + \beta_q x_q + \epsilon \text{ with } \epsilon \sim N(0,\, \Sigma)$$
   is a multivariate multiple linear regression model.

   (2) Parameters
   Let $B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1p} \\ \vdots & \ddots & \vdots \\ \beta_{q1} & \cdots & \beta_{qp} \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}$. Then the model in (1) can be written as $y = B'x + \epsilon$.
   This model has two parameter matrices $B \in R^{q \times p}$ and $\Sigma \in R^{p \times p}$.
   The model is multivariate multiple linear regression model since Multivariate: The response $y \in R^p$ is multivariate with components $y_1, ..., y_p$.
   Multiple: The model has multiple predictors $x_1, ..., x_q$ in $x \in R^q$.
   Linear: The regression $E(y) = B'x$ is a linear function of unknown parameter matrix $B \in R^{q \times p}$.

   (3) Infinite number of populations
   $y = B'x + \epsilon \sim N(B'x,\, \Sigma)$ represents infinite number of p-variant normal populations indexed by $x \in R^q$. With different values of x, the model represents different populations. These populations share a common but unknown variance-covariance matrix Σ.

   (4) A relation
   The model implies p univariate multiple regression models for each component of y.
   The regression for y_i is $y_i = \beta_{1i} x_1 + \cdots + \beta_{qi} x_q + \epsilon_i$ with $\epsilon_i \sim N(0,\, \sigma_i^2)$ where regression coefficients $\beta_{1i}, ..., \beta_{qi}$ are from the ith column of B.
   Note that the ith column of B is only associated with the univariate regression for y_i.
   The set of p univariate models does not imply the original multivariate model since the univariate models do not specify $\text{cov}(y_i,\, y_j) = \text{cov}(\epsilon_i,\, \epsilon_j) = \sigma_{ij}$ when $i \neq j$.

   Comment: B can be estimated from p univariate regression models. But Σ must be estimated from the original multivariate regression model.

   Ex1: When one predictor is 1, we have $y = \beta_0 1 + \beta_1 x_1 + \cdots + \beta_{q-1} x_{q-1} + \epsilon$. For this model the regression as a function of $x_1, .., x_{q-1}$, $E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{q-1} x_{q-1}$, has intercept β_0 and hence the model is called the model with intercept. But for model in (1), the regression $E(y) = \beta_1 x_1 + \cdots + \beta_q x_q$ does not have an intercept and hence is the model without intercept.

   Ex2: The model with intercept but without other predictors is $y = \beta_0 + \epsilon \sim N(\beta_0,\, \sigma^2 I_p)$ is a one population model $N(\beta_0,\, \sigma^2 I)$.

2. Samples and distributions

   (1) Data matrices and random error matrix
   From the regression $y = B'x + \epsilon$, we observed $\mathbf{y}_i = B'\mathbf{x}_i + \epsilon_i$, $i = 1, ..., n$.
   Let $Y' = (\mathbf{y}_1, ..., \mathbf{y}_n) \in R^{p \times n}$, $X' = (\mathbf{x}_1, ..., \mathbf{x}_q) \in R^{q \times n}$ and $\mathcal{E} = (\epsilon_1, .., \epsilon_n) \in R^{p \times n}$. Then
   $$Y' = B'X' + \mathcal{E}$$
   where $Y \in R^{n \times p}$ and $X \in R^{n \times q}$ are two data matrices of observed $\mathbf{y}$ and $\mathbf{x}$. But $\mathcal{E} \in R^{p \times n}$ is unobserved random error matrix.

   (2) A relation
   Let e_i be the ith column of I_p. Then
   $$Y' = B'X' + \mathcal{E} \Longleftrightarrow Y = XB + \mathcal{E}' \Longrightarrow Ye_i = XBe_i + \mathcal{E}'e_i, i = 1, ..., p$$

The ith row of Y' is the ith column of Y, Ye_i, contains n observations on the ith component of y, y_i. The ith column of B, Be_i, contains q coefficients for the regression model for y_i. So $Ye_i = XBe_i + \mathcal{E}'e_i$, $i = 1, ..., p$ represents data on p univariate regression models.

(3) Distributions

Note that $\mathcal{E} = (\epsilon_1, .., \epsilon_n) \sim N_{p\times n}(0, \Sigma, I_n) \iff \mathcal{E}' \sim N_{n\times p}(0, I_n, \Sigma)$
$\implies \mathcal{E}'e_i \sim N_{n\times 1}(0, I_n, \sigma_i^2) = N(0, \sigma_i^2 I_n)$.

So $Y' = B'X' + \mathcal{E} \sim N_{p\times n}(B'X', \Sigma, I_n)$, $\quad Y = XB + \mathcal{E}' \sim N_{n\times p}(XB, I_n, \Sigma)$
and $Ye_i = XBe_i + \mathcal{E}'e_i \sim N(XBe_i, \sigma_i^2 I_n)$, $i = 1, ..., p$.

3. MLEs for B and Σ

(1) MLEs for B and Σ

From $Y' \sim N_{p\times n}(B'X', \Sigma, I_n)$ where $\text{rank}(X) = p$,

$$L(B, \Sigma) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\{-\frac{1}{2}\text{tr}[\Sigma^{-1/2}(Y - XB)'(Y - XB)\Sigma^{-1/2}]\}.$$

Let $\widehat{B} = (X'X)^{-1}X'Y$, error matrix $E = (Y - X\widehat{B})'(Y - X\widehat{B}) = Y'[I - X(X'X)^{-1}X']Y$ and $\widehat{\Sigma} = \frac{E}{n}$. Then

$$L(B, \Sigma) \le L(\widehat{B}, \widehat{\Sigma}) = \left(\frac{n}{2\pi e}\right)^{np/2} |E|^{-n/2}.$$

Thus $\widehat{B}$ and $\widehat{\Sigma}$ are MLEs for B and Σ respectively.

(2) Sampling distributions

$\widehat{B} = (X'X)^{-1}X'Y$, $Y \sim N_{n\times p}(XB, I_n, \Sigma) \Longrightarrow \widehat{B} \sim N_{q\times p}(B, (X'X)^{-1}, \Sigma)$.
So $E(\widehat{B}) = B$, i.e., $\widehat{B}$ is an UE for B.

$$\left\{\begin{array}{l} E = Y'AY,\ A = I - X(X'X)^{-1}X' = A' = A^2 \\ Y' \sim N_{p\times n}(B'X', \Sigma, I_n),\ \text{tr}(A) = n - p \\ \text{ and } (XB)[I - X(X'X)^{-1}X'](XB) = 0 \end{array}\right. \Longrightarrow E \sim W_{p\times p}(0, \Sigma, n-p) = W_{p\times p}(\Sigma, n-p).$$

So $E(E) = (n-p)\Sigma$. Thus $E(\widehat{\Sigma}) = \frac{n-p}{n}\Sigma$, i.e., $\widehat{\Sigma}$ is a biased estimator for Σ.

$Y \sim N_{n\times p}(XB, I_n, \Sigma)$ and $(X'X)^{-1}X'IA = 0$, So $\widehat{B}$ and AY are independent. Thus $\widehat{B}$ and E are independent.

(3) Computation for $\widehat{B}$.

$\widehat{B} = (X'X)^{-1}X'Y \iff \widehat{B}e_i = (X'X)^{-1}X'Ye_i$, $i = 1, ..., p$.
So $\widehat{B}$ can be obtained by computing $\widehat{B}e_i$, $i = 1, ..., p$, in the univariate model for the y_i.

Ex3: $y = B'\begin{pmatrix}1\\ x1\\ x2\end{pmatrix} + \epsilon$

```
proc reg;
  model y1 y2=x1 x2;
  run;
```

$\Rightarrow$

y1 parameter	value	y2 parameter	value
intercept	1.111	intercept	-1.11
x1	2.222	x1	-2.22
x2	3.333	x2	-3.33

$\Rightarrow \widehat{B} = \begin{pmatrix}1.111 & -1.11\\ 2.222 & -2.22\\ 3.333 & -3.33\end{pmatrix}$

Ex4: $y = B'\begin{pmatrix}x1\\ x2\end{pmatrix} + \epsilon$

```
proc reg;
  model y1 y2=x1 x2/noint;
  run;
```