L16: t-tests in linear model

1. Recall t-interval in linear model

   (1) Two basic statistics in linear model
   In linear model $y = X\beta + e$, $e \sim N(0, \sigma^2 I_n)$, $X \in R^{n\times p}$ has full column rank. Then the MLE and LSE for β are equal, and is a BLUE for β.

   $$\widehat{\beta} = X^+ y = (X'X)^{-1}X'y \sim N(\beta, \sigma^2(X'X)^{-1}).$$

   The MLE of σ^2 is $\frac{SSE}{n}$, the UE of σ^2 is MSE $= \frac{\text{SSE}}{n-p}$ where SSE $\sim \sigma^2\chi^2(n-p)$ is independent to $\widehat{\beta}$.

   (2) A variable with t-distribution
   Based on $l'\widehat{\beta} \sim N(l'\beta, \sigma^2_{l'\widehat{\beta}})$ and $\frac{\text{SSE}}{\sigma^2} \sim \chi^2(n-p)$, by their independence and the fact that $\sigma^2_{l'\widehat{\beta}} = \sigma^2 l'(X'X)^{-1}l$ is estimated by $s^2_{l'\widehat{\beta}} = \text{MSE}\, l'(X'X)^{-1}l$, it can be seen that

   $$\frac{l'\widehat{\beta} - l'\beta}{s_{l'\widehat{\beta}}} \sim t(n-p).$$

   (3) t-confidence interval
   $l'\widehat{\beta} \pm t_{\alpha/2}(n-p)s_{l'\widehat{\beta}}$ is a $1-\alpha$ CI for $l'\beta$.
   $(-\infty,\ l'\widehat{\beta} + t_\alpha(n-p)s_{l'\widehat{\beta}})$ is a $1-\alpha$ lower-sided CI for $l'\beta$.
   $(l'\widehat{\beta} - t_\alpha(n-p)s_{l'\widehat{\beta}},\ \infty)$ is a $1-\alpha$ upper-sided CI for $l'\beta$.
   Comment: CIs for β_i are special cases with $l = e_i$, the ith column of I_p.

2. t-tests in linear model

   (1) α-level t-tests

   $H_0: l'\beta = c$ versus $H_a: l'\beta \neq c$
   Test statistic: $t = \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}}$
   Reject H_0 if $t < -t_{\alpha/2}(n-p)$ or $t > t_{\alpha/2}(n-p)$

   $H_0: l'\beta \le c$ versus $H_a: l'\beta > c$
   Test statistic: $t = \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}}$
   Reject H_0 if $t > t_\alpha(n-p)$

   $H_0: l'\beta \ge c$ versus $H_a: l'\beta < c$
   Test statistic: $t = \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}}$
   Reject H_0 if $t < -t_\alpha(n-p)$

   Proof Consider the last one.
   By H_a, $l'\beta < c$. Replacing $l'\beta$ by its estimated value, when H_a is true, we expect $l'\widehat{\beta} - c$ to be negative. So it is reasonable to reject H_0 when t is less than a negative number $-d$. We show that with $-d$ in the test scheme, the significance level of the test is no more than α.

   $$\begin{aligned} P(\text{Type I error}) &= P(\text{Rejecting } H_0 \mid H_0 \text{ is true}) \\ &= P\left(\frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}} < -t_\alpha(n-p) \mid l'\beta \ge c\right) \\ &\le P\left(\frac{l'\widehat{\beta}-l'\beta}{s_{l'\widehat{\beta}}} < -t_\alpha(n-p)\right) = P(t(n-p) < -t_\alpha(n-p)) = \alpha. \end{aligned}$$

(2) Observed significance level

$H_0: l'\beta = c$ versus $H_a: l'\beta \neq c$ Test statistic: $t = \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}}$ p-value: $2 \times P(t(n-p) > \lvert t_{0b}\rvert)$

$H_0: l'\beta \leq c$ versus $H_a: l'\beta > c$ Test statistic: $t = \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}}$ p-value: $P(t(n-p) > t_{ob})$	$H_0: l'\beta \geq c$ versus $H_a: l'\beta < c$ Test statistic: $t = \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}}$ p-value: $P(t(n-p) < t_{ob})$

Proof p-value is a probability calculated from observed sample such that α-level test rejects H_0 if and only if p-vlue$< \alpha$. Here α is significance level and p-value is the observed significance level. We show that in the last case the defined p-value is indeed the observed significance level.

$$\begin{aligned} & \alpha\text{-level rejects } H_0 \Longleftrightarrow t_{ob} < -t_\alpha(n-p) \\ \Longleftrightarrow \quad & P(t(n-p) < t_{ob}) < P(t(n-p) < -t_\alpha(n-p)) = \alpha \end{aligned}.$$

Comment: In SAS proc reg will display $\widehat{\beta}_i$, $s_{\widehat{\beta}_i}$, $t = \frac{\widehat{\beta}_i}{s_{\widehat{\beta}_i}}$ and $2 \times P(t(n-p) > |t_{ob}|)$ for $i = 1, .., p$.

3. Relations of $1-\alpha$ CI and α-level tests

(1) Relations

$$\begin{array}{lcl} c \text{ is in a } 1-\alpha \text{ CI for } l'\beta & \Longleftrightarrow & \alpha\text{-level test fails to reject } H_0: l'\beta = c. \\ c \text{ is in a } 1-\alpha \text{ upper-sided CI for } l'\beta & \Longleftrightarrow & \alpha\text{-level test fails to reject } H_0: l'\beta \leq c. \\ c \text{ is in a } 1-\alpha \text{ lower-sided CI for } l'\beta & \Longleftrightarrow & \alpha\text{-level test fails to reject } H_0: l'\beta \geq c. \end{array}$$

Proof We show the last one.

$$\begin{aligned} & c \text{ is in a } 1-\alpha \text{ lower-sided CI for } l'\beta \Longleftrightarrow c \in (-\infty,\ l'\widehat{\beta} + t_\alpha(n-p)s_{l'\widehat{\beta}}) \\ \Longleftrightarrow \quad & c < l'\widehat{\beta} + t_\alpha(n-p)s_{l'\widehat{\beta}} \Longleftrightarrow \frac{l'\widehat{\beta}-c}{s_{l'\widehat{\beta}}} > -t_\alpha(n-p) \\ \Longleftrightarrow \quad & \alpha\text{-level test fails to reject } H_0: l'\beta \geq c. \end{aligned}$$

(2) Smallest confidence coefficient

Confidence intervals with higher confidence coefficient are wider. But the confidence intervals with lower confidence coefficient are narrower. Thus one might be interested in finding the smallest confidence coefficient such that the confidence interval covers c.

Test on $H_0: l'\beta = c$ produced p-value: p, then the smallest confidence coefficient such that c is in the interval is $1-\alpha > 1-p$.

Test on $H_0: l'\beta \leq c$ produced p-value: p, then the smallest confidence coefficient such that c is in the upper-sided interval is $1-\alpha > 1-p$.

Test on $H_0: l'\beta \geq c$ produced p-value: p, then the smallest confidence coefficient such that c is in the lower-sided interval is $1-\alpha > 1-p$.

Proof: We show the last one.

$$\begin{aligned} & c \text{ is in the } 1-\alpha \text{ lower sided CI for } l'\beta \\ \Longleftrightarrow \quad & \alpha\text{-level test on } H_0: l'\beta \geq c \text{ fails to reject } H_0 \Longleftrightarrow p > \alpha \\ \Longleftrightarrow \quad & 1-\alpha > 1-p. \end{aligned}$$

L17: F-test in linear model

1. Model, hypothesis and likelihood ratio

   (1) A linear model
   Linear model $y = X\beta + e$ with $e \sim N(0, \sigma^2 I_n)$ and full column rank $X \in R^{n\times p}$ has two parameters $\beta \in R^p$ and $\sigma^2 > 0$.
   The MLE of β and the LSE of β are equal and is the BLUE $\widehat{\beta} = X^+ y = (X'X)^{-1}X'y \sim N(\beta, \sigma^2(X'X)^{-1})$.
   The SS (Sum of squares) from the error is
   $\text{SSE} = \|y - X\widehat{\beta}\|^2 = y'(I - XX^+)y \sim \sigma^2\chi^2(n-p)$ where $n - p = \text{rank}(I - XX^+) = \text{DF}$.
   The MLE of σ^2 is $\widehat{\sigma}^2 = \frac{\text{SSE}}{n}$ and $\text{MSE} = \frac{\text{SSE}}{n-p}$ is an UE for σ^2.
   SSE and $\widehat{\beta}$ are independent. $L(\beta, \sigma^2) \le L(\widehat{\beta}, \widehat{\sigma}^2) = \left(\frac{n}{2\pi e}\right)^{n/2} \text{SSE}^{-n/2}$.

SSE	DF	MSE
$y'(I - XX^+)y$	$n - p$	SSE/(n-p)

   (2) A hypothesis and a reduced model
   Let $H_0 : H\beta = b \in R^q$ be a consistent hypothesis where $H \in R^{q\times p}$ has full row rank q.
   $H\beta = b \Longrightarrow \beta = H^+b + \mathcal{N}(H) = H^+b + \mathcal{R}(I - H^+H)$.
   Thus under H_0 the model is reduced to $y - XH^+b = X(I - H^+H)\gamma + e$.
   $y = X\beta + e \Longrightarrow y = X[H^+b + (I - H^+H)\gamma] + e \Longrightarrow y - XH^+b = X(I - H^+H)\gamma + e$.
   By (1), $\text{SSE}_r = (y - XH^+b)'\{I - [X(I - H^+H)][X(I - H^+H)]^+\}(y - XH^+b)$ with DF$= n - \text{rank}[X(I_p - H^+H)] = n - (p - q)$, and $\max[L(\beta, \sigma^2) : H_0] = \left(\frac{n}{2\pi e}\right)^{n/2} \text{SSE}_r^{-n/2}$.

SSE_r	DF
$(y - XH^+b)'\{I - [X(I - H^+H)][X(I - H^+H)]^+\}(y - XH^+b)$	$n - (p - q)$

   (3) Likelihood raio
   $\text{LR} = \frac{\max[L(\beta,\sigma^2):]}{\max[L(\beta,\sigma^2):H_0]} = \left(\frac{\text{SSE}_r}{\text{SSE}}\right)^{n/2}$ is an increasing function of $\frac{\text{SSE}_r}{\text{SSE}}$.
   $\text{SSE} = y'(I - XX^+)y = (y - XH^+b)'(I - XX^+)(y - XH^+b)$.

2. An SS table and LRT

   (1) SSH
   Let SSH be the difference between SSE_r and SSE caused by the hypothesis H_0
   $\text{SSH} = SSE_r - SSE = (y - XH^+b)'\{XX^+ - [X(I - H^+H)][X(I - H^+H)]^2\}(y - XH^+b)$,
   $$\begin{aligned} \text{DF} &= \text{rank}\{XX^+ - [X(I - H^+H)][X(I - H^+H)]^+\} \\ &= \text{rank}(X) - \text{rank}[X(I - H^+H)] = p - \text{rank}(I_p - H^+H) \\ &= p - [p - \text{rank}(H)] = p - (p - q) = q. \end{aligned}$$

   (2) An SS table

SS	DF	MS	F
SSH	q	MSH	MSH/MSE
SSE	n-p	MSE	
SSE_r	n-p+q		

   where $SSE_r = SSH + SSE$ and (DF of SSE_r)=(DF of SSH)+(DF of SSE).

(3) LRT
Likelihood ratio is an increasing function of
$\frac{SSE_r}{SSE} = \frac{SSE_r - SSE}{SSE} + 1 = 1 + \frac{SSH}{SSE} = 1 + \frac{MSH \cdot q}{MSE \cdot (n-p)} = 1 + \frac{q}{n-p} F$.
Thus LR is an increasing function of $F = MSH/MSE$. So

$$\boxed{\begin{array}{l} H_0 : H\beta = b \text{ versus } H_a : H\beta \neq b \\ \text{Test statistic } F = \frac{MSH}{MSE} \\ \text{Reject } H_0 \text{ if } F > c \end{array}}$$

is a LRT scheme.

3. Expression of SSH

(1) Let $M = [X(I - H^+H),\, X(X'X)^{-1}H']$. Then $XX^+ = MM^+$.

Proof Note $M = X[I - H^+H,\, (X'X)^{-1}H']$. So $\mathcal{R}(M) \subset \mathcal{R}(X)$. But

$$\begin{aligned} \dim[\mathcal{R}(M)] &= \operatorname{rank}(M) = \operatorname{rank}[X(I - H^+H)] + \operatorname{rank}[X(X'X)^{-1}H'] \\ &= \operatorname{rank}(I_p - H^+H) + \operatorname{rank}(H') = p - \operatorname{rank}(H) + \operatorname{rank}(H') \\ &= p = \operatorname{rank}(X) = \dim[\mathcal{R}(X)]. \end{aligned}$$

Thus $\mathcal{R}(X) = \mathcal{R}(M)$. So are their projection matrices. Therefore $XX^+ = MM^+$.

(2) Let $M_1 = X(I - H^+H)$ and $M_2 = X(X'X)^{-1}$. Then $XX^+ = M_1M_1^+ + M_2M_2^+$

Proof $M = (M_1,\, M_2)$. But $M_1'M_2 = 0$. So $M^+ = \begin{pmatrix} M_1^+ \\ M_2^+ \end{pmatrix}$.

Thus $XX^+ = MM^+ = M_1M_1^+ + M_2M_2^+$.

(3) SSH$= [M_2'(y - XH^+b)]'(M_2'M)^+[M_2'(y - XH^+b)]$

Proof Let $z = y - XH^+b$.

$$\begin{aligned} SSH &= SSE_r - SSE = z'(XX^+ - M_1M_1^+)z = z'M_2M_2^+z \\ &= z'M_2(M_2'M_2)^+M_2'z = (M_2'z)'(M_2'M_2)^+(M_2'z) \\ &= [M_2'(y - XH^+b)]'(M_2'M)^+[M_2'(y - XH^+b)]. \end{aligned}$$