**L16: $l'\beta$ in regression**

1. Confidence interval for $E(y)$

   (1) Confidence interval for $E[y(x_0)]$.
   For regression $y \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1},\, \sigma^2)$ with data $Y \sim N(X\beta,\, \sigma^2 I_n)$, $E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1}$ is the regression function. With $x_0 = (1,\, x_{01}, ...x_{0,p-1})'$, $E[y(x_0)] = \beta_0 + \beta_1 x_{01} + \cdots + \beta_{p-1}x_{0,p-1} = x_0'\beta$ is of the type of $l'\beta$ and has $1-\alpha$ CI

   $$E[y(x_0)] = x_0'\beta \in x_0'\widehat{\beta} \pm t_{\alpha/2}(n-p)S_{x_0'\widehat{\beta}} = \widehat{y}(x_0) \pm t_{\alpha/2}(n-p)S_{\widehat{y}(x_0)}$$

   where $\widehat{y}(x_0) = x_0'\widehat{\beta} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_{p-1}x_{0,p-1}$, $\widehat{\beta} = X^+Y = (X'X)^{-1}X'Y$, $S^2_{\widehat{y}(x_0)} = S^2_{x_0'\widehat{\beta}} = \text{MSE }x_0'(X'X)^{-1}x_0$.

   (2) Computation by SAS

   **Ex1:** Suppose for $y \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2,\, \sigma^2)$ we need 90% confidence interval for $E[y(x_0)]$ where $x_0 = (1,\, 3,\, -2)'$.

```
data a; infile "D:\ex.dat"; input y x1 x2;
data b; input y x1 x2; datalines;
. 3 -2
;
data c; set a b;
proc reg;
     model y=x1 x2/p alpha=0.10 clm;
     run;
```

   The output displays $y_i$, $\widehat{y}_i$, $y_i - \widehat{y}_i$, $S_{\widehat{y}_i}$ and 90% CI for $E(y_i)$ for all $i = 1, ..., n$. For $x_0 = (1,\, 3,\, -2)'$, $\widehat{y}(x_0)$, $S_{\widehat{y}(x_0)}$ and 90% CI for $E[y(x_0)]$ are displayed.

2. Prediction intervals

   (1) Definitions
   Two different concepts
   Suppose $y_f = y(x_0)$ is a future response with mean $E[y(x_0)] = x_0'\beta$ where vector $x_0$ is given but $y(x_0)$ has not been observed yet. Suppose $L < U$ are two statistics and we predict that $y(x_0) \in (L,\, U)$. Then $(L,\, U)$ is called a prediction interval for $y(x_0)$. If $P(L < y(x_0) < U) \geq 1-\alpha$, then $(L,\, U)$ is a prediction interval for $y(x_0)$ with confidence coefficient $1-\alpha$.

   (2) Predictors and estimators
   Recall Statistic $\widehat{y}$ is an UP for $y(x_0) \Longleftrightarrow$ Statistic $\widehat{y}$ is an UE for $E[y(x_0)] = x_0'\beta$.
   If $y_f = y(x_0)$ is independent to the data vector $Y$, then
   Statistic $\widehat{y}$ is a BLUP for $y(x_0) \Longleftrightarrow$ Statistic $\widehat{y}$ is a BLUE for $E[y(x_0)] = x_0'\beta$.

   (3) Prediction interval
   Suppose $y_f = y(x_0)$ is independent to data vector $Y$, then

   $$y(x_0) \in \widehat{y}(x_0) \pm t_{\alpha/2}(n-p)S_{y(x_0)-\widehat{y}(x_0)}$$

   is a $1-\alpha$ PI for $y(x_0)$ where $\widehat{y}(x_0) = x_0'\widehat{\beta} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_{p-1}x_{0,p-1}$, $\widehat{\beta} = (X'X)^{-1}X'Y$, $S^2_{y(x_0)-\widehat{y}(x_0)} = S^2_{y(x_0)} + S^2_{\widehat{y}(x_0)} = \text{MSE }\left[1 + x_0'(X'X)^{-1}x_0\right]$.

**Proof.** $y(x_0) \sim N(x_0'\beta,\, \sigma^2)$ and $\widehat{y}(x_0) = x_0'\widehat{\beta} \sim N(x_0'\beta,\, \sigma^2 x_0'(X'X)^{-1}x_0)$ are independent. So $y(x_0) - \widehat{y}(x_0) \sim N(0,\, \sigma^2 + \sigma^2 x_0'(X'X)^{-1}x_0)$ has the variance $\sigma^2_{y(x_0)-\widehat{y}(x_0)} = \sigma^2\left[1 + x_0'(X'X)^{-1}x_0\right]$ estimated by $S^2_{y(x_0)-\widehat{y}(x_0)} = \text{MSE}\left[1 + x_0'(X'X)^{-1}x_0\right]$. Here $S^2_{y(x_0)-\widehat{y}(x_0)} = \sigma^2_{y(x_0)-\widehat{y}(x_0)}\,\dfrac{\text{MSE}}{\sigma^2}$.

Note that $\frac{y(x_0)-\widehat{y}(x_0)}{\sigma_{y(x_0)-\widehat{y}(x_0)}} \sim N(0,\, 1^2)$ and $\dfrac{\text{SSE}}{\sigma^2} \sim \chi^2(n-p)$ are independent.
Thus $t = \dfrac{\frac{y(x_0)-\widehat{y}(x_0)}{}}{\sigma_{y(x_0)-\widehat{y}(x_0)}\sqrt{\frac{\text{SSE}}{\sigma^2(n-p)}}} \sim t(n-p)$, i.e., $\frac{y(x_0)-\widehat{y}(x_0)}{S_{y(x_0)-\widehat{y}(x_0)}} \sim t(n-p)$. Therefore

$$
\begin{aligned}
1-\alpha &= P(-t_{\alpha/2}(n-p) < t(n-p) < t_{\alpha/2}(n-p)) \\
&= P\left(-t_{\alpha/2}(n-p) < \tfrac{y(x_0)-\widehat{y}(x_0)}{S_{y(x_0)-\widehat{y}(x_0)}} < t_{\alpha/2}(n-p)\right) \\
&= P\left(\widehat{y}(x_0) - t_{\alpha/2}(n-p)S_{y(x_0)-\widehat{y}(x_0)} < y(x_0) < \widehat{y}(x_0) + t_{\alpha/2}(n-p)S_{y(x_0)-\widehat{y}(x_0)}\right).
\end{aligned}
$$

Hence $\widehat{y}(x_0) \pm t_{\alpha/2}(n-p)S_{y(x_0)-\widehat{y}(x_0)}$ is a $1-\alpha$ PI for $y(x_0)$.

(4) SAS

**Ex2:** For the model and $y(x_0)$ in Ex1, find 90% prediction interval for $y(x_0)$.

```
proc reg;
     model y=x1 x2/p alpha=0.10 cli;
     run;
```

The output displays $y_i$, $\widehat{y}_i$, $y_i - \widehat{y}_i$, $S_{\widehat{y}_i}$ and 90% PI for $y_i$ for all $i = 1, ..., n$. For $x_0 = (1,\, 3,\, -2)'$, $\widehat{y}(x_0)$, $S_{\widehat{y}(x_0)}$ and 90% PI for $y(x_0)$ are displayed.

3. $F$-test on $l'\beta$

(1) $F$-test on $l'\beta$
To implement the test

$H_0 : l'\beta = 0$ vs $H_a : l'\beta \neq b$
Test Statistic: $F = \dfrac{(l'\widehat{\beta}-b)'[l'(X'X)^{-1}l]^{-1}(l'\widehat{\beta}-b)}{\text{MSE}}$
$p$-value: $P(F(1,\, n-p) > F_{ob})$

we need a computation table

| | MS | DF | F | p |
|---|---|---|---|---|
| Numerator | $(l'\widehat{\beta}-b)'[l'(X'X)^{-1}l]^{-1}(l'\widehat{\beta}-b)$ | 1 | $F_{ob}$ | $p$-value |
| Denominator | MSE | $n-p$ | | |

(2) SAS

**Ex3:** Suppose we need to test $H_0 : 2\beta_0 - 3\beta_1 + \beta_2 = -2$ vs $H_a : 2\beta_0 - 3\beta_1 + \beta_2 \neq -2$. The output of SAS code below will display the computation table

```
proc reg;
     model y=x1 x2;
     test 2*intercept-3*x1+x2=-2;
     run;
```

**L17: ANOVA table**

1. SSE

   (1) Model M
   For model M: $Y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_n)$, $\beta$ is estimated by its LSE $\widehat{\beta}$ that satisfies $\|Y - X\widehat{\beta}\|^2 \le \|Y - X\beta\|^2$ for all $\beta$. Then $E(Y) = X\beta$ is estimated by its BLUE $\widehat{Y} = X\widehat{\beta} = XX^+Y$. So $\|Y - \widehat{Y}\|^2 = \|Y - X\widehat{\beta}\|^2 = \|Y - XX^+Y\|^2$ is minimized $\|Y - X\beta\|^2$.

   (2) Notation
   $\|Y - \widehat{Y}\|^2 = \sum_i (y_i - \widehat{y}_i)^2$ is a sum of squares (SS). This SS measures the error of the Model M and hence is denoted as SSE. So SSE is the variation in $Y$ unexplained by the Model M.

   (3) DF
   SSE $= \|Y - XX^+Y\|^2 = \|(I - XX^+)Y\|^2 = Y'(I - XX^+)Y$ is a quadratic form of $Y$ with matrix $I - XX^+$. The rank of this matrix is called the DF of SSE. But $\text{rank}(I - XX^+) = n - \text{rank}(X) = n - r$. So we have

| Source | SS | DF |
|---|---|---|
| Error | SSE$= Y'(I - XX^+)Y$ | $n - r$ |

   **Ex1:** For model $Y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2\Sigma)$, $\beta$ is estimated by its GLSE that satisfies $\|Y - X\widehat{\beta}\|^2_{\Sigma^{-1}} \le \|Y - X\beta\|^2_{\Sigma^{-1}}$ for all $\beta$. Then $E(Y) = X\beta$ is estimated by its BLUE $\widehat{Y} = X\widehat{\beta} = X(\Sigma^{-1/2}X)^+\Sigma^{-1}Y$. Thus

$$\begin{aligned}\text{SSE} &= \|Y - \widehat{Y}\|^2_{\Sigma^{-1}} = \|Y - X(\Sigma^{-1/2}X)^+\Sigma^{-1/2}Y\|^2_{\Sigma^{-1}} \\ &= \|(\Sigma^{-1/2}Y) - (\Sigma^{-1/2}X)(\Sigma^{-1/2}X)(\Sigma^{-1/2}Y)\|^2 \\ &= (\Sigma^{-1}Y)'\left[I - (\Sigma^{-1/2}X)(\Sigma^{-1/2}X)^+\right](\Sigma^{-1/2}Y).\end{aligned}$$

   with DF$= n - r$.

2. ANOVA table for Model M with $\mathcal{R}(1_n) \subset \mathcal{R}(X)$.

   (1) C.SSTO
   $\mathcal{R}(1_n) \subset \mathcal{R}(X)$ implies that Model $M_1 : Y = 1_n\mu + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_n)$ is a special case of Model M. Thus there exists a hypothesis $H_0$ under which the reduced model is Model $M_1$. For this reduced model $\text{SSE}_0 = \|Y - 1_n\widehat{\mu}\|^2 = \|Y - 1_n\overline{Y}\|^2 = \|Y - 1_n 1_n^+ Y\|^2$.
   $\text{SSE}_0 = \|Y - 1_n\overline{Y}\|^2 = \sum_i (y_i - \overline{Y})^2$ is the CSS of $Y$.
   CSS measures the total variation in $Y$ and is denoted as C.SSTO.
   C.SSTO$= \|Y - 11^+Y\|^2 = Y'(I - 11^+)Y$ with DF $\text{rank}(I - 11^+) = n - 1$.

   (2) SSD
   The difference between the estimated mean of $Y$ in the Models M and $M_1$ is
   $\text{SSD} = \|XX^+Y - 11^+Y\|^2 = \|(XX^+ - 11^+)Y\|^2$.
   The symmetric matrix $XX^+ - 11^+$ is idempotent as shown below.
   $\text{R}(1) \subset \mathcal{R}(X) \Longrightarrow 1_n = Xh$ for some $h \Longrightarrow XX^+11^+ = XX^+Xh1^+ = Xh1^+ = 11^+$.
   So $11^+XX^+ = (XX^+11^+)' = (11^+)' = 11^+$. Thus
   $$(XX^+ - 11^+)(XX^+ - 11^+) = XX^+ - 11^+ - 11^+ + 11^+ = XX^+ - 11^+.$$
   Therefore $\text{SSD} = \|(XX^+ - 11^+)Y\|^2 = Y'(XX^+ - 11^+)Y$.

   (3) SSM
   Note that

$$\begin{aligned} &\text{Total variation in } Y - \text{Variation unexplained by Model M} = \text{C.SSTO-SSE} \\ = \ & Y'(I - 11^+)Y - Y'(I - XX^+)Y = Y'(XX^+ - 11^+)Y = \text{SSD}. \end{aligned}$$

Tus SSD is the variation in $Y$ explained by Model M and hence is denoted as SSM. Clearly the DF of SSM is rank$(XX^+ - 11^+) = r - 1$. So we have ANOVA table

| Source | SS | DF |
|---|---|---|
| Model | SSM= $Y'(XX^+ - 11^+)Y$ | $r-1$ |
| Error | SSE= $Y'(I - XX^+)Y$ | $n-r$ |
| C.Total | C.SSTO= $Y'(I - 11^+)Y$ | $n-1$ |

**Ex2:** For $M_1 : Y = 1_n\mu + \epsilon$, $\epsilon \sim N(0, \sigma^2\Sigma)$,

$$\begin{aligned} \text{SSE}_0 &= \|Y - 1_n\widehat{\mu}\|^2_{\Sigma^{-1}} = \|Y - 1_n(\Sigma^{-1/2}1_n)^+(\Sigma^{-1/2}Y)\|^2_{\Sigma^{-1}} \\ &= \|(\Sigma^{-1/2}Y) - (\Sigma^{-1/2}1)^+(\Sigma^{-1/2}1)(\Sigma^{-1/2}Y)\|^2 \\ &= (\Sigma^{-1/2}Y)'\left[I - (\Sigma^{-1/2}1_n)(\Sigma^{-1/2}1_n)^+\right](\Sigma^{-1/2}Y). \end{aligned}$$

So one can have ANOVA table

| Source | SS | DF |
|---|---|---|
| Model | SSM= $Y'\left[(\Sigma^{-1/2}X)(\Sigma^{-1/2}X)^+ - (\Sigma^{-1/2}1)(\Sigma^{-1/2}1)^+\right](\Sigma^{-1/2}Y)$ | $r-1$ |
| Error | SSE= $(\Sigma^{-1/2}Y)'\left[I - (\Sigma^{-1/2}X)(\Sigma^{-1/2}X)^+\right](\Sigma^{-1/2}Y)$ | $n-r$ |
| C.Total | C.SSTO= $(\Sigma^{-1/2}Y)'\left[I - (\Sigma^{-1/2}1)(\Sigma^{-1/2}1)^+\right](\Sigma^{-1/2}Y)$ | $n-1$ |

3. ANOVA table for Model M, $\mathcal{R}(1_n) \not\subset \mathcal{R}(X)$

(1) U.SSTO

Model M $Y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_n)$ where $\mathcal{R}(1_n) \not\subset \mathcal{R}(X)$, under $H_0 : \beta = 0$ is reduced to Model $M_0 : Y = 0 + \epsilon$, $\epsilon \sim N(0, \sigma^2 I_n)$. For this reduced model $\text{SSE}_0 = \|Y - 0\|^2 = \sum_i y_i^2$ is the uSS of $Y$.

This uss gives the total variation in $Y$ and is denoted as u.SSTO.

C.SSTO= $\|Y\|^2 = Y'I_nY$ has DF rank$(I) = n$.

(2) SSD

The difference between the estimated mean of $Y$ in the Models M and $M_0$ is

SSD $= \|XX^+Y - 0\|^2 = \|XX^+Y\|^2 = Y'XX^+Y$. Note that

$$\begin{aligned} &\text{Total variation in } Y - \text{Variation unexplained by Model M} = \text{U.SSTO-SSE} \\ = \ & Y'I_nY - Y'(I - XX^+)Y = Y'XX^+Y = \text{SSD}. \end{aligned}$$

Thus SSD is the variation in $Y$ explained by Model M and hence is denoted as SSM. Clearly the DF of SSM is rank$(XX^+) = r$. So we have ANOVA table

| Source | SS | DF |
|---|---|---|
| Model | SSM= $Y'XX^+Y$ | $r$ |
| Error | SSE= $Y'(I - XX^+)Y$ | $n-r$ |
| C.Total | U.SSTO= $Y'I_nY$ | $n$ |

**Ex3:** The regression model without intercept, $y = \beta_1x_1 + \cdots + \beta_px_p + \epsilon$, $\epsilon \sim N(0, \sigma^2)$ has data $Y \sim N(X\beta, \sigma^2 I_n)$ where $\mathcal{R}(1_n) \not\subset \mathcal{R}(X)$. So it has ANOVa table in (2).