L10: ANOVA table

1. ANOVA table: variation in y
   $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$ –(1) and $y = \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$–(2)
   with sample is unified as $y \sim N(X\beta, \sigma^2 I)$.

For model (1)

Source	SS	DF
Model	SSM	k
Error	SSE	$n-(k+1)$
C.Total	SSTO	$n-1$

For model (2)

Source	SS	DF
Model	SSM	k
Error	SSE	$n-k$
U.Total	SSTO	n

(1) SSE
$\|y-\widehat{y}\|^2 = \|y - X\widehat{\beta}\|^2 = \|y-Hy\|^2 = y'(I-H)y = \sum(y_i - \widehat{y}_i)^2 = \sum(e_i^2) = SSE$
(Sum of squares of Error) is the variation in y unexplained by $x_1, ..., x_k$.
(DF of SSE)= rank$(I-H) = \text{tr}(I-H) = n-p$.
$H = X(X'X)^{-1}X' \in R^{n\times n}$, $X \in R^{n\times p}$ and $p = \begin{cases} k+1 & (1) \\ k & (2) \end{cases}$

(2) SSTO
SSE after removing $x_1, ..., x_k$ out is SSTO, the total variation in y.
For (1) SSTO $= y'\left(I - \frac{1_n 1_n'}{n}\right) y = \sum(y_i - \overline{y})^2 = CSS$ (Corrected Sum of Squares of y)
with $DF = \text{rank}\left(I - \frac{1_n 1_n'}{n}\right) = \text{tr}\left(I_n - \frac{1_n 1_n'}{n}\right) = n-1$.
For (2) SSTO $= y' I_n y = \sum y_i^2 = USS$ (Uncorrected Sum of Squares of y) with
DF $= \text{rank}(I_n) = \text{tr}(I_n) = n$.

(3) SSM
SSM $= SSTO - SSE$ is the variation in y explained by $x_1, ..., x_k$
For (1), $SSM = y'\left(H - \frac{1_n 1_n'}{n}\right) y = \sum(\widehat{y}_i - \overline{y})^2$ with
DF $= \text{rank}\left(H - \frac{1_n 1_n'}{n}\right) = \text{tr}\left(H - \frac{1_n 1_n'}{n}\right) = p-1 = k$
For (2), $SSM = y'Hy = \sum \widehat{y}_i^2$ with DF $= \text{rank}(H) = \text{tr}(H) = p = k$

2. F-distribution in ANOVA table
   Consider $H_0 : \beta_1 = \cdots = \beta_k = 0$

(1) Recall: $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$.

(2) $\frac{SSM}{\sigma^2} \overset{H_0}{\sim} \chi^2(k)$

Proof. Show (1) only. For (1) $\frac{SSM}{\sigma^2} = y'Ay$ where $A = \frac{H - \frac{1_n 1_n'}{n}}{\sigma^2}$, $y \overset{H_0}{\sim} N(1_n\beta_0, \sigma^2 I)$.
Now $A\sigma^2 A = A$, $(1_n\beta_0)'A(1_n\beta_0) = 0$ and $\text{tr}(A\sigma^2 I) = \text{tr}\left(H - \frac{1_n 1_n'}{n}\right) = k$.
So $\frac{SSM}{\sigma^2} \overset{H_0}{\sim} \chi^2(k)$.

(3) $F = \frac{MSM}{MSE} = \frac{SSM/k}{SSE/(n-p)} \overset{H_0}{\sim} F(k, n-p)$.

Proof. Show (1) only. $\frac{SSM}{\sigma^2} \overset{H_0}{\sim} \chi^2(k)$; $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$.
$SSM = y'Ay$ where $A = H - \frac{1_n 1_n'}{n}$, $SSE = y'By$ where $B = I - H$.

But $y \sim N(X\beta,\, \sigma^2 I_n)$ and $A\sigma^2 IB = 0$. So SSM and SSE are independent.
Thus $F = \frac{SSM/\sigma^2}{k} \div \frac{SSE/\sigma}{n-p} \overset{H_0}{\sim} F(k,\, n-p)$, i.e., $F = \frac{MSM}{MSE} \sim F(k,\, n-p)$ under H_0.

3. Global F-test in ANOVA table

(1) α-level LRT
Below is an α-level LRT.

$H_0:\ \beta_i = 0$ for all $i = 1, ..., k$ versus $H_a:\ \beta_i \neq 0$ for some $i = 1, ..., k$
Test Statistic: $F = \frac{MSM}{MSE}$
Reject H_0 if $F > F_\alpha(k,\, n-p)$

Proof. $\Lambda = \frac{\max[L(\beta,\sigma^2)\colon \beta,\sigma^2]}{\max[L(\beta,\sigma^2)\colon H_0]} = \frac{\left(\frac{n}{2\pi e}\right)^{n/2} SSE^{-n/2}}{\left(\frac{n}{2\pi e}\right)^{n/2} SSTO^{-n/2}} = \left(\frac{SSTO}{SSE}\right)^{n/2} = \left(1 + \frac{SSM}{SSE}\right)^{n/2}$

$= \left(1 + F \cdot \frac{k}{n-p}\right)^{n/2}$ is an increasing function of $F = \frac{MSM}{MSE}$.
Thus a test that rejects H_0 for large value of F is a LRT.

$$\begin{aligned} P(\text{Type I error}) &= P(\text{Rejecting } H_0 | H_0 \text{ is true}) = P(F > F_\alpha(k,\, n-p) | H_0 \text{ is ture}) \\ &= P(F(k,\, n-p) > F_\alpha(k,\, n-p)) = \alpha. \end{aligned}$$

(2) Test using p-value

$H_0:\ \beta_i = 0$ for all $i = 1, ..., k$ versus $H_a:\ \beta_i \neq 0$ for some $i = 1, ..., k$.
Test Statistic: $F = \frac{MSM}{MSE}$
p-value: $P(F(k,\, n-p) > F_{ob})$

Comment:

$$\begin{aligned} \text{Rejecting } H_0 \text{ at } \alpha &\iff F_{ob} > F_\alpha(k,\, n-p) \\ &\iff P(F(k,\, n-p) > F_{ob}) < P(F(k,\, n-p) > F_\alpha(k,\, n-p)) \\ &\iff p\text{-value} < \alpha. \end{aligned}$$

Ex: With ANOVA on p105, write out the model and test its usefulness

Source	DF	SS	MS	F	$Pr > F$
Model	2	8768.754	4384.38	46.7715	< 0.0001
Error	22	2062.286	93.74		
C.Total	24	10831.040			

(i) Model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, $\epsilon \sim N(0,\, \sigma^2)$.

(ii) $H_0:\ \beta_i = 0$ for all $i = 1, 2$ vs $H_a:\ \beta_i \neq 0$ for some $i = 1, 2$
Test statistic: $F = \frac{MSM}{MSE}$
p-value: $P(F(2, 22) > F_{ob})$

$F = 46.78$, p-value: $P(F(2, 22) > 46.78) < 0.0001$
Reject H_0. The model is useful.

L11 t-intervals

1. Coefficient of determination

   (1) Coefficient of determination
   $R^2 = \frac{SSM}{SSTO}$ gives proportion of total variation in y explained by $x_1, ..., x_k$.
   $0 \le R^2 \le 1$. The large the R^2, the better the model.
   $F = \frac{MSM}{MSE} = \frac{SSM}{SSE} \cdot \frac{n-p}{k} = \frac{SSM}{SSTO-SSM} \cdot \frac{n-p}{k} = \frac{R^2}{1-R^2} \cdot \frac{n-p}{k}$.

   (2) Adjusted R^2
   For given y by increasing k, SSM is increasing. Consequently R^2 is increasing.
   To penalize increasing k adjusted coefficient of determination is proposed.
   $R^2 = \frac{SSM}{SSTO} = SSTO - SSESSTO = 1 - \frac{SSE}{SSTO}$.
   Define $R_a^2 = 1 - \frac{MSE}{MSTO}$ where $MSE = \frac{SSE}{n-p}$ and $MSTO = \begin{cases} \frac{SSTO}{n-1} & \text{With intercept} \\ \frac{SSTO}{n} & \text{Without intercept} \end{cases}$

$$\begin{aligned} F &= \frac{MSM}{MSE} = \frac{SSM}{MSE} \cdot \frac{1}{k} = \frac{SSTO-SSE}{MSE} \cdot \frac{1}{k} = \frac{MSTO \cdot m - MSE \cdot (n-p)}{MSM} \frac{1}{k} = \frac{m-(n-p)(1-R_a^2)}{1-R_a^2} \frac{1}{k} \\ &= \frac{m/k}{1-R_a^2} - \frac{n-p}{k} \end{aligned}$$

   where $m = (\text{DF of SSTO}) = \begin{cases} n-1 & \text{With intercept} \\ n & \text{Without intercept} \end{cases}$.

   (3) Caution: Adjusted R^2 could assume a negative value!

2. t-intervals

   (1) Framework
   If $\frac{\widehat{\theta}-\theta}{s} \sim t(df)$, then

$\widehat{\theta} \pm t_{\alpha/2}(df)s$	is $1-\alpha$ C.I. for θ
$(-\infty, \widehat{\theta} + t_\alpha(df)s)$	is $1-\alpha$ lower sided C.I. for θ
$(\widehat{\theta} - t_\alpha(df)s, \infty)$	is $1-\alpha$ upper-sided C.I. for θ

   (2) t-intervals for β_i
   From $\frac{\widehat{\beta}_i-\beta_i}{S_{\widehat{\beta}_i}} \sim t(n-p)$

$\widehat{\beta}_i \pm t_{\alpha/2}(n-p)S_{\widehat{\beta}_i}$	is $1-\alpha$ C.I. for β_i
$(-\infty, \widehat{\beta}_i + t_\alpha(n-p)S_{\widehat{\beta}_i})$	is $1-\alpha$ lower sided C.I. for β_i
$\widehat{\beta}_i - t_\alpha(n-p)S_{\widehat{\beta}_i}, \infty)$	is $1-\alpha$ upper-sided C.I. for β_i

   while $\widehat{\beta}_i = e_i'(X'X)^{-1}X'y$ and $S_{\widehat{\beta}_i} = \sqrt{MSE\ e_i'(X'X)^{-1}e_i}$ where e_i is the ith column of I, the values of the two statistics can be looked up from SAS output.

   (3) t-intervals for $E(y(x_0))$

   $y(x_0) = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} + \epsilon$ is the future value of y at $x_0 = \begin{pmatrix} 1 \\ x_{01} \\ \vdots \\ x_{ok} \end{pmatrix}$,

   $E(y(x_0)) = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} = x_0'\beta$ is estimated by
   $\widehat{y}(x_0) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_k x_{ok} = x_0'\widehat{\beta}$
   $S_{\widehat{y}(x_0)} = \sqrt{MSE\ x_0'(X'X)^{-1}x_0}$ and $\frac{\widehat{y}(x_0)-E(y(x_0))}{S_{\widehat{y}(x_0)}} \sim t(n-p)$. So

$\widehat{y}(x_0) \pm t_{\alpha/2}(n-p)S_{\widehat{y}(x_0)}$	is $1-\alpha$ C.I. for β_i
$(-\infty, \widehat{y}(x_0) + t_\alpha(n-p)S_{\widehat{y}(x_0)})$	is $1-\alpha$ lower sided C.I. for β_i
$\widehat{y}(x_0) - t_\alpha(n-p)S_{\widehat{y}(x_0)}, \infty)$	is $1-\alpha$ upper-sided C.I. for β_i

3. SAS

(1) Put x_0 into SAS data set without y.

```
data a;
   infile "D:\Table32.txt";
   input y x1 x2 @@;
```

```
data b;
   input y x1 x2;
   datalines;
   . 10 300
   ;
```

```
data c;
   set a b;
proc print;
   run;
```

(2) Display $\widehat{y}_i$, $i = 1, .., n$, and $\widehat{y}(x_0)$

```
proc reg;
   model y=x1 x2/noint p;
   run;
```

(3) Display $\widehat{y}_i$, $S_{\widehat{y}_i}$, $\widehat{y}_i - t_{0.025}(n-p)S_{\widehat{y}_i}$, $\widehat{y}_i + t_{0.025}(n-p)S_{\widehat{y}_i}$, $i = 1, ..., n$, and $\widehat{y}(x_0)$, $S_{\widehat{y}(x_0)}$, $\widehat{y}(x_0) - t_{0.025}(n-p)S_{\widehat{y}(x_0)}$, $\widehat{y}(x_0) + t_{0.025}(n-p)S_{\widehat{y}(x_0)}$

```
proc reg;
    model y=x1 x2/noint p clm;
    run;
```

(4) Display $\widehat{y}_i$, $S_{\widehat{y}_i}$, $\widehat{y}_i - t_{0.1}(n-p)S_{\widehat{y}_i}$, $\widehat{y}_i + t_{0.1}(n-p)S_{\widehat{y}_i}$, $i = 1, ..., n$, and $\widehat{y}(x_0)$, $S_{\widehat{y}(x_0)}$, $\widehat{y}(x_0) - t_{0.1}(n-p)S_{\widehat{y}(x_0)}$, $\widehat{y}(x_0) + t_{0.1}(n-p)S_{\widehat{y}(x_0)}$

```
proc reg;
    model y=x1 x2/noint p clm alpha=0.2;
    run;
```