L03 ANOVA table

1. ANOVA table for model with intercept
   Sample $y \sim N(X\beta,\, \sigma^2 I)$ is from model $y = \beta_0 + \beta_1 x + \epsilon$. $H_0 : \beta_1 = 0$ is a null hypothesis. Under H_0 the sample $y \sim N(\beta_0 1_n,\, \sigma^2 I)$.

   (1) SSTO=SSM+SSE
   Under $H_0 : \beta_1 = 0$, $\sum(y_i - \overline{y})^2 = Syy$ measures the total variation in y denoted as SSTO, Sum of Squares due to Total variation in y, also called Corrected Sum of Squares of y (CSS of y). $SSTO = \left[\left(I_n - \frac{1_n1_n'}{n}\right)y\right]'\left[\left(I_n - \frac{1_n1_n'}{n}\right)y\right] = y'\left(I_n - \frac{1_n1_n'}{n}\right)y$.

   Call $H = X(X'X)^{-1}X'$ hat-matrix since for $y \sim N(X\beta,\, \sigma^2 I)$, $E(y) = X\beta$ is estimated by $\widehat{y} = X\widehat{\beta} = Hy$. $\sum(y_i - \widehat{y}_i)^2 = Syy - \frac{(Sxy)^2}{Sxx}$ is the variation in y not explained by the model denoted as SSE, Sum of Squares due to Error. $SSE = y'(I - H)y$.

   $SSTO - SSE = \frac{(Sxy)^2}{Sxx} = y'\left(H - \frac{1_n1_n'}{n}\right)y = \sum_i(\widehat{y}_i - \overline{y})^2$ is a Sum of Squares representing the variation in y explained by the Model denoted as SSM.
   Then SSTO=SSM+SSE.

   (2) (DF of SSTO)=(DF of SSM)+(DF of SSE)
   Define DF (degrees of freedom) of SSTO as rank $\left(I - \frac{11'}{n}\right) = n - 1$.
   DF of SSE as rank$(I_n - H) = n - 2$.
   DF of SSM as rank $\left(H - \frac{1_n1_n'}{n}\right) = 2 - 1 = 1$.
   Then (DF of SSTO)=(DF of SSM)+(DF of SSE)

   (3) Define $MSM = \frac{SSM}{\text{DF of SSM}} = \frac{SSM}{1}$ and $MSE = \frac{SSE}{\text{DF of SSE}} = \frac{SSE}{n-2}$.

   (4) Under $H_0 : \beta_1 = 0$, $\frac{\text{SSM}}{\sigma^2} \sim \chi^2(1) = \chi^2(\text{DF of SSM}) = \chi^2(1)$

   Proof. Under H_0, $y \sim N(1_n\beta_0,\, \sigma^2 I_n)$. Note that $\frac{SSM}{\sigma^2} = y'Ay$ where $A = \frac{H - \frac{11'}{n}}{\sigma^2}$.
   But $A' = A = A\sigma^2 IA$, $(1_n\beta_0)'\left(H - \frac{11'}{n}\right)(1_n\beta_0) = 0$ and rank$(A) = 2 - 1 = 1$.
   By Theorem II in L02, $\frac{SSM}{\sigma^2} \sim \chi^2(0,\, 1) = \chi^2(1)$.

   (5) Under$H_0 : \beta_1 = 0$, $F = \frac{MSM}{MSE} \sim F(1,\, n-2)$

   Proof. Definition: If $W_1 \sim \chi^2(m)$ is independent to $W_2 \sim \chi^2(n)$, $\frac{W_1/m}{W_2/n} \sim F(m,\, n)$.
   Recall that $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$ and $\frac{SSM}{\sigma^2} \overset{H_0}{\sim} \chi^2(1)$.
   SSM and SSE are independent since $SSE = y'Ay$ with $A = I - H$, $SSM = y'By$ with $B = H - \frac{11'}{n}$, $y \overset{H_0}{\sim} N(1_n\beta_0,\, \sigma^2 I)$ and $A\sigma^2 B = 0$.
   Thus $\frac{SSM/(\sigma^2 \times 1)}{SSE/(\sigma^2 \times (n-2))} \sim F(1,\, n-2)$, i.e., $\frac{MSM}{MSE} \overset{H_0}{\sim} F(1,\, n-2)$.

   (6) All above are summarized in ANOVA table

Source	SS	DF	MS	F	p
Model	SSM	1	MSM	MSM/MSE	$P(F(1, n-2) > F_{ob})$
Error	SSE	$n-2$	MSE		
C.Total	SSTO	$n-1$			

2. ANOVA table for model without intercept

Sample $y \sim N(X\beta, \sigma^2 I)$ is from model $y = \beta x + \epsilon$. $H_0: \beta = 0$ is a null hypothesis. Under H_0 the sample $y \sim N(0, \sigma^2 I)$.

(1) SSTO=SSM+SSE

Under $H_0: \beta = 0$, $\sum y_i^2 = y' I_n y$ is the total variation in y denote as SSTO. It is Uncorrected Sum of Squares (USS) of y.

With hat-matrix H, $\sum(y_i - \widehat{y}_i)^2 = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2} = y'(I-H)y$ is the variation in y unexplained by the model denoted as SSE, the Sum of Squares due to Error.

SSTO-SSE=$\frac{(\sum xy)^2}{\sum x^2} = \sum_i \widehat{y}_i^2 = y'Hy$ is a Sum of Squares representing the variation in y explained by the model denoted as SSM.

Clearly SSTO=SSM+SSE.

(2) (DF of SSTO)=(DF of SSM)+(DF of SSE)

Define DF of SSTO as rank$(I_n) = n$, DF of SSM as rank$(H) = 1$ and DF of SSE as rank$(I - H) = n - 1$.

Then (DF of SSTO)=(DF of SSM)+(DF of SSE)

(3) Define $MSM = \frac{SSM}{\text{DF of SSM}} = \frac{SSM}{1}$ and $MSE = \frac{SSE}{\text{DF of SSE}} = \frac{SSE}{n-1}$.

(4) Under $H_0: \beta = 0$, $\frac{SSM}{\sigma^2}\chi^2(1)$

Proof. Under H_0, $y \sim N(0, \sigma^2 I)$. $\frac{SSM}{\sigma^2} = y'Ay$ where $A = \frac{H}{\sigma^2}$. But $A' = A = A\sigma^2 IA$, $0'A0 = 0$ and $\text{tr}(A\sigma^2 I) = 1$. By Theorem II in L02, under H_0, $\frac{SSM}{\sigma^2} \overset{H_0}{\sim} \chi^2(1)$.

(5) Under $H_0: \beta = 0$, $F = \frac{MSM}{MSE} \sim F(1, n-1)$

Proof. $\frac{SSE}{\sigma^2} \sim \chi^2(n-1)$. Under H_0, $\frac{SSM}{\sigma^2} \sim \chi^2(1)$. SSE and SSM are independent since $SSE = y'Ay$, $SSM = y'By$, $A = I - H$, $B = H$ and $A\sigma^2 IB = 0$.

$$F = \frac{MSM}{MSE} = \frac{SSM/\sigma^2}{SSE/\sigma^2(n-1)} \overset{H_0}{\sim} \frac{\chi^2(1)}{\chi^2(n-1)/(n-1)} = F(1, n-1).$$

(6) All above are summarized in ANOVA table

Source	SS	DF	MS	F	p
Model	SSM	1	MSM	MSM/MSE	$P(F(1, n-1) > F_{ob})$
Error	SSE	$n-1$	MSE		
U.Total	SSTO	n			

3. An example

A sample for the model without intercept produced statistics below. Fill out ANOVA table.

$$n = 4, \sum x = 2, \sum y = 14, \sum x^2 = 6, \sum y^2 = 54, \sum xy = 4.$$

$SSM = \frac{(\sum xy)^2}{\sum x^2} = 2.6667$ $U.SSTO = \sum y^2 = 54$ $SSE = 54 - 2.6667 = 51.3333$

Source	SS	DF	MS	F	p
Model	2.6667	1	2.6667	0.156	0.719
Error	51.3333	3	17.1111		
U.Total	54	4			

$P(F(1, 3) > 0.156) = 0.719$ is produced by F-distribution calculator APP with link on class webpage

L04 Test for significance of regression

1. For $y = \beta_0 + \beta_1 x + \epsilon$, using rejection region

$$\boxed{\begin{array}{l} H_0: \beta_1 = 0 \text{ versus } H_a: \beta_1 \neq 0 \\ \text{Test Statistic: } F = \frac{MSM}{MSE} \\ \text{Reject } H_0 \text{ if } F > F_\alpha(1,\, n-2) \end{array}}$$

is an α-level likelihood ratio test (LRT) for significance of regression.

(1) Hypotheses
The null hypothesis $H_0: \beta_1 = 0$ means that the model is useless.
Determining if H_0 is true is the first question facing researchers after selecting model and collecting data.

(2) LRT statistic

$$\begin{aligned} \Lambda &= \frac{\max[L(\beta,\sigma^2)]}{\max[L(\beta,\sigma^2)\colon H_0]} = \frac{\left(\frac{n}{2\pi e}\right)^{n/2}(SSE)^{-n/2}}{\left(\frac{n}{2\pi e}\right)^{n/2}(SSTO)^{-n/2}} = \left(\frac{SSTO}{SSE}\right)^{n/2} = \left(\frac{SSM}{SSE}+1\right)^{n/2} \\ &= \left(\frac{MSM}{MSE}\cdot\frac{1}{n-2}+1\right)^{n/2} = \left(F\cdot\frac{1}{n-2}+1\right)^{n/2} \end{aligned}$$

is the likelihood ratio.
If H_0 is rejected when $\Lambda > c_1$, then the test scheme is a LRT scheme.
But Λ is an increasing function of F. So $\Lambda > c_1 \Longrightarrow F > c_2$
Thus F can be used as LRT statistic and H_0 is rejected for $F > c_2$.

(3) α-level test
In the rejection rule, $F_\alpha(1,\, n-2)$ is a constant such that

$$P\left(F(1,\, n-2) > F_\alpha(1,\, n-2)\right) = \alpha.$$

The value of $F_\alpha(1,\, n-2)$ can be looked up by F-distribution APP.

$$\begin{aligned} P(\text{Type I error}) &= P\left(\text{Rejecting } H_0 | H_0 \text{ is true}\right) = P\left(F > F_\alpha(1,\, n-2) | \beta_1 = 0\right)\right) \\ &= P(F(1,\, n-2) > F_\alpha(1,\, n-2) = \alpha. \end{aligned}$$

Here pre-selected α controls the probability of Type I error and is called the significance level of the test.

Ex1: Design a test for significance of simple linear regression with intercept at significance level $\alpha = 0.05$ with a sample of size 25.

$F_\alpha(1,\, n-2) = F_{0.05}(1,\, 23) = 4.279$. So $\boxed{\begin{array}{l} H_0: \beta_1 = 0 \text{ versus } H_a: \beta_1 \neq 0 \\ \text{Test statistic: } F = \frac{MSM}{MSE} \\ \text{Reject } H_0 \text{ if } F > 4.279 \text{ for } \alpha = 0.05 \end{array}}$

2. Using p-value

$$\boxed{\begin{array}{l} H_0: \beta_1 = 0 \text{ versus } H_a: \beta_1 \neq 0 \\ \text{Test Statistic } F = \frac{MSM}{MSE} \\ \text{p-value: } P(F(1,\, n-2) > F_{ob}) \end{array}}$$

is the test for the significance of the model $y = \beta_0 + \beta_1 x + \epsilon$.

(1) P-value
The test produces p-value calculated from sample and is called the observed significance level. All statistical software on tests produce p-values in stead of rejection regions.

(2) Usage
False H_0 produces large F_{ob} and small p-value.
So p-value shows the consistency of data with H_0.
Thus the universal rule for all tests is: Reject H_0 for small p-value.

(3) Significance level α and observed significance level p.

$$\begin{aligned}
&\text{Reject } H_0 \text{ if } p < \alpha \Longleftrightarrow \text{Reject } H_0 \text{ if } P\left(F(1,\, n-2) > F_{ob}\right) < \alpha \\
&\Longleftrightarrow \text{Reject } H_0 \text{ if } P\left(F(1,\, n-2) > F_{ob}\right) < P\left(F(1,\, n-2) > F_\alpha(1,\, n-2)\right) \\
&\Longleftrightarrow \text{Reject } H_0 \text{ if } F_{ob} > F_\alpha(1,\, n-2).
\end{aligned}$$

Ex2: A sample from a simple linear regression with intercept produced ANOVA table.

Source	SS	DF	MS	F	p
Model	153	1	153	165	< 0.0001
Error	17	18	0.9		
C.Total	170	19			

Write your report on the test on the usefulness of the model by rejection region and by p-value.

$H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ Test statistic: $F = \frac{MSM}{MSE}$ Reject H_0 if $F > 4.414$ for $\alpha = 0.05$ $F_{ob} = 165$ reject H_0. The model is useful.	$H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ Test statistic: $F = \frac{MSM}{MSE}$ p-value: $P(1,\, n-2) > F_{0b}$ $F_{ob} = 165$ and p-value< 0.0001 Reject H_0. The model is useful

3. Model without intercept

(1) Test on $H_0 : \beta = 0$ using rejection region

$H_0 : \beta = 0$ versus $H_a : \beta \neq 0$
Test Statistic: $F = \frac{MSM}{MSE}$
Reject H_0 if $F > F_\alpha(1,\, n-1)$

(2) Test on $H_0 : \beta = 0$ using p-value

$H_0 : \beta = 0$ versus $H_a : \beta \neq 0$
Test Statistic: $F = \frac{MSM}{MSE}$
p-value: $P(F(1,\, n-1) > F_{ob})$

Comment: To report on a test, first write out three line test scheme followed by the value of your calculated statistics. Then state the conclusion. Skip detailed computations.