STAT-1302; Lecture 16; March 12, '24

Refer to the drugs example in Lecture 15.

Often, ANOVA results are summarized in an "ANOVA Table".

ANOVA

Source	df	SS	MS	F
Between	$k-1$	SSB	SSB/(k-1)	MSB/MSW
Within	$n-k$	SSW	SSW/(n-k)	
Total	$n-1$	SST		

$$SST = SSB + SSW$$

df = degrees of freedom

SS = Sum of Squares

MS = Mean Square

SST = Total Sum of Squares

SSB = Between Sum of Squares,

SSW = Within Sum of Squares.

ANOVA

Source	df	SS	MS	F
Between	2	2645.7	1322.85	9.54
Within	12	1664.03	138.67	
Total	14	4309.73		

Ex. An experiment is conducted to determine the soil moisture deficit resulting from varying amounts of residual timber left after cutting trees in a forest. Interest centres on comparing three groups. In "Group 1", there is no timber left. In "Group 2", 2,000 broad feet are left. In "Group 3", there are 8,000 broad feet left. The data are as follows.

Moisture Deficit in Soil

Group 1	Group 2	Group 3
1.52	1.63	2.56
1.38	1.82	3.32
1.29	1.35	2.76
1.48	1.03	2.63
1.63	2.30	2.12
$T_1 = 7.30$	1.45	2.78
	$T_2 = 9.58$	$T_3 = 16.17$
$\bar{y}_1 = 1.460$	$\bar{y}_2 = 1.597$	$\bar{y}_3 = 2.695$

Test whether the group means are different.
Let $\alpha = 0.05$.

$H_0 : \mu_1 = \mu_2 = \mu_3$ vs.

H_1: Not all μ_i's are equal.

$$\sum x = T_1 + T_2 + T_3 = 33.05$$

$$n_1 = 5, \quad n_2 = 6, \quad n_3 = 6 \quad \Rightarrow \quad n = n_1 + n_2 + n_3 = 17.$$

$$\sum x^2 = 71.3047$$

$$SSB = \sum_{i=1}^{3} \frac{T_i^2}{n_i} - \frac{\left(\sum x\right)^2}{n}$$

$$= \left(\frac{7.30^2}{5} + \frac{9.58^2}{6} + \frac{16.17^2}{6}\right) - \frac{33.05^2}{17}$$

$$= 69.5322 - 64.2531 = 5.2791.$$

$$\left(\sum x^2 = 1.5^2 + 1.38^2 + \dots + 2.12^2 + 2.78^2 \right.$$
$$\left. = 71.3047 \right)$$

$$SSW = \sum x^2 - \sum_{i=1}^{3} \frac{T_i^2}{n_i}$$

$$= 71.3047 - 69.5322 = 1.7725$$

$$MSB = SSB/(k-1) = 5.2791/2 = 2.640$$

$$MSW = SSW/(n-k) = 1.7725/14 = 0.127$$

($k = 3$ groups ; $n = 17$; $n-k = 14$; $n-1 = 16$)

$$F = \frac{MSB}{MSW} = \frac{2.640}{0.127} = 20.79.$$

ANOVA

Source	df	SS	MS	F
Between	2	5.2791	2.640	20.79
Within	14	1.7725	0.127	
Total	16			

df for SSB $= k-1 = 3-1 = 2$

df for SSW $= n-k = 17-3 = 14$

df for SST $= n-1 = 17-1 = 16$

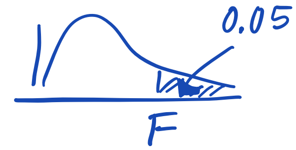
Compare $F_{obs.} = 20.79$ with $F_{2,14;0.05} = 3.74$

Decision rule: Reject H_0 if $F_{obs.} > 3.74$.

Since $20.79 > 3.74$, we reject H_0.

Conclusion: We are 95% confident that not all group means are equal.

Table VII
The F- Distribution Table



Df for Numerator

Df for Denominator

	1	2	3	. . . . .
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14		$\boxed{3.74}$		

$$F_{2,14;0.05} = 3.74.$$

Assumptions:

Three independent groups from a normal population all with the same variance σ^2.

MCQ

The observations in each group are independent.

Rule of Thumb:

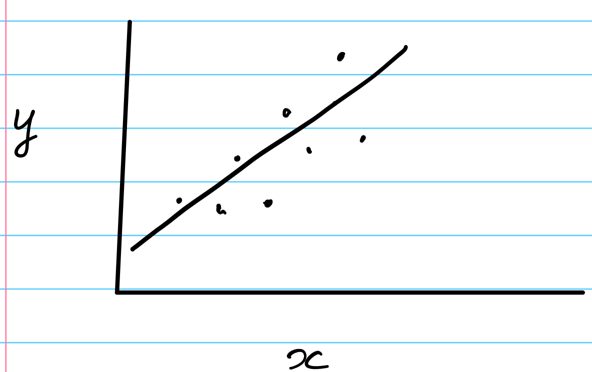
(Used to verify constant variance σ^2 for all groups.)

TBC .....

Ch. 13. Simple Linear Regression

Objective: Study the linear relationship between a quantitative outcome variable (y) and an explanatory variable (x).

e.g. monthly income (x) food expenditure (y)



Want to use x to predict y based on the linear relationship observed in the scatterplot.

We want to regress Y against X.

Notes:

1. Y (e.g. food expenditure) $\Leftrightarrow$ response variable $\Leftrightarrow$ outcome variable $\Leftrightarrow$ dependent variable.

2. X (e.g. monthly expenditure) $\Leftrightarrow$ explanatory variable $\Leftrightarrow$ predictor $\Leftrightarrow$ independent variable.

3. A scatterplot of Y against X reveals a linear trend. We want to fit the "best" straight line through these data points.

(High School: $Y = \underbrace{m}_{\text{slope}} X + \underbrace{b}_{\text{intercept}}$ is the equation of a straight line.)