

## ANOVA Example

$k = 3$ treatments		Drug A	Drug B	Drug C
column #	$j$	1	2	3
sample size	$n_j$	7	8	10
mean	$\bar{Y}_{\bullet j}$	80	88	90
variance	$S_j^2$	5.2	4.8	5.4

Are these drugs different?

## ANOVA Finding totals and averages

<i>treatment</i>	Drug A	Drug B	Drug C
$n_j$	7	8	10
$\bar{Y}_{\bullet j}$	80	88	90
$S_j^2$	5.2	4.8	5.4

The total for each treatment is given by

$$T_{\bullet j} = n_j \bar{Y}_{\bullet j}$$

With grand total  $T_{\bullet\bullet} = \sum T_{\bullet j}$ :

$$T_{\bullet\bullet} = \sum_j n_j \bar{Y}_{\bullet j} = 7 * 80 + 8 * 88 + 10 * 90$$

$$= 560 + 704 + 900 = 2164$$

$$Y_{\bullet\bullet} = \frac{1}{n} T_{\bullet\bullet} = \frac{2164}{25} = 86.56$$

## Finding $SS_{Tr}$ , $MS_{Tr}$

The treatment sum of squares is

$$SS_{Tr} = \sum_i n_j (\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2$$

This number tends to be big if the treatments are different. It is small when the treatments have no effect.

$$\begin{aligned} SS_{Tr} &= 7 * (80 - 86.56)^2 \\ &\quad + 8 * (88 - 86.56)^2 \\ &\quad + 10 * (90 - 86.56)^2 \\ &= 436.16 \end{aligned}$$

The number of degrees of freedom is the number of treatments minus 1:

$$df = k - 1 = 2$$

The formula for  $MS_{Tr}$  is

$$MS_{Tr} := \frac{SS_{Tr}}{k - 1} = \frac{436.16}{2} = 218.08$$

## Finding SSE, MSE

The error sum of squares is

$$\begin{aligned}SS_E &= \sum_j \sum_i (Y_{ij} - \bar{Y}_{\bullet j})^2 \\ &= \sum_j (n_j - 1)S_j^2\end{aligned}$$

This number measures random errors and variability of data. It tell us nothing about the treatments (drugs in this case).

$$SS_E = 6(5.2) + 7(4.8) + 9(5.4) = 113.4$$

The degrees of freedom is

$$df = \sum_j (n_j - 1) = n - k = 25 - 3 = 22$$

So the mean square error is:

$$MS_E = \frac{SS_E}{n - k} = \frac{113.4}{22} = 5.15$$

## ANOVA, F-test

The test statistic is

$$F_{(k-1, n-k)} = \frac{MS_{Tr}}{MS_E}$$

$$F_{(2,22)} = \frac{218.08}{5.15} = 42.3$$

In ANOVA the F-test is **always right tailed**. When  $F$  is large we conclude that there is a significant difference between the drugs. This is because the numerator measures the difference between treatments.

The critical value is

$$F_{2,22,.95} = 3.44$$

Since the data value of  $F$  is much larger than critical we conclude that the drugs are different. But we don't know if they make people better or worse!

## Summary of results

The traditional way to summarize the results is by the following chart with either the critical  $F$  value or the  $p$ -value in the last column. The  $p$ -value is from the command “=Fdist(42.3,2,22)” in Excel.

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
<i>Treatment</i>	436.16	2	218.08	42.3	$2.9 \times 10^{-8}$
<i>Error</i>	113.4	22	5.15		
<i>Total</i>	549.56	24			

The conclusion is that at least one of the drugs is different from the other two. We will do additional tests to see which one is different.