L20: Chapter 11 ANOVA

Department of Statistics, University of South Carolina

Stat 205: Elementary Statistics for the Biological and Life Sciences

Comparing more than two means

- In Chapter 7 we had two groups and tested $H_0 : \mu_1 = \mu_2$.
- In Chapter 11 we will have I groups and test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$.
- We are still interested in whether the population means are the same across groups, there's just more than two.
- The alternative hypothesis is H_A : one or more of $\mu_1, \mu_2, \ldots, \mu_I$ are different.
- Let's look at an example where $I = 5$.

Example 11.1.1

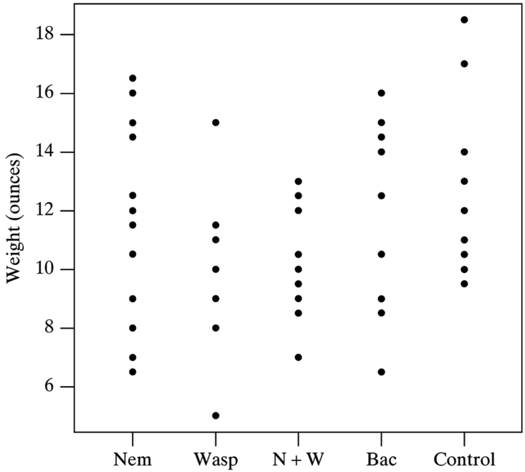
When growing sweet corn, can organic methods be used successfully to control harmful insects and limit their effect on the corn? In a study of this question researchers compared the weights of ears of corn under five conditions in an experiment in which sweet corn was grown using organic methods. In one plot of corn a beneficial soil nematode was introduced. In a second plot a parasitic wasp was used. A third plot was treated with both the nematode and the wasp. In a fourth plot a bacterium was used. Finally, a fifth plot of corn acted as a control; no special treatment was applied here. Thus, the treatments were

- Treatment 1: Nematodes
- Treatment 2: Wasps
- Treatment 3: Nematodes and wasps
- Treatment 4: Bacteria
- Treatment 5: Control

Table 11.1.1 Weights (ounces) of ears of sweet corn

	Treatment				
	1	2	3	4	5
	16.5	11.0	8.5	16.0	13.0
	15.0	15.0	13.0	14.5	10.5
	11.5	9.0	12.0	15.0	11.0
	12.0	9.0	10.0	9.0	10.0
	12.5	11.5	12.5	10.5	14.0
	9.0	11.0	8.5	14.0	12.0
	16.0	9.0	9.5	12.5	11.0
	6.5	10.0	7.0	9.0	9.5
	8.0	9.0	10.5	9.0	18.5
	14.5	8.0	10.5	9.0	17.0
	7.0	8.0	13.0	6.5	10.0
	10.5	5.0	9.0	8.5	11.0
Mean	11.5	9.6	10.3	11.1	12.3
SD	3.5	2.4	2.0	3.1	2.9
n	12	12	12	12	12

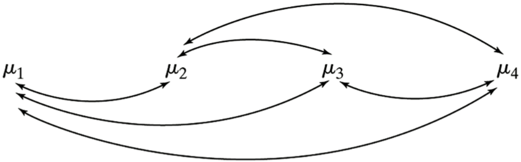
Weights of ears of corn receiving five treatments



- Its natural to ask: why not now just compare all possible pairs $\mu_1 - \mu_2, \mu_1 - \mu_3, \mu_2 - \mu_3$, etc., each with a t-test (or conf. interval)?
- DEFN: The PROBLEM of MULTIPLE COMPARISONS occurs when the same data set is used to make multiple inferences on $I > 2$ associated parameters.

Comparing four population means requires six comparisons

$$H_0: \mu_1 = \mu_2 \quad H_0: \mu_1 = \mu_3 \quad H_0: \mu_1 = \mu_4$$
$$H_0: \mu_2 = \mu_3 \quad H_0: \mu_2 = \mu_4 \quad H_0: \mu_3 = \mu_4$$



Type I error

Table 11.1.2 Overall risk of Type I error in using repeated t tests at $\alpha = 0.05$

I	Overall risk
2	0.05
3	0.12
4	0.20
6	0.37
8	0.51
10	0.63

Building the test

- We will test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$ via *analysis of variance* (ANOVA).
- ANOVA compares how variable the sample means $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_I$ are to how variable observations are *around each mean*.
- Assumptions: Observations in each group are indepedently normally distributed with the same variance σ^2.
- The data in different groups are also independent.

Sums of Squares

- DEFN: SUMS OF SQUARES (S.S.) are sums of squared deviations from a central value.
- the WITHIN GROUPS S.S.
- the BETWEEN GROUPS S.S.
- the TOTAL S.S.
- Note: TOTAL S.S. = BETWEEN S.S. + WITHIN S.S.

ANOVA formulae

ANOVA Quantities with Formulas

Source	df	SS (Sum of Squares)	MS (Mean Square)
Between groups	$I - 1$	$\sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$	SS/df
Within groups	$n. - I$	$\sum_{i=1}^{I} (n_i - 1)s_i^2$	SS/df
Total	$n. - 1$	$\sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$	

Lamb weights

Table 11.2.1 Weight gains of lambs (lb)*

	Diet 1	Diet 2	Diet 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Sum $= \sum_{j=1}^{n_i} y_{ij}$	33	75	48
Mean $= \bar{y}_i$	11.000	15.000	12.000
SD $= s_i$	4.359	4.950	4.967

*Extra digits are reported for accuracy of subsequent calculations.

ANOVA table

- DEFN: A MEAN SQUARE (MS) is the average of the squared deviations from a central value. It is a Sum of Squares (SS) divided by the number of informative values in the SS. called "degrees of freedom", or df
- We collect S.S., Mean squares, df, etc in a table called an ANOVA table.

ANOVA table

Table 11.2.3 ANOVA table for lamb weight gains

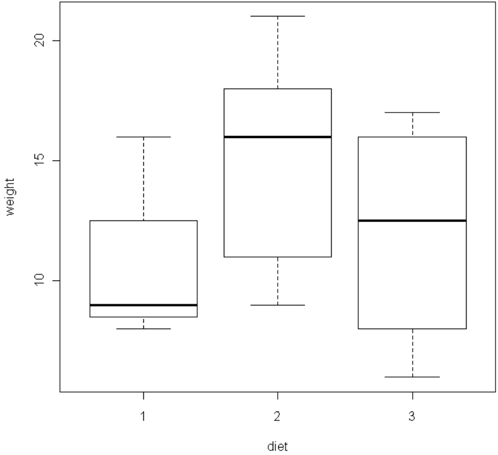
Source	df	SS	MS
Between diets	2	36	18.00
Within diets	9	210	23.33
Total	11	246	

- overall estimate of $\sigma^2 = s^2_{pool} = MS(within)$.

F test

- $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$
- Test statistic: $F = MS(Between)/MS(Within)$
  which has an $F(df_{between}, df_{within})$ distribution if H_0 is true.
- DEFN: The F-DISTRIBUTION with ν_1 and ν_2 degrees of freedom is the distn of the ratio of two (indep.) mean squares.
  NOTATION: $F \sim F(\nu_1, \nu_2)$
- In R 'Within S.S.' is called 'Residual S.S.'
- Obtain P-value from R. Reject H_0 if ....

Lamb data



R code for lamb diet data

```
> weight=c( 8,16, 9, 9,16,21,11,18,15,10,17, 6)
> diet  =c( 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3)
> diet=factor(diet)
> fit=aov(weight~diet)
> summary(fit)
            Df Sum Sq Mean Sq F value Pr(>F)
diet         2     36  18.000  0.7714 0.4907
Residuals    9    210  23.333
```

R code for corn growth data

```
weight=c(16.5,11.0, 8.5,16.0,13.0,15.0,15.0,13.0,14.5,10.5,
         11.5, 9.0,12.0,15.0,11.0,12.0, 9.0,10.0, 9.0,10.0,
         12.5,11.5,12.5,10.5,14.0, 9.0,11.0, 8.5,14.0,12.0,
         16.0, 9.0, 9.5,12.5,11.0, 6.5,10.0, 7.0, 9.0, 9.5,
          8.0, 9.0,10.5, 9.0,18.5,14.5, 8.0,10.5, 9.0,17.0,
          7.0, 8.0,13.0, 6.5,10.0,10.5, 5.0, 9.0, 8.5,11.0)
treat= c(1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,
         1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,1,2,3,4,5,
         1,2,3,4,5,1,2,3,4,5)
> treat=factor(treat)
> fit=aov(weight~treat)
> summary(fit)
            Df Sum Sq Mean Sq F value Pr(>F)
treat        4  52.31 13.0771  1.6461 0.1758
Residuals   55 436.94  7.9443
```

MAO activity in schizophrenia

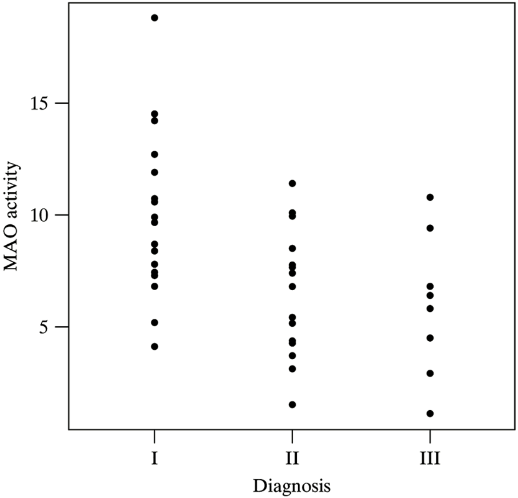


Figure 1.1.2 MAO activity in schizophrenic patients

MAO activity (Fig. 1.1.2)

```
> x=read.table("mao.txt")
> y=x[,1]
> diagnosis=factor(x[,2])
> fit= aov(y~diagnosis)
> summary(fit)
            Df Sum Sq Mean Sq F value   Pr(>F)
diagnosis    2 136.12  68.059  6.3461 0.004111 **
Residuals   39 418.25  10.724
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```