# Stat 324 – Introduction to Statistics for Engineers

LECTURE 15: ANOVA AND KRUSKAL-WALLIS, TESTS OF CENTER FOR 2+ GROUPS

SECTIONS 8.2-8.4, 8.6, 9.3-9.5

## Comparison of Multiple Groups

In the article "Review of Development and Application of CRSTER and MPTER Models" (R.Wilson *Atmospheric Environment*), several measurements of the maximum hourly concentrations of S02 are presented for each of four power plants. The results are as follows.

We would like to evaluate whether all four plants have similar maximum hourly concentrations of S02 (in the populations of S02s based on this sample data).

| Plant 1 | | | | | |
|---|---|---|---|---|---|
| 438 | 619 | 732 | 638 | | |

| Plant 2 | | | | | |
|---|---|---|---|---|---|
| 857 | 1014 | 1153 | 883 | 1053 | |

| Plant 3 | | | | | |
|---|---|---|---|---|---|
| 925 | 786 | 1179 | 786 | | |

| Plant 4 | | | | | |
|---|---|---|---|---|---|
| 893 | 891 | 917 | 695 | 675 | 595 |
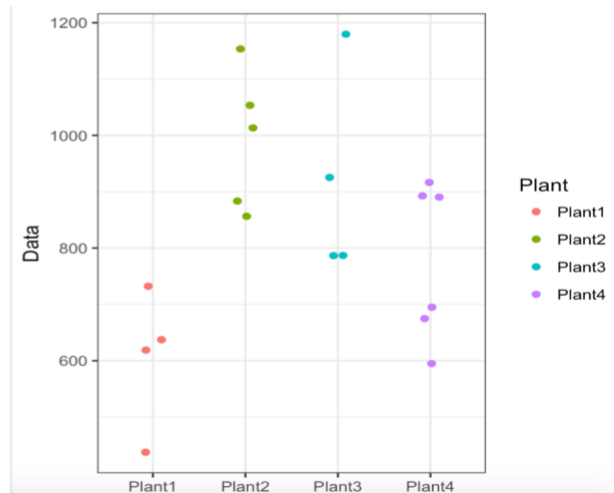
# Comparison of Multiple Groups

We know how to compare pairs of populations – we could use :

*Independant, so need to use 2 sample t test, wilcoxn rank sum (mann-whitney U) permutation, bootstrap*

The pairs (or more) of populations and their measurements are distinguished by **levels of a factor.**
The factor is a **categorical** variable. In experiments, the different levels are often different **treatments.**

E.g.     The factor in the article is "Plant" and it has levels 1, 2, 3, and 4.

The experimental units on which the treatments are assigned and measured are called **replicates.**

E.g.     The Plant1 level has 4 replicates, Plant2 level as 5 replicates, etc.

To compare centers across populations, we need to evaluate _____shifts_____ relative to the _____variability_____ we see in the replicates.

# Comparison of Multiple Groups Using ANOVA

**One-Way Analysis of Variance (ANOVA):** When we only have 1 factor and 1 quantitative response and we are interested in whether the factor level means differ

Null: $H_o: \mu_1 = \mu_2 = \ldots = \mu_T$ *where $\mu_i$ is the population mean of the i'th group.* Vs
Alternative: $H_A$: _at least one of the $\mu_i$ differ from another_

Our test statistic is built by **comparing the variability**

(A)  *between group means*                 to variability                 (B) *within the groups*-pooled variance
     (~SD_of group means_____)                                                                  across groups

If larger variability between group means compared to small variability within groups,
the _stronger_ the evidence against the null  (and therefore a larger test statistics and smaller p-value)

An example:        *(same # of obs)*

*The treatment means*
*are equal across both graphs,*
*but there is more overlap between*
*groups in the Right graph.*
*Left graph should give us stronger*
*evidence of difference in means (evidence against null)*
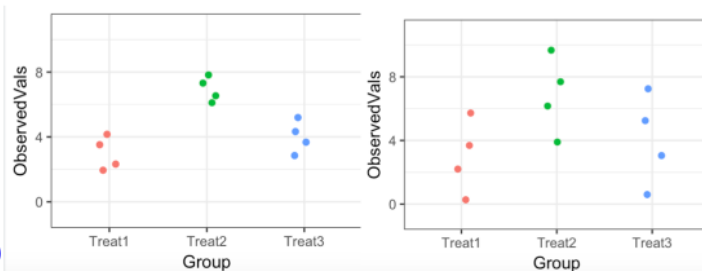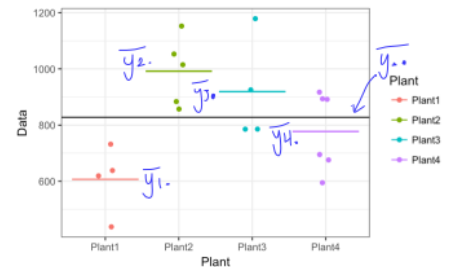
One-Way Analysis of Variance (ANOVA) [nitty gritty details]

The notation we will use:

- $t$ be the number of treatments

- $i$ index the treatments, i from 1:4

- $n_i$ number of replicates in treatment i, $n_1 = 4$

- $N = \sum_{i=1} n_i$ is total sample size, $N = 19$

- $y_{ij}$ is the *jth* observation in the *ith* group

| $T_i$ | | | | | | $n_i$ | $\bar{y}_{i.}$ | $s_i$ |
|---|---|---|---|---|---|---|---|---|
| Plant 1 | | | | | | | | |
| 438 | 619 | 732 | 638 | | | 4 | 606.8 | 122.9 |
| Plant 2 | | | | | | | | |
| 857 | 1014 | 1153 | 883 | 1053 | | 5 | 992.0 | 122.7 |
| Plant 3 | | | | | | | | |
| 925 | 786 | 1179 | 786 | | | 4 | 919.0 | 185.3 |
| Plant 4 | | | | | | | | |
| 893 | 891 | 917 | 695 | 675 | 595 | 6 | 777.7 | 138.8 |
| | | | | Overall: | $N = 19, \bar{y}_{..} = 827.8$ | | | |

- $\overline{y_{i.}} = \frac{1}{n_i}\sum_{j=1} y_{ij}$ sample mean of the ith group, $\overline{y_{1.}} = 606.8$

  ↑ average over

- $\bar{y}_{..} = \frac{1}{N}\sum_i (\sum_{j=1} y_{ij})$ sample mean grand mean , $\bar{y}_{..} = 827.8$



---

One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Each observation's value $y_{ij}$ can be thought of as a sum of pieces

| Observation = | Grand Mean + | Deviation due + to Treatment | Deviation due to Obs within group (Residual) |
|---|---|---|---|
| $y_{ij}$ | $= \quad \bar{y}_{..} +$ | $(\overline{y_{i.}} - \bar{y}_{..}) +$ | $(y_{ij} - \overline{y_{i.}})$ |

Trt Group 1:

$438 = 827.8 + (\overset{\overline{y_{i.}}}{606.8} - \overset{\overline{y_{..}}}{827.8}) + (\overset{\overline{y_i}}{438} - \overset{\overline{y_{i.}}}{606.8})$
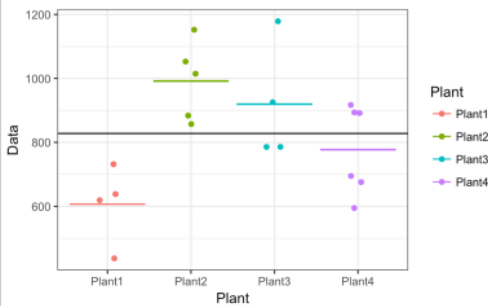
$438 = 827.8 + (-221) + (-168.8)$

Trt Group 3:

$1179 = 827.8 + (919 - 827.8) + (1179 - 919)$

$1179 = 827.8 + 91.2 + 260$

## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Observation =  Grand Mean +  Deviation due + to Treatment  Deviation due to Obs within group (Residual)

$$\begin{bmatrix} 438 & 619 & 732 & 638 & \\ 857 & 1014 & 1153 & 883 & 1053 \\ 925 & 786 & 1179 & 786 & \\ 893 & 891 & 917 & 695 & 675 & 595 \end{bmatrix}$$

=

$$\begin{bmatrix} 827.8 & 827.8 & 827.8 & 827.8 & \\ 827.8 & 827.8 & 827.8 & 827.8 & 827.8 \\ 827.8 & 827.8 & 827.8 & 827.8 & \\ 827.8 & 827.8 & 827.8 & 827.8 & 827.8 & 827.8 \end{bmatrix}$$

+

$$\begin{bmatrix} -221 & -221 & -221 & -221 & \\ 164.2 & 164.2 & 164.2 & 164.2 & 164.2 \\ 91.2 & 91.2 & 91.2 & 91.2 & \\ -50.1 & -50.1 & -50.1 & -50.1 & -50.1 & -50.1 \end{bmatrix}$$

$$\begin{bmatrix} -168.8 & 12.2 & 125.2 & 31.2 & \\ -135 & 22 & 161 & -109 & 61 \\ 6 & -133 & 260 & -133 & \\ 115.3 & 113.3 & 139.3 & -82.7 & -102.7 & -182.7 \end{bmatrix}$$



$438 = 827.8 - 221 - 168.8$

*Slightly off due to Rounding

## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

(Observation  ---  Grand Mean) =  Deviation due + to Treatment  Deviation due to Obs within group

$$(y_{ij} \quad - \quad \overline{y..}) = \quad (\overline{y_{i.}} - \overline{y..}) + \quad (y_{ij} - \overline{y_{i.}})$$

$$\begin{bmatrix} 438 & 619 & 732 & 638 & \\ 857 & 1014 & 1153 & 883 & 1053 \\ 925 & 786 & 1179 & 786 & \\ 893 & 891 & 917 & 695 & 675 & 595 \end{bmatrix} -$$

$$\begin{bmatrix} 827.8 & 827.8 & 827.8 & 827.8 & \\ 827.8 & 827.8 & 827.8 & 827.8 & 827.8 \\ 827.8 & 827.8 & 827.8 & 827.8 & \\ 827.8 & 827.8 & 827.8 & 827.8 & 827.8 & 827.8 \end{bmatrix}$$

=

$$\begin{bmatrix} -221 & -221 & -221 & -221 & \\ 164.2 & 164.2 & 164.2 & 164.2 & 164.2 \\ 91.2 & 91.2 & 91.2 & 91.2 & \\ -50.1 & -50.1 & -50.1 & -50.1 & -50.1 & -50.1 \end{bmatrix}$$

+

$$\begin{bmatrix} -168.8 & 12.2 & 125.2 & 31.2 & \\ -135 & 22 & 161 & -109 & 61 \\ 6 & -133 & 260 & -133 & \\ 115.3 & 113.3 & 139.3 & -82.7 & -102.7 & -182.7 \end{bmatrix}$$

Expression holds for each cell in the arrays,

$$\begin{bmatrix} -389.8 & -208.8 & -95.8 & -189.8 & \\ 29.2 & 186.2 & 325.2 & 55.2 & 225.2 \\ 97.2 & -41.8 & 351.2 & -41.8 & \\ 65.2 & 63.2 & 89.2 & -132.8 & -152.8 & -232.8 \end{bmatrix} =$$

$$\begin{bmatrix} -221 & -221 & -221 & -221 & \\ 164.2 & 164.2 & 164.2 & 164.2 & 164.2 \\ 91.2 & 91.2 & 91.2 & 91.2 & \\ -50.1 & -50.1 & -50.1 & -50.1 & -50.1 & -50.1 \end{bmatrix}$$

+

$$\begin{bmatrix} -168.8 & 12.2 & 125.2 & 31.2 & \\ -135 & 22 & 161 & -109 & 61 \\ 6 & -133 & 260 & -133 & \\ 115.3 & 113.3 & 139.3 & -82.7 & -102.7 & -182.7 \end{bmatrix}$$

And each array sums to zero! $\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y..}) = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\overline{y_{i.}}-\overline{y..}) + \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{i.}})$

## One-Way Analysis of Variance (ANOVA) [nitty gritty details] ya, no shit

(Observation --- Grand Mean)^2 = (Deviation due to Treatment)^2 + (*Deviation due to Obs within group)^2
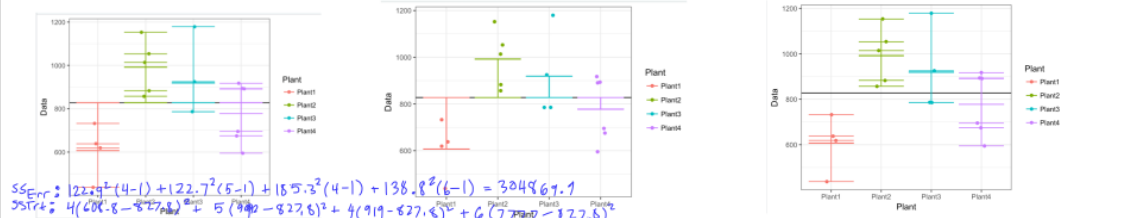
(Residual^2)

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{..}})^2 = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\overline{y_{i.}}-\overline{y_{..}})^2 + \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{i.}})^2 =$$

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{..}})^2 = \sum_{i=1}^{t}n_i(\overline{y_{i.}}-\bar{y}_{..})^2 + \sum_{i=1}^{t}(n_i-1)s_i^2 =$$

$$s_i^2 = \frac{(y_{ij}-\overline{y_{i.}})^2}{n_j-1}$$

**Total** Sum of Squares = **Treatment** Sum of Squares (between) + **Error** Sum of Squares (within)

SSTOT= SSTreat+SSError    683448.5 = 378610.4+ 304838.1  (Leaving Values Exact)



$SS_{Err}$: $122.9^2(4-1) + 122.7^2(5-1) + 185.3^2(4-1) + 138.8^2(6-1) = 304869.1$
$SSTrt$: $4(608.8-827.8)^2 + 5(992-827.8)^2 + 4(919-827.8)^2 + 6(779.0-827.8)^2$
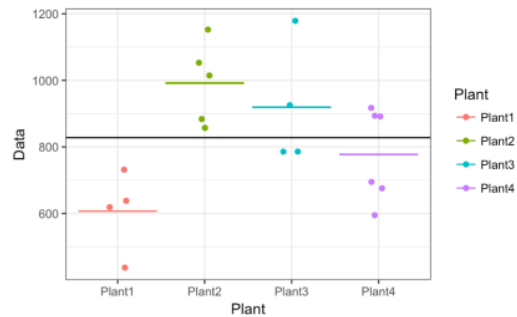
## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

We need to divide the squared errors by the appropriate degrees of freedom (df) so they are proper variances.



$$f_{Tot}=N-1 \quad MSTot = \frac{SSTot}{N-1} = \frac{1}{N-1}\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{..}})^2$$

$$f_{trt}=t-1 \quad MSTrt = \frac{SSTrt}{t-1} = \frac{1}{t-1}\sum_{i=1}^{t}n_i(\overline{y_{i.}}-\bar{y}_{..})^2$$

$$f_{Err}=N-t \quad MSE = \frac{SSError}{N-t} = \frac{1}{N-t}\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\overline{y_{i.}})^2$$

Anova Table

| Source | Sum of Squares | df | MS* |
|---|---|---|---|
| Treatment | SSTrt | t-1 | SSTrt/(t-1) |
| Error | SSError | N-t | SSErr/(N-t) |
| Total | SSTot | N-1 | |

Anova Table

| Source | Sum of Squares | df | MS* |
|---|---|---|---|
| Treatment | 378610.4 | 4-1=3 | 378610.4/3 |
| Error | 304838.1 | 19-4=15 | 304838.1/15 |
| Total | 683448.5 | or 19-1=18 | |

## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Under the null, our test statistic (F) is the ratio of between variability to the within variability

$$F = \frac{MSTrt}{MSE} = \frac{SSTrt/(t-1)}{SSErr/(N-t)}$$

If $H_o: \mu_1 = \mu_2 = \cdots \mu_t$ vs $\qquad H_A$: *at least one of the $\mu_i$ are different from another*

If MSTrt is large relative to MSE, then F is ____Bigger____ and we have ___more___ evidence against the null (and ____more____ evidence of a treatment effect).

If F is small, there is a lot of variability _unaccounted for by the group means_ which suggests there is no real difference between groups.

$$F_{obs} = \frac{MSTrt_{obs}}{MSE_{obs}} = \frac{378610.4/3}{304838.1/15} = 6.210024$$

But what are sizes of the values of F that we would expect to see even if the null is true?
_We need to look at F distribution to see what "extreme" is_

---

## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Under the null of our test statistic (and another big assumption we'll discuss soon...)

$$F_{(v_1=t-1, v_2=N-t)} = \frac{MSTrt}{MSE} = \frac{SSTrt/(t-1)}{SSError/(N-t)} = \frac{between\ treatment\ variability}{within\ treatment\ variability}$$

The sizes of the values of F that we would expect to see even if the null is true is determined by the Trt and Error degrees of freedom

One-Sided Test always? _yes, we will only find area above F_

Looking at an F table....

$$F_{obs} = \frac{MSTrt_{obs}}{MSE_{obs}} = \frac{378610.4/3}{304838.1/15} = 6.210024$$

$P(F_{(3,15)} > 6.21) =$ value between _.01 and .001_

Conclusion? _strong evidence that not all group means are equal_

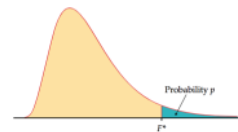Table entry for p is the critical value F* with probability p lying to its right.

**TABLE E**
**F critical values**

| | p | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| | .100 | 39.86 | 49.50 | 53.59 | 55.83 | 57.24 | 58.20 | 58.91 | 59.44 | 59.86 |
| | .050 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 | 240.54 |
| 1 | .025 | 647.79 | 799.50 | 864.16 | 899.58 | 921.85 | 937.11 | 948.22 | 956.66 | 963.28 |
| | .010 | 4052.2 | 4999.5 | 5403.4 | 5624.6 | 5763.6 | 5859.0 | 5928.4 | 5981.1 | 6022.5 |
| | .001 | 405284 | 500000 | 540379 | 562500 | 576405 | 585937 | 592873 | 598144 | 602284 |
| | .100 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 |
| | .050 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 |
| 2 | .025 | 38.51 | 39.00 | 39.17 | 39.25 | 39.30 | 39.33 | 39.36 | 39.37 | 39.39 |
| | .010 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 |
| | .001 | 998.50 | 999.00 | 999.17 | 999.25 | 999.30 | 999.33 | 999.36 | 999.37 | 999.39 |

15 .01 .001

## One-Way Analysis of Variance (ANOVA)

In R:

```
> anova<-aov(Data~Plant, data=Plantdata)
> Plantanova<-aov(Data~Plant, data=Plantdata)
> summary(Plantanova)
             Df Sum Sq Mean Sq F value  Pr(>F)
Plant         3 378610  126203    6.21 0.00592 **
Residuals    15 304838   20323
```
(Error)

```
> Plantlm<-lm(Data~Plant, data=Plantdata)
> anova(Plantlm)
Analysis of Variance Table

Response: Data
             Df Sum Sq Mean Sq F value   Pr(>F)
Plant         3 378610  126203    6.21 0.005917 **
Residuals    15 304838   20323
```
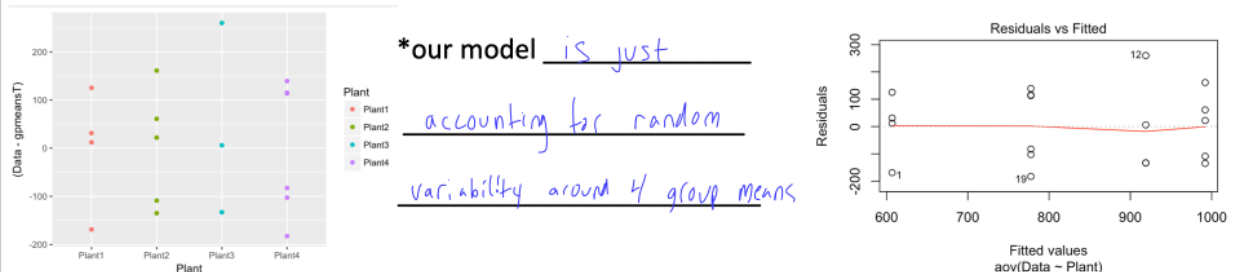
## One-Way Analysis of Variance (ANOVA)

Assumptions we are making:
(1) Groups are independent and observations within groups are independent/randomly selected
(2) MSE is like a pooled variance between the groups, so we need to make the assumption that the variances are approximately equal in each group.

      a. Just like we did for the T-test, we can graphically evaluate this by looking at dotplots of data by group and/or
      b. Dot plots of within-group error (residuals)

Residual = Actually Observed Value - Value The *Model Predicted It Should Have Been



*our model _is just_ _accounting for random_ _variability around 4 group means_

## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Assumptions we are making:
(2) MSE is like a pooled variance between the groups, so we need to make the assumption that the variances are ___equal___ in each population.

The F test is so powerful because pooling of variances. If we believe the treatment variances are equal, we get a much more accurate estimate of the within variability if we use all the data rather than just the data from two treatments at a time, and that estimate will be consistent over all pairs. The accuracy is reflected in the increased df of an F compared to each of the possible two-sample t test.

*We can also look at ratio of SDs :
      if ratio of smallest sample SD to largest sample SD is ___within___
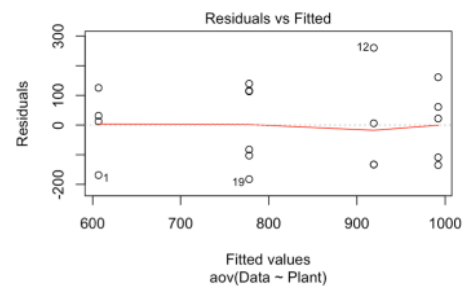___a factor of 2 (1/2)___, then assumption is reasonable.

Trt1 SD: 122.9
Trt2 SD: 122.7
Trt3 SD: 185.3
Trt4 SD: 138.8

$122.7/185.3 = 0.662$


Residuals vs Fitted
aov(Data ~ Plant)

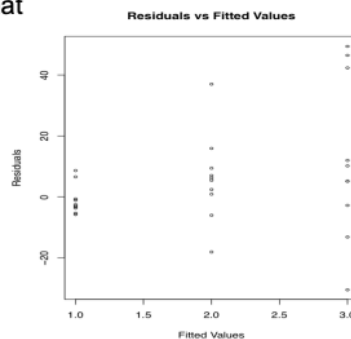## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Assumptions we are making:
(2) MSE is like a pooled variance between the groups, so we need to make the assumption that the variances are equal in each populations.

The reason we plot the residuals vs the fitted values is that it is often the case that variability will ___increase___ with increasing fitted values.

*As an extreme example, the SD of weights of 10 elephants will tend to be larger than the SD of weights of 10 hamsters. So, we like to plot against the fitted values to make it easier to see changes in variability of this type.

Here are some examples of residuals vs fitted plots that would indicate violation of constant variance:



"___Funnel___ pattern"
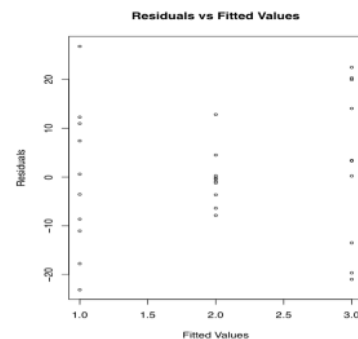
Unequal variance

## One-Way Analysis of Variance (ANOVA) [nitty gritty details]

Assumptions we are making:

(3) The distribution of $\frac{MSTrt}{MSE} = \frac{SSTrt/(t-1)}{SSError/(N-t)} = \frac{between\ treatment\ variability}{within\ treatment\ variability}$ is only $F_{(v_1=t-1, v2=N-t)}$

if the population of errors (residuals) is ___Normal___

> We check this again with the QQplot (and/or histogram) of our residuals

> Does this assumption seem to be well met? *Not great because we have a heavy tail, we will learn rank-based next time*
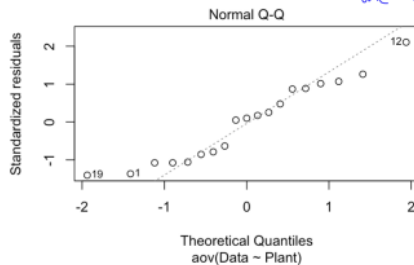


## Multiple Comparisons Following Significant ANOVA

Conclusions we can draw:

(1) If the F test is not significant, we do not reject the null and _we're done_.

(2) If the F test is significant, we can only say _at least two of the groups are significantly different_
(It doesn't tell us which groups though or by how much)

(3) One way to figure these things out is with _post-hoc pairwise tests_ that uses the pooled variance
(The F test is more powerful than regular two-sample t tests because if treatment variances are actually equal, we get a much more accurate estimate of the shared variance).

To compare treatments a and b we can:

A) compare the test statistic $t = \dfrac{\overline{y_a} - \overline{y_b}}{\sqrt{MSE\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}$ to a $t_{df_E}$ or

B) Make a $100(1-\alpha)\%$ CI on the difference of means: $\overline{y_a} - \overline{y_b} \pm t_{df_E, \alpha/2} \sqrt{MSE\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}$

*notice, you'll need to compute a different SE for the difference if there are different sample sizes in each group, but you'll get to use the t multiplier with the $df_E$

## Multiple Comparisons Following Significant ANOVA – Fisher's Least Significant Difference (LSD)

(2) Fisher's LSD method for finding CI's for the difference in group means (after a significant ANOVA)

$df_{Err} = 15$ and MSE=20322.54 and choose $\alpha = 0.05$, t multiplier = $\underline{t_{(0.025, 15)} = 2.13}$

For $\mu_1 - \mu_2$: $606.8 - 992.0 \pm 2.13 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{5}\right)} = -385.2 \pm 203.831 = (-589.03, -181.369)$ *0 not included

For $\mu_1 - \mu_3$: $606.8 - 919.0 \pm 2.13 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{4}\right)} = 312.2 \pm 214.86 = (-527.06, -97.34)$ *0 not included

For $\mu_1 - \mu_4$: $606.8 - 777.7 \pm 2.13 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{6}\right)} = -170.9 \pm 196.13 = (-367.04, 25.24)$ *0 included

For $\mu_2 - \mu_4$: $992.0 - 777.7 \pm 2.13 \sqrt{20322.54 \left(\frac{1}{5} + \frac{1}{6}\right)} = 214.3 \pm 183.99 = (30.31, 398.29).$ * 0 not included

For $\mu_2 - \mu_3$: $992.0 - 919.0 \pm 2.13 \sqrt{20322.54 \left(\frac{1}{5} + \frac{1}{4}\right)} = 73 \pm 203.831 = (-130.83, ,276.83)$ * 0 included

For $\mu_3 - \mu_4$: $919.0 - 777.7 \pm 2.13 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{6}\right)} = 141.3 \pm 196.13 = (-54.84, 337.43)$ *0 included

Anova Table

| Source | Sum of Squares | df | MS* |
|---|---|---|---|
| Treatment | 378610.4 | 4-1 | 378610.4 /3= 126203.5 |
| Error | 304838.1 | 19-4 | 304838.1 /15=20322.54 |
| Total | 683448.5 | 19-1 | |

| Plant 1 | | | | | $n_i$ | $\bar{y}_{i.}$ | $s_i$ |
|---|---|---|---|---|---|---|---|
| 438 | 619 | 732 | 638 | | 4 | 606.8 | 122.9 |
| Plant 2 | | | | | | | |
| 857 | 1014 | 1153 | 883 | 1053 | 5 | 992.0 | 122.7 |
| Plant 3 | | | | | | | |
| 925 | 786 | 1179 | 786 | | 4 | 919.0 | 185.3 |
| Plant 4 | | | | | | | |
| 893 | 891 | 917 | 695 | 675 595 | 6 | 777.7 | 138.8 |

Overall: $N = 19, \bar{y}_{..} = 827.8$

Ack
P1

## Multiple Comparisons Following Significant ANOVA – Fisher's Least Significant Difference (LSD)

(3) We often use a lettering system to show groups that are[not] significantly different from one another (Though this make us look how different/not different they are)

Increasing means and groups not significantly different share a letter.

| Treatment | Mean | Group (*Based on Fisher's LSD with $\alpha = 0.05$ for each CI) |
|---|---|---|
| Plant 1 | 606.8 | A |
| Plant 4 | 777.7 | A  B |
| Plant 3 | 919.0 | B  C |
| Plant 2 | 992.0 | C |

Intervals that contained 0: $\mu_1 - \mu_4$; $\mu_2 - \mu_3$; $\mu_3 - \mu_4$

Intervals that did not contain 0: $\mu_1 - \mu_2$; $\mu_1 - \mu_3$; $\mu_2 - \mu_4$

## Multiple Comparisons and Family-Wise Error Rates

Doing a large collection of tests greatly inflates our probability of making a __typa l error__ somewhere in the collection. (Think green jelly bean comic example)

- If any single test has probability α of making a Type I error (we call this the **Comparison-Wise Error Rate**, or **CW** for short), then when doing m independent tests, the probability of a Type I error on at least one test is __$1-(1-\alpha)^m$__ (we call this the **Family-Wise Error Rate**, or **FWER**).

- At a CW rate of 0.05, family-wise rate increases quickly:

| m | $\alpha_{FW}$ |
|---|---|
| 1 | 0.0500 |
| 2 | 0.0975 |
| 5 | 0.2262 |
| 10 | 0.4013 |

- If we want maximum power for detecting a difference and we are ok with a larger FWER than 5%, we can just leave the tests or CIs as they are.

If we want to bound our family-wise error rate, two common methods are the

a) Bonferroni Method        and      b) Tukey Method.

## One-Way Analysis of Variance (ANOVA) – Bounding Family Error Rates

1) **Bonferroni Method**

This method can be used for *any* collection of m tests, not just pairwise tests following a significant ANOVA.

Simply set the individual test comparison rate to $\frac{1}{m}th$ of the desired family error rate, ie $\alpha_{CW} =$ __$\frac{\alpha_{FW}}{m}$__

Practically, we can use this value in a few different ways:
- If using CI's, use the multiplier, $t_{df_E,(\frac{\alpha}{m})/2}$.
- If using p values, compute them as usual, but determine significance by comparing to __$\frac{\alpha_{FW}}{m}$__

- If using p values, multiply all p-values by m, and then determine significance by comparing to __$\alpha_{FW}$__

$$\text{CI for difference of } \mu_i - \mu_j \text{ is: } \overline{X_{i.}} - \overline{X_{j.}} \pm t_{df_E, (\frac{\alpha}{m})/2} * \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

In our case, the multiplier is t with $df_E = 19 - 4 = 15$ and

$\frac{\frac{\alpha}{m}}{2} = \frac{.05}{6} * \frac{1}{2} = 0.0042$ so we use the multiplier 3.03

```
> qt(p=.0042, df=15, lower.tail=FALSE)
[1] 3.032375
```

Bonferroni is easy, but if m is very large, the power can be very __small__ (CI very __large__ )

## One-Way Analysis of Variance (ANOVA) -Bounding Family Error Rates

### (1) Bonferonni Method for finding (multiple comparison) adjusted CI's for the difference in group means

MSE=20322.54, $df_E = 19 - 4 = 15$ and $\frac{\frac{\alpha}{m}}{2} = \frac{.05}{6} * \frac{1}{2} = 0.0042$ so we use the multiplier 3.03

For $\mu_1 - \mu_2$: $606.8 - 992.0 \pm 3.03 \sqrt{20322.54\left(\frac{1}{4} + \frac{1}{5}\right)} = -385.2 \pm 289.76 = (-674.96, -95.44)$ * no 0

For $\mu_1 - \mu_3$: $606.8 - 919.0 \pm 3.03 \sqrt{20322.54\left(\frac{1}{4} + \frac{1}{4}\right)} = -312.2 \pm 305.43 = (-617.63, -6.77)$ * no 0

For $\mu_1 - \mu_4$: $606.8 - 777.7 \pm 3.03 \sqrt{20322.54\left(\frac{1}{4} + \frac{1}{6}\right)} = -170.9 \pm 278.82 = (-449.72, 107.92)$ * has 0

For $\mu_2 - \mu_4$: $992.0 - 777.7 \pm 3.03 \sqrt{20322.54\left(\frac{1}{5} + \frac{1}{6}\right)} = 214.3 \pm 261.55 = (-47.25, 475.85)$ * has 0

For $\mu_2 - \mu_3$: $992.0 - 919.0 \pm 3.03 \sqrt{20322.54\left(\frac{1}{5} + \frac{1}{4}\right)} = 73 \pm 289.76 = (-216.76, 362.76)$ * has 0

For $\mu_3 - \mu_4$: $919.0 - 777.7 \pm 3.03 \sqrt{20322.54\left(\frac{1}{4} + \frac{1}{6}\right)} = 141.3 \pm 278.82 = (-137.52, 420.12)$ * has 0

| Treatment | Mean | | Group (*Based on Bonferroni adjusted CI) |
|---|---|---|---|
| Plant 1 | 606.8 | A | -Notice we are seeing wider CI than |
| Plant 4 | 777.7 | A B | unadjusted, and __fewer__ groupings. Groups |
| Plant 3 | 919.0 | B | __2 , 4__ are no longer significantly different. |
| Plant 2 | 992.0 | B | |

## One-Way Analysis of Variance (ANOVA) -Bounding Family Error Rates

### (2) Tukey Method for finding (multiple comparison) adjusted CI's for the difference in group means is a little more work, but better if we want to look at all pairwise tests (typically. more __powerful__ than Bonferroni while still making adjustment.)

For a Tukey CI,

we use a multiplier of $\frac{Q_{\alpha FW, t, df_E}}{\sqrt{2}}$ where Q is called the Studentized Range Distribution.

We will use an R function to get $Q_{\alpha FW, t, df_E}$ for us:

```
> qtukey(p=(1-0.05), nmeans=4, df=(19-4)) / sqrt(2)
[1] 2.882149
```

**CI for difference of $\mu_i - \mu_j$ is:** $\overline{X_{i.}} - \overline{X_{j.}} \pm \frac{q_{T,N-T,\alpha}}{\sqrt{2}} * \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

Reject $H_o: \mu_i - \mu_j = 0$ $iff$ the interval doesn't contain 0 or $|\overline{X_{i.}} - \overline{X_{j.}}| > \frac{q_{T,N-T,\alpha}}{\sqrt{2}} * \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

## One-Way Analysis of Variance (ANOVA) -Bounding Family Error Rates

(2) Tukey Method for finding (multiple comparison) adjusted CI's for the difference in group means

For $\mu_1 - \mu_2$: $606.8 - 992.0 \pm \frac{4.075974}{\sqrt{2}} \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{5}\right)} = -385.2 \pm 274.403 = (-659.6, -110.60)$ *no 0

For $\mu_1 - \mu_3$: $606.8 - 919.0 \pm 2.88 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{4}\right)} = -312.2 \pm 289.25 = (-601.44, -22.95)$ * no 0

For $\mu_1 - \mu_4$: $606.8 - 777.7 \pm 2.88 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{6}\right)} = -170.9 \pm 264.04 = (-434.94, 93.14)$ * has 0

For $\mu_2 - \mu_4$: $992.0 - 777.7 \pm 2.88 \sqrt{20322.54 \left(\frac{1}{5} + \frac{1}{6}\right)} = 214.3 \pm 247.696 = (-33.40, 462.0)$ * has 0

For $\mu_2 - \mu_3$: $992.0 - 919.0 \pm 2.88 \sqrt{20322.54 \left(\frac{1}{5} + \frac{1}{4}\right)} = 73 \pm 274.4034 = (-201.40, 347.40)$ * has 0

For $\mu_3 - \mu_4$: $919.0 - 777.7 \pm 2.88 \sqrt{20322.54 \left(\frac{1}{4} + \frac{1}{6}\right)} = 141.3 \pm 264.04 = (-122.74, 405.34)$ * has 0

| Treatment | Mean | | Group (*Based on Tukey adjusted CI) |
|-----------|------|------|-------------------------------------|
| Plant 1 | 606.8 | A | -Notice we are seeing slightly _narrower_ CI |
| Plant 4 | 777.7 | A B | but same groupings as Bonferroni. |
| Plant 3 | 919.0 | B | |
| Plant 2 | 992.0 | B | |

## One-Way Analysis of Variance (ANOVA) -Bounding Family Error Rates

Checking our work in R:

Tukey Groupings:

| Treatment | Mean | Group |
|-----------|------|-------|
| Plant 1 | 606.8 | A |
| Plant 4 | 777.7 | AB |
| Plant 3 | 919.0 | B |
| Plant 2 | 992.0 | B |

```
> TukeyHSD(Plantanova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Data ~ Plant, data = Plantdata)

$Plant
                 diff        lwr       upr     p adj
Plant2-Plant1  385.2500  109.62944 660.87056 0.0053783
Plant3-Plant1  312.2500   21.72042 602.77958 0.0332173
Plant4-Plant1  170.9167  -94.29934 436.13268 0.2866336
Plant3-Plant2  -73.0000 -348.62056 202.62056 0.8695632
Plant4-Plant2 -214.3333 -463.12800  34.46134 0.1036269
Plant4-Plant3 -141.3333 -406.54934 123.88268 0.4420141
```

## Kruskal-Wallis (Rank-based comparison of centers)

If the residuals give strong evidence group populations aren't normal, we again use the ranks to do an overall test of difference between groups, called a Kruskal-Wallis. If this is significant, we follow it by pairwise Wilcoxon ___Rank Sum___/Mann Whitney tests.

Assumptions:
1. n1 observations randomly selected from population 1,
   n2 observations randomly selected from population 2, etc

2. The populations are identical except for a possible difference in shift parameter. (have same shape and spread)

Computation of test statistic is complicated. As for Mann-Whitney, start by ranking all the data, regardless of treatment. Again assign average ranks if there are ties. The test statistic is based on how similar the average ranks are in each of the groups.

We'll use the kruskal.test()
Command in R.

```
> kruskal.test(Data~Plant, data=Plantdata)

        Kruskal-Wallis rank sum test

data:  Data by Plant
Kruskal-Wallis chi-squared = 9.2913, df = 3, p-value = 0.02566
```

## Kruskal-Wallis (Rank-based comparison of centers)

Since the overall Kruskal-Wallis was significant, we have evidence that at least 1 population is shifted from another population.

### Lets do Wilcoxon Rank Sum pairwise comparisons

```
        Wilcoxon rank sum test

data:  Plant1 and Plant2
W = 0, p-value = 0.01587

data:  Plant1 and Plant3
W = 0, p-value = 0.0294

data:  Plant1 and Plant4
W = 5, p-value = 0.1714

data:  Plant2 and Plant3
W = 13, p-value = 0.5386

data:  Plant2 and Plant4
W = 24, p-value = 0.1255

data:  Plant3 and Plant4
W = 18, p-value = 0.2395
```

Notice, when we compare each to p value to 0.05, we find have evidence of a significant difference in shift between Plants _1 and 2_ and between Plants _1 and 3_ (similar to Bonferonni and Tukey before).

When we adjust for the multiple comparisons and use the Bonferroni adjustment (compare each p value to _0.5/6 = 0.0083_), ____none____ of the pairwise comparisons are significant.

Because Wilxocon is only using ranks, with the small sample sizes, it has ___low___ power to find a difference between the groups.