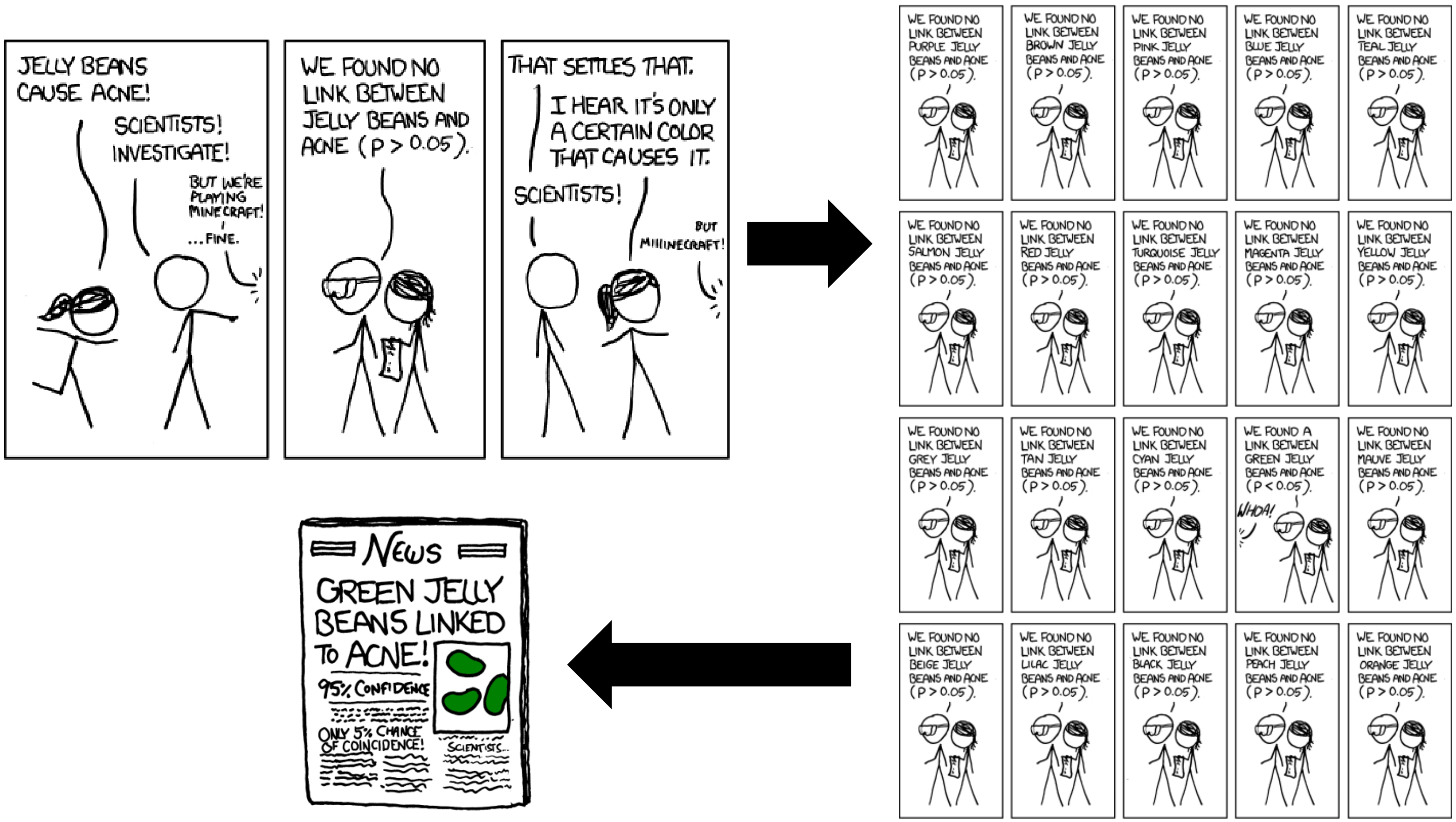
# Multi-Way Analysis of Variance (ANOVA)




“An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem”

John Tukey (American Mathmetician)

# Multi-way ANOVA

- Just like one-way ANOVA but with more than one treatment

- Each treatment may still have many levels (e.g. VARIETY – A,B,C and FARM – farm1, farm2)

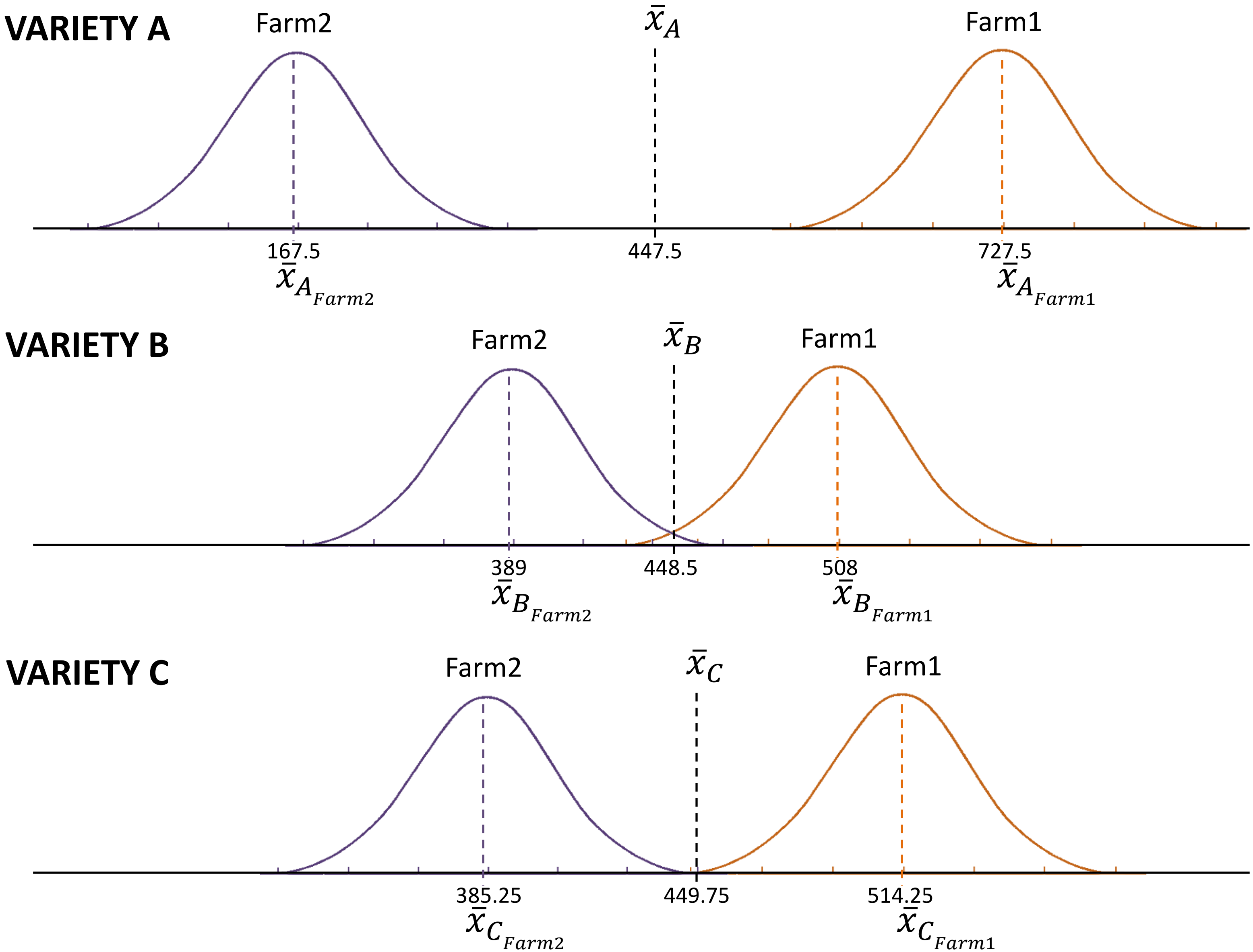
- We do not only look at the effect of each treatment – *but* we must also look at the <u>interaction between treatments</u>

- Like the one-way ANOVA we use the F-statistic

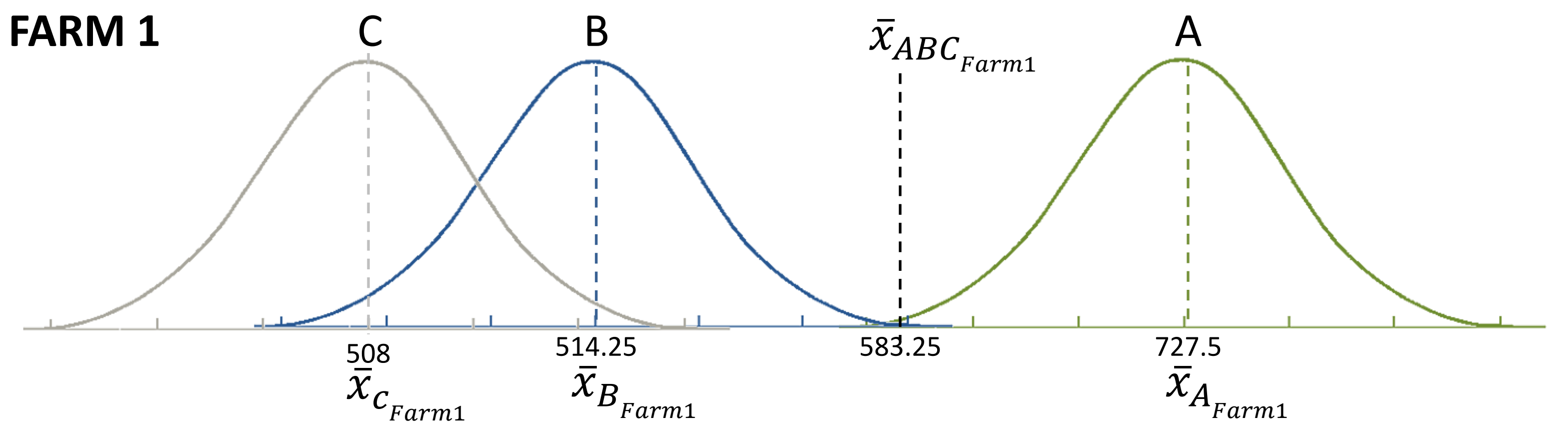
# Multi-way ANOVA

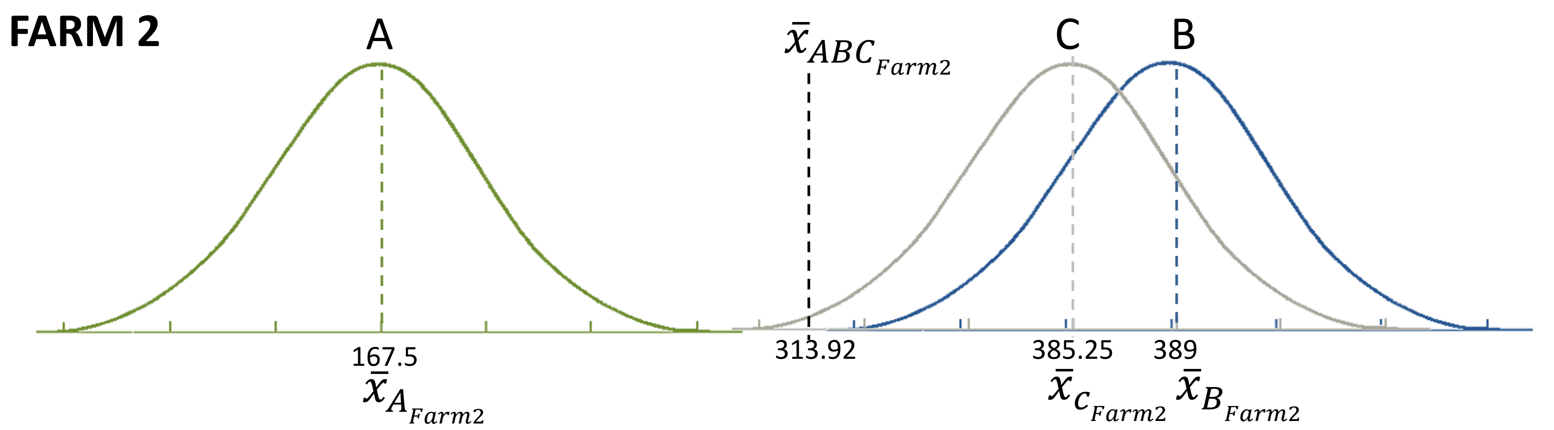
| Source of Variation | *df* | Sum of Squares | Mean Squares | F-value |
|---|---|---|---|---|
| A (e.g. VARIETY) | $t_A$-1 | $SS_A=MS_A*df_A$ | $MS_A$=signal$_A$ | $MS_A/MS_{ERROR}$ |
| B (e.g. FARM) | $t_B$-1 | $SS_B=MS_B*df_B$ | $MS_B$=signal$_B$ | $MS_B/MS_{ERROR}$ |
| A X B (interaction) | $(t_A-1)*(t_B-1)$ | $SS_{AXB}=MS_{AXB}*df_{AXB}$ | $MS_{AXB}$=signal$_{AXB}$ | $MS_{AXB}=MS_{ERROR}$ |
| Error | $n-(t_A*t_B)$ | $SS_{ERROR}=MS_{ERROR}*df_{ERROR}$ | $MS_{ERROR}$=noise | |

# Lentil Example – Treatment A



# Lentil Example – Treatment B



# Lentil Example

| VARIETY\FARM | FARM 1 | FARM2 | Marginal Mean |
|---|---|---|---|
| **A** | 720, 690, 740, 760 (727.5) | 163, 176, 163, 168 (167.5) | 447.5 |
| **B** | 515, 480, 545, 492 (508) | 375, 389, 405, 387 (389) | 448.5 |
| **C** | 540, 502, 510, 505 (514.25) | 375, 385, 381, 400 (385.25) | 449.75 |
| **Marginal Mean** | 583.25 | 313.9167 | 448.5833 |

# Lentil Example – Main Effects

$$MS_A = \frac{r * t_B * \sum_i^{t_A} \left(\bar{x}_{t_{Ai}} - \bar{x}_{ALL}\right)^2}{t_A - 1}$$

$$MS_B = \frac{r * t_A * \sum_i^{t_B} \left(\bar{x}_{t_{Bi}} - \bar{x}_{ALL}\right)^2}{t_B - 1}$$

| VARIETY\FARM | FARM1 | FARM2 | Marginal Mean |
|---|---|---|---|
| A | 720, 690, 740, 760 (727.5) | 163, 176, 163, 168 (167.5) | 447.5 |
| B | 515, 480, 545, 492 (508) | 375, 389, 405, 387 (389) | 448.5 |
| C | 540, 502, 510, 505 (514.25) | 375, 385, 381, 400 (385.25) | 449.75 |
| Marginal Mean | 583.25 | 313.9167 | 448.5833 |

### Treatment A - VARIETY

$$MS_{VARIETY} = \frac{4 * 2 * [(447.5 - 448.5833)^2 + (448.5 - 448.5833)^2 + (449.75 - 448.5833)^2]}{3 - 1}$$

$$MS_{VARIETY} = \frac{4 * 2 * 2.5416667}{2} = \mathbf{10.1667}$$

### Treatment B - FARM

$$MS_{FARM} = \frac{4 * 3 * [(583.25 - 448.5833)^2 + (313.9167 - 448.5833)^2]}{2 - 1}$$

$$MS_{FARM} = \frac{4 * 3 * 36270.21324445}{1} = \mathbf{435242.559}$$

# Lentil Example – Interaction

$$MS_{AXB} = \frac{r * \sum_i^{t_A} \sum_j^{t_B} \left(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}_{ALL}\right)^2}{(t_A - 1) * (t_B - 1)}$$

| VARIETY\FARM | FARM1 | FARM2 | Marginal Mean |
|---|---|---|---|
| A | 720, 690, 740, 760 (727.5) | 163, 176, 163, 168 (167.5) | 447.5 |
| B | 515, 480, 545, 492 (508) | 375, 389, 405, 387 (389) | 448.5 |
| C | 540, 502, 510, 505 (514.25) | 375, 385, 381, 400 (385.25) | 449.75 |
| Marginal Mean | 583.25 | 313.9167 | 448.5833 |

## Interaction – VARIETY x FARM

$$MS_{VARIETY\ x\ FARM} = \frac{4 * \begin{bmatrix} (727.5 - 447.5 - 583.25 + 448.5833)^2 + (167.5 - 447.5 - 313.9167 + 448.5833)^2 \\ +(508 - 448.5 - 583.25 + 448.5833)^2 + (389 - 448.5 - 313.9167 + 448.5833)^2 \\ +(514.25 - 449.75 - 583.25 + 448.5833)^2 + (385.25 - 449.75 - 313.9167 + 448.5833)^2 \end{bmatrix}}{(3-1) * (2-1)}$$

$$MS_{VARIETY\ x\ FARM} = \frac{4 * 63390.33333335}{2} = \mathbf{126780.66667}$$

# Lentil Example – Error

$$MS_{ERROR} = \frac{\sum_i^{t_A} \sum_j^{t_B} \sum_k^{r} \left(x_{ijk} - \bar{x}_{ij}\right)^2}{n - (t_A * t_B)}$$

| VARIETY\FARM | FARM1 | FARM2 | Marginal Mean |
|---|---|---|---|
| A | 720, 690, 740, 760 (727.5) | 163, 176, 163, 168 (167.5) | 447.5 |
| B | 515, 480, 545, 492 (508) | 375, 389, 405, 387 (389) | 448.5 |
| C | 540, 502, 510, 505 (514.25) | 375, 385, 381, 400 (385.25) | 449.75 |
| Marginal Mean | 583.25 | 313.9167 | 448.5833 |

**Error**

$$MS_{ERROR} = \frac{\begin{bmatrix} (720-727.5)^2 + (690-727.5)^2 + (740-727.5)^2 + (760-727.5)^2 \\ +(163-167.5)^2 + (176-167.5)^2 + (163-167.5)^2 + (168-167.5)^2 \\ +(515-508)^2 + (480-508)^2 + (545-508)^2 + (492-508)^2 \\ +(375-389)^2 + (385-389)^2 + (405-389)^2 + (387-389)^2 \\ +(540-514.25)^2 + (502-514.25)^2 + (510-514.25)^2 + (505-514.25)^2 \\ +(375-385.25)^2 + (385-385.25)^2 + (381-385.25)^2 + (400-385.25)^2 \end{bmatrix}}{24 - (3 * 2)}$$

$$MS_{ERROR} = \frac{2282058}{18} = \mathbf{126781}$$

# How to report results from a Multi-way ANOVA

R Console

```
> anova(lm(data$YIELD~data$VARIETY*data$FARM))
Analysis of Variance Table

Response: data$YIELD
                        Df Sum Sq Mean Sq   F value    Pr(>F)    
data$VARIETY             2     20      10    0.0263    0.9741    
data$FARM                1 435243  435243 1125.7085 < 2.2e-16 ***
data$VARIETY:data$FARM   2 253561  126781  327.9046 6.928e-15 ***
Residuals               18   6959     387                        
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

| Source of Variation | *df* | Sum of Squares | Mean Squares | F-value | P-value |
|---|---|---|---|---|---|
| Variety (A) | 2 | 20 | 10 | 0.0263 | 0.9741 |
| Farm (B) | 1 | 435243 | 43243 | 1125.7085 | <0.05 |
| Variety x Farm (AxB) | 2 | 253561 | 126781 | 327.9046 | <0.05 |
| Error | 18 | 6959 | 387 | | |

**Multi-way ANOVA in R:**

```
anova(lm(YIELD~VARIETY*FARM))
anova(lm(YIELD~VARIETY+FARM)+VARIETY:FARM)
```

# How to report results from a Multi-way ANOVA

R Console

```
> anova(lm(data$YIELD~data$VARIETY*data$FARM))
Analysis of Variance Table

Response: data$YIELD
                        Df Sum Sq Mean Sq   F value    Pr(>F)
data$VARIETY             2     20      10    0.0263    0.9741
data$FARM                1 435243  435243 1125.7085 < 2.2e-16 ***
data$VARIETY:data$FARM   2 253561  126781  327.9046 6.928e-15 ***
Residuals               18   6959     387
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

| Source of Variation | *df* | Sum of Squares | Mean Squares | F-value | P-value |
|---|---|---|---|---|---|
| Variety (A) | 2 | 20 | 10 | 0.0263 | 0.9741 |
| Farm (B) | 1 | 435243 | 43243 | 1125.7085 | <0.05 |
| Variety x Farm (AxB) | 2 | 253561 | 126781 | 327.9046 | <0.05 |
| Error | 18 | 6959 | 387 | | |

P-values:
$pt(F_A, df_A, df_{ERROR})$
$pt(F_B, df_B, df_{ERROR})$
$pt(F_{AxB}, df_{AxB}, df_{ERROR})$

F-values:
$MS_A/MS_{ERROR}$
$MS_B/MS_{ERROR}$
$MS_{AxB}/MS_{ERROR}$

Sum of Squares:
$MS_A * df_A$
$MS_B * df_B$
$MS_{AxB} * df_{AxB}$

Error Sum of Squares:
$MS_{ERROR} * df_{ERROR}$

# How to report results from a Multi-way ANOVA

R Console

```
> anova(lm(data$YIELD~data$VARIETY*data$FARM))
Analysis of Variance Table

Response: data$YIELD
                        Df Sum Sq Mean Sq   F value    Pr(>F)
data$VARIETY             2     20      10    0.0263    0.9741
data$FARM                1 435243  435243 1125.7085 < 2.2e-16 ***
data$VARIETY:data$FARM   2 253561  126781  327.9046 6.928e-15 ***
Residuals               18   6959     387
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

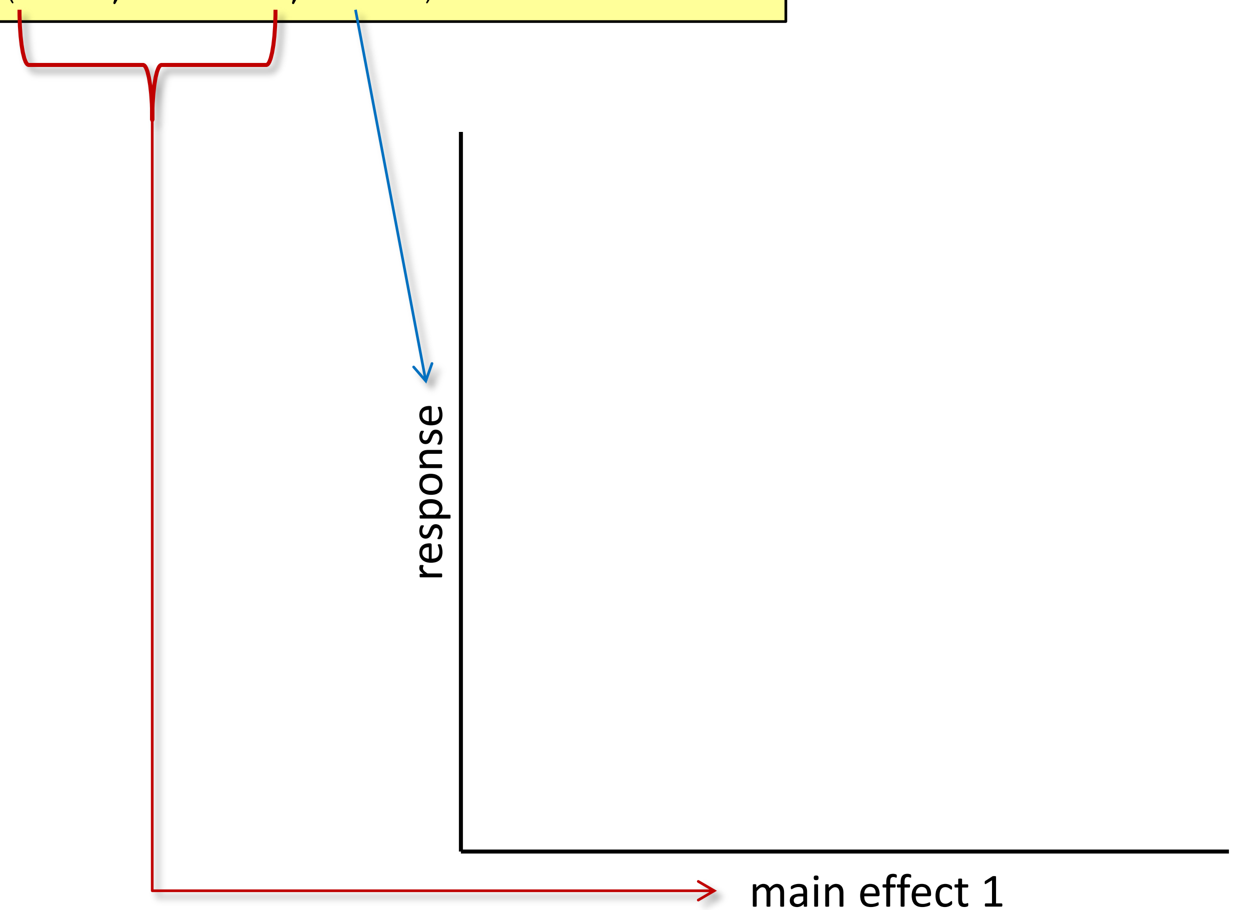
| Source of Variation | *df* | Sum of Squares | Mean Squares | F-value | P-value |
|---|---|---|---|---|---|
| Variety (A) | 2 | 20 | 10 | 0.0263 | 0.9741 |
| Farm (B) | 1 | 435243 | 43243 | 1125.7085 | <0.05 |
| Variety x Farm (AxB) | 2 | 253561 | 126781 | 327.9046 | <0.05 |
| Error | 18 | 6959 | 387 | | |

**If the interaction is significant – you should ignore the main effects because the story is not that simple!**

# Interaction plots – Different story under different conditions

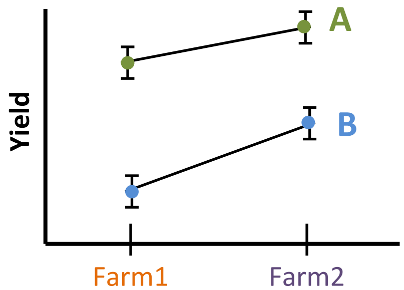
Interaction plot in R:

```
interaction.plot(mainEffect1,mainEffect2,response)
interaction.plot(FARM,VARIETY,YIELD)
```



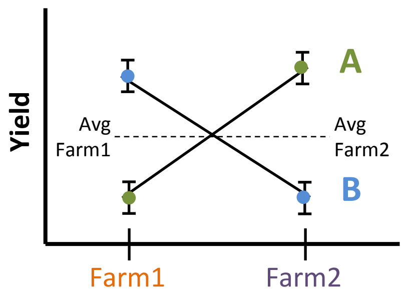
# Interaction plots – Different story under different conditions

1.



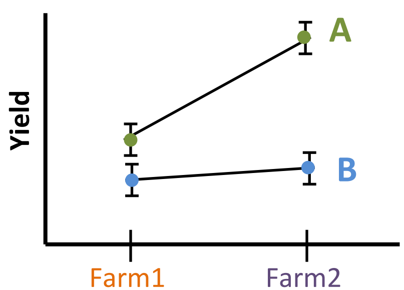

- VARIETY is significant (*)
- FARM is significant (*)
- FARM2 has better yield than FARM1
- **No Interaction**

2.



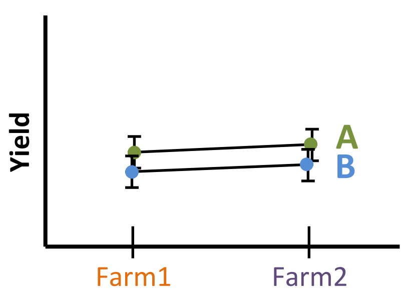

- VARIETY is not significant
- FARM is significant (*)
- VARIETY A is better on FARM2 and VARIETY B is better on FARM1
- **Significant Interaction**

3.




- VARIETY is significant (*)
- FARM is significant (*) – small difference
- Main effects are significant, BUT hard to interpret with overall means
- **Significant Interaction**

4.




- VARIETY is not significant
- FARM is not significant
- Cannot distinguish a difference between VARIETY or FARM
- **No Interaction**

# **Interaction plots – Different story under different conditions**

- An interaction detects non-parallel lines
- Difficult to interpret interaction plots for more than a 2-WAY ANOVA
- If the interaction effect is NOT significant then you can just interpret the main effects
- BUT if you find a significant interaction you don’t want to interpret main effects because the combination of treatment levels results in different outcomes

# Pairwise comparisons – What to do when you have an interaction

a.k.a Pairwise t-tests

Lentil Example: 3 VARITIES (A, B, and C)

$$C = \frac{t(t-1)}{2} = \frac{3(2)}{2} = \mathbf{3}$$

- A – B
- A – C
- B – C

Number of comparisons:

$$C = \frac{t(t-1)}{2}$$

$$t = number\ of\ treatment\ levels$$

Probability of making a Type I Error in **at least one** comparison = 1 – probability of making no Type I Error at all

Experiment-wise Type I Error for $\alpha$ = 0.05:

$$probability\ of\ Type\ I\ Error = 1 - 0.95^C$$

Lentil Example:

$$\begin{aligned} probability\ of\ Type\ I\ Error &= 1 - 0.95^3 \\ &= 1 - 0.87 \\ &= \mathbf{0.13} \end{aligned}$$

Significantly increased probability of making an error!

**Therefore pairwise comparisons leads to compromised experiment-wise $\alpha$-level**

# Pairwise comparisons – What to do when you have an interaction

a.k.a Pairwise t-tests

Another Example in R:

Untitled - R Editor

```
# Generate 2 sets of random numbers from the Normal Distribution
A=rnorm(10)
B=rnorm(10)

# Calculate if there is a difference between them
t.test(A,B)
```

**There should NOT be a significant difference between these 2 groups**
**Did anyone get a significant difference?**

# Pairwise comparisons – What to do when you have an interaction

a.k.a Pairwise t-tests

Another Example in R:

- If we have $\alpha$ = 0.05, this indicates that you will get a difference as big or bigger 1 out of every 20 times

- Now out of the 24 people in this room (C=24), what is the probability of getting at least one significant p-value (false positive)?

HARD to Answer

- What is the probability of NOT getting at least one significant p-value (false positive)?

EASIER to Answer

Class Example:

$$probability\ of\ Type\ I\ Error = 1 - 0.95^{24}$$
$$= 1 - 0.29$$
$$= \mathbf{0.71}$$

**That's a 71% chance of making an error!!!!**

**We need to adjust our $\alpha$ and p-values to correct for this bias!**

# Benferroni Adjustment – Adjust α-level for multiple comparisons

Benferroni Adjustment:

$$\alpha_{adj} = \frac{\alpha}{C}$$

- New α accounts for multiple comparisons
- Our new cutoff for significance

Untitled - R Editor

```r
# Generate 2 sets of random numbers from the Normal Distribution
A=rnorm(10)
B=rnorm(10)

# Calculate if there is a difference between them
t.test(A,B)
```

Class Example:

$$\alpha_{adj} = \frac{0.05}{24} = 0.0208\bar{3}$$

But now we evaluate significance at this value

**Now at this new significance level…Did anyone get a significant difference?**

# Pairwise comparisons – Tukey Honest Differences (Test)

a.k.a Pairwise t-tests with adjusted p-values

Pairwise comparisons in R:

```
lentil.model=aov (lm(YIELD~VARIETY*FARM))
TukeyHSD(lentil.model)
```

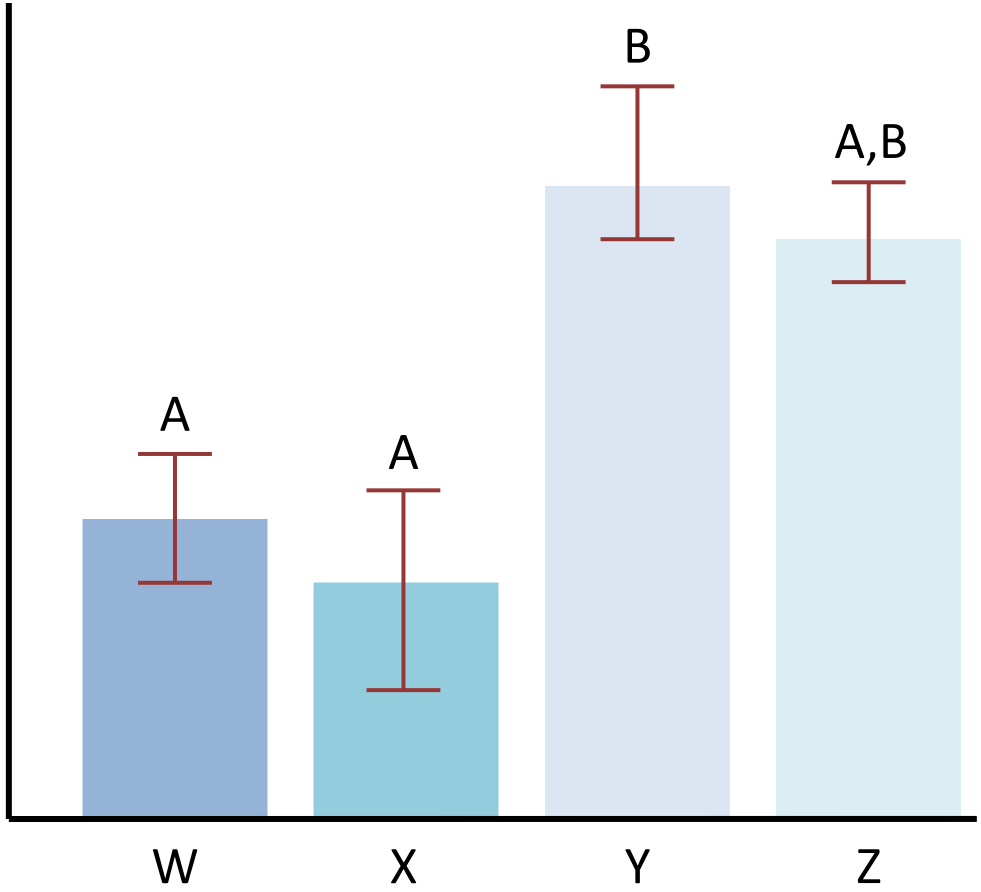
R Console

```
> lentil.model <- aov(lm(YIELD~FARM*VARIETY))
> TukeyHSD(lentil.model)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = lm(YIELD ~ FARM * VARIETY))

$FARM
                 diff       lwr       upr p adj
Farm2-Farm1 -269.3333 -286.1984 -252.4683     0

$VARIETY
    diff       lwr      upr     p adj
B-A 1.00 -24.09176 26.09176 0.9943144
C-A 2.25 -22.84176 27.34176 0.9715857
C-B 1.25 -23.84176 26.34176 0.9911321

$`FARM:VARIETY`
                   diff        lwr        upr     p adj
Farm2:A-Farm1:A -560.00 -604.18719 -515.81281 0.0000000
Farm1:B-Farm1:A -219.50 -263.68719 -175.31281 0.0000000
Farm2:B-Farm1:A -338.50 -382.68719 -294.31281 0.0000000
Farm1:C-Farm1:A -213.25 -257.43719 -169.06281 0.0000000
Farm2:C-Farm1:A -342.25 -386.43719 -298.06281 0.0000000
Farm1:B-Farm2:A  340.50  296.31281  384.68719 0.0000000
Farm2:B-Farm2:A  221.50  177.31281  265.68719 0.0000000
Farm1:C-Farm2:A  346.75  302.56281  390.93719 0.0000000
Farm2:C-Farm2:A  217.75  173.56281  261.93719 0.0000000
Farm2:B-Farm1:B -119.00 -163.18719  -74.81281 0.0000012
Farm1:C-Farm1:B    6.25  -37.93719   50.43719 0.9972898
Farm2:C-Farm1:B -122.75 -166.93719  -78.56281 0.0000008
Farm1:C-Farm2:B  125.25   81.06281  169.43719 0.0000006
Farm2:C-Farm2:B   -3.75  -47.93719   40.43719 0.9997694
Farm2:C-Farm1:C -129.00 -173.18719  -84.81281 0.0000004

>
```

<u>**If**</u> we have NO significant interaction effect – we can just look at the main effects

<u>**If**</u> we have a significant interaction effect – use these values

# How to report a significant difference in a graph




Create a matrix of significance and use it to code your graph

| | W | X | Y | Z |
|---|---|---|---|---|
| W | - | NS | * | NS |
| X | | - | * | NS |
| Y | | | - | NS |
| Z | | | | - |

Same letter = non significant
Different letter = significant