



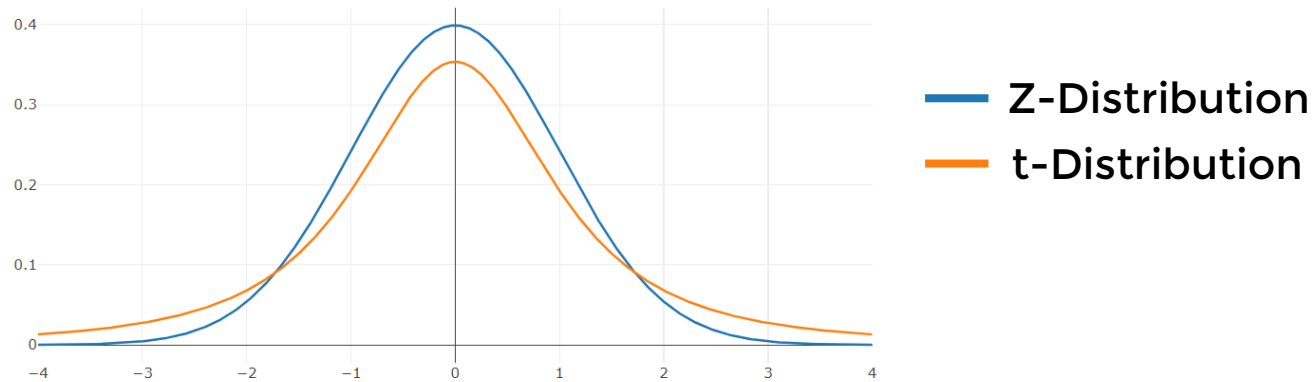
Probability and Statistics for Business and Data

PART 5 - ANOVA



Analysis of Variance

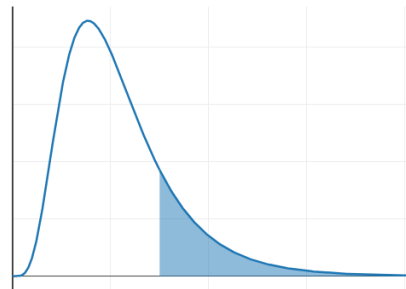
- In the previous section we used Z- and t-Distributions to answer the question “*What is the probability that two samples come from the same population?*”





Analysis of Variance

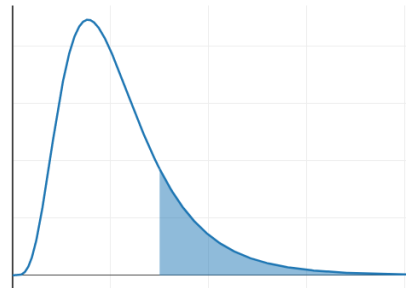
- In this section we introduce a new distribution – the F-Distribution
- Used to answer the question
“What is the probability that two samples come from populations that have the same variance?”





Analysis of Variance

- In this section we introduce a new distribution – the F-Distribution
- Can also answer the question
*“What is the probability that **three or more** samples come from the same population?”*





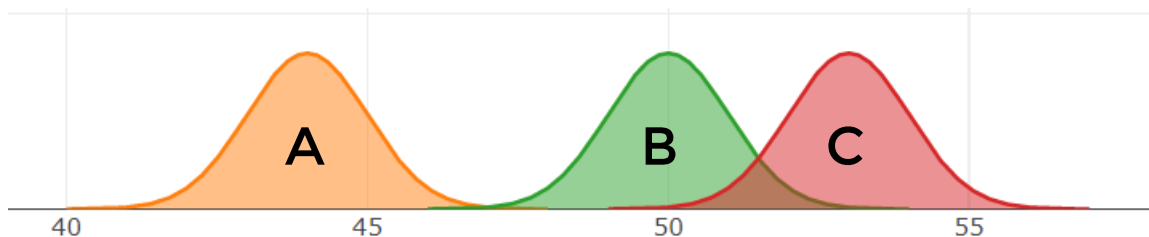
ANOVA

Analysis of Variance



ANOVA

- In the previous section we tested two samples to see if they likely came from the same parent population.
- What if we had three (or more) samples?
- Could we do the same thing?

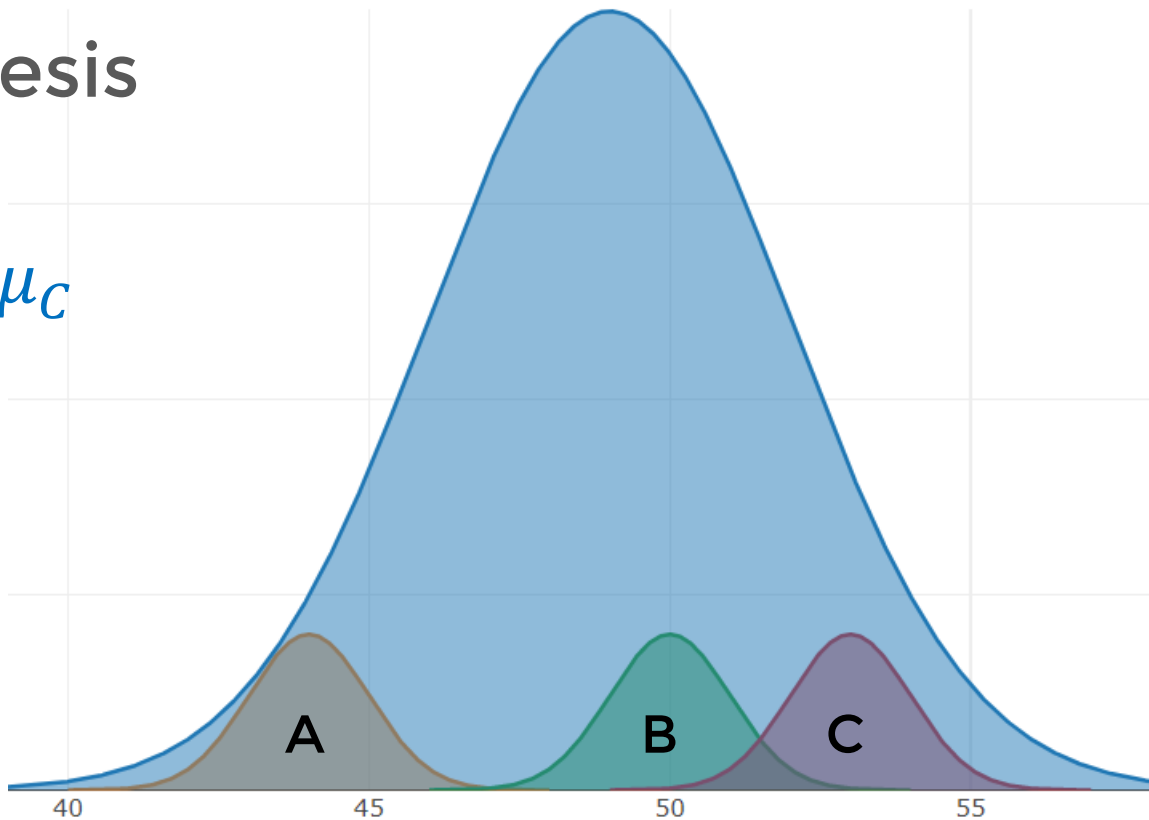




ANOVA

- Our null hypothesis would look like:

$$H_0: \mu_A = \mu_B = \mu_C$$





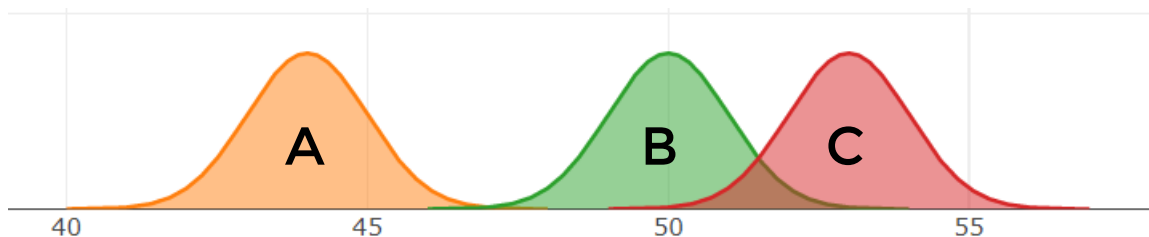
ANOVA

- We *could* test each pair:

$$H_0: \mu_A = \mu_B \quad \alpha = 0.05$$

$$H_0: \mu_A = \mu_C \quad \alpha = 0.05$$

$$H_0: \mu_B = \mu_C \quad \alpha = 0.05$$





ANOVA

- The problem is, our overall confidence drops:

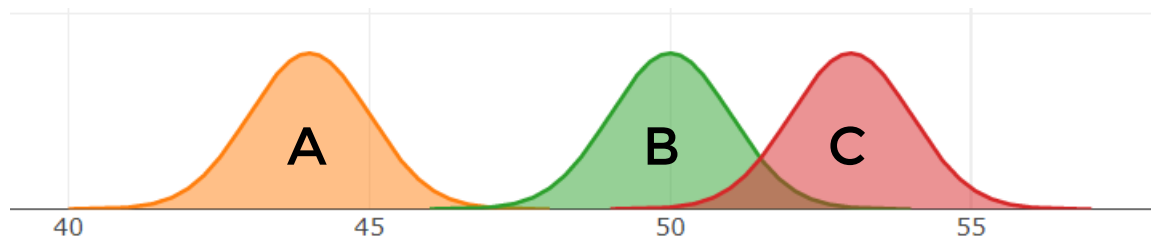
$$H_0: \mu_A = \mu_B \quad \alpha = 0.05$$

$$H_0: \mu_A = \mu_C \quad \alpha = 0.05$$

$$H_0: \mu_B = \mu_C \quad \alpha = 0.05$$

$$.95 \times .95 \times .95 = .857$$

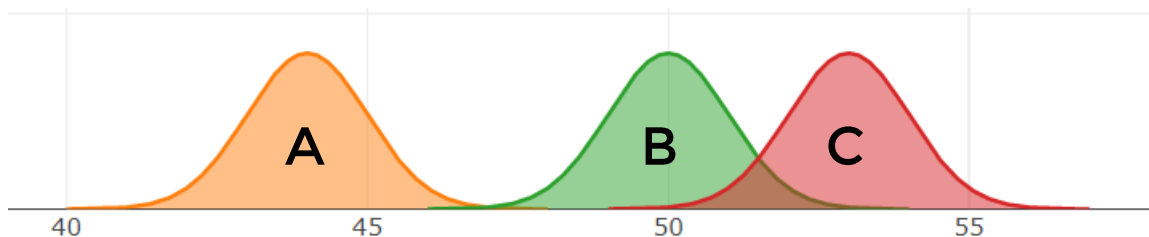
85.7% confidence level





ANOVA

- This is where ANOVA comes in!
- We compute an **F value**, and compare it to a critical value determined by our **degrees of freedom** (the number of groups, and the number of items in each group)

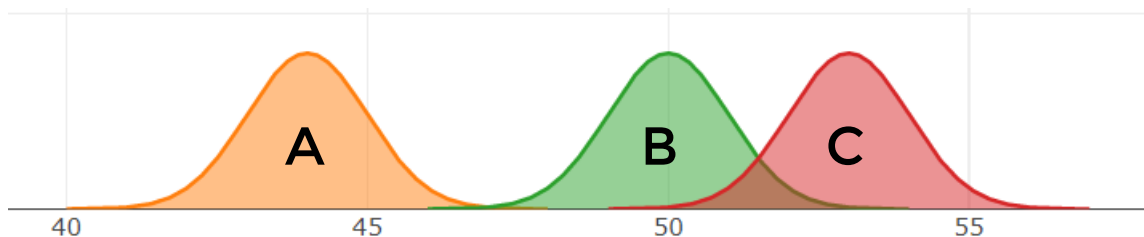




ANOVA

Let's work with some data:

GroupA	GroupB	GroupC
37	62	50
60	27	63
52	69	58
43	64	54
40	43	49
52	54	52
55	44	53
39	31	43
39	49	65
23	57	43





ANOVA

First calculate the sample means

Next calculate the overall mean

	GroupA	GroupB	GroupC
	37	62	50
	60	27	63
	52	69	58
	43	64	54
	40	43	49
	52	54	52
	55	44	53
	39	31	43
	39	49	65
	23	57	43
$\mu_{A,B,C}$	44	50	53
μ_{TOT}	49		



ANOVA

ANOVA considers two types of **variance**:

Between Groups

how far group means stray
from the total mean

Within Groups

how far individual values stray
from their respective group mean



ANOVA

The F value we're trying to calculate is simply the ratio between these two variances!

$$F = \frac{\textit{Variance Between Groups}}{\textit{Variance Within Groups}}$$



ANOVA

Recall the equation for variance:

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{SS}{df}$$

Here $\Sigma(x - \bar{x})^2$ is the “sum of squares” SS
and $n - 1$ is the “degrees of freedom” df



ANOVA

So the formula for the F value becomes:

$$F = \frac{\text{Variance Between Groups}}{\text{Variance Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

SSG = Sum of Squares Groups

SSE = Sum of Squares Error

df_{groups} = degrees of freedom (groups)

df_{error} = degrees of freedom (error)



ANOVA

$$SSG = 420$$

Sum of Squares Groups

$$(\mu_A - \mu_{TOT})^2 = (44 - 49)^2 = 25$$

$$(\mu_B - \mu_{TOT})^2 = (50 - 49)^2 = 1$$

$$(\mu_C - \mu_{TOT})^2 = (53 - 49)^2 = 16$$

42

Multiply by the number of items in each group:

$$42 \times 10 = 420$$

GroupA	GroupB	GroupC
37	62	50
60	27	63
52	69	58
43	64	54
40	43	49
52	54	52
55	44	53
39	31	43
39	49	65
23	57	43
$\mu_{A,B,C}$	44	50
μ_{TOT}	49	53



ANOVA

$$SSG = 420$$

$$df_{groups} = 2$$

Degrees of Freedom Groups

$$df_{groups} = n_{groups} - 1$$

$$= 3 - 1$$

$$= 2$$

	GroupA	GroupB	GroupC
	37	62	50
	60	27	63
	52	69	58
	43	64	54
	40	43	49
	52	54	52
	55	44	53
	39	31	43
	39	49	65
	23	57	43
$\mu_{A,B,C}$	44	50	53
μ_{TOT}	49		



ANOVA

$$\begin{aligned}SSG &= 420 \\df_{groups} &= 2 \\SSE &= 3300\end{aligned}$$

Sum of Squares Error

$(x_A - \mu_A)^2$	$(x_A - \mu_A)^2$	$(x_B - \mu_B)^2$	$(x_B - \mu_B)^2$	$(x_C - \mu_C)^2$	$(x_C - \mu_C)^2$
49	64	144	16	9	1
256	121	529	36	100	0
64	25	361	361	25	100
1	25	196	1	1	144
16	441	49	49	16	100
1062			1742		496

TOTAL

3300

GroupA	GroupB	GroupC
37	62	50
60	27	63
52	69	58
43	64	54
40	43	49
52	54	52
55	44	53
39	31	43
39	49	65
23	57	43
$\mu_{A,B,C}$	44	50
μ_{TOT}	49	

$$\begin{aligned}(37-44)^2 \\&= (-7)^2 \\&= 49\end{aligned}$$



ANOVA

$$SSG = 420$$

$$df_{groups} = 2$$

$$SSE = 3300$$

$$df_{error} = 27$$

Degrees of Freedom Error

$$df_{error} = (n_{rows} - 1) * n_{groups}$$

$$= (10 - 1) * 3$$

$$= 27$$

	GroupA	GroupB	GroupC
	37	62	50
	60	27	63
	52	69	58
	43	64	54
	40	43	49
	52	54	52
	55	44	53
	39	31	43
	39	49	65
	23	57	43
$\mu_{A,B,C}$	44	50	53
μ_{TOT}	49		



ANOVA

$$\begin{aligned}SSG &= 420 \\df_{groups} &= 2 \\SSE &= 3300 \\df_{error} &= 27\end{aligned}$$

Plug these into our formula:

$$F = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}} = \frac{\frac{420}{2}}{\frac{3300}{27}} = \frac{210}{122.22} = \mathbf{1.718}$$

GroupA	GroupB	GroupC
37	62	50
60	27	63
52	69	58
43	64	54
40	43	49
52	54	52
55	44	53
39	31	43
39	49	65
23	57	43
44	50	53
49		



ANOVA with Excel Data Analysis

	A	B	C	D	E		
1	Anova: Single Factor						
2							
3	SUMMARY						
4	Groups	Count	Sum	Average	Variance		
5	GroupA	10	440	44	118		
6	GroupB	10	500	50	193.5555556		
7	GroupC	10	530	53	55.11111111		
8							
9							
10	ANOVA						
11	Source of Variation	SS	df	MS	F	P-value	F crit
12	Between Groups	420	2	210	1.718181818	0.198430533	3.354130829
13	Within Groups	3300	27	122.2222			
14							
15	Total	3720	29				
16							

Data Analysis

Analysis Tools

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram

OK

Cancel

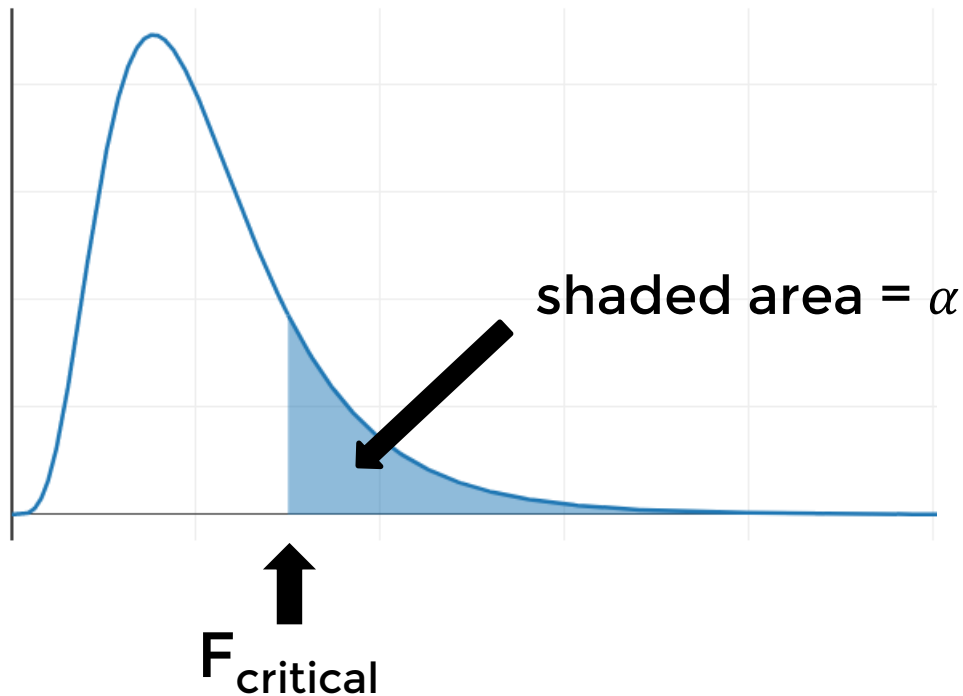
Help



F Distribution



F-Distribution





F-Distribution

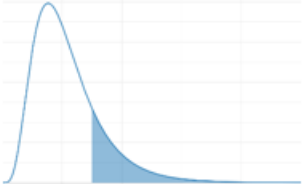
Look up our critical value from an F-table

use a table set for
95% confidence

find numerator df

find denominator df

critical value = 3.35

 denominator df		F-Table Upper Tail Area of 0.05				
		Numerator df				
		1	2	3	4	5
25	25	4.24	3.39	2.99	2.76	2.60
26	26	4.23	3.37	2.98	2.74	2.59
27	27	4.21	3.35	2.96	2.73	2.57
28	28	4.20	3.34	2.95	2.71	2.56
29	29	4.18	3.33	2.93	2.70	2.55
30	30	4.17	3.32	2.92	2.69	2.53



F-Scores in MS Excel

- In Microsoft Excel, the following function returns an F-score:

α	df1	df2	Formula	Output Value
0.05	2	27	=FINV(A2,B2,C2)	3.3541308285292



F-Scores in Python

```
>>> from scipy import stats  
>>> stats.f.ppf(1-.05,dfn=2,dfd=27)  
3.3541308285291986
```



ANOVA

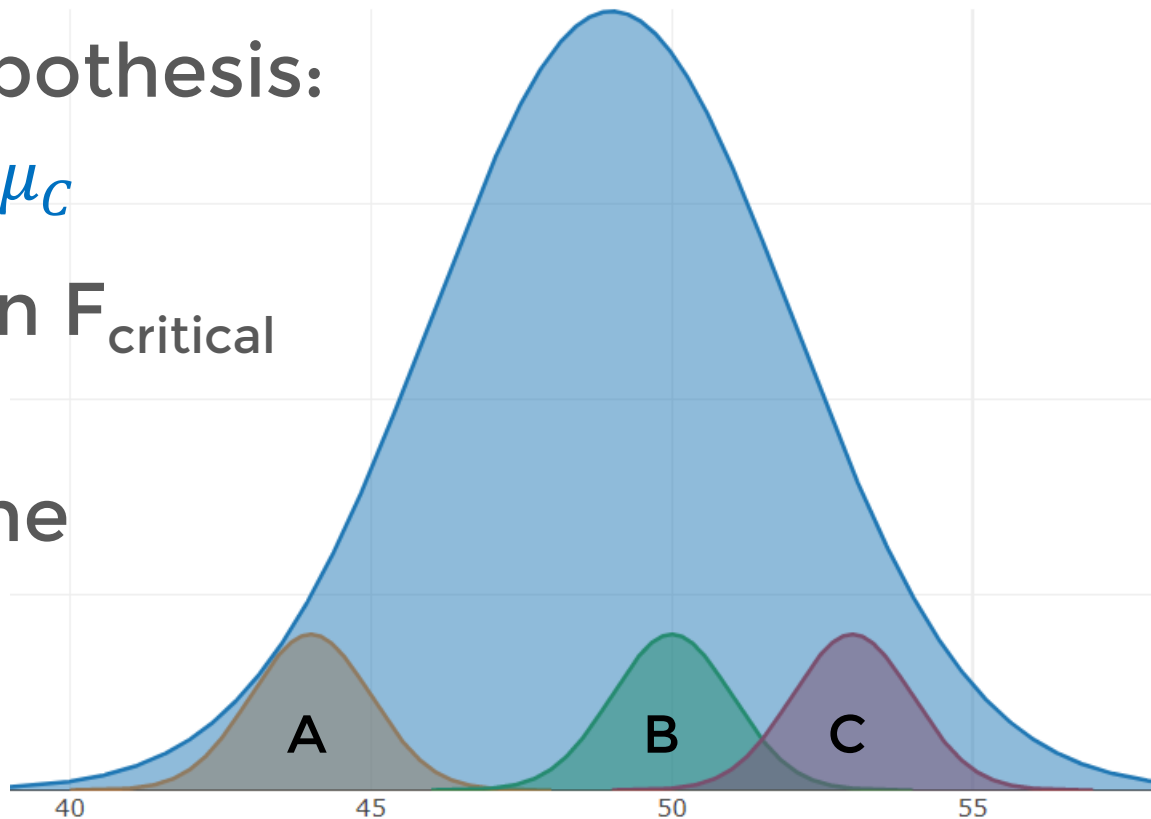
Recall our null hypothesis:

$$H_0: \mu_A = \mu_B = \mu_C$$

Since F is less than F_{critical}

$$1.718 < 3.354$$

we fail to reject the null hypothesis!





ANOVA Exercise #1



- In an effort to receive faster payment of invoices, a company introduces two discount plans
- One set of customers is given a 2% discount if they pay their invoice early
- Another set is offered a 1% discount
- A third set is not offered any incentive



ANOVA Exercise #1

- The results are as follows:
- Using ANOVA, can we say that the offers result in faster payments?



2% disc	1% disc	no disc
11	21	14
16	15	11
9	23	18
14	10	16
10	16	21



ANOVA Exercise #1

1. Calculate the means



	2% disc	1% disc	no disc
	11	21	14
	16	15	11
	9	23	18
	14	10	16
	10	16	21
$\mu_{2,1,0}$	12	17	16
μ_{TOT}	15		



ANOVA Exercise #1

$$SSG = 70$$



2. Find Sum of Squares Groups

$$(\mu_2 - \mu_{TOT})^2 = (12 - 15)^2 = 9$$

$$(\mu_1 - \mu_{TOT})^2 = (17 - 15)^2 = 4$$

$$(\mu_0 - \mu_{TOT})^2 = (16 - 15)^2 = 1$$

14

Multiply by the number of items in each group:

$$14 \times 5 = 70$$

	2% disc	1% disc	no disc
	11	21	14
	16	15	11
	9	23	18
	14	10	16
	10	16	21
$\mu_{2,1,0}$	12	17	16
μ_{TOT}	15		



ANOVA Exercise #1

$$SSG = 70$$
$$df_{groups} = 2$$



3. Degrees of Freedom Groups

$$df_{groups} = n_{groups} - 1$$
$$= 3 - 1$$
$$= 2$$

	2% disc	1% disc	no disc
	11	21	14
	16	15	11
	9	23	18
	14	10	16
	10	16	21
$\mu_{2,1,0}$	12	17	16
μ_{TOT}	15		



ANOVA Exercise #1

$$\begin{aligned}SSG &= 70 \\df_{groups} &= 2 \\SSE &= 198\end{aligned}$$



4. Sum of Squares Error

$(x_2 - \mu_2)^2$	$(x_1 - \mu_1)^2$	$(x_0 - \mu_0)^2$
1	16	4
16	4	25
9	36	4
4	49	0
4	1	25
34	106	58
TOTAL		198

	2% disc	1% disc	no disc
	11	21	14
	16	15	11
	9	23	18
	14	10	16
	10	16	21
$\mu_{2,1,0}$	12	17	16
μ_{TOT}	15		



ANOVA Exercise #1

$$\begin{aligned}SSG &= 70 \\df_{groups} &= 2 \\SSE &= 198 \\df_{error} &= 12\end{aligned}$$



5. Degrees of Freedom Error

$$\begin{aligned}df_{error} &= (n_{rows} - 1) * n_{groups} \\&= (5 - 1) * 3 \\&= 12\end{aligned}$$

	2% disc	1% disc	no disc
	11	21	14
	16	15	11
	9	23	18
	14	10	16
	10	16	21
$\mu_{2,1,0}$	12	17	16
μ_{TOT}	15		



ANOVA Exercise #1

$$SSG = 70$$

$$df_{groups} = 2$$

$$SSE = 198$$

$$df_{error} = 12$$



6. Calculate F value:

$$F = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}} = \frac{\frac{70}{2}}{\frac{198}{12}} = \frac{35}{16.5} = 2.121$$

7. Look up $F_{critical}$: 3.885

	2% disc	1% disc	no disc
	11	21	14
	16	15	11
	9	23	18
	14	10	16
	10	16	21
$\mu_{2,1,0}$	12	17	16
μ_{TOT}	15		



ANOVA Exercise #1

$$SSG = 70$$

$$df_{groups} = 2$$

$$SSE = 198$$

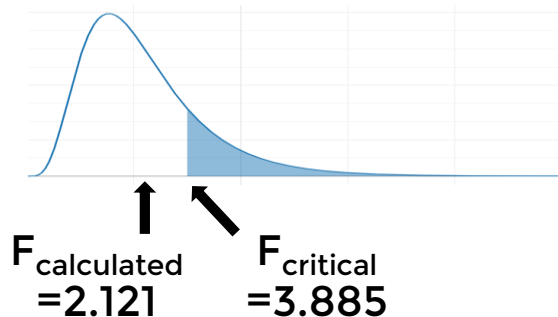
$$df_{error} = 12$$



Since F falls to the left of $F_{critical}$

$$2.121 < 3.885$$

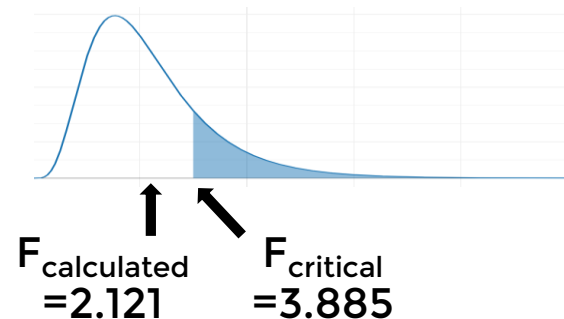
we fail to reject the
null hypothesis!





ANOVA Exercise #1

We don't have enough to support the idea that our offers changed the average number of days that customers took to pay their invoices!





Two-Way ANOVA



One-Way vs Two-Way ANOVA

- In the previous examples we used one-way ANOVA to test one independent variable.
- For the invoice problem, the independent variable was the incentive offered.
- The dependent variable was the time it took to receive payment.



One-Way vs Two-Way ANOVA

- Two-Way ANOVA lets us test two independent variables at the same time
- For the invoice example, we might also consider the amount due
- We would have 3 invoices for \$50, 3 for \$100, etc. and offer different incentives at each dollar amount.



One-Way vs Two-Way ANOVA

- The resulting data might look like this:
- Here, each row or dollar amount is called a **block**.
- Essentially, we want to isolate and remove any variance contributed by the blocks, to better understand the variance in the groups.

	2% disc	1% disc	no disc
\$50	16	23	21
\$100	14	21	16
\$150	11	16	18
\$200	10	15	14
\$250	9	10	11



One-Way vs Two-Way ANOVA

- So how do we do that?

	2% disc	1% disc	no disc
\$50	16	23	21
\$100	14	21	16
\$150	11	16	18
\$200	10	15	14
\$250	9	10	11



Two-Way ANOVA

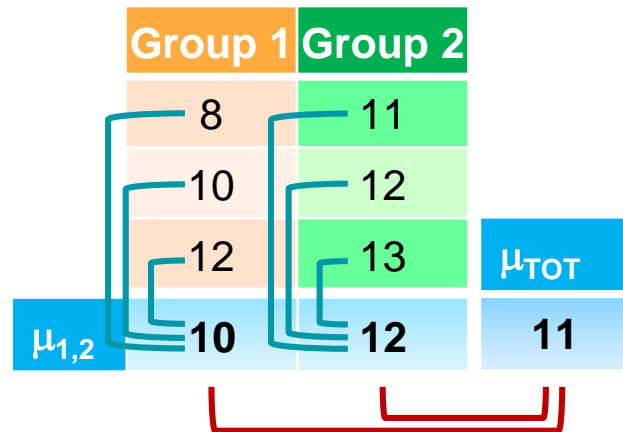
- The goal of ANOVA is to separate different aspects of the total variance.
- In the previous examples we had only

Sum of Squares Groups (SSG)

» between groups

and **Sum of Squares Error (SSE)**

» within groups

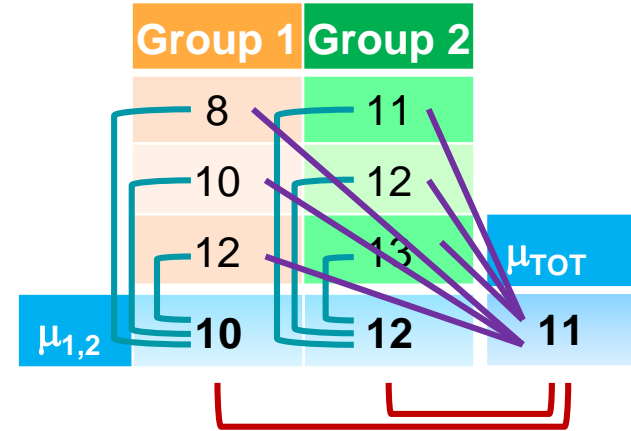




Two-Way ANOVA

- These two variances
SSG and **SSE** add up to
our total variance

Sum of Squares Total (SST)



Sum of Squares Groups (SSG)

» between groups

and Sum of Squares Error (SSE)

» within groups



Two-Way ANOVA

- Now we'll look at variance between rows, or **blocks**

	Group 1	Group 2	
Block A	8	11	
Block B	10	12	
Block C	12	13	μ_{TOT}
$\mu_{1,2}$	10	12	11

Sum of Squares Groups (SSG)

» between groups

and **Sum of Squares Error (SSE)**

» within groups



Two-Way ANOVA

- First calculate the block means

- Then calculate the

Sum of Squares Blocks (SSB)

Sum of Squares Groups (SSG)

and **Sum of Squares Error (SSE)**

» between blocks

» between groups

» within groups

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11



Two-Way ANOVA

- ANOVA still considers the relationship between the SSG and the SSE

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

Sum of Squares Blocks (SSB)

» between blocks

Sum of Squares Groups (SSG)

» between groups

and Sum of Squares Error (SSE)

» within groups



Two-Way ANOVA

- By calculating the SSB, we remove some of the variance in SSE

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

Sum of Squares Blocks (SSB)

» between blocks

Sum of Squares Groups (SSG)

» between groups

and Sum of Squares Error (SSE)

» within groups



Two-Way ANOVA

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{\text{groups}}}}{\frac{SSE}{df_{\text{error}}}}$$

Sum of Squares Groups (SSG)

$$(\mu_1 - \mu_{TOT})^2 = (10 - 11)^2 = 1$$

$$(\mu_2 - \mu_{TOT})^2 = (12 - 11)^2 = 1$$

2

multiply by the number
of items in each group: $2 \times 3 = 6$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

SSG = 6



Two-Way ANOVA

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{\text{groups}}}}{\frac{SSE}{df_{\text{error}}}}$$

Sum of Squares Blocks (SSB)

$$(\mu_A - \mu_{TOT})^2 = (9.5 - 11)^2 = 2.25$$

$$(\mu_B - \mu_{TOT})^2 = (11 - 11)^2 = 0$$

$$(\mu_C - \mu_{TOT})^2 = (12.5 - 11)^2 = 2.25$$

4.5

multiply by the number
of items in each block:

$$4.5 \times 2 = 9$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$



Two-Way ANOVA

Sum of Squares Total (SST)

$$\begin{aligned} &(8 - 11)^2 + (11 - 11)^2 + \\ &(10 - 11)^2 + (12 - 11)^2 + \\ &(12 - 11)^2 + (13 - 11)^2 = 16 \end{aligned}$$

no need to multiply since
every item is represented

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{\text{groups}}}}{\frac{SSE}{df_{\text{error}}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$



Two-Way ANOVA

Sum of Squares Error (SSE)

$$\begin{aligned} SSE &= SST - SSG - SSB \\ &= 16 - 6 - 9 = 1 \end{aligned}$$

no need to multiply since we're working with totals already

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{\text{groups}}}}{\frac{SSE}{df_{\text{error}}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$\begin{aligned} SSG &= 6 \\ SSB &= 9 \\ SST &= 16 \\ SSE &= 1 \end{aligned}$$



Two-Way ANOVA

So how do we calculate F?

Degrees of Freedom Groups is unchanged:

$$\begin{aligned}df_{groups} &= n_{groups} - 1 \\&= 2 - 1 \\&= 1\end{aligned}$$

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$



Two-Way ANOVA

So how do we calculate F?

Degrees of Freedom Error has changed:

$$\begin{aligned}df_{error} &= (n_{blocks} - 1)(n_{groups} - 1) \\&= (3 - 1)(2 - 1) \\&= 2\end{aligned}$$

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$



Two-Way ANOVA

So how do we calculate F?

$$F = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}} = \frac{\frac{6}{1}}{\frac{1}{2}} = 12$$

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$



Two-Way ANOVA

$F_{groups} = 12$ feels like a high value.

However, in a two-way ANOVA, $F_{critical}$ is found for groups and blocks separately!

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$



Two-Way ANOVA

$F_{groups} = 12$ feels like a high value.

For groups, with 1 df in the numerator and 2 df in the denominator,

$$F_{critical} = 18.5$$

$$F = \frac{\text{Var. Between Groups}}{\text{Var. Within Groups}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

	Group 1	Group 2	$\mu_{A,B,C}$
Block A	8	11	9.5
Block B	10	12	11
Block C	12	13	12.5
$\mu_{1,2}$	10	12	11

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$



ANOVA Exercise #2



- Let's go back to the invoice problem, and add a new independent variable
- Here each **block** represents an invoice amount
- The dependent variable is still days elapsed until payment

	2% disc	1% disc	no disc
\$50	16	23	21
\$100	14	21	16
\$150	11	16	18
\$200	10	15	14
\$250	9	10	11



ANOVA Exercise #2

1. Calculate the group means, the block means, and the total mean

	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15





ANOVA Exercise #2

2. Sum of Squares Groups

$$(\mu_2 - \mu_{TOT})^2 = (12 - 15)^2 = 9$$

$$(\mu_1 - \mu_{TOT})^2 = (17 - 15)^2 = 4$$

$$(\mu_0 - \mu_{TOT})^2 = (16 - 15)^2 = 1$$

14

Multiply by the number of items in each group:

$$14 \times 5 = 70$$



	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$



ANOVA Exercise #2

3. Degrees of Freedom Groups

$$\begin{aligned} df_{groups} &= n_{groups} - 1 \\ &= 3 - 1 \\ &= 2 \end{aligned}$$

	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$

$$df_{groups} = 2$$





ANOVA Exercise #2

4. Sum of Squares Blocks

$$(\mu_{50} - \mu_{TOT})^2 = (20 - 15)^2 = 25$$

$$(\mu_{100} - \mu_{TOT})^2 = (17 - 15)^2 = 4$$

$$(\mu_{200} - \mu_{TOT})^2 = (15 - 15)^2 = 0$$

$$(\mu_{200} - \mu_{TOT})^2 = (13 - 15)^2 = 4$$

$$(\mu_{250} - \mu_{TOT})^2 = (10 - 15)^2 = 25$$

$$58 \times 3 = 174$$

58

	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

SSG = 70

SSB = 174

$df_{\text{groups}} = 2$





ANOVA Exercise #2

5. Sum of Squares Total

$(x_2 - \mu_{tot})^2$	$(x_1 - \mu_{tot})^2$	$(x_0 - \mu_{tot})^2$
1	64	36
1	36	1
16	1	9
25	0	1
36	25	16
79	126	63
TOTAL		268

	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$df_{groups} = 2$$






ANOVA Exercise #2

6. Sum of Squares Error

$$\begin{aligned}SSE &= SST - SSG - SSB \\ &= 268 - 70 - 174 = 24\end{aligned}$$



	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

$$df_{\text{groups}} = 2$$



ANOVA Exercise #2

7. Degrees of Freedom Error

$$\begin{aligned}df_{error} &= (n_{blocks} - 1)(n_{groups} - 1) \\&= (5 - 1)(3 - 1) \\&= 8\end{aligned}$$



	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

$$df_{groups} = 2$$

$$df_{error} = 8$$



ANOVA Exercise #2

8. Calculate F

$$F = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}} = \frac{\frac{70}{2}}{\frac{24}{8}} = \frac{35}{3} = 11.67$$

	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

$$df_{groups} = 2$$

$$df_{error} = 8$$

$$F = 11.67$$





ANOVA Exercise #2

9. Find F_{critical}

$$\alpha = 0.05$$

$$df_{\text{numerator}} = 2$$

$$df_{\text{denominator}} = 8$$

$$F_{\text{critical}} = 4.46$$



	2% disc	1% disc	no disc	μ_{block}
\$50	16	23	21	20
\$100	14	21	16	17
\$150	11	16	18	15
\$200	10	15	14	13
\$250	9	10	11	10
μ_{col}	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

$$df_{\text{groups}} = 2$$

$$df_{\text{error}} = 8$$

$$F = 11.67$$

$$F_{\text{critical}} = 4.46$$

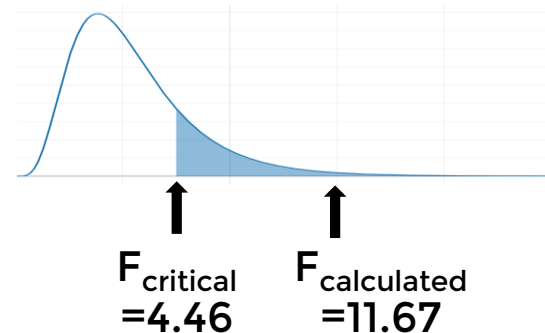


ANOVA Exercise #1

Since F falls to the right of F_{critical}

$$4.46 < 11.67$$

we reject the null hypothesis!



$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

$$df_{\text{groups}} = 2$$

$$df_{\text{error}} = 8$$

$$F = 11.67$$

$$F_{\text{critical}} = 4.46$$



2-way ANOVA in Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Anova: Two-Factor Without Replication											
2												
3	SUMMARY	Count	Sum	Average	Variance							
4	Row 1	3	60	20	13							
5	Row 2	3	51	17	13							
6	Row 3	3	45	15	13							
7	Row 4	3	39	13	7							
8	Row 5	3	30	10	1							
9												
10	Column 1	5	60	12	8.5							
11	Column 2	5	85	17	26.5							
12	Column 3	5	80	16	14.5							
13												
14												
15	ANOVA											
16	Source of Variation	SS	df	MS	F	P-value	F crit					
17	Rows	174	4	43.5	14.5	0.000974668	3.837853355					
18	Columns	70	2	35	11.666667	0.004249458	4.458970108					
19	Error	24	8	3								
20												
21	Total	268	14									

Data Analysis

Analysis Tools

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram

OK

Cancel

Help



Two-Way ANOVA with Replication



Without vs With Replication

without replication

	GroupA	GroupB	GroupC
Block1	16	23	21
Block2	14	21	16
Block3	11	16	18
Block4	10	15	14
Block5	9	10	11
Block6	8	8	10

with replication

	GroupA	GroupB	GroupC
Block1	16	23	21
	14	21	16
	11	16	18
Block2	10	15	14
	9	10	11
	8	8	10

Samples have multiple values
Samples have a mean value



Two-Way ANOVA with Replication

- Introduces the concept of **sample means** and **sample variance**
- Introduces the concept of **interactions**



Two-Way ANOVA with Replication

- As with our previous 2-way ANOVA, we consider two independent variables organized into groups and blocks
- We sample every block/group combination
- With replication, block/group samples have multiple measurements



Two-Way ANOVA with Replication

- Consider an experiment that measures the height of plants
- We apply three types of fertilizer A, B & C – these are our Groups
- Plants are kept at two temperatures (warm & cold) – these are our Blocks
- We assign 3 plants to each sample





Two-Way ANOVA

- First calculate the mean for each 3-item sample
- Calculate column means
- Calculate block means
- Calculate the overall mean

Fertilizer:				
	A	B	C	
Warm	13	21	18	16
	14	19	15	
	12	17	15	
Cold	16	14	15	14
	18	11	13	
	17	14	8	
Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15



Two-Way ANOVA

- As before, calculate the Sum of Squares Blocks
 $(16 - 15)^2 + (14 - 15)^2 = 2$
 $\times 9 \text{ items per block} = 18$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

SSB = 18



Two-Way ANOVA

- As before, calculate the Sum of Squares Columns

$$\begin{aligned} & (15 - 15)^2 + (16 - 15)^2 + \\ & (14 - 15)^2 = 2 \\ & \times 6 \text{ items per column} = 12 \end{aligned}$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

SSB = 18 SSC = 12



Two-Way ANOVA

- As before, calculate the Degrees of Freedom Columns

$$df_{columns} = (3 - 1) = 2$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	14
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

$$SSB = 18 \quad SSC = 12 \quad df_{columns} = 2$$



Two-Way ANOVA

- We have a new statistic:
SS Interactions
- For each sample mean, subtract the matching block and column means, add back the overall mean, square the result

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

$$SSB = 18 \quad SSC = 12 \quad df_{\text{columns}} = 2$$



Two-Way ANOVA

$$\begin{aligned}
 & (13 - 16 - 15 + 15)^2 + \\
 & (19 - 16 - 16 + 15)^2 + \\
 & (16 - 16 - 14 + 15)^2 + \\
 & (17 - 14 - 15 + 15)^2 + \\
 & (13 - 14 - 16 + 15)^2 + \\
 & (12 - 14 - 14 + 15)^2 = 28 \\
 & \times 3 \text{ items per sample} = 84
 \end{aligned}$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	14
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

$$\begin{aligned}
 \text{SSB} &= 18 & \text{SSC} &= 12 & \text{df}_{\text{columns}} &= 2 \\
 \text{SSI} &= 84
 \end{aligned}$$



Two-Way ANOVA

- Calculate the Sum of Squares Total

4	36	9	
1	16	0	
9	4	0	
1	1	0	
9	16	4	
4	1	49	
			164

Fertilizer:	A	B	C	
Warm	13	21	18	16
	14	19	15	
	12	17	15	
Cold	16	14	15	14
	18	11	13	
	17	14	8	

Sample Means	13	19	16
	17	13	12
Column Means	15	16	14
	15		

SSB = 18 SSC = 12 $df_{\text{columns}} = 2$

SSI = 84

SST = 164



Two-Way ANOVA

- Calculate the Sum of Squares Error by subtracting the other values from the SST:

$$164 - 18 - 12 - 84 = 50$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

$$SSB = 18 \quad SSC = 12 \quad df_{\text{columns}} = 2$$

$$SSI = 84 \quad SSE = 50$$

$$SST = 164$$



Two-Way ANOVA

- Degrees of Freedom Error

$$\text{blocks} \times \text{columns} \times (\text{items} - 1)$$

$$= 2 \times 3 \times (3 - 1) = 12$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	14
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

$$\text{SSB} = 18 \quad \text{SSC} = 12 \quad \text{df}_{\text{columns}} = 2$$

$$\text{SSI} = 84 \quad \text{SSE} = 50 \quad \text{df}_{\text{error}} = 12$$

$$\text{SST} = 164$$



Two-Way ANOVA

- Calculate F

$$F = \frac{\frac{SSC}{df_{columns}}}{\frac{SSE}{df_{error}}} = \frac{\frac{12}{2}}{\frac{50}{12}} = \mathbf{1.44}$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

SSB = 18 SSC = 12 $df_{columns} = 2$

SSI = 84 SSE = 50 $df_{error} = 12$

SST = 164



Two-Way ANOVA

$$F = 1.44$$

- Look up F_{critical}

$$F_{(0.05, 2, 12)} = 3.885$$

Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	
	18	11	13	
	17	14	8	

Sample Means	13	19	16	
	17	13	12	
Column Means	15	16	14	15

$$\text{SSB} = 18 \quad \text{SSC} = 12 \quad \text{df}_{\text{columns}} = 2$$

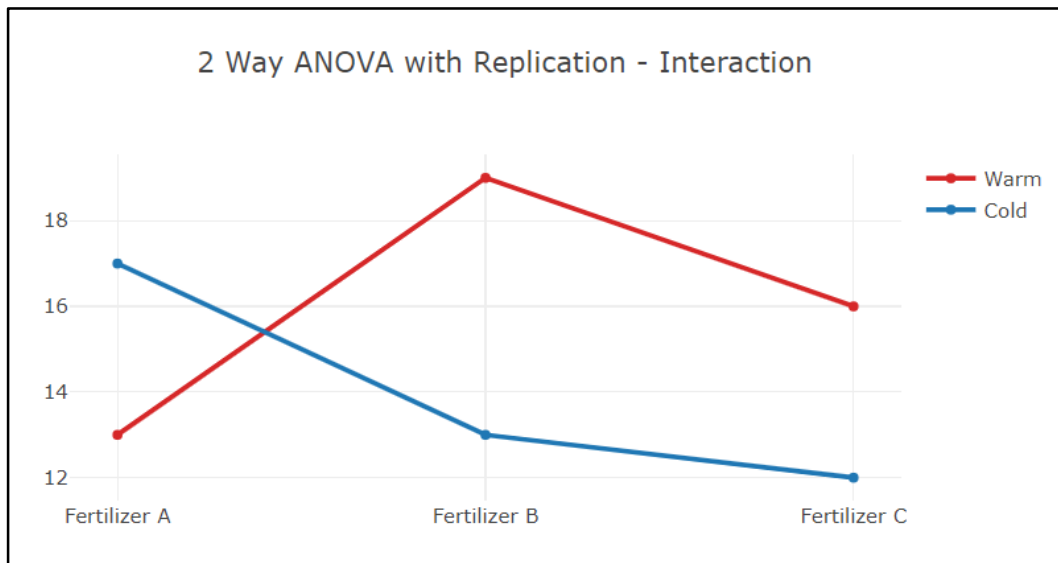
$$\text{SSI} = 84 \quad \text{SSE} = 50 \quad \text{df}_{\text{error}} = 12$$

$$\text{SST} = 164$$



Two-Way ANOVA

- A look at Interaction:



Fertilizer:	A	B	C	Block Means
Warm	13	21	18	
	14	19	15	
	12	17	15	
Cold	16	14	15	14
	18	11	13	
	17	14	8	

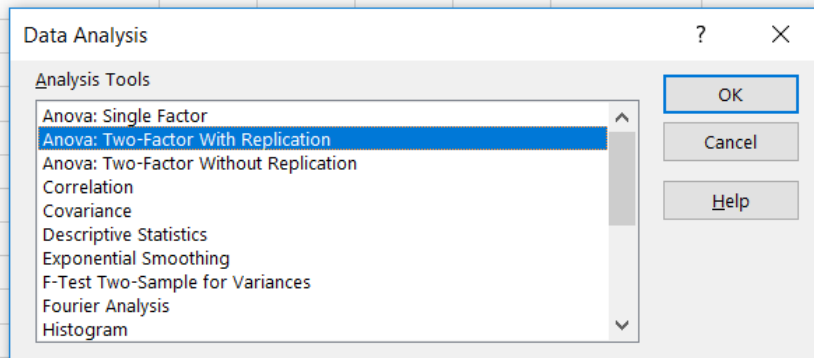
Sample Means	13	19	16
	17	13	12
Column Means	15	16	14
	15	16	14

$$\begin{aligned}
 \text{SSB} &= 18 & \text{SSC} &= 12 & \text{df}_{\text{columns}} &= 2 \\
 \text{SSI} &= 84 & \text{SSE} &= 50 & \text{df}_{\text{error}} &= 12 \\
 \text{SST} &= 164
 \end{aligned}$$



2-way with Replication in Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Anova: Two-Factor With Replication											
2												
3	SUMMARY	Fertilizer A	Fertilizer B	Fertilizer C	Total							
4	WARM											
5	Count	3	3	3	9							
6	Sum	39	57	48	144							
7	Average	13	19	16	16							
8	Variance	1	4	3	8.75							
9												
10	COLD											
11	Count	3	3	3	9							
12	Sum	51	39	36	126							
13	Average	17	13	12	14							
14	Variance	1	3	13	9.5							
15												
16	Total					ANOVA						
17	Count	6	6	6		Source of Variation	SS	df	MS	F	P-value	F crit
18	Sum	90	96	84		Sample	18	1	18	4.32	0.059785686	4.747225347
19	Average	15	16	14		Columns	12	2	6	1.44	0.275086887	3.885293835
20	Variance	5.6	13.6	11.2		Interaction	84	2	42	10.08	0.002698928	3.885293835
21						Within	50	12	4.16667			





Next Up: REGRESSION