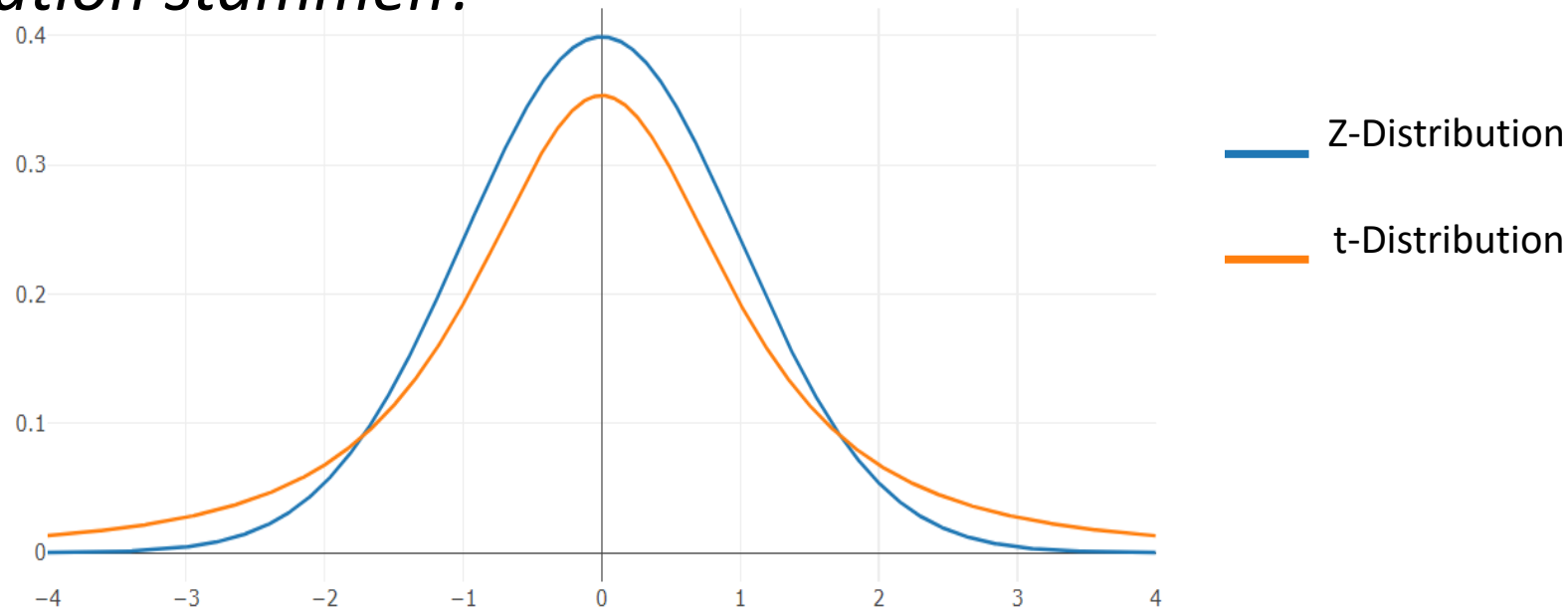


# Teil 5: ANOVA

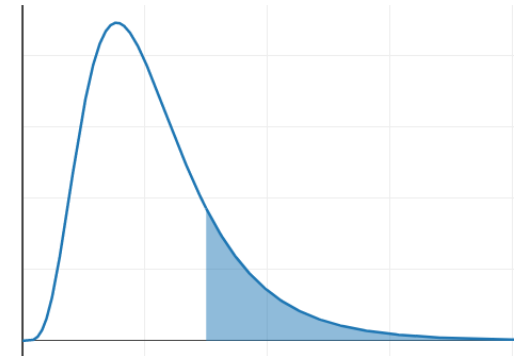
# Varianzanalyse

- In der vorherigen Lektion haben wir Z- und t-Verteilungen verwendet, um die Frage zu beantworten:  
*"Wie groß ist die Wahrscheinlichkeit, dass zwei Proben aus derselben Population stammen?"*



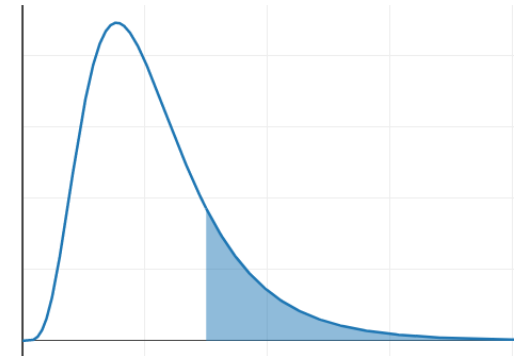
# Varianzanalyse

- In dieser Lektion stellen wir eine neue Verteilung vor:  
die F-Verteilung
- Sie wird verwendet, um die Frage zu beantworten:  
*"Wie groß ist die Wahrscheinlichkeit, dass zwei Stichproben aus Populationen mit derselben Varianz stammen?"*



# Varianzanalyse

- In dieser Lektion stellen wir eine neue Verteilung vor:  
die F-Verteilung
- Sie kann ebenfalls die Frage beantworten:  
*"Wie groß ist die Wahrscheinlichkeit, dass **drei oder mehr** Stichproben aus der selben Population stammen? "*

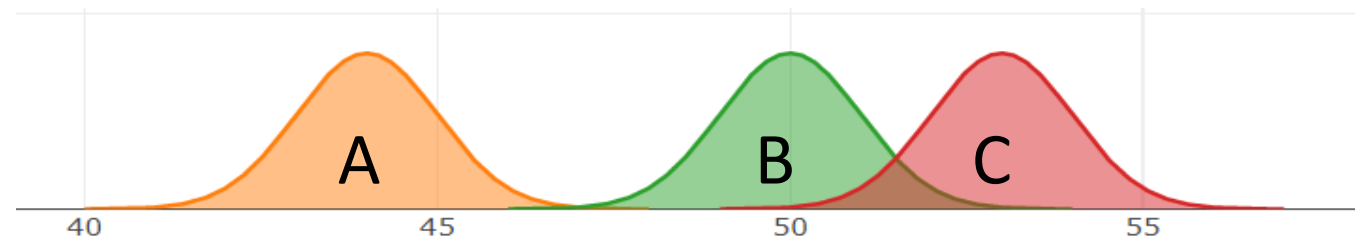


# ANOVA

## Varianzanalyse

# ANOVA

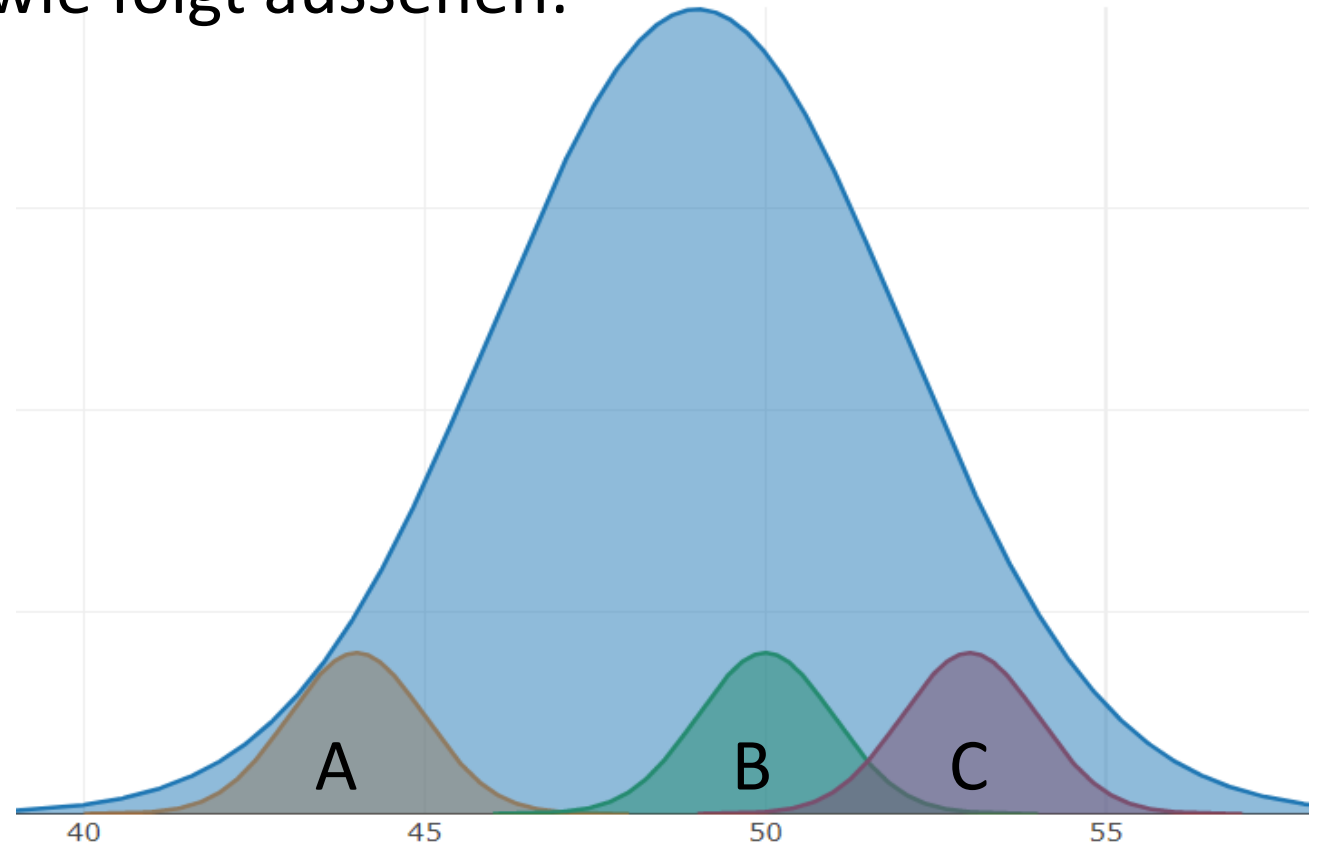
- In dem letzten Kapitel haben wir zwei Stichproben getestet, um zu sehen, ob sie wahrscheinlich von derselben Grundgesamtheit (parent population) stammen.
- Was wäre, wenn wir drei (oder mehr) Stichproben hätten?
- Könnten wir hier gleich vorgehen?



# ANOVA

- Unsere Nullhypothese würde wie folgt aussehen:

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



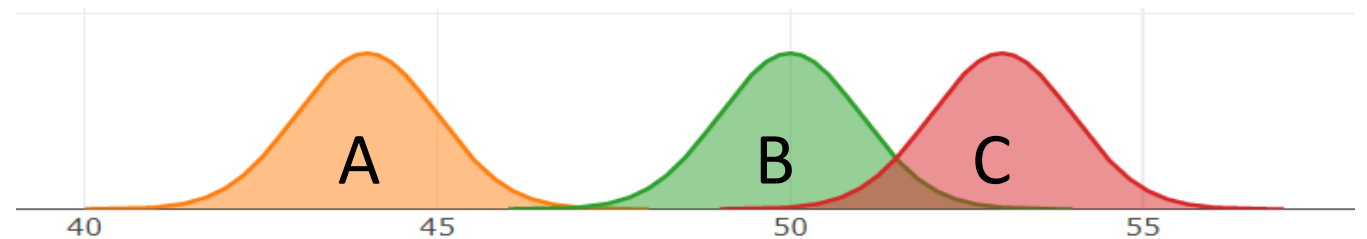
# ANOVA

- Wir *könnten* jedes Paar testen:

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

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

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





# ANOVA

- Das Problem ist, dass unser Konfidenzintervall zurückgeht:

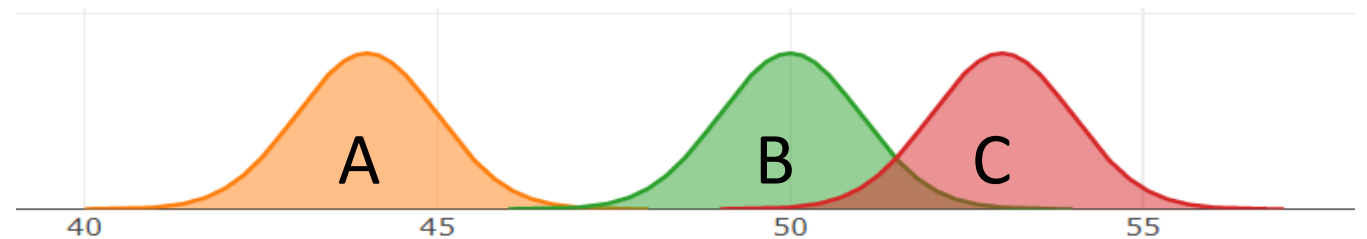
$$H_0: \mu_A = \mu_B \quad \alpha = 0.05$$

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

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

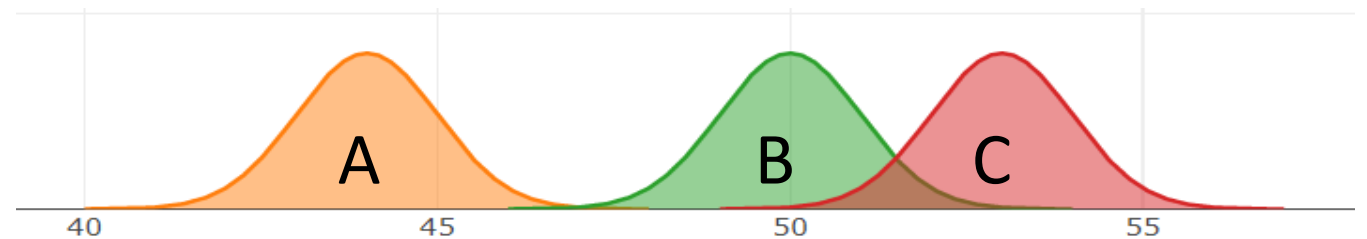
$$.95 \times .95 \times .95 = 0.857$$

*85.7% confidence level*



# ANOVA

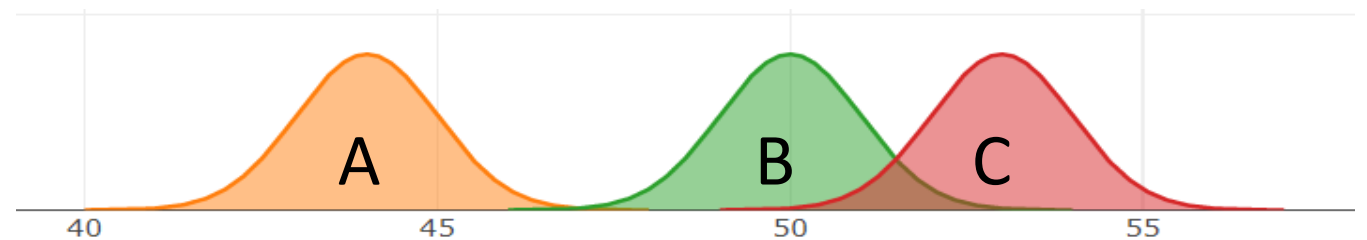
- Hier kommt ANOVA ins Spiel!
- Wir berechnen einen **F-Wert** und vergleichen ihn mit einem kritischen Wert, der durch unsere **Freiheitsgrade** bestimmt wird (die Anzahl der Gruppen und die Anzahl der Elemente in jeder Gruppe).



# ANOVA

- Lass uns mit einigen Daten arbeiten:

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

- Kalkuliere zuerst das Stichprobenmittel
- Anschließend das Mittel über alle Gruppen

	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
$\mu_{TOT}$	49		

# ANOVA

- ANOVA berücksichtigt zwei Arten von **Varianz**:
  - **Zwischen Gruppen**  
wie weit die Gruppe von der Gesamtsumme abweicht
  - **Innerhalb von Gruppen**  
wie weit einzelne Werte von ihrer jeweiligen Gruppe abweichen

# ANOVA

- Der F-Wert, den wir zu berechnen versuchen, ist einfach das Verhältnis zwischen diesen beiden Varianzen!

$$F = \frac{\textit{Varianz zwischen den Gruppen}}{\textit{Varianz innerhalb der Gruppe}}$$

# ANOVA

- Erinnern wir uns an die Gleichung für die Varianz:

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

- Hier ist  $\Sigma(x - \bar{x})^2$  die „Summe der Quadrate“ **SQ** und  $n-1$  sind die "Freiheitsgrade" **df**

# ANOVA

- Somit wird die Formel für den F-Wert:

$$F = \frac{\text{Varianz zwischen den Gruppen}}{\text{Varianz innerhalb der Gruppe}} = \frac{\frac{SSG}{df_{groups}}}{\frac{SSE}{df_{error}}}$$

$SSG$  = Sum of Squares Groups (quadratische Mittelwerte der Gruppen)

$SSE$  = Sum of Squares Error (gesamte Abweichung von den Mittelwerten in den Gruppen)

$df_{groups}$  = degrees of freedom [groups] Freiheitsgrade der Gruppen

$df_{error}$  = degrees of freedom [error] Freiheitsgrade des Fehlers



# ANOVA

$$SSG = 420$$

Quadratsumme Gruppen

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

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

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

42

Multipliziere mit der Anzahl  
der Elemente in jeder Gruppe:

$$42 \times 10 = 420$$

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
44	50	53
49		

# ANOVA

Freiheitsgrade Gruppen

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

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

	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
$\mu_{TOT}$	49		

# ANOVA

Quadratsumme Abweichungen

$(x_A - \mu_A)^2$	$(x_A - \mu_A)^2$	$(x_B - \mu_B)^2$	$(x_B - \mu_B)^2$	$(x_C - \mu_C)^2$	$(x_C - \mu_C)^2$
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

$$SSG = 420$$

$$df_{groups} = 2$$

$$SSE = 3300$$

$$(37 - 44)^2$$

$$= (-7)^2$$

$$= 49$$

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

Freiheitsgrade Abweichungen

$$\begin{aligned}df_{error} &= (n_{rows} - 1) * n_{groups} \\&= (10 - 1) * 3 \\&= 27\end{aligned}$$

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

	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
$\mu_{TOT}$	49		

# ANOVA

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

Setze nun alles in die Formel ein:

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

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 mit 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

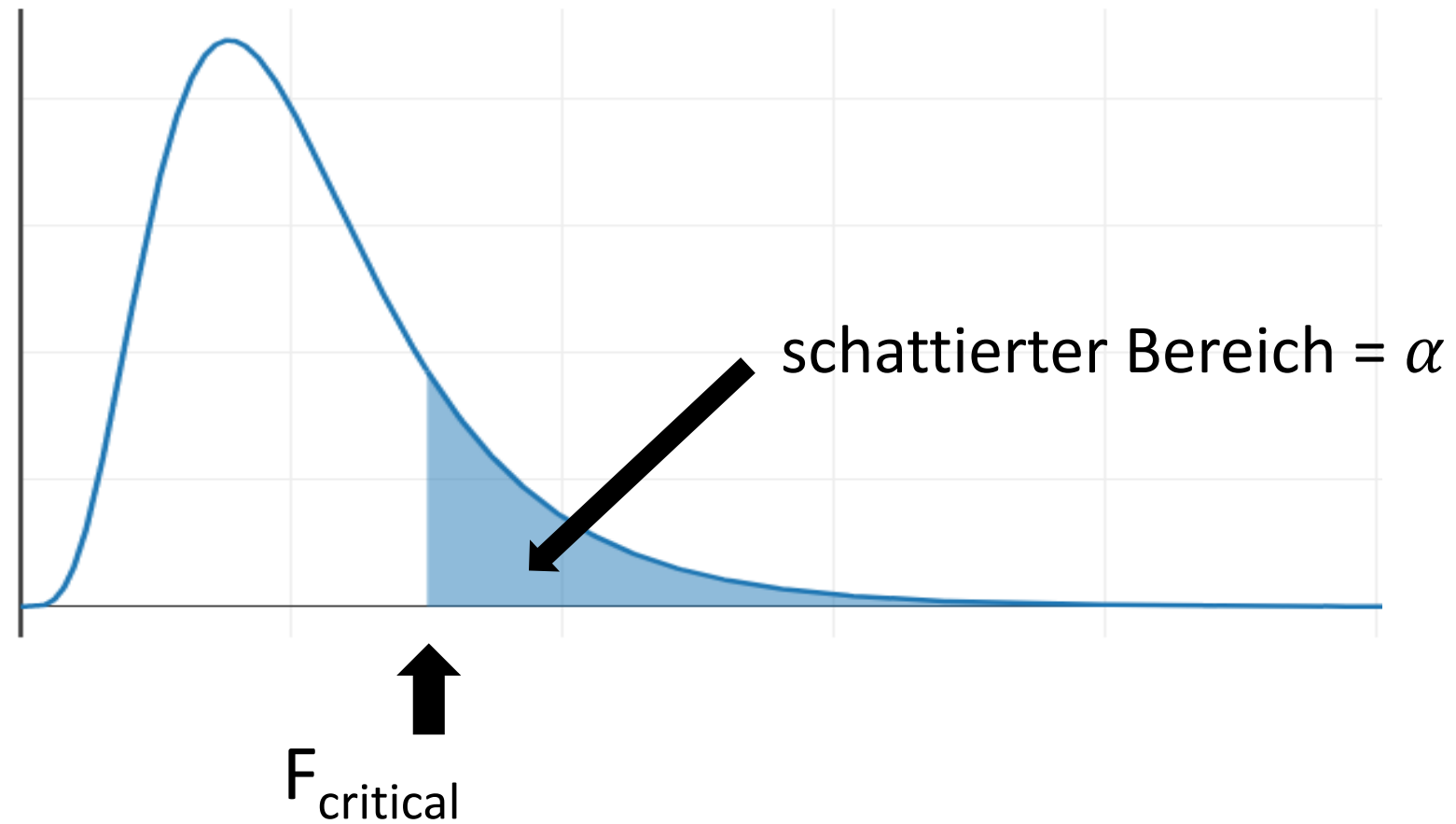
OK

Cancel

Help

# F-Verteilung

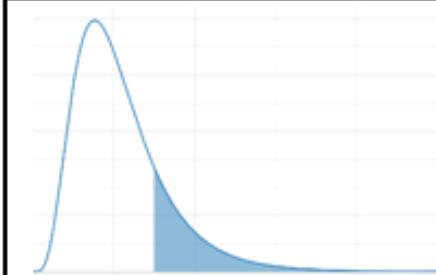
# F-Verteilung





# F-Verteilung

- In der F-Tabelle findest du den kritischen Wert
- Verwende eine Tabelle mit einer Konfidenz von 95%
- Finde den Zähler df
- Und den Nenner df
- kritischer Wert = 3,35



The image shows a graph of an F-distribution curve. The area under the curve to the right of a certain point is shaded in blue, representing the upper tail area. This visualizes the concept of finding a critical value from the upper tail of the distribution.

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

# F-Werte in Excel

- In Microsoft Excel gibt die folgende Funktion einen F-Wert aus:

$\alpha$	df1	df2	Formel	Output
0.05	2	27	=FINV(A2,B2,C2)	3.3541308285292

# F-Werte in Python

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

# ANOVA

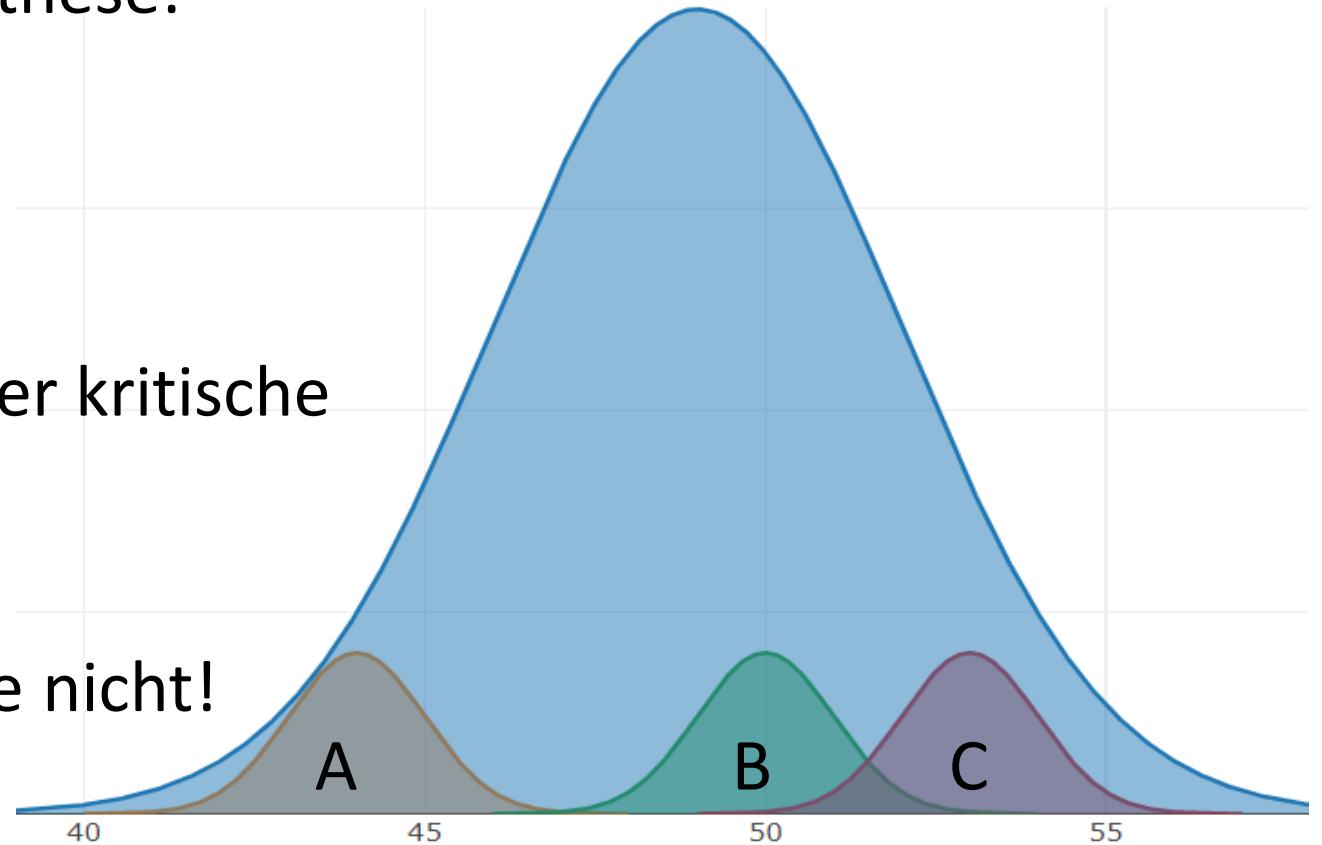
- Erinnere dich an die Nullhypothese:

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

- Da der F-Wert kleiner ist als der kritische Wert ( $F_{\text{critical}}$ )

$$1.718 < 3.354$$

verwerfen wir die Nullhypothese nicht!



# ANOVA Übung #1



- Um eine schnellere Zahlung von Rechnungen zu erhalten, führt ein Unternehmen zwei Rabattpläne ein
- Eine Gruppe von Kunden erhält 2% Rabatt, wenn sie ihre Rechnung vorzeitig bezahlen
- Einer anderen Gruppe wird 1% angeboten
- Eine dritte Kundengruppe erhält keine Vergünstigungen

# ANOVA Übung #1



- Die Ergebnisse sehen wie folgt aus:
- Können wir mit ANOVA herausfinden, ob die Angebote schnellere Zahlungen zur Folge haben?

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

# ANOVA Übung #1



1. Berechne die Mittelwerte

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
---------------	----	----	----

$\mu_{TOT}$	15
-------------	----

$$SSG = 70$$



# ANOVA Übung #1

2. Ermittle die Quadratsumme Gruppen

$$\begin{aligned} (\mu_2 - \mu_{TOT})^2 &= (12 - 15)^2 = 9 \\ (\mu_1 - \mu_{TOT})^2 &= (17 - 15)^2 = 4 \\ (\mu_0 - \mu_{TOT})^2 &= (16 - 15)^2 = 1 \\ \hline &= 14 \end{aligned}$$

Multipliziere mit der Anzahl der Items in jeder Gruppe:

$$14 \times 5 = 70$$

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
---------------	----	----	----

$\mu_{TOT}$	15
-------------	----



# ANOVA Übung #1

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



## 3. Freiheitsgrade Gruppen

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

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
---------------	----	----	----

$\mu_{TOT}$	15
-------------	----

# ANOVA Übung #1

$$SSG = 70$$

$$df_{groups} = 2$$

$$SSE = 198$$



## 4. Quadratsumme Abweichungen

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

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

$\mu_{2,1,0}$	<b>12</b>	<b>17</b>	<b>16</b>
$\mu_{TOT}$	<b>15</b>		

# ANOVA Übung #1

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



## 5. Freiheitsgrade Abweichungen

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

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
$\mu_{TOT}$	15		

# ANOVA Übung #1

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



6. Berechne nun den F-Wert

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

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
---------------	----	----	----

$\mu_{TOT}$	15
-------------	----

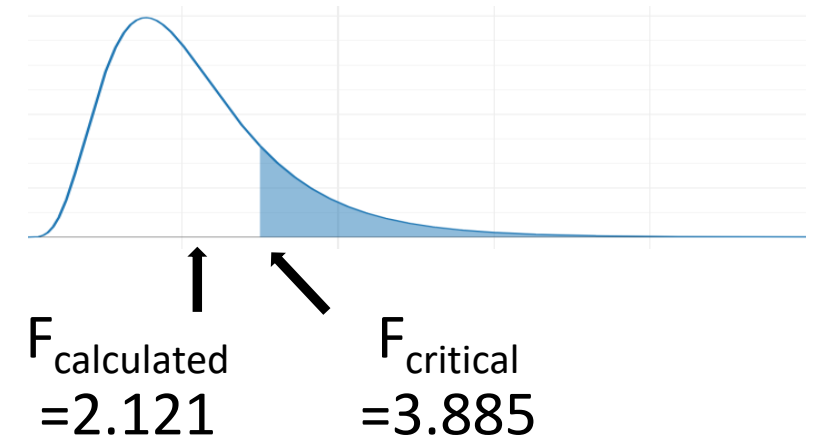
7. Finde  $F_{critical}$ : 3.885

# ANOVA

- Da  $F$  links von  $F_{\text{critical}}$  liegt

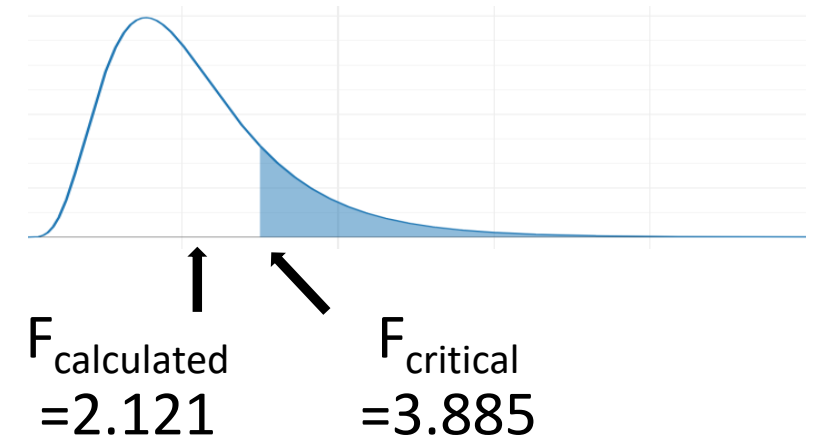
$$2.121 < 3.885$$

verwerfen wir die Nullhypothese nicht!



# ANOVA

Wir haben also nicht genug, um die Idee weiter zu verfolgen, dass sich mit unseren Rabatten etwas daran ändert, wie viele Tage die Kunden im Durchschnitt zur Begleichung ihrer Rechnungen benötigen.



# Zwei-Wege-ANOVA

# Ein-Wege-ANOVA vs. Zwei-Wege-ANOVA

- In den vorherigen Beispielen verwendeten wir eine Ein-Wege-ANOVA, um eine unabhängige Variable zu testen.
- Für das Rechnungsproblem war die unabhängige Variable der angebotene Rabatt.
- Die abhängige Variable war die Zeit, die für die Zahlung benötigt wurde.



# Ein-Wege-ANOVA vs. Zwei-Wege-ANOVA

- Mit der Zwei-Wege-ANOVA können wir zwei unabhängige Variablen gleichzeitig testen
- Für das Rechnungsbeispiel könnten wir zusätzlich den fälligen Betrag berücksichtigen
- Wir haben zum Beispiel 3 Rechnungen für 50€, 3 Rechnungen für 100€ usw. und bieten verschiedene Anreize für jeden Betrag.

# Ein-Wege-ANOVA vs. Zwei-Wege-ANOVA

- Die resultierenden Daten könnten folgendermaßen aussehen:
- Hier wird jeder Zeilen- oder Eurobetrag als **Block** bezeichnet.
- Im Wesentlichen wollen wir jede Varianz der Blöcke isolieren und entfernen, um die Varianz in den Gruppen besser zu verstehen.

	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

# Ein-Wege-ANOVA vs. Zwei-Wege-ANOVA

- Nun, wie gehen wir dabei vor?

	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

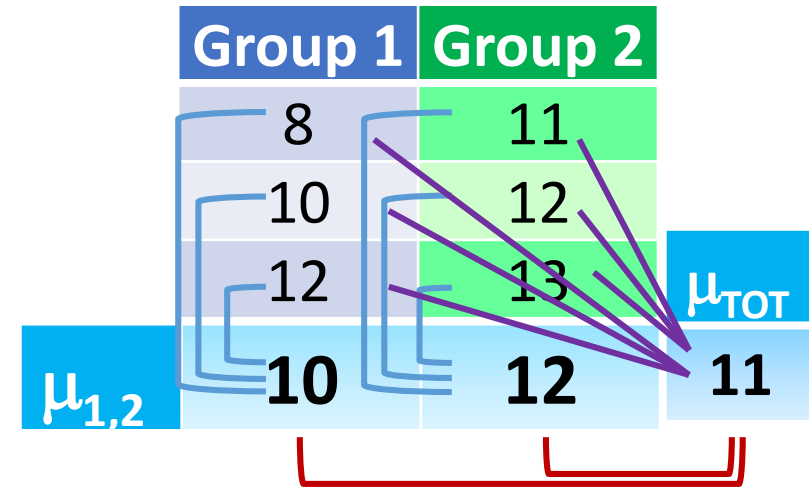
# Zwei-Wege-ANOVA

- Das Ziel einer ANOVA ist es, verschiedene Aspekte der Gesamtvarianz zu trennen.
- In den vorherigen Beispielen hatten wir nur die
- **Quadratsumme Gruppen (SSG)** » zwischen Gruppen
- **Quadratsumme Abweichungen (SSE)** » innerhalb von Gruppen

	Group 1	Group 2	
	8	11	
	10	12	
	12	13	
$\mu_{1,2}$	10	12	$\mu_{TOT}$
			11

# Zwei-Wege-ANOVA

- Diese zwei Varianten **SSG** und **SSE** addieren sich zu unserer Gesamtvarianz auf **Quadratsumme Total (SST)**



- **Quadratsumme Gruppen (SSG)** » zwischen Gruppen
- **Quadratsumme Abweichungen (SSE)** » innerhalb von Gruppen

# Zwei-Wege-ANOVA

- Schauen wir uns die Varianzen zwischen den Zeilen bzw. **Blöcken** an

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

- **Quadratsumme Gruppen (SSG)**
- **Quadratsumme Abweichungen (SSE)**

» **zwischen** Gruppen

» **innerhalb von** Gruppen

# Zwei-Wege-ANOVA

- Zuerst kalkulieren wir den Mittelwert der Blöcke
- Im Anschluss die
- **Quadratsumme Blöcke (SSB)**
- **Quadratsumme Gruppen (SSG)**
- **Quadratsumme Abweichungen (SSE)**

	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

- » zwischen Blöcken
- » zwischen Gruppen
- » innerhalb von Gruppen

# Zwei-Wege-ANOVA

- ANOVA berücksichtigt immer noch die Beziehung zwischen der SSG und SSE.

	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

- **Quadratsumme Blöcke (SSB)**
- **Quadratsumme Gruppen (SSG)**
- **Quadratsumme Abweichungen (SSE)**

- » **zwischen** Blöcken
- » **zwischen** Gruppen
- » **innerhalb von** Gruppen

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



# Zwei-Wege-ANOVA

- Durch die Berechnung des SSB **entfernen** wir einen Teil der Varianz in SSE

	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

- **Quadratsumme Blöcke (SSB)**
- **Quadratsumme Gruppen (SSG)**
- **Quadratsumme Abweichungen (SSE)**

- » **zwischen** Blöcken
- » **zwischen** Gruppen
- » **innerhalb von** Gruppen

# Zwei-Wege-ANOVA

- **Quadratsumme Gruppen (SSG)**

$$\begin{aligned}(\mu_1 - \mu_{TOT})^2 &= (10 - 11)^2 = 1 \\(\mu_2 - \mu_{TOT})^2 &= (12 - 11)^2 = 1 \\ \hline &2\end{aligned}$$

Multipliziert mit der Anzahl der Elemente in jeder Gruppe:

$$2 \times 3 = 6$$

	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
$\mu_{1,2}$	<b>10</b>	<b>12</b>	<b>11</b>


$$SSG = 6$$

# Zwei-Wege-ANOVA

- **Quadratsumme Blöcke (SSB)**

$$\begin{aligned}(\mu_A - \mu_{TOT})^2 &= (9.5 - 11)^2 = 2.25 \\(\mu_B - \mu_{TOT})^2 &= (11 - 11)^2 = 0 \\(\mu_C - \mu_{TOT})^2 &= (12.5 - 11)^2 = 2.25 \\ \hline &4.5\end{aligned}$$

Multipliziere mit der Anzahl der Elemente in jedem 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

$SSG = 6$
$SSB = 9$

# Zwei-Wege-ANOVA

- **Quadratsumme Total (SST)**

$$(8 - 11)^2 + (11 - 11)^2 + (10 - 11)^2 + (12 - 11)^2 + (12 - 11)^2 + (13 - 11)^2 = 16$$

keine Notwendigkeit zu multiplizieren, da jedes Item repräsentiert ist.

	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

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

# Zwei-Wege-ANOVA

- Quadratsumme Abweichungen (SSE)

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

keine Notwendigkeit zu multiplizieren,  
da wir bereits mit Summen arbeiten

	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

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

# Zwei-Wege-ANOVA

- Also, wie berechnen wir nun F ?
- Die Freiheitsgrade der Gruppen bleiben unverändert:

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

	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
$\mu_{1,2}$	<b>10</b>	<b>12</b>	<b>11</b>

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

# Zwei-Wege-ANOVA

- Also, wie berechnen wir nun F ?
- Die Freiheitsgrade der Abweichungen haben sich **verändert**:

	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
$\mu_{1,2}$	<b>10</b>	<b>12</b>	<b>11</b>

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

$$\begin{aligned}SSG &= 6 \\SSB &= 9 \\SST &= 16 \\SSE &= 1 \\df_{groups} &= 1 \\df_{error} &= 2\end{aligned}$$

# Zwei-Wege-ANOVA

- Also, wie berechnen wir nun F ?

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

	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

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$



# Zwei-Wege-ANOVA

$F_{groups} = 12$  fühlt sich an wie ein hoher Wert

In einer Zwei-Wege-ANOVA wird  $F_{critical}$  jedoch für Gruppen und Blöcke separat ermittelt!

	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

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$

# Zwei-Wege-ANOVA

$F_{groups} = 12$  fühlt sich an wie ein hoher Wert

	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

Für Gruppen mit 1 df im Zähler und 2 df im Nenner

$$F_{critical} = 18.5$$

$$SSG = 6$$

$$SSB = 9$$

$$SST = 16$$

$$SSE = 1$$

$$df_{groups} = 1$$

$$df_{error} = 2$$

# Zwei-Wege-ANOVA

## Beispielübung

# ANOVA Übung #2



- Lasst uns das Problem mit den Rechnungen in eine Zwei-Wege-ANOVA umwandeln!
- Unsere Nullhypothese lautet, dass das Anbieten von Rabatten die Anzahl der Tage bis zur Zahlung der Rechnung nicht verändert.

	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 Übung #2



- Jetzt werden wir zudem Blöcke zu den Gruppen hinzufügen.
- Mal sehen, ob wir die gleichen Ergebnisse wie beim letzten Mal erhalten!

	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 Übung #2



- Gehen wir zurück zum Rechnungsproblem und fügen eine neue unabhängige Variable hinzu
- Hier repräsentiert jeder **Block** einen Rechnungsbetrag
- Die abhängige Variable ist die Zahl der Tage, die bis zur Zahlung verstrichen sind

	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 Übung #2



1. Berechne nun das Gruppenmittel, den Mittelwert der Blöcke und den Gesamtmittelwert

	2% disc	1% disc	no disc	$\mu_{\text{block}}$
€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
$\mu_{\text{col}}$	12	17	16	15

# ANOVA Übung #2

## 2. Quadratsumme Gruppe

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

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

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

14

Multipliziert mit der Anzahl der Items je Gruppe:

$$14 \times 5 = 70$$



	2% disc	1% disc	no disc	$\mu_{\text{block}}$
€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
$\mu_{\text{col}}$	12	17	16	15

$$SSG = 70$$



# ANOVA Übung #2



## 3. Quadratsumme Freiheitsgrade

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

	2% disc	1% disc	no disc	$\mu_{block}$
€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
$\mu_{col}$	12	17	16	15

$$SSG = 70$$

$$df_{groups} = 2$$

# ANOVA Übung #2



## 4. Quadratsumme Blöcke

$$\begin{aligned}(\mu_{50} - \mu_{TOT})^2 &= (20 - 15)^2 = 25 \\(\mu_{100} - \mu_{TOT})^2 &= (17 - 15)^2 = 4 \\(\mu_{150} - \mu_{TOT})^2 &= (15 - 15)^2 = 0 \\(\mu_{200} - \mu_{TOT})^2 &= (13 - 15)^2 = 4 \\(\mu_{250} - \mu_{TOT})^2 &= (10 - 15)^2 = 25 \\&\hline &58\end{aligned}$$

$$58 \times 3 = 174$$

	2% disc	1% disc	no disc	$\mu_{\text{block}}$
€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
$\mu_{\text{col}}$	12	17	16	15

$$\begin{aligned}SSG &= 70 \\SSB &= 174\end{aligned}$$

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

# ANOVA Übung #2

## 5. Quadratsumme Total

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

	2% disc	1% disc	no disc	$\mu_{\text{block}}$
€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
$\mu_{\text{col}}$	12	17	16	<b>15</b>

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

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



# ANOVA Übung #2



## 6. Quadratsumme Abweichungen

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

	2% disc	1% disc	no disc	$\mu_{\text{block}}$
€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
$\mu_{\text{col}}$	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

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

# ANOVA Übung #2



## 7. Freiheitsgrade Abweichungen

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

	2% disc	1% disc	no disc	$\mu_{block}$
€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
$\mu_{col}$	12	17	16	15

$SSG = 70$   
 $SSB = 174$   
 $SST = 268$   
 $SSE = 24$

$df_{groups} = 2$   
 $df_{error} = 8$

# ANOVA Übung #2



8. Kalkuliere nun F

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

	2% disc	1% disc	no disc	$\mu_{block}$
€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
$\mu_{col}$	12	17	16	15

$SSG = 70$   
 $SSB = 174$   
 $SST = 268$   
 $SSE = 24$

$df_{groups} = 2$   
 $df_{error} = 8$   
 $F = 11.67$

# ANOVA Übung #2



9. Suche nun  $F_{\text{critical}}$

$$\alpha = 0.05$$

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

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

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

	2% disc	1% disc	no disc	$\mu_{\text{block}}$
€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
$\mu_{\text{col}}$	12	17	16	15

$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

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

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

$$F = 11.67$$

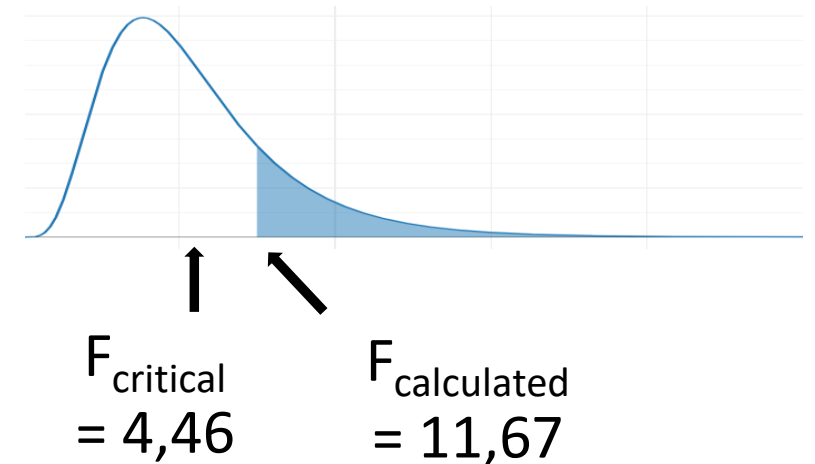
# ANOVA Übung #2



- Da  $F$  rechts von  $F_{\text{critical}}$  liegt

$$4,46 < 11,67$$

verwerfen wir die Nullhypothese!



$$SSG = 70$$

$$SSB = 174$$

$$SST = 268$$

$$SSE = 24$$

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

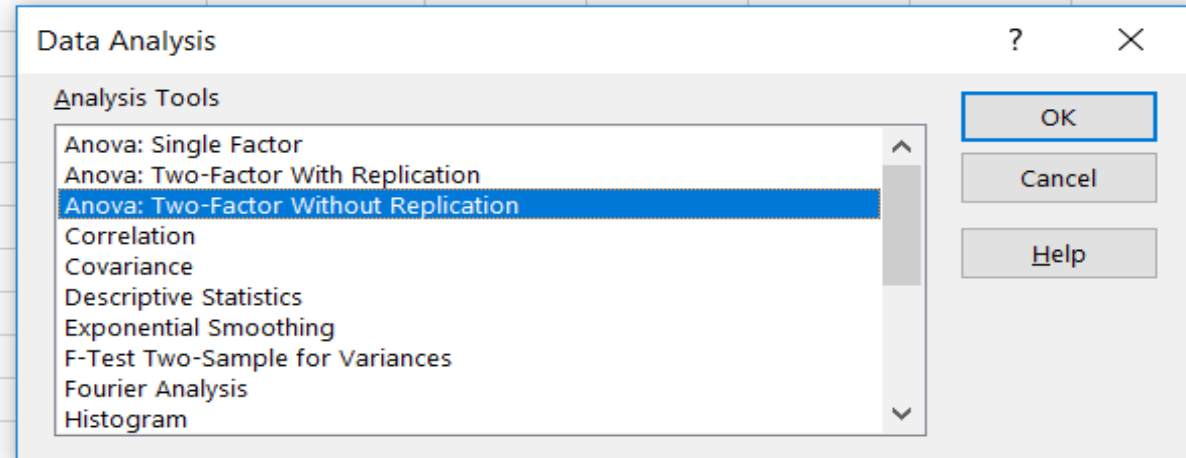
$$df_{\text{error}} = 8$$

$$F = 11.67$$



# Die Zwei-Wege-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									



# Zweifaktorielle ANOVA mit Messwiederholung

# Ohne und mit Wiederholung

Ohne Wiederholung

	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

Mit Wiederholung

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

Stichprobe hat mehrere Messwerte

Stichprobe hat einen Mittelwert

# Zweifaktorielle ANOVA mit Messwiederholung

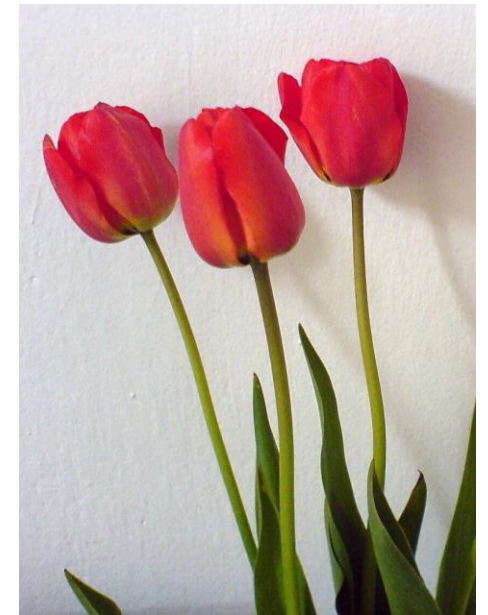
- Führt das Konzept der **Stichprobenmittel** und **Stichprobenvarianz** ein
- Stellt das Konzept der **Interaktionseffekte** vor

# Zweifaktorielle ANOVA mit Messwiederholung

- Wie bei unserer vorherigen Zwei-Wege-ANOVA betrachten wir zwei unabhängige Variablen, die in Gruppen und Blöcken organisiert sind
- Wir probieren jede Block- / Gruppenkombination aus
- Bei der Messwiederholung haben Block- / Gruppenstichproben mehrere Messungen

# Zweifaktorielle ANOVA mit Messwiederholung

- Wir betrachten ein Experiment, das die Größe von Pflanzen misst
- Wir verwenden drei Arten von Dünger A, B & C - das sind unsere Gruppen
- Die Pflanzen werden bei zwei unterschiedlichen Temperaturen gehalten (warm & kalt) - das sind unsere Blöcke
- Wir ordnen jeder Probe 3 Pflanzen zu



# Zwei-Wege-ANOVA

- Berechnen wir nun zunächst den Mittelwert für jede 3-Werte-Stichprobe
- Kalkuliere den Spaltenmittelwert
- Kalkuliere den Blockmittelwert
- Kalkuliere den Gesamtmittelwert

Dünger:		A	B	C	Blockmittel
Warm		13	21	18	
		14	19	15	
		12	17	15	
Kalt		16	14	15	
		18	11	13	
		17	14	8	
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

# Zwei-Wege-ANOVA

- Nun berechnen wir wie bisher die Summe der Quadrate pro Block

$$(\textcolor{red}{16} - \textcolor{green}{15})^2 + (\textcolor{blue}{14} - \textcolor{green}{15})^2 = 2$$

$$\times 9 \text{ Werte pro Block} = \mathbf{18}$$

Dünger:		A	B	C	
Warm	13	21	18	16	B l o c k m i t t e l
	14	19	15		
	12	17	15		
Kalt	16	14	15	14	
	18	11	13		
	17	14	8		
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

$$\text{SSB} = 18$$



# Zwei-Wege-ANOVA

- Nun berechnen wir wie bisher die Summe der Quadrate pro Spalte

$$\begin{aligned}
 & (15 - 15)^2 + (16 - 15)^2 + \\
 & (14 - 15)^2 = 2 \\
 & \times 6 \text{ Werte pro Spalte} = 12
 \end{aligned}$$

Dünger:	A	B	C		
Warm	13	21	18		B l o c k m i t t e l
	14	19	15	16	
	12	17	15		
Kalt	16	14	15		
	18	11	13	14	
	17	14	8		
Stichprobenmittel	13	19	16		
	17	13	12		
Spaltenmittel	15	16	14	15	

SSB = 18

SSC = 12

# Zwei-Wege-ANOVA

- Nun berechnen wir wie bisher die Freiheitsgrade pro Spalte

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

Dünger:		A	B	C	Blockmittel
Warm		13	21	18	
		14	19	15	
		12	17	15	
Kalt		16	14	15	
		18	11	13	
		17	14	8	
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

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

# Zwei-Wege-ANOVA

- Wir haben eine neue Statistik:  
**SS Interactions (Interaktionseffekt)**
- Für jeden Stichprobenmittelwert subtrahieren wir die übereinstimmenden Block- und Spaltenmittel  
addieren den Gesamtmittelwert,  
und quadrieren das Ergebnis

Dünger:		A	B	C	Blockmittel
Warm		13	21	18	
		14	19	15	
		12	17	15	
Kalt		16	14	15	
		18	11	13	
		17	14	8	
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

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

# Zwei-Wege-ANOVA

$$\begin{aligned}
 & (13 - 16 - 15 + 15)^2 + \\
 & (19 - 16 - 16 + 15)^2 + \\
 & (16 - 16 - 14 + 15)^2 + \\
 & (17 - 14 - 15 + 15)^2 + \\
 & (13 - 14 - 16 + 15)^2 + \\
 & (12 - 14 - 14 + 15)^2 = 28 \\
 & \times 3 \text{ Werte pro Stichprobe} = \mathbf{84}
 \end{aligned}$$

Dünger:		A	B	C		
Warm	13	21	18	16	B l o c k m i t t e l	
	14	19	15			
	12	17	15			
Kalt	16	14	15	14		
	18	11	13			
	17	14	8			
Stichprobenmittel		13	19	16		
		17	13	12		
Spaltenmittel		15	16	14	15	

SSB = 18

SSC = 12

df<sub>columns</sub> = 2

SSI = 84

# Zwei-Wege-ANOVA

- Kalkuliere nun die Gesamtquadratsumme

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

Dünger:	A	B	C	
Warm	13	21	18	<b>16</b>
	14	19	15	
	12	17	15	
Kalt	16	14	15	<b>14</b>
	18	11	13	
	17	14	8	
Stichprobenmittel				
	13	19	16	
	17	13	12	

B  
l  
o  
c  
k  
m  
i  
t  
t  
e  
l

Spaltenmittel	<b>15</b>	<b>16</b>	<b>14</b>	<b>15</b>
---------------	-----------	-----------	-----------	-----------

$$SSB = 18$$

$$SSC = 12$$

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

$$SSI = 84$$

$$SST = 164$$

# Zwei-Wege-ANOVA

- Kalkuliere nun die Quadratsumme der Abweichungen, indem du die anderen Werte von SST (Gesamtquadratsumme) abziehst:

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

Dünger:		A	B	C	Blockmittel
Warm		13	21	18	
		14	19	15	
		12	17	15	
Kalt		16	14	15	
		18	11	13	
		17	14	8	
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

$$SSB = 18$$

$$SSC = 12$$

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

$$SSI = 84$$

$$SSE = 50$$

$$SST = 164$$

# Zwei-Wege-ANOVA

- Abweichungen der Freiheitsgrade

$$\text{Bl\"ocke} \times \text{Spalten} \times (\text{Werte} - 1) \\ = 2 \times 3 \times (3 - 1) = \mathbf{12}$$

Dünger:		A	B	C	Blockmittel
Warm		13	21	18	
		14	19	15	
		12	17	15	
Kalt		16	14	15	
		18	11	13	
		17	14	8	
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

$$SSB = 18$$

$$SSI = 84$$

$$SST = 164$$

$$SSC = 12$$

$$SSE = 50$$

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

$$df_{\text{error}} = 12$$

# Zwei-Wege-ANOVA

- Kalkuliere nun F

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

Dünger:	A	B	C		
Warm	13	21	18		B l o c k m i t t e l
	14	19	15	16	
	12	17	15		
Kalt	16	14	15		
	18	11	13	14	
	17	14	8		
Stichprobenmittel	13	19	16		
	17	13	12		
Spaltenmittel	15	16	14	15	

SSB = 18

SSC = 12

df<sub>columns</sub> = 2

SSI = 84

SSE = 50

df<sub>error</sub> = 12

SST = 164



# Zwei-Wege-ANOVA

$$F = 1.44$$

- Suche nun nach dem kritischen Wert  $F_{\text{critical}}$

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

Dünger:		A	B	C	Blockmittel
Warm		13	21	18	
		14	19	15	
		12	17	15	
Kalt		16	14	15	
		18	11	13	
		17	14	8	
Stichprobenmittel		13	19	16	
		17	13	12	
Spaltenmittel		15	16	14	15

$$SSB = 18$$

$$SSC = 12$$

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

$$SSI = 84$$

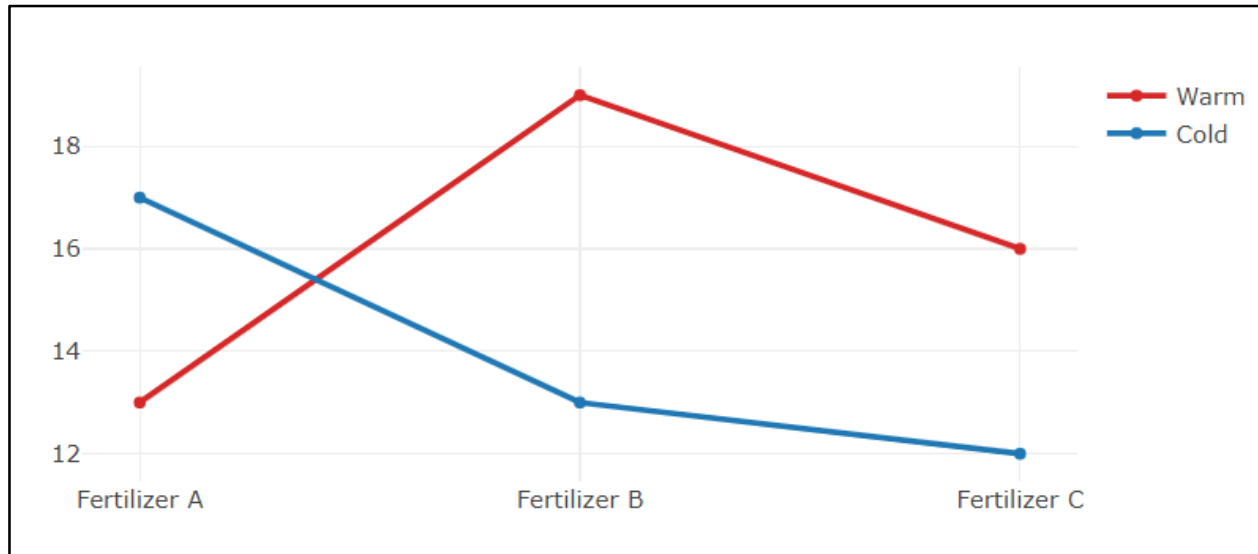
$$SSE = 50$$

$$df_{\text{error}} = 12$$

$$SST = 164$$

# Zwei-Wege-ANOVA

- 2-Wege-ANOVA mit Wiederholen  
- Interaktion:



Dünger:	A	B	C	Blockmittel
Warm	13	21	18	
	14	19	15	
	12	17	15	
Kalt	16	14	15	
	18	11	13	
	17	14	8	
Stichprobenmittel				
Spaltenmittel				

Stichprobenmittel

Spaltenmittel

$$SSB = 18$$

$$SSI = 84$$

$$SST = 164$$

$$SSC = 12$$

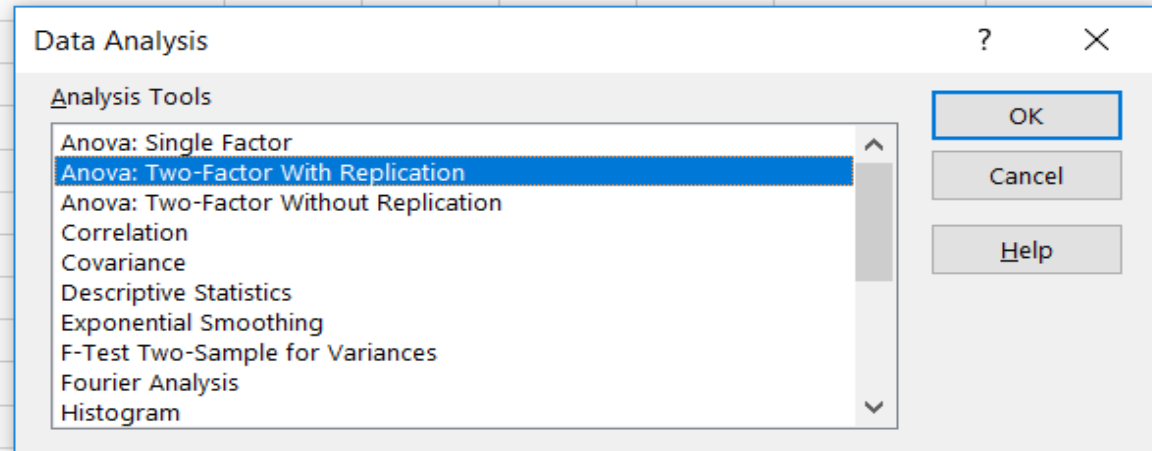
$$SSE = 50$$

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

$$df_{\text{error}} = 12$$

# Zwei-Wege-ANOVA 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			



Als nächstes: Regressionen