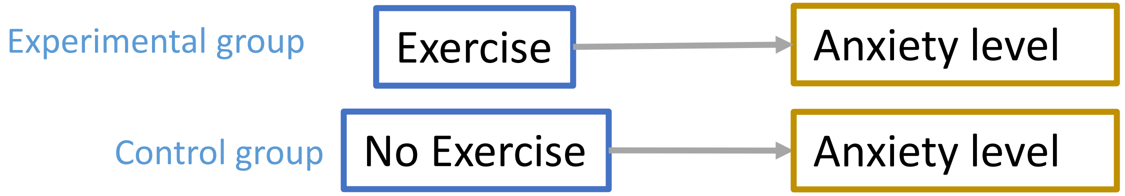
ANOVA

BRSM

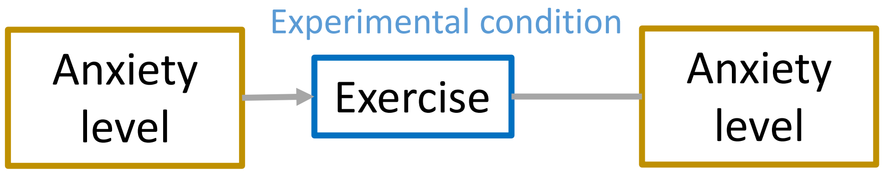
DV: Anxiety level

IV: Exercise

People who exercise have lower levels of anxiety



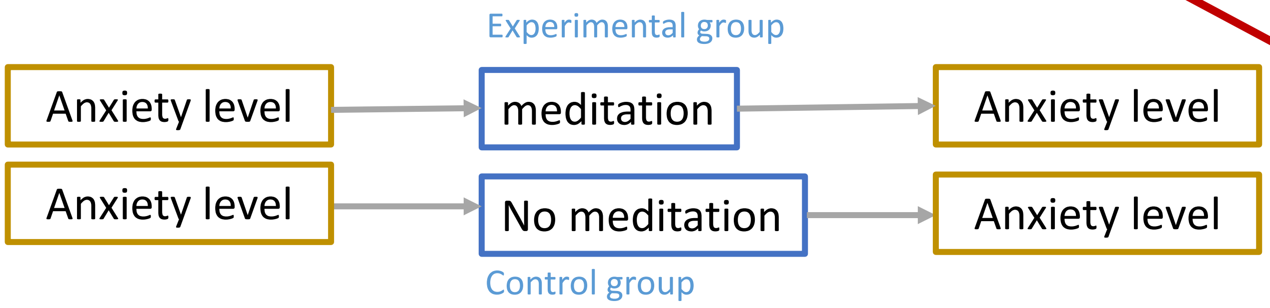
Between groups
(this does not allow you to measure change)



Within group/Repeated measures
(crossover design)

- Participant fatigue
- Longer experimental duration
- Carry over effects

Exercise lowers anxiety



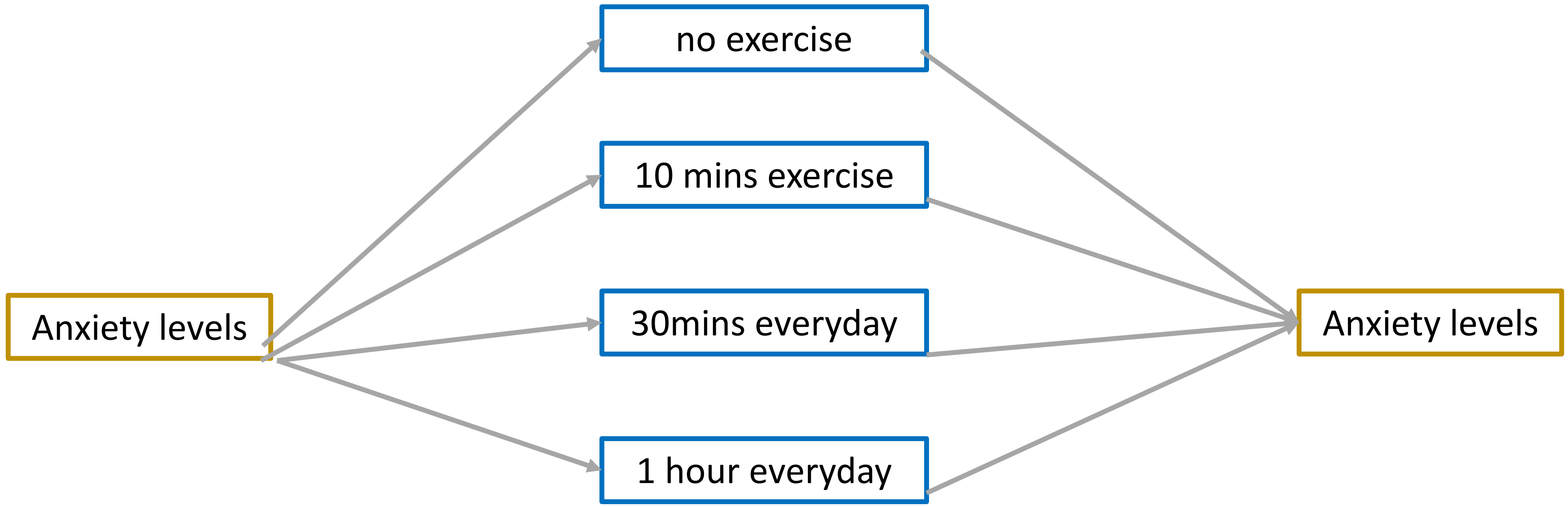
Mixed design
Between groups & Within groups

DV: Anxiety level

IV: Exercise

1 factor, 4 levels

IV – 4 levels



Can we perform multiple t-tests?

Can you do a t-test on this data?

Factor = Independent variable

2 Independent Variables - 2 levels each

Exercise – exercise vs control

Time of Day – morning vs evening

Two factorial design

Exercise-morning	Control-morning
Exercise-evening	Control-evening

2x2 factorial design

2 Independent Variables – different levels

Exercise – 30mins, 1 hour, 2 hours

Time of Day – morning vs evening

Two factorial design

30 mins-morning	1 hr-morning	2 hrs - morning
30 mins-evening	1 hr-evening	2 hrs - evening

3x2 factorial design

Can we perform multiple t-tests?

Why not just perform multiple t-tests?

To avoid Type I error – false positive

- For ’k’ independent groups there are $k(k-1)/2$ possible t-tests
  - For 5 groups = 5(5-1)/2= 10 t-tests
  - For 4 groups = 4(4-1)/2= 6 t-tests
  - For 3 groups = 3(3-1)/2= 3 t-tests

- Using too many t-test comparisons increases the chances of finding random significant effects which may be due to chance. In reality there may be no difference between the groups/conditions

Risk of family-wise error – Increase Type I error

Bonferroni Correction

$$\frac{0.05\ (\alpha)}{\text{No. of comparisons}} = 0.05/3 = 0.0167 \text{ (new } \alpha \text{ value)}$$

Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

s^2 = sample variance

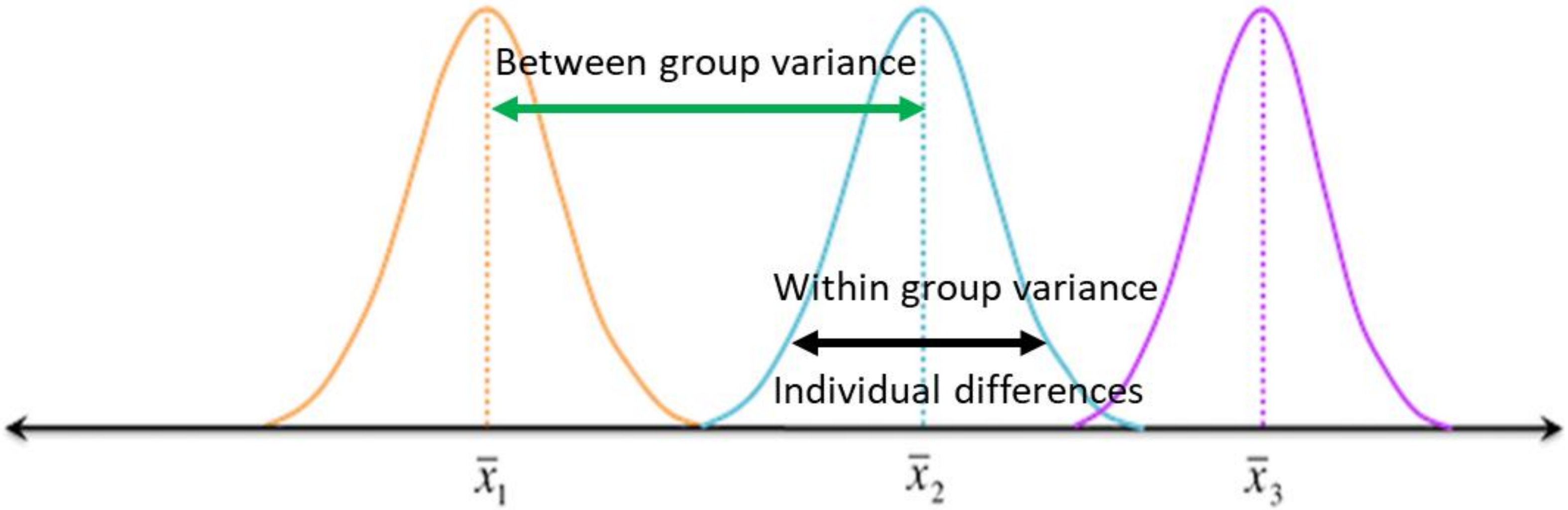
x_i = value of i^{th} element

$\bar{x}$ = sample mean

n = sample size

ANOVA

ANalysis Of VAriance

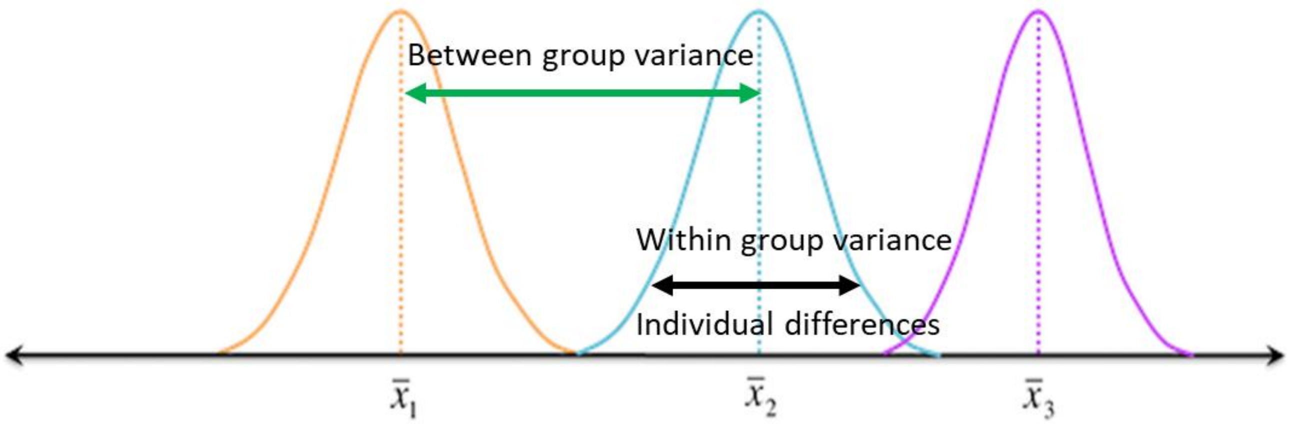


ANOVA performs all three comparisons simultaneously in one test.

No matter how many different means are being compared, ANOVA uses one test with one alpha level to evaluate the difference in variance

$$F = \frac{\text{variance (differences) between sample means}}{\text{variance (differences) within sample}}$$

$$= \frac{\text{difference due to treatment/experimental condition}}{\text{individual differences in each condition}}$$



Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$$

s^2 = sample variance

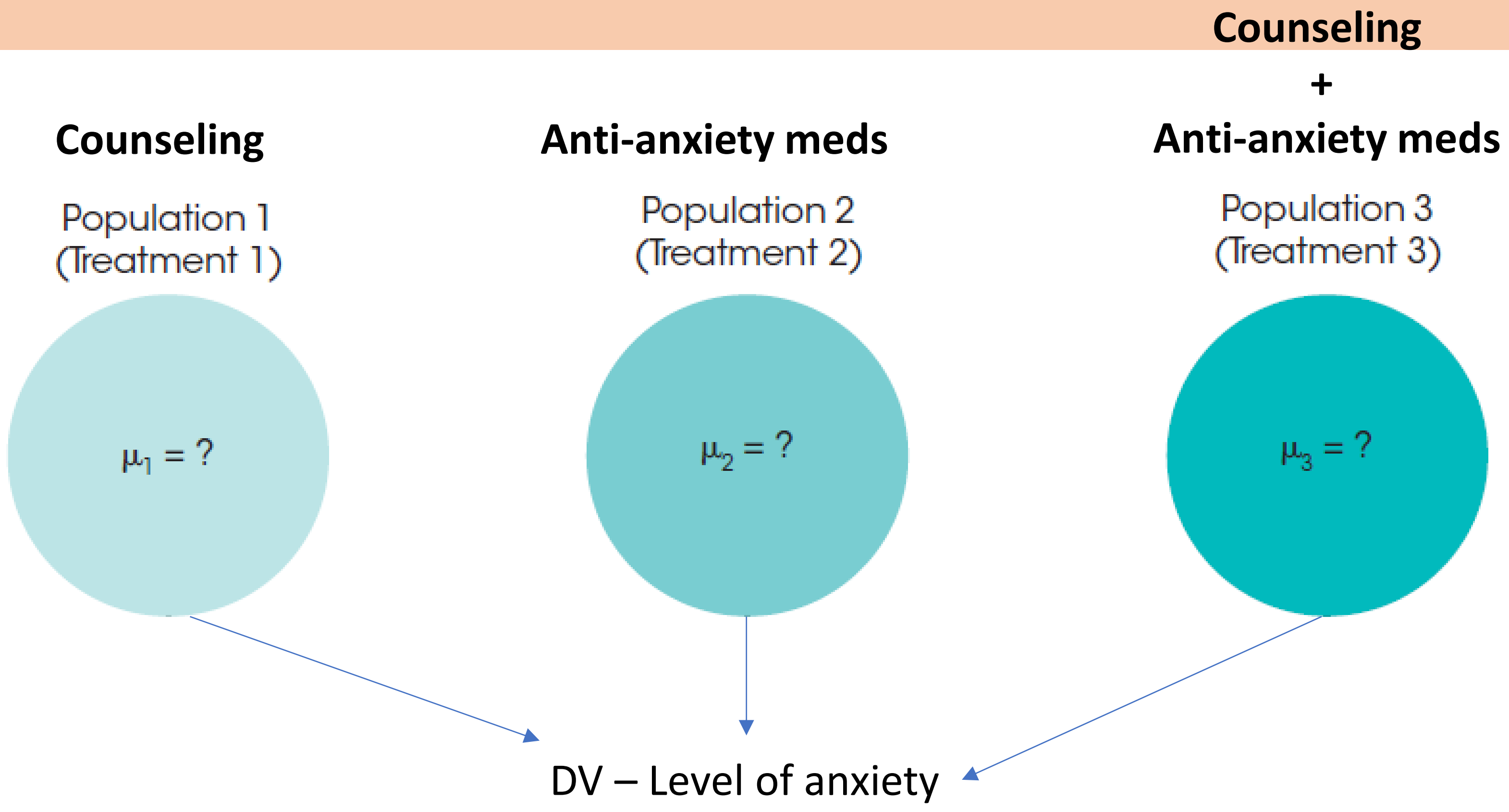
x_i = value of i^{th} element

$\bar{x}$ = sample mean

n = sample size

One Way ANOVA

IV (factor) – Type of treatment



H0 - Null hypothesis – anxiety levels are equal across all groups after treatment (no difference between treatments)

H1 - Alternate hypothesis ?

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}} = \frac{\text{systematic treatment effects} + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$ ≥ 1 (treatment had an effect)

$$F = \frac{0 + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$ ≤ 1 (Treatment had no effect)

	Source	**SS** (SS)	**df**	**s^2** (MS)	F
Effect of treatment	Between treatments	$\sum n_i(\bar{X}_i - \bar{\bar{X}})^2$	$k-1$	$\frac{SS_b}{df_b}$ = MSB	F = $\frac{MSB}{MSW}$
Random differences/error	Within treatments	$\sum (X_{ij} - \bar{X}_i)^2$	$N-k$	$\frac{SS_w}{df_w}$ = MSW	
	Total	$\sum (X_{ij} - \bar{\bar{X}})^2$	$N-1$		

X_{ij} = an individual observation

$\bar{X}_i$ = the mean of the ith group

$\bar{\bar{X}}$ = the grand mean

k = the number of groups

n_i = the number of subjects in the ith group

N = the number of subjects total

Source	SS	df	s^2 (MS)	F
Between treatments	$\sum n_i(\bar{X}_i - \bar{\bar{X}})^2$	$k-1$	$\frac{SS_b}{df_b}$	s_b^2 / s_w^2
Within treatments	$\sum (X_{ij} - \bar{X}_i)^2$	$N-k$	$\frac{SS_w}{df_w}$	
Total	$\sum (X_{ij} - \bar{\bar{X}})^2$	$N-1$		

X_{ij} = an individual observation

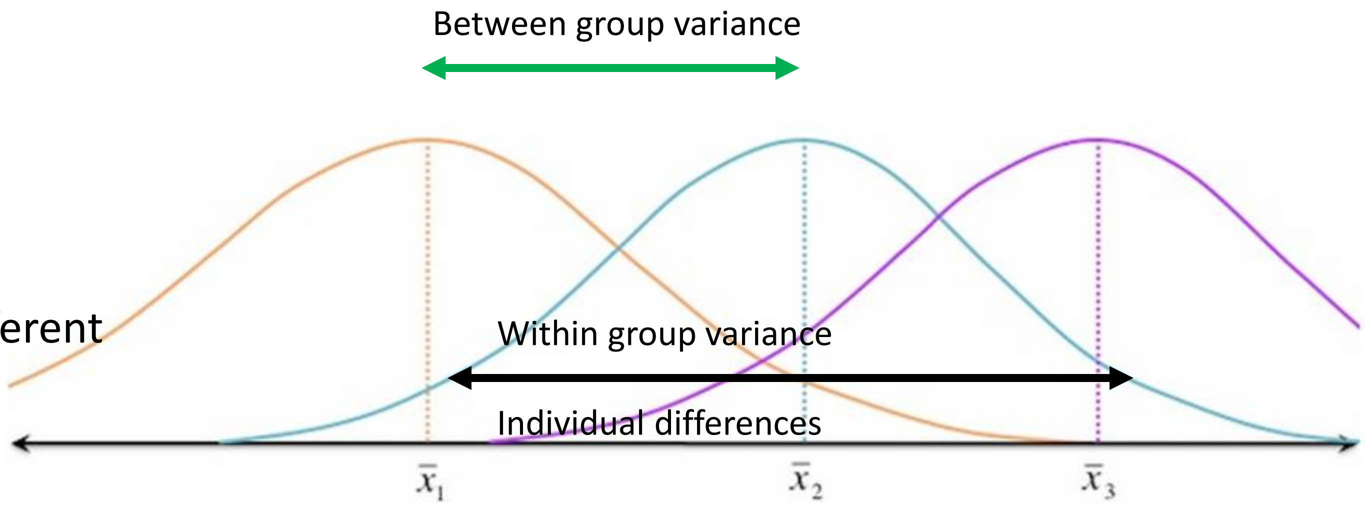
$\bar{X}_i$ = the mean of the ith group

$\bar{\bar{X}}$ = the grand mean

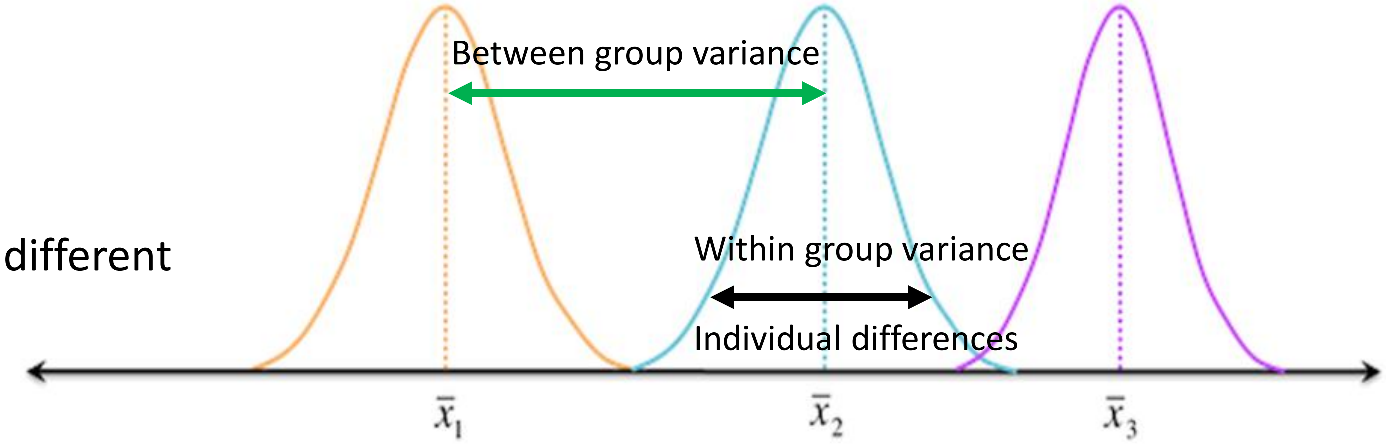
k = the number of groups

n_i = the number of subjects in the ith group

N = the number of subjects total



Group/Treatments are less likely to be significantly different



Groups/Treatments are more likely to be significantly different

Calculating variances

Source	SS	df	s^2 (MS)	F
Between treatments	$\sum n_i(\bar{X}_i - \bar{\bar{X}})^2$	$k-1$	$\frac{SS_b}{df_b}$	F = $\frac{MSB}{MSW}$
Within treatments	$\sum (X_{ij} - \bar{X}_i)^2$	$N-k$	$\frac{SS_w}{df_w}$	
Total	$\sum (X_{ij} - \bar{\bar{X}})^2$	$N-1$		

groups	scores	means	diff	diff_squared
T1	20	11	4	16
T1	11	11	4	16
T1	2	11	4	16
T2	6	5	-2	4
T2	2	5	-2	4
T2	7	5	-2	4
T3	2	5	-2	4
T3	11	5	-2	4
T3	2	5	-2	4
Sums	63	63	0	72
Means	7	7	0	8

SSbetween

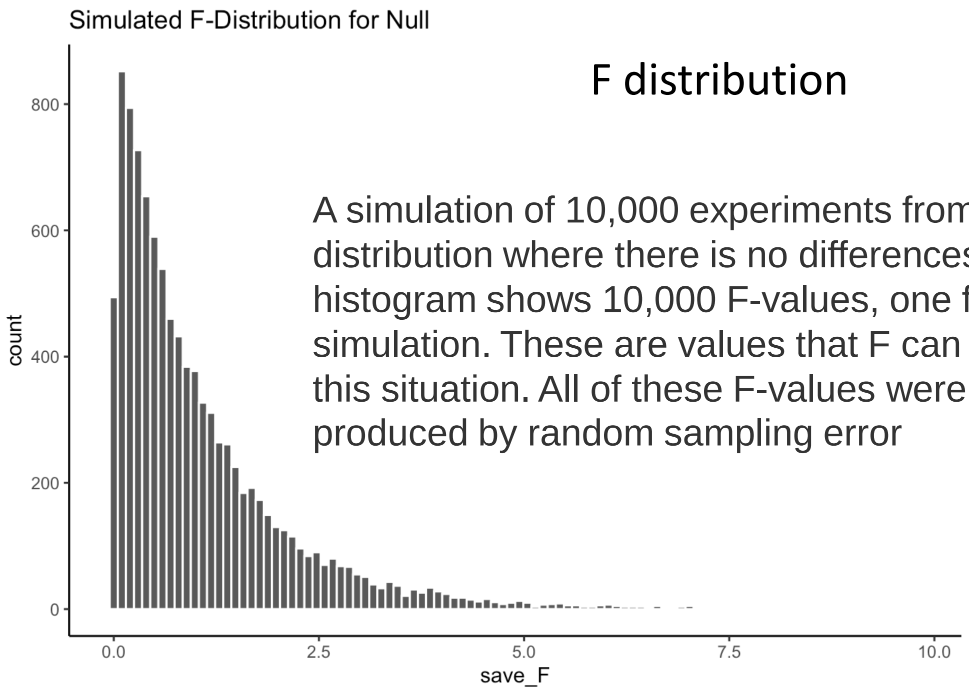
groups	scores	diff	diff_squared
T1	20	13	169
T1	11	4	16
T1	2	-5	25
T2	6	-1	1
T2	2	-5	25
T2	7	0	0
T3	2	-5	25
T3	11	4	16
T3	2	-5	25
Sums	63	0	302
Means	7	0	33.556

SStotal = SSbetween + SSwithin

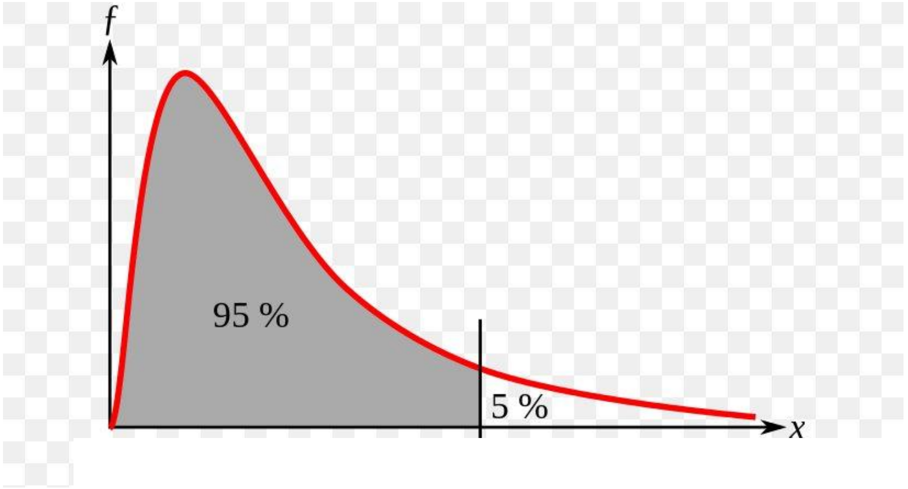
groups	scores	means	diff	diff_squared
T1	20	11	-9	81
T1	11	11	0	0
T1	2	11	9	81
T2	6	5	-1	1
T2	2	5	3	9
T2	7	5	-2	4
T3	2	5	3	9
T3	11	5	-6	36
T3	2	5	3	9
Sums	63	63	0	230
Means	7	7	0	25.556

SSwithin

F distribution



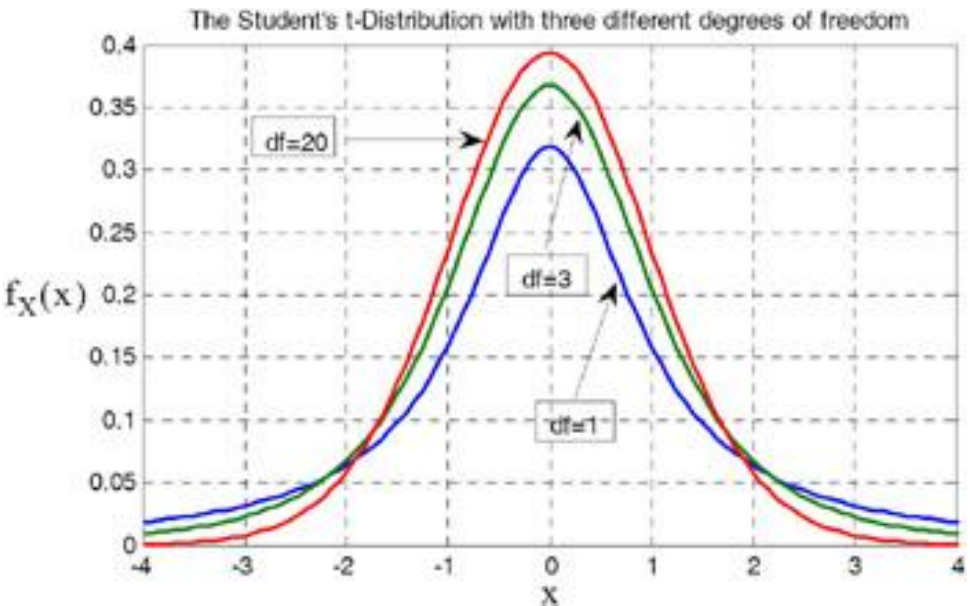
A simulation of 10,000 experiments from a null distribution where there is no differences. The histogram shows 10,000 F-values, one for each simulation. These are values that F can take in this situation. All of these F-values were produced by random sampling error



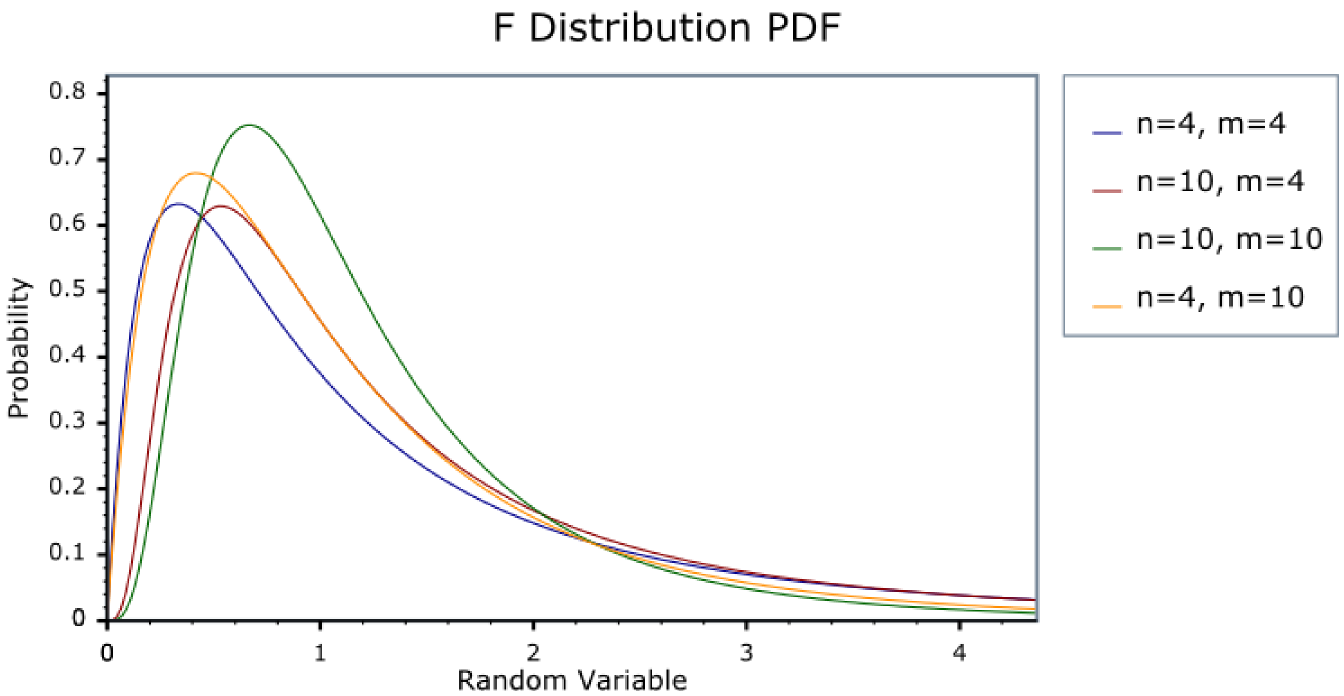
Why is F distribution positively skewed?

F distribution has only one tail – can only tell whether there is difference between the groups or not. Does not tell which group is better or worse.

$$F = t^2$$

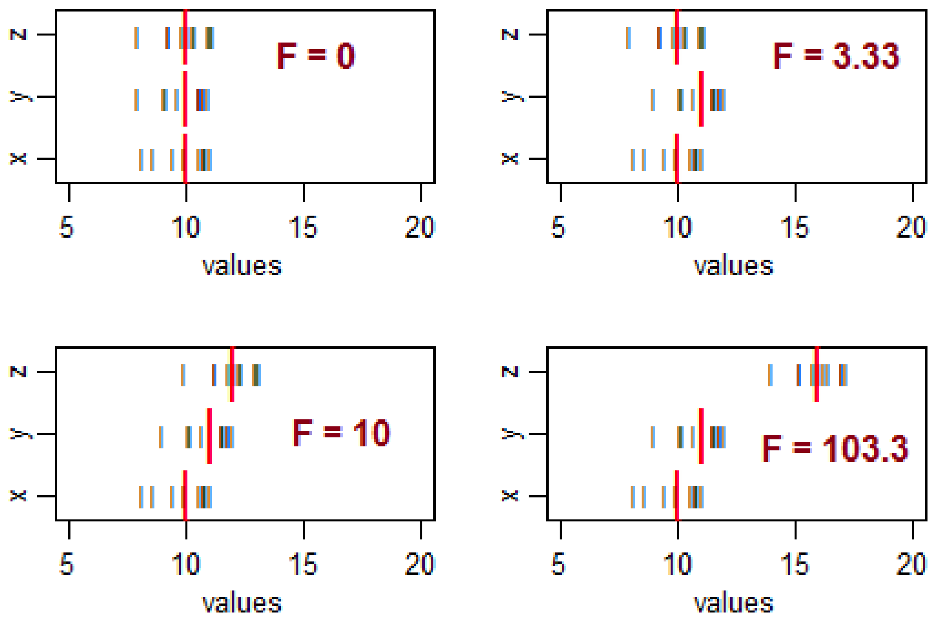


t distribution
(can be one tailed or two-tailed)

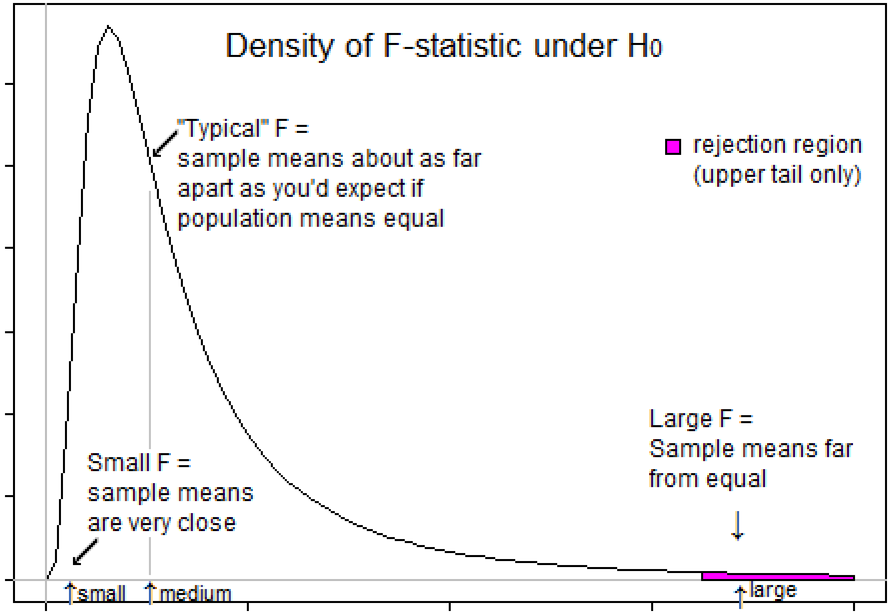


F < 1 – no effect

F > 1 – there might be an effect



Groups	*Count*	*Sum*	*Average*	*Variance*
T1	3	33	11	81
T2	3	15	5	7
T3	3	15	5	27



$$F = \frac{SSB}{SSW}$$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	72	2	36	0.93913	0.441736	5.143253
Within Groups	230	6	38.33333			
Total	302	8				

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	72	2	36	0.93913	0.441736	$\boxed{5.143253}$
Within Groups	230	6	38.33333			
Total	302	8				

$$F = \frac{SSB}{SSW}$$

of the *F* Distribution

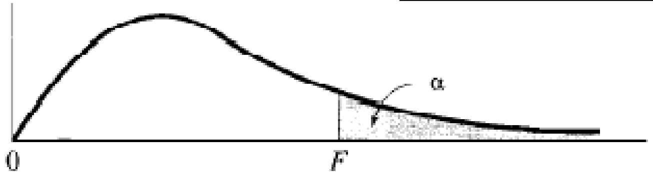


Table 1 $\alpha = 0.05$

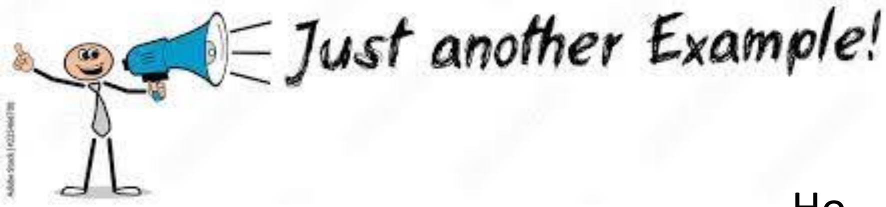
Degrees of Freedom for Numerator

ees of Freedom for Denominator

	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44
6	5.99	$\boxed{5.14}$	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91

One way ANOVA showed that type of treatment had no effect on the level of anxiety $F_{(2,6)} = 0.93$, $p=0.44$

F < 1 (no effect)



One way (One factor, One IV) ANOVA

Ho – exam performance not affected by type of schooling

H1 – Type of schooling affects exam performance

FAKE DATA

Exam performance		
Home school	Boarding school	Regular Day school
89	85	91
75	78	88
49	59	84
87	77	81
84	63	91
68	88	75
88	71	69
78	73	93
77	69	95
93	80	85
67	72	87
79	68	84
69	66	83
88	59	80
91	70	77

Groups	*Count*	*Sum*	*Average*	*Variance*
Home school	15	1182	78.8	141.1714
Boarding school	15	1078	71.86667	73.98095
Regular Day school	15	1263	84.2	50.45714

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1146.711	2	573.3556	6.475922	0.003537	3.219942
Within Groups	3718.533	42	88.53651			
Total	4865.244	44				

$F_{(2,42)} = 6.47, p<0.01$

or

$F_{(2,42)} = 6.47, p=.003$

Effect size for ANOVA

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}} = \frac{1146.711}{4865.244} = 0.236$$

Eta-squared

Table 1 Values of Effect Sizes and Their Interpretation

Kind of Effect Size	*Small*	*Medium*	*Large*
r	.10	.30	.50
d	0.20	0.50	0.80
η^2_p	.01	.06	$\boxed{.14}$
f^2	.02	.15	.35

Source: Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

$F_{(2,42)} = 6.47$, $p=.003$, $\eta^2 = .24$

Type of schooling explains 24% of variance in exam performance

We know there is difference between the groups, but which groups perform better or worse?

Why not just perform multiple t-tests?

To avoid Type I error – false positive

- For 'k' independent groups there are $k(k-1)/2$ possible t-tests
  - For 5 groups = 5(5-1)/2= 10 t-tests
  - For 4 groups = 4(4-1)/2= 6 t-tests
  - For 3 groups = 3(3-1)/2= 3 t-tests

- Using too many t-test comparisons increases the chances of finding random significant effects which may be due to chance. In reality there may be no difference between the groups/conditions

SOLUTION?

Hypothesis Driven
(like a one-tailed test)

Option 1 (Planned Contrasts): Pre-planned, therefore limited no. of comparisons.

You are not comparing all groups to one another, very specific comparisons (so the risk of Type 1 error is controlled)

Exploratory
(like a two-tailed test)

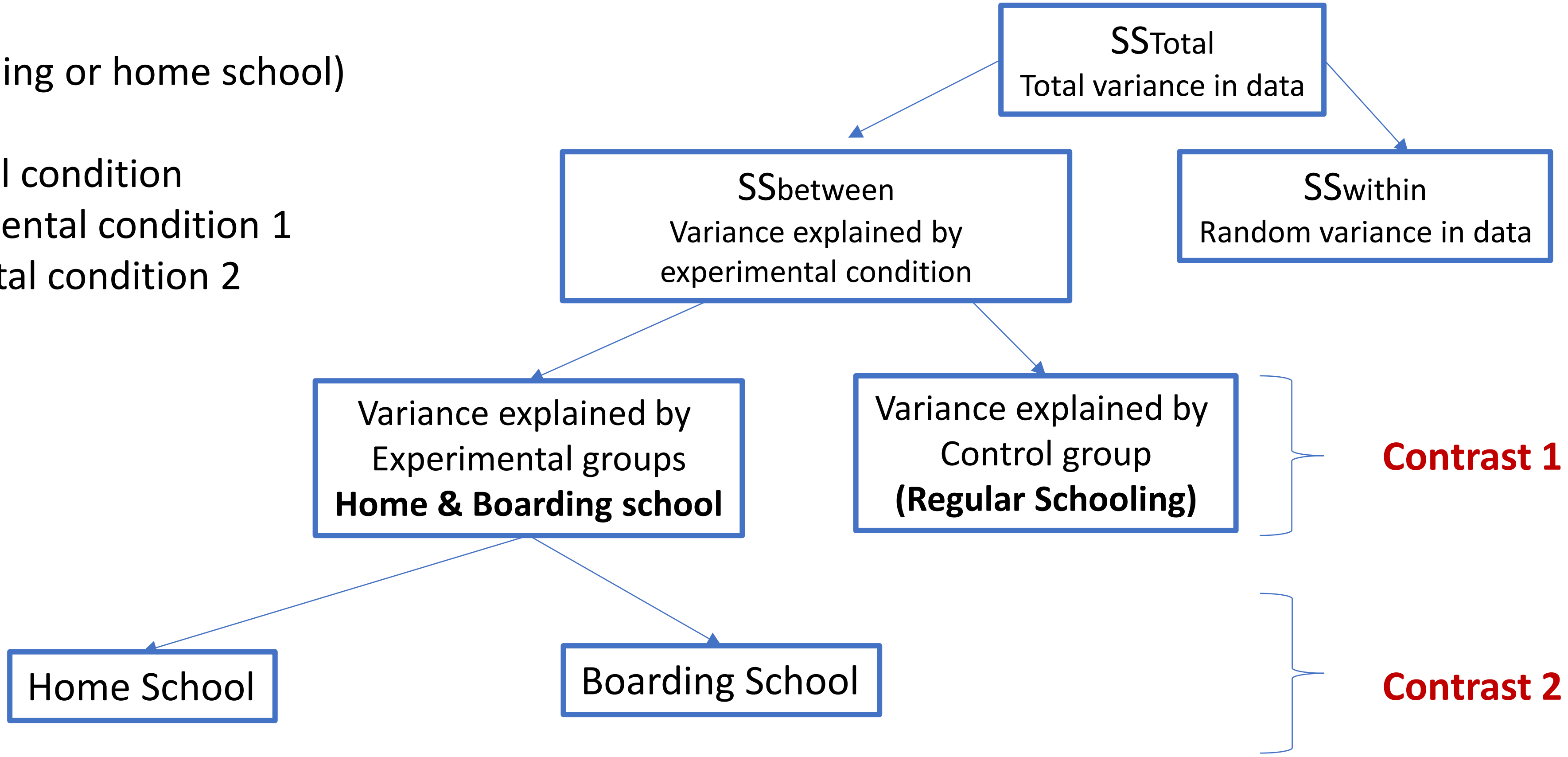
Option 2 (Post Hoc tests) – All possible comparisons can be using special tests (to avoid Type I error).

Planned Comparison

Planned comparison (contrast) – prior to experiment (based on the literature)

Regular schooling > (boarding or home school)

Regular schooling – control condition
Boarding school – Experimental condition 1
Home school – Experimental condition 2



But as the no. of planned comparisons increase (>2 comparisons), the alpha level has to adjusted/Bonferroni correction, again to avoid Type I error.

Bonferroni correction

Adjust the α level by the no. of comparisons

For 3 comparisons, $\alpha/3 = 0.05/3 = 0.0167 \sim 0.01$

Conduct 3 t-tests with $\alpha = 0.01$

Groupwise comparisons	T-test	Bonferroni correction corrected p value
Home vs Boarding	0.07780999	0.016666667
Boarding vs Regular	0.00019644	
Regular vs Home	0.14204177	

Good for small no. of comparisons, else risk of Type II error

Post-hoc test (Tukey's)

Tukey's test requires that the sample size, *n*, be the same for all treatments.

Groups	*Count*	*Sum*	*Average*	*Variance*
Home school	15	1182	78.8	141.1714
Boarding school	15	1078	71.86667	73.98095
Regular Day school	15	1263	84.2	50.45714

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1146.711	2	573.3556	6.475922	0.003537	3.219942
Within Groups	3718.533	42	88.53651			
Total	4865.244	44				

Tukey's $HSD = q\sqrt{\frac{MS_{\text{within}}}{n}}$ = 3.44 √(88.53/15) = 6.95

q – studentized range statistic

The mean difference between any two samples must be more than 6.95 (at α=0.05) to be significant.

$M_{Home} - M_{Regular}$ = |78.8 – 84.2| = 5.4 (not significantly different)
$M_{Boarding} - M_{Regular}$ = |71.866 – 84.2| = 12.33 (significantly different)
$M_{Home} - M_{Boarding}$ = |78.8 – 71.87| = 6.93 (not significantly different)

TABLE B.5 The Studentized Range Statistic (*q*)*

*The critical values for *q* corresponding to α = .05 (lightface type) and α = .01 (boldface type).

df for Error Term	*k* = Number of Treatments 2	3	4	5	6	7	8	9	10	11	12
5	3.64 **5.70**	4.60 **6.98**	5.22 **7.80**	5.67 **8.42**	6.03 **8.91**	6.33 **9.32**	6.58 **9.67**	6.80 **9.97**	6.99 **10.24**	7.17 **10.48**	7.32 **10.70**
6	3.46 **5.24**	4.34 **6.33**	4.90 **7.03**	5.30 **7.56**	5.63 **7.97**	5.90 **8.32**	6.12 **8.61**	6.32 **8.87**	6.49 **9.10**	6.65 **9.30**	6.79 **9.48**
7	3.34 **4.95**	4.16 **5.92**	4.68 **6.54**	5.06 **7.01**	5.36 **7.37**	5.61 **7.68**	5.82 **7.94**	6.00 **8.17**	6.16 **8.37**	6.30 **8.55**	6.43 **8.71**
8	3.26 **4.75**	4.04 **5.64**	4.53 **6.20**	4.89 **6.62**	5.17 **6.96**	5.40 **7.24**	5.60 **7.47**	5.77 **7.68**	5.92 **7.86**	6.05 **8.03**	6.18 **8.18**
9	3.20 **4.60**	3.95 **5.43**	4.41 **5.96**	4.76 **6.35**	5.02 **6.66**	5.24 **6.91**	5.43 **7.13**	5.59 **7.33**	5.74 **7.49**	5.87 **7.65**	5.98 **7.78**
10	3.15 **4.48**	3.88 **5.27**	4.33 **5.77**	4.65 **6.14**	4.91 **6.43**	5.12 **6.67**	5.30 **6.87**	5.46 **7.05**	5.60 **7.21**	5.72 **7.36**	5.83 **7.49**
11	3.11 **4.39**	3.82 **5.15**	4.26 **5.62**	4.57 **5.97**	4.82 **6.25**	5.03 **6.48**	5.20 **6.67**	5.35 **6.84**	5.49 **6.99**	5.61 **7.13**	5.71 **7.25**
12	3.08 **4.32**	3.77 **5.05**	4.20 **5.50**	4.51 **5.84**	4.75 **6.10**	4.95 **6.32**	5.12 **6.51**	5.27 **6.67**	5.39 **6.81**	5.51 **6.94**	5.61 **7.06**
13	3.06 **4.26**	3.73 **4.96**	4.15 **5.40**	4.45 **5.73**	4.69 **5.98**	4.88 **6.19**	5.05 **6.37**	5.19 **6.53**	5.32 **6.67**	5.43 **6.79**	5.53 **6.90**
14	3.03 **4.21**	3.70 **4.89**	4.11 **5.32**	4.41 **5.63**	4.64 **5.88**	4.83 **6.08**	4.99 **6.26**	5.13 **6.41**	5.25 **6.54**	5.36 **6.66**	5.46 **6.77**
15	3.01 **4.17**	3.67 **4.84**	4.08 **5.25**	4.37 **5.56**	4.59 **5.80**	4.78 **5.99**	4.94 **6.16**	5.08 **6.31**	5.20 **6.44**	5.31 **6.55**	5.40 **6.66**
16	3.00 **4.13**	3.65 **4.79**	4.05 **5.19**	4.33 **5.49**	4.56 **5.72**	4.74 **5.92**	4.90 **6.08**	5.03 **6.22**	5.15 **6.35**	5.26 **6.46**	5.35 **6.56**
17	2.98 **4.10**	3.63 **4.74**	4.02 **5.14**	4.30 **5.43**	4.52 **5.66**	4.70 **5.85**	4.86 **6.01**	4.99 **6.15**	5.11 **6.27**	5.21 **6.38**	5.31 **6.48**
18	2.97 **4.07**	3.61 **4.70**	4.00 **5.09**	4.28 **5.38**	4.49 **5.60**	4.67 **5.79**	4.82 **5.94**	4.96 **6.08**	5.07 **6.20**	5.17 **6.31**	5.27 **6.41**
19	2.96 **4.05**	3.59 **4.67**	3.98 **5.05**	4.25 **5.33**	4.47 **5.55**	4.65 **5.73**	4.79 **5.89**	4.92 **6.02**	5.04 **6.14**	5.14 **6.25**	5.23 **6.34**
20	2.95 **4.02**	3.58 **4.64**	3.96 **5.02**	4.23 **5.29**	4.45 **5.51**	4.62 **5.69**	4.77 **5.84**	4.90 **5.97**	5.01 **6.09**	5.11 **6.19**	5.20 **6.28**
24	2.92 **3.96**	3.53 **4.55**	3.90 **4.91**	4.17 **5.17**	4.37 **5.37**	4.54 **5.54**	4.68 **5.69**	4.81 **5.81**	4.92 **5.92**	5.01 **6.02**	5.10 **6.11**
30	2.89 **3.89**	3.49 **4.45**	3.85 **4.80**	4.10 **5.05**	4.30 **5.24**	4.46 **5.40**	4.60 **5.54**	4.72 **5.65**	4.82 **5.76**	4.92 **5.85**	5.00 **5.93**
40	2.86 **3.82**	3.44 **4.37**	3.79 **4.70**	4.04 **4.93**	4.23 **5.11**	4.39 **5.26**	4.52 **5.39**	4.63 **5.50**	4.73 **5.60**	4.82 **5.69**	4.90 **5.76**
60	2.83 **3.76**	3.40 **4.28**	3.74 **4.59**	3.98 **4.82**	4.16 **4.99**	4.31 **5.13**	4.44 **5.25**	4.55 **5.36**	4.65 **5.45**	4.73 **5.53**	4.81 **5.60**
120	2.80 **3.70**	3.36 **4.20**	3.68 **4.50**	3.92 **4.71**	4.10 **4.87**	4.24 **5.01**	4.36 **5.12**	4.47 **5.21**	4.56 **5.30**	4.64 **5.37**	4.71 **5.44**
∞	2.77 **3.64**	3.31 **4.12**	3.63 **4.40**	3.86 **4.60**	4.03 **4.76**	4.17 **4.88**	4.28 **4.99**	4.39 **5.08**	4.47 **5.16**	4.55 **5.23**	4.62 **5.29**

Table 29 of E. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, 2nd ed. New York: Cambridge University Press, 1958. Adapted and reprinted with permission of the Biometrika trustees.

Other post-hoc tests

Games-Howell Test

For unequal variance (unequal sample size)

Calculations are the same as Tukey's but df is calculated using the formula used for unequal sample t-test (Slide 1)

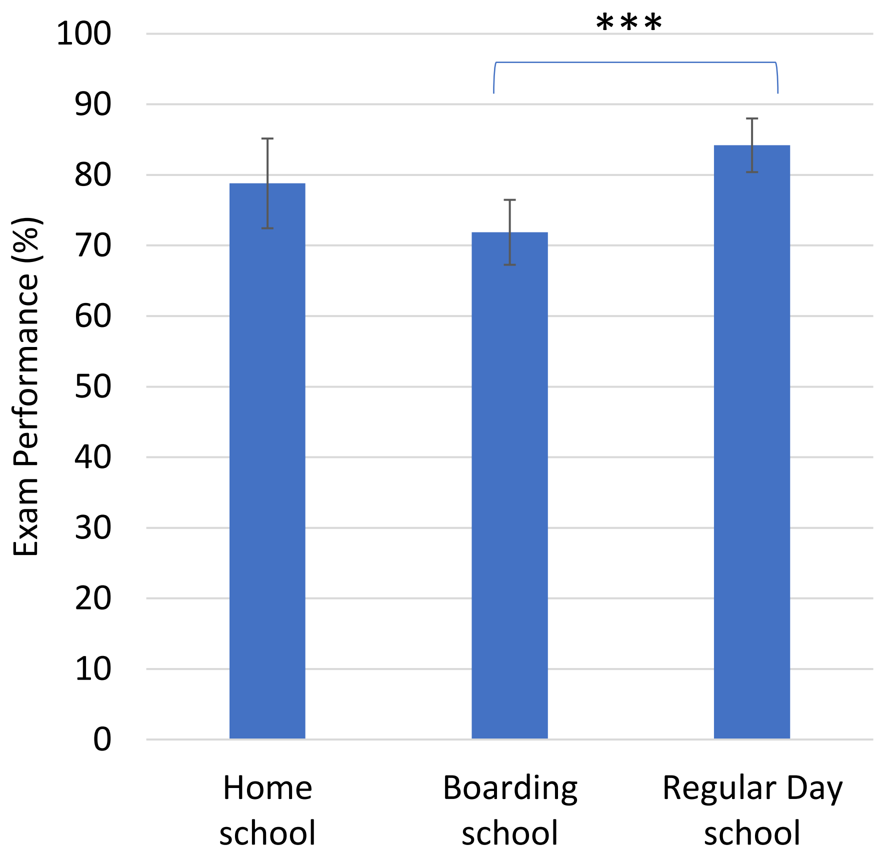
For unequal sample size

$$\text{degrees of freedom, df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Reporting results

- Using a one way ANOVA we observed that the schooling method has a significant effect on exam performance $F_{(2,42)} = 6.47$, $p=.003$, $\eta^2 = .24$

- Using Bonferroni post-hoc test, we found that regular school resulted in better exam performance than boarding school ($p<.001$). There was no significant difference between the other groups.



* p<0.05
** p<0.01
*** p<0.001

Error bars denote confidence intervals

ANOVA assumptions

1. The populations from which the samples are selected must be normal **(parametric vs non-parametric)**
– Shapiro-Wilk test / Kolmogorov-Smirnov test

If violated – use Kruskal –Wallis Test

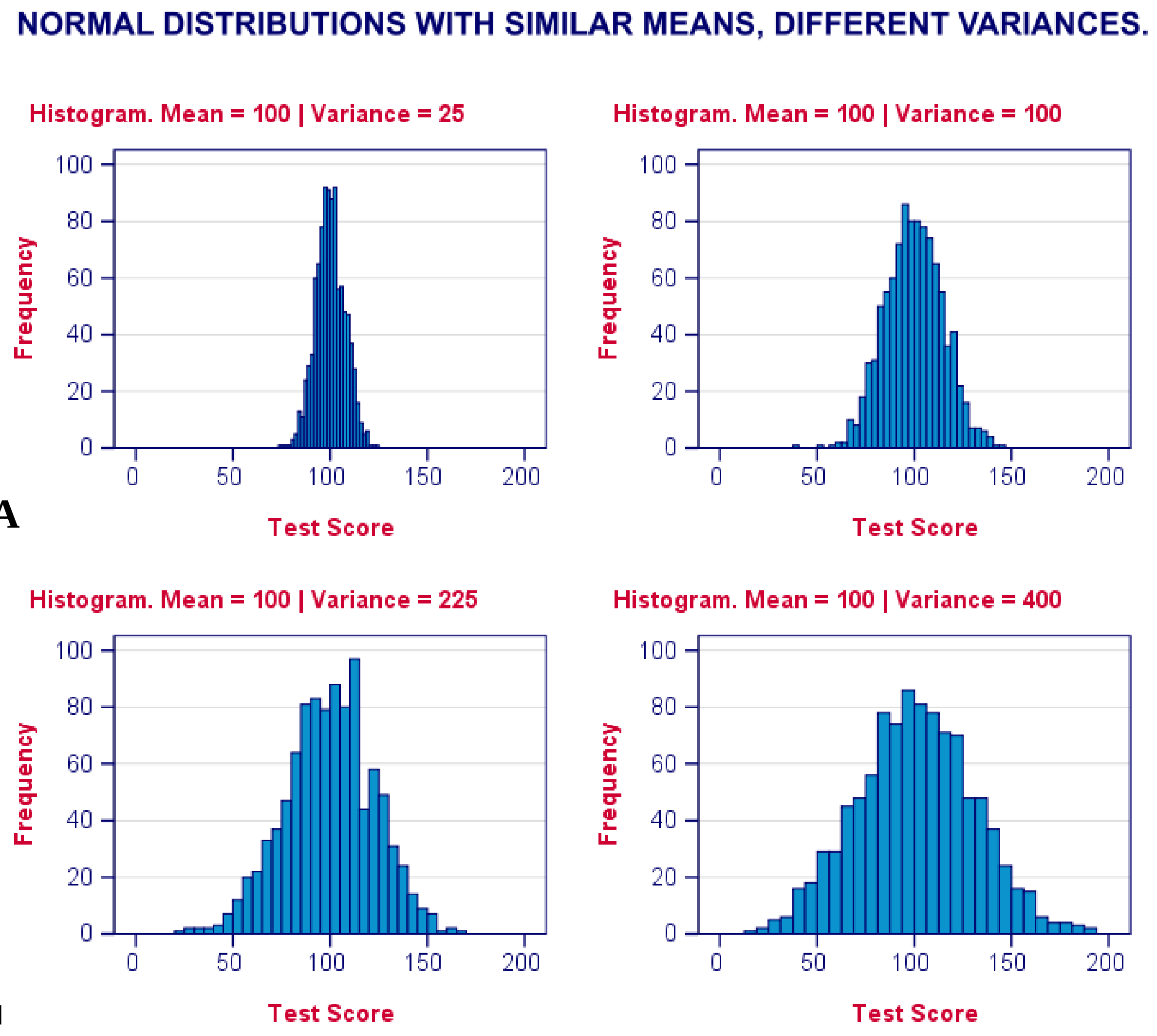
Typically for n>25 in each group, normality can be overlooked in ANOVA

2. The populations from which the samples are selected must have equal variances* **(homogeneity of variance).**
– Levene's or Hartley's *F*-max test for homogeneity of variance

If violated, solution -

1. Collect more samples and equate samples in all groups.

2. Data transformations - natural log or square root transformations
Consequences -- once you transform a variable and conduct your analysis, you can only interpret the transformed variable. You cannot provide an interpretation of the results based on the untransformed variable values.



ANOVA is a robust statistical test, slight violations of assumptions has minor effects on the test outcomes. As long as the largest variance < 4-5 times smallest variance, ANOVA results are valid

Homogeneity of variances (homoscedasticity)

$$F\text{-max} = \frac{s^2(\text{largest})}{s^2(\text{smallest})}$$

F-max ~ 1.00 → sample variances are similar and homogenous
Look up df=(n-1), k in the Fmax table

If your calculated value < table value, the variance is homogeneous.
If your calculated value > table value, the variance is not homogeneous.

TABLE B.3 Critical Values for the *F*-Max Statistic*

*The critical values for $\alpha = .05$ are in lightface type, and for $\alpha = .01$, they are in boldface type.

	k = Number of Samples										
$n-1$	2	3	4	5	6	7	8	9	10	11	12
4	9.60 **23.2**	15.5 **37.**	20.6 **49.**	25.2 **59.**	29.5 **69.**	33.6 **79.**	37.5 **89.**	41.4 **97.**	44.6 **106.**	48.0 **113.**	51.4 **120.**
5	7.15 **14.9**	10.8 **22.**	13.7 **28.**	16.3 **33.**	18.7 **38.**	20.8 **42.**	22.9 **46.**	24.7 **50.**	26.5 **54.**	28.2 **57.**	29.9 **60.**
6	5.82 **11.1**	8.38 **15.5**	10.4 **19.1**	12.1 **22.**	13.7 **25.**	15.0 **27.**	16.3 **30.**	17.5 **32.**	18.6 **34.**	19.7 **36.**	20.7 **37.**
7	4.99 **8.89**	6.94 **12.1**	8.44 **14.5**	9.70 **16.5**	10.8 **18.4**	11.8 **20.**	12.7 **22.**	13.5 **23.**	14.3 **24.**	15.1 **26.**	15.8 **27.**
8	4.43 **7.50**	6.00 **9.9**	7.18 **11.7**	8.12 **13.2**	9.03 **14.5**	9.78 **15.8**	10.5 **16.9**	11.1 **17.9**	11.7 **18.9**	12.2 **19.8**	12.7 **21.**
9	4.03 **6.54**	5.34 **8.5**	6.31 **9.9**	7.11 **11.1**	7.80 **12.1**	8.41 **13.1**	8.95 **13.9**	9.45 **14.7**	9.91 **15.3**	10.3 **16.0**	10.7 **16.6**
10	3.72 **5.85**	4.85 **7.4**	5.67 **8.6**	6.34 **9.6**	6.92 **10.4**	7.42 **11.1**	7.87 **11.8**	8.28 **12.4**	8.66 **12.9**	9.01 **13.4**	9.34 **13.9**
12	3.28 **4.91**	4.16 **6.1**	4.79 **6.9**	5.30 **7.6**	5.72 **8.2**	6.09 **8.7**	6.42 **9.1**	6.72 **9.5**	7.00 **9.9**	7.25 **10.2**	7.48 **10.6**
15	2.86 **4.07**	3.54 **4.9**	4.01 **5.5**	4.37 **6.0**	4.68 **6.4**	4.95 **6.7**	5.19 **7.1**	5.40 **7.3**	5.59 **7.5**	5.77 **7.8**	5.93 **8.0**
20	2.46 **3.32**	2.95 **3.8**	3.29 **4.3**	3.54 **4.6**	3.76 **4.9**	3.94 **5.1**	4.10 **5.3**	4.24 **5.5**	4.37 **5.6**	4.49 **5.8**	4.59 **5.9**
30	2.07 **2.63**	2.40 **3.0**	2.61 **3.3**	2.78 **3.5**	2.91 **3.6**	3.02 **3.7**	3.12 **3.8**	3.21 **3.9**	3.29 **4.0**	3.36 **4.1**	3.39 **4.2**
60	1.67 **1.96**	1.85 **2.2**	1.96 **2.3**	2.04 **2.4**	2.11 **2.4**	2.17 **2.5**	2.22 **2.5**	2.26 **2.6**	2.30 **2.6**	2.33 **2.7**	2.36 **2.7**

Table 31 of E. Pearson and H.O. Hartley, Biometrika Tables for Statisticians, 2nd ed. New York: Cambridge University Press, 1958. Adapted and reprinted with permission of the Biometrika trustees.

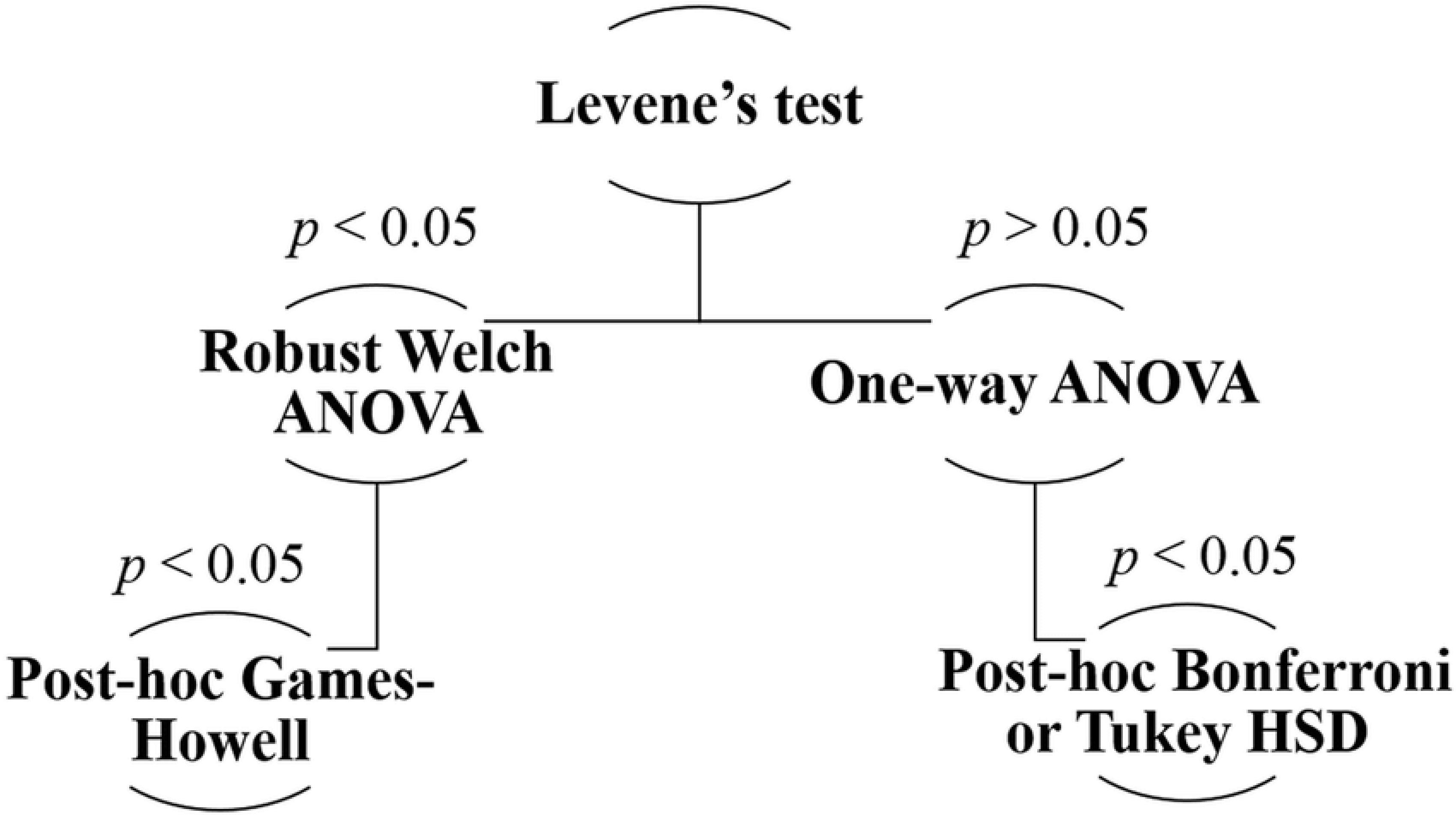
Levene's Test (more robust)

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2}$$

where

- k is the number of different groups to which the sampled cases belong,
- N_i is the number of cases in the ith group,
- N is the total number of cases in all groups,
- Y_{ij} is the value of the measured variable for the jth case from the ith group,
- $Z_{ij} = |Y_{ij} - \bar{Y}_{i\cdot}|$, $\bar{Y}_{i\cdot}$ is a mean of the i-th group,
- $Z_{i\cdot} = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij}$ is the mean of the Z_{ij} for group i,
- $Z_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{N_i} Z_{ij}$ is the mean of all Z_{ij}.

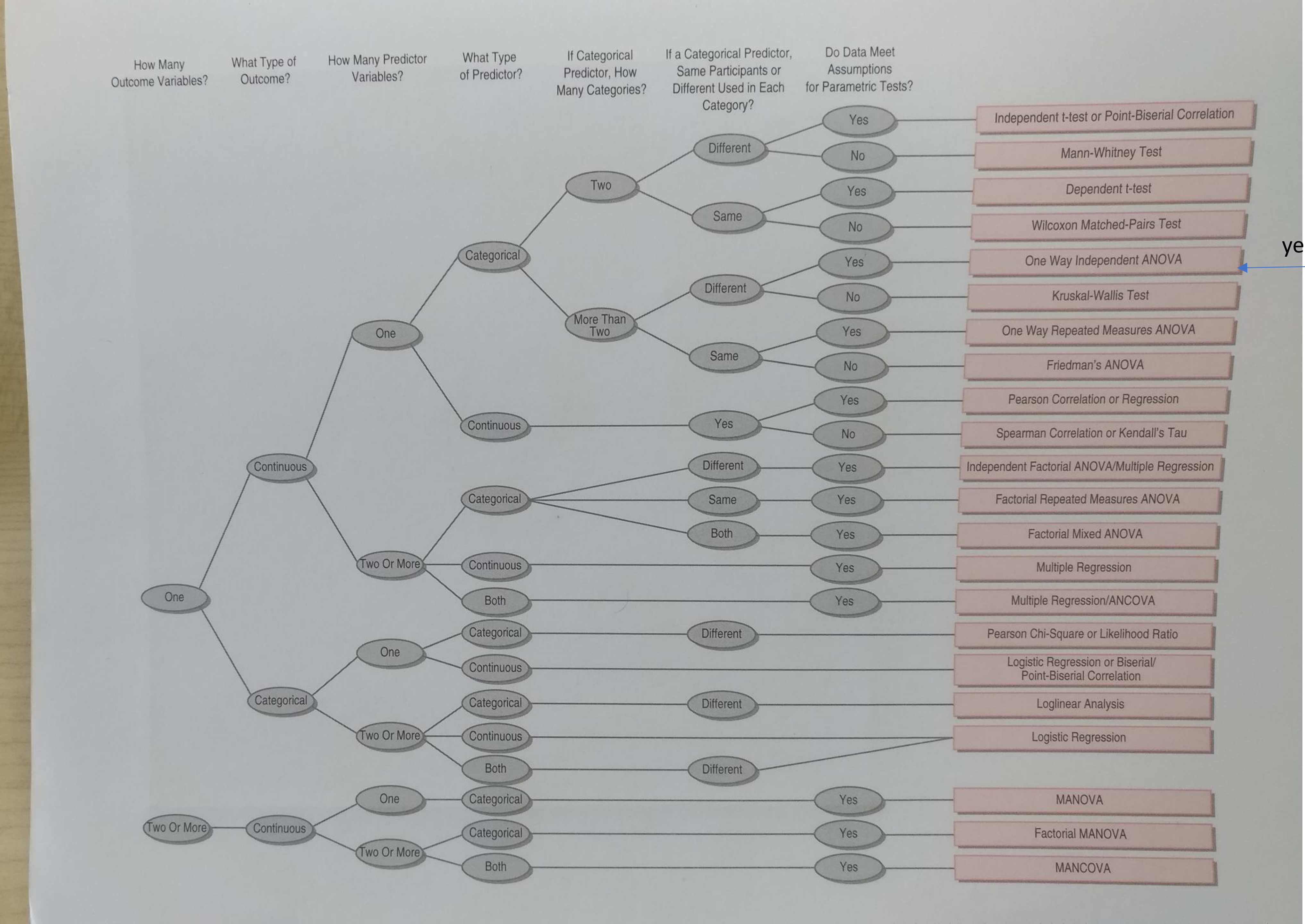
https://www.youtube.com/watch?v=gL7K4vZq0Z4



Levene's test
p < 0.05
p > 0.05
Robust Welch ANOVA
One-way ANOVA
p < 0.05
p < 0.05
Post-hoc Games-Howell
Post-hoc Bonferroni or Tukey HSD

Steps in ANOVA

- Check for normality (parametric vs non-parametric tests)
- Check homogeneity of variances (spread of the data is equal across groups or not)
- Choose the appropriate test (ANOVA, Kruskal-Wallis, Welch)
- Only if main effect (F) significant, use a post-hoc test
- Report effect size for significant effects
- Plot analyzed data



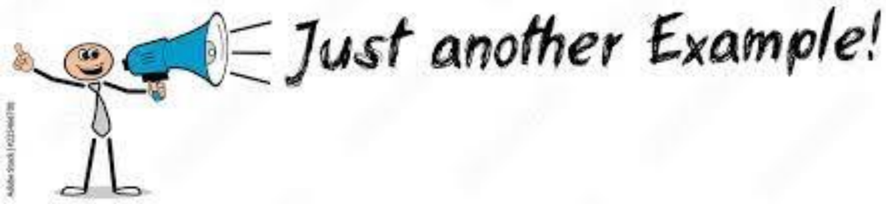
Welch ANOVA Test

no

yes

homogenous

Andy Field
Discovering Stats



One way (One factor, One IV) ANOVA

Testing homogeneity

Ho – variance across groups is equal

H1 – variance across groups is unequal

Exam performance		
Home school	Boarding school	Regular Day school
89	85	91
75	78	88
49	59	84
87	77	81
84	63	91
68	88	75
88	71	69
78	73	93
77	69	95
93	80	85
67	72	87
79	68	84
69	66	83
88	59	80
91	70	77

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
exam_performance	Based on Mean	1.675	2	42	.200
	Based on Median	1.648	2	42	.205
	Based on Median and with adjusted df	1.648	2	35.806	.207
	Based on trimmed mean	1.690	2	42	.197

Tests of Normality

<50 samples

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	school_code	Statistic	df	Sig.	Statistic	df	Sig.
exam_performance	home school	.155	15	.200*	.905	15	.115
	boarding school	.114	15	.200*	.969	15	.847
	regular scchool	.100	15	.200*	.975	15	.923

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

(When normality is violated)

Kruskal-Wallis Test

$$H = \left(\frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j}\right) - 3(N+1)$$

Where, N = Total observation in all groups (total sample size); k = Number of groups; n_j= sample size for jth group, and R_j is the sum of ranks of jth group

(a) Original Numerical Scores

I	II	III	
14	2	26	$N = 15$
3	14	8	
21	9	14	
5	12	19	
16	5	20	
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	

(b) Ordinal Data (Ranks)

I	II	III	
9	1	15	$N = 15$
2	9	5	
14	6	9	
3.5	7	12	
11	3.5	13	
$T_1 = 39.5$	$T_2 = 26.5$	$T_3 = 54$	
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	

$$H = \frac{12}{N(N+1)}\left(\Sigma\frac{T^2}{n}\right) - 3(N+1)$$

$$\begin{aligned} H &= \frac{12}{15(16)}\left(\frac{39.5^2}{5} + \frac{26.5^2}{5} + \frac{54^2}{5}\right) - 3(16) \\ &= 0.05(312.05 + 140.45 + 583.2) - 48 \\ &= 0.05(1035.7) - 48 \\ &= 51.785 - 48 \\ &= 3.785 \end{aligned}$$

Use chi-square distribution

With df = (k-1) = 2

$H_{critical}$ = 5.99 for α = .05

H (3.785) < 5.99

Accept H_0.

Since the data were not normally distributed, Kruskal-Wallis test for non-parametric data was used to evaluate differences among the three treatments. The outcome of the test indicated no significant differences among the treatment conditions, $H = 3.785$ $(2, N = 15)$, $p > .05$.

(When homogeneity is violated)

Welch ANOVA Test

Welch ANOVA effect size $$est.\,\omega^2 = \frac{df_{bet}(F-1)}{df_{bet}(F-1)+N_T}$$

$$F = \frac{\frac{1}{k-1}\sum_{j=1}^{k} w_j(\bar{x}_j - \bar{x}')^2}{1 + \frac{2(k-2)}{k^2-1}\sum_{j=1}^{k}\left(\frac{1}{n_j-1}\right)\left(1-\frac{w_j}{w}\right)^2}$$

$$w_j = \frac{n_j}{s_j^2} \qquad w = \sum_{j=1}^{k} w_j \qquad \bar{x}' = \frac{\sum_{j=1}^{k} w_j \bar{x}_j}{w}$$

Robust Tests of Equality of Means

exam_performance

	Statistic[a]	df1	df2	Sig.
Welch	8.954	2	26.995	.001

a. Asymptotically F distributed.

$$F \sim F(k-1, df)$$

$$df = \frac{k^2-1}{3\sum_{j=1}^{k}\left(\frac{1}{n_j-1}\right)\left(1-\frac{w_j}{w}\right)^2}$$

The degrees of freedom for Welch's t-test takes into account the difference between the two standard deviations.

df with decimal places → round off to look up in the F table

https://www.real-statistics.com/one-way-analysis-of-variance-anova/welchs-procedure/

One way Repeated Measures ANOVA

Advantage?

Disadvantage?

subjects	T1 counseling	T2 Anti-anxiety meds	T3 both
1	20	11	2
2	6	2	7
3	2	11	2

Experimental Design is equally important

subjects	T1 counseling	T2 counselling	T3 counselling
1	20	11	2
2	6	2	7
3	2	11	2

One way Repeated Measures ANOVA

SStotal = SSbetween + SSwithin

SSTotal = SSbetween + SSsubjects + SSerror

	Dependent variable (DV)		
Participants	**Timepoint 1**	**Timepoint 2**	**Timepoint 3**
1	20	11	2
2	6	2	7
3	2	11	2

Source of variance	**SS**	**df**	**MS**	**F**
Between	SS*between*	k-1	MS*between* = $\frac{SSbetween}{k-1}$	F = $\frac{MSbetween}{MSerror}$
Within	SS*within*			
Subjects	SS*subjects*	n-1		
Error (left-over error)	SS*error* = SS*within* - SS*subjects*	(k-1)(n-1)	MS*error* = $\frac{SSerror}{(k-1)(n-1)}$	
Total	SS*total*	N-1		

subjects	conditions	scores	means	diff	diff_squared
1	A	$\boxed{20}$ 28/3	9.33	2.33	5.4289
2	A	11	8	1	1
3	A	2	3.66	-3.34	11.1556
1	B	$\boxed{6}$	9.33	2.33	5.4289
2	B	2	8	1	1
3	B	7	3.66	-3.34	11.1556
1	C	$\boxed{2}$	9.33	2.33	5.4289
2	C	11	8	1	1
3	C	2	3.66	-3.34	11.1556
Sums		63	62.97	-0.0299	52.75
Means		7	6.997	-0.0033	5.8615

SSsubjects

SSwithin = SSsubjects + Sserror

SSerror = 230 – 52.75 = 177.25

subjects	conditions	scores	diff	diff_squared
1	A	$\boxed{20}$	13	169
2	A	11	4	16
3	A	2	-5	25
1	B	6	-1	1
2	B	2	-5	25
3	B	7	0	0
1	C	2	-5	25
2	C	11	4	16
3	C	2	-5	25
Sums		63	0	302
Means		$\boxed{7}$	0	33.556

SStotal

subjects	conditions	scores	means	diff	diff_squared
1	A	20	$\boxed{11}$	4	16
2	A	11	11	4	16
3	A	2	11	4	16
1	B	6	5	-2	4
2	B	2	5	-2	4
3	B	7	5	-2	4
1	C	2	5	-2	4
2	C	11	5	-2	4
3	C	2	5	-2	4
Sums		63	63	0	72
Means		7	$\boxed{7}$	0	8

SSbetween

subjects	conditions	scores	means	diff	diff_squared
1	A	$\boxed{20}$	$\boxed{11}$	-9	81
2	A	11	11	0	0
3	A	2	11	9	81
1	B	6	5	-1	1
2	B	2	5	3	9
3	B	7	5	-2	4
1	C	2	5	3	9
2	C	11	5	-6	36
3	C	2	5	3	9
Sums		63	63	0	230
Means		7	7	0	25.556

SSwithin

Source of variance	SS	df	MS	F	*p*
Between	*52.67*	2	26.33	0.594	.59
Error (left-over error)	177.33	4	44.33		

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	52.667	2	26.333	.594	.594
	Greenhouse-Geisser	52.667	1.814	29.032	.594	.584
	Huynh-Feldt	52.667	2.000	26.333	.594	.594
	Lower-bound	52.667	1.000	52.667	.594	.521
Error(factor1)	Sphericity Assumed	177.333	4	44.333		
	Greenhouse-Geisser	177.333	3.628	48.877		
	Huynh-Feldt	177.333	4.000	44.333		
	Lower-bound	177.333	2.000	88.667		

Using a one way repeated measures ANOVA we observed that there was no difference in scores in the 3 the timepoints $F_{(2,4)} = 0.594$, p=.59.

Mauchly's sphericity test

Sphericity → condition where the variances of related groups (levels T1, T2 , T3) are equal.

- Analogous to homogeneity of variances
- Used specifically in repeated measures testing

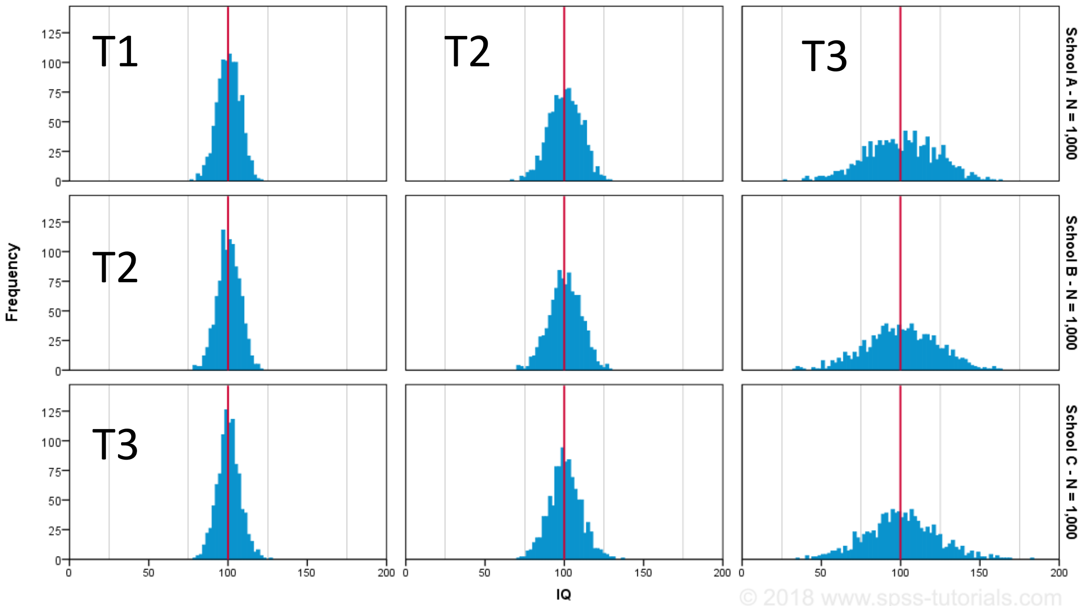
Df*between OR* $df_{time/condition} = (k-1)$

$$df_{error} = (k-1)(n-1)$$

$$\boxed{\begin{aligned} df_{time/condition} &= \hat{\varepsilon}(k-1) \\ df_{error} &= \hat{\varepsilon}(k-1)(n-1) \end{aligned}}$$

Non-homogenous

homogenous



Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Lower-bound
factor1	.898	.108	2	.947	$\boxed{.907}$	1.000	.500

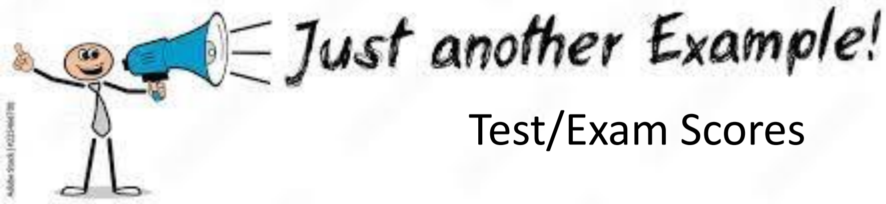
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: factor1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Steps in ANOVA

- Check for normality
- Check homogeneity of variances (different participants in different groups, Factorial ANOVA)
- Check for sphericity of variances (same participants across groups, repeated measures ANOVA)
- Choose the appropriate test
- Only if main effect (F) significant, use a post-hoc test
- Report effect size for significant effects
- Plot analyzed data



One way (One DV) repeated measures ANOVA

Test/Exam Scores

Student	Reread	Answer Prepared Questions	Create and Answer Questions
A	2	5	8
B	3	9	6
C	8	10	12
D	6	13	11
E	5	8	11
F	6	9	12

Mauchly's Test of Sphericity[a]

Measure: test_score

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt
factor1	.372	3.957	2	.138	.614	.712

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportio
to an identity matrix.

a. Design: Intercept
Within Subjects Design: factor1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the
Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: test_score

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
factor1	Sphericity Assumed	84.000	2	42.000	19.091	<.001	.792
	Greenhouse-Geisser	84.000	1.228	68.380	19.091	.004	.792
	Huynh-Feldt	84.000	1.424	58.991	19.091	.002	.792
	Lower-bound	84.000	1.000	84.000	19.091	.007	.792
Error(factor1)	Sphericity Assumed	22.000	10	2.200			
	Greenhouse-Geisser	22.000	6.142	3.582			
	Huynh-Feldt	22.000	7.120	3.090			
	Lower-bound	22.000	5.000	4.400			

Tests of Normality

<50 samples

	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
reread	.176	6	.200*	.955	6	.783
answer_questions	.184	6	.200*	.957	6	.799
create_answer_questions	.325	6	.047	.827	6	.101

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Using a one way repeated measures ANOVA we observed that strategy for studying in preparation for a test had an effect on exam score $F_{(2,10)} = 19.09$, $p<.001$, $\eta^2 = .79$

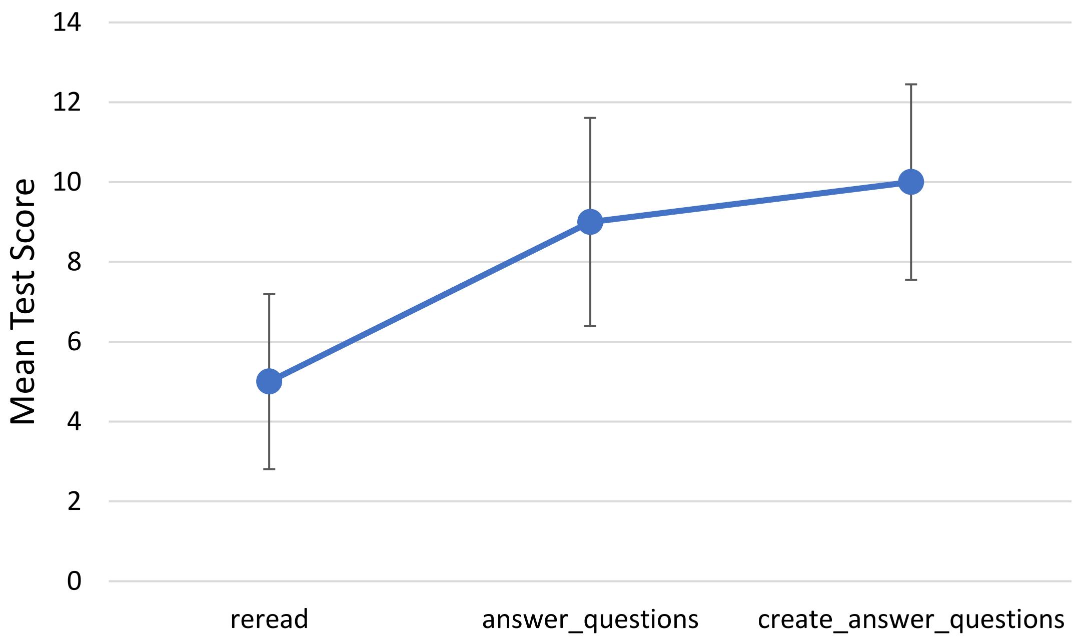
Tests of Within-Subjects Contrasts

Measure: test_score

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
factor1	Linear	75.000	1	75.000	93.750	<.001	.949
	Quadratic	9.000	1	9.000	2.500	.175	.333
Error(factor1)	Linear	4.000	5	.800			
	Quadratic	18.000	5	3.600			

Using a one way repeated measures ANOVA we observed that strategy for studying in preparation for a test had an effect on exam score $F_{(2,10)} = 19.09$, $p<.001$, $\eta^2 = .79$

Within-subjects contrasts revealed that there was a linear trend $F_{(1,5)} = 93.75$, $p<.001$, $\eta^2 = .95$



Error bars denote standard deviations

Friedman's Test (non-normal repeated measures)

$$M = \frac{12}{Nk(k+1)} \sum R_i^2 - 3N(k+1)$$

Where, k = number of columns (treatments)

n = number of rows (blocks)

R_j = sum of the ranks

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	6
Test Statistic	9.333
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.009

Student	Reread	Answer Prepared Questions	Create and Answer Questions
A	2	5	8
B	3	9	6
C	8	10	12
D	6	13	11
E	5	8	11
F	6	9	12

Rank		
1	2	3
1	3	2
1	2	3
1	3	2
1	2	3
1	2	3
Sum = 6	14	16

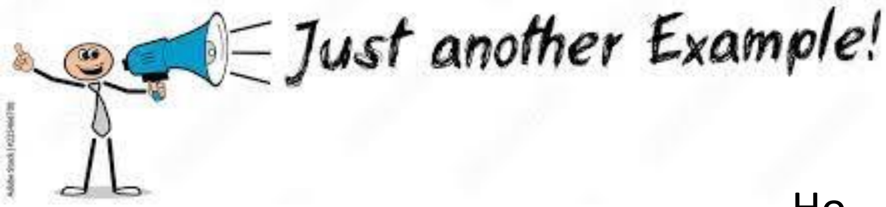
IV – categorical DV – continuous (interval, ratio)

		Independent factor 1 IV > 2 groups	Dependent (Related) Samples 1 DV > 2 timepoints
Parametric (normal)	Homogeneous	One way ANOVA	Repeated measures ANOVA
	Non homogenous	Welch ANOVA	Sphericity correction
Non-parametric		*Kruskal-Wallis ANOVA*	Friedman's ANOVA

In class activity

Exam performance		
Home school	Boarding school	Regular Day school
89	85	91
75	78	88
49	59	84
87	77	81
84	63	91
68	88	75
88	71	69
78	73	93
77	69	95
93	80	85
67	72	87
79	68	84
69	66	83
88	59	80
91	70	77

- Check for normality
- Check homogeneity of variances (different participants in different groups, Factorial ANOVA)
- Check for sphericity of variances (same participants across groups, repeated measures ANOVA)
- Choose the appropriate test
- Only if main effect (F) significant, use a post-hoc test
- Report effect size for significant effects
- Plot analyzed data



One way (One factor, One IV) ANOVA

Ho – exam performance not affected by type of schooling

H1 – Type of schooling affects exam performance

FAKE DATA

Exam performance		
Home school	Boarding school	Regular Day school
89	85	91
75	78	88
49	59	84
87	77	81
84	63	91
68	88	75
88	71	69
78	73	93
77	69	95
93	80	85
67	72	87
79	68	84
69	66	83
88	59	80
91	70	77

Groups	*Count*	*Sum*	*Average*	*Variance*
Home school	15	1182	78.8	141.1714
Boarding school	15	1078	71.86667	73.98095
Regular Day school	15	1263	84.2	50.45714

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1146.711	2	573.3556	6.475922	0.003537	3.219942
Within Groups	3718.533	42	88.53651			
Total	4865.244	44				

$F_{(2,42)} = 6.47, p<0.01$

or

$F_{(2,42)} = 6.47, p=.003$

- Check for normality
- Check homogeneity of variances (different participants in different groups)
- Choose the appropriate test
- Only if main effect (F) significant, use a post-hoc test
- Report effect size for significant effects (eta/partial eta squared)
- Plot analyzed data

Effect size for ANOVA

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}} = \frac{1146.711}{4865.244} = 0.236$$

Eta-squared

Table 1 Values of Effect Sizes and Their Interpretation

Kind of Effect Size	*Small*	*Medium*	*Large*
r	.10	.30	.50
d	0.20	0.50	0.80
η^2_p	.01	.06	$\boxed{.14}$
f^2	.02	.15	.35

Source: Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi:10.1037/0033-2909.112.1.155

$F_{(2,42)} = 6.47$, $p=.003$, $\eta^2 = .24$

Type of schooling explains 24% of variance in exam performance

We know there is difference between the groups, but which groups perform better or worse?