PSYC214: Statistics
Lecture 3 – Assumptions of ANOVA and follow-up procedures – Part I

Michaelmas Term
Dr Richard Philpot
r.philpot@lancaster.ac.uk

Assumptions of ANOVA and follow-up procedures

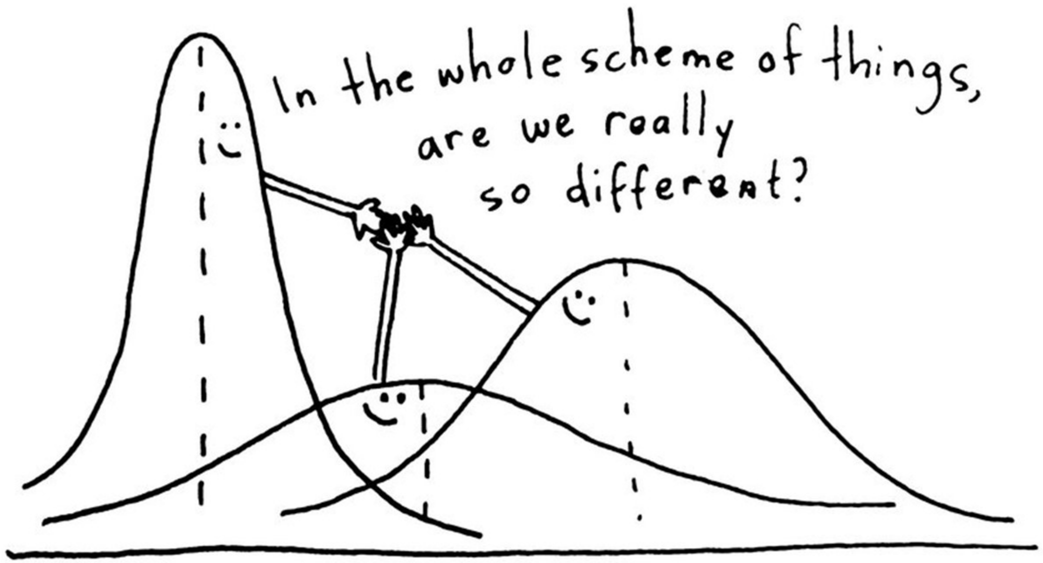
Lancaster University

Agenda/Content for Lecture 3

- Assumptions of ANOVA
  - Assumption of independence
  - Assumption of normality
  - Assumption of homogeneity of variance
- Data transformations
- Pairwise between-level comparisons
  - Planned comparisons
  - Post-hoc tests

The assumptions of ANOVA

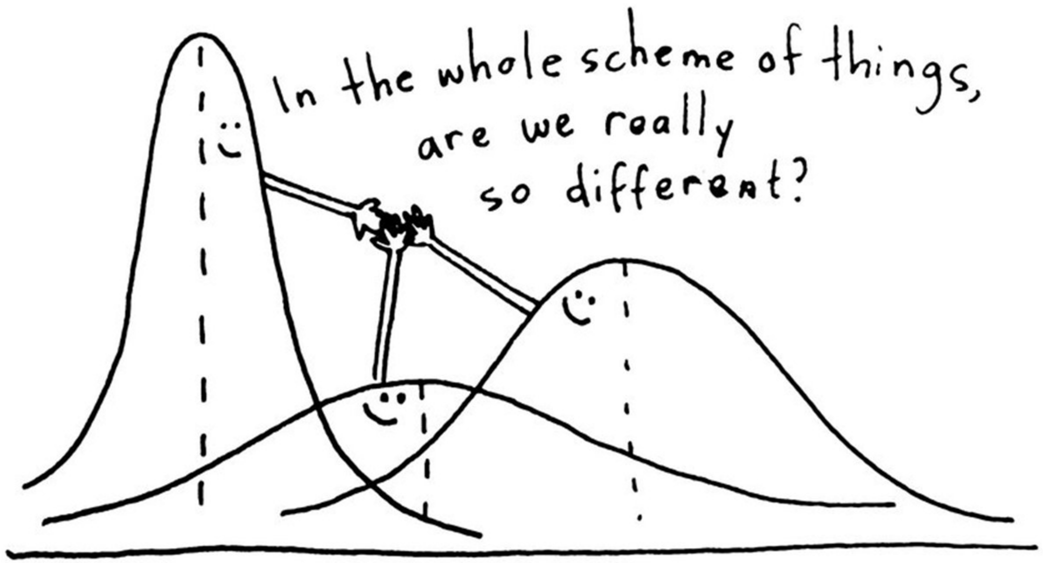
- The analysis of variance (ANOVA) is a parametric test
- ANOVAs have a set of assumptions, which should be met
- These are often ignored by researchers, because ANOVAs are typically very robust!
- Even small/moderate deviations



Source: Questionpro

The assumptions of ANOVA

- It is unlikely that highly significant results, e.g., $p < .01$, will drastically change because of small violations
- Marginally significant results, i.e., those around $p = .05$ value, however, may be affected by even small violations!



Source: Questionpro

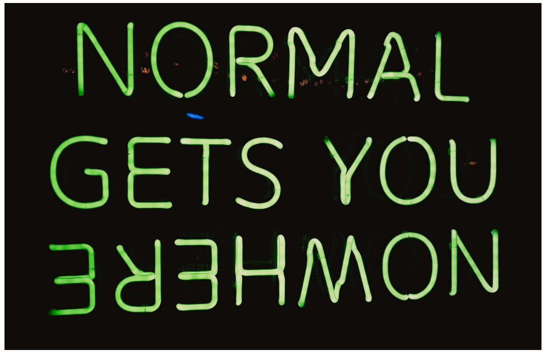
In a perfect world…

- You would have equal number of participants per level (e.g., per condition)
- Your data would be on an interval/ratio scale

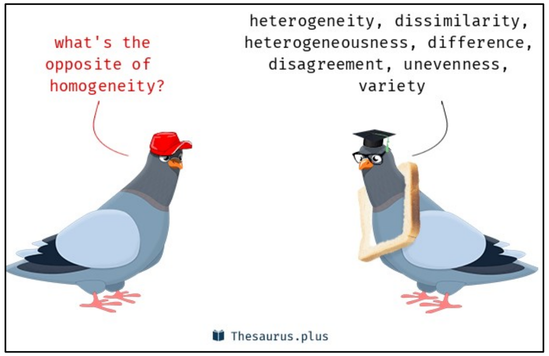
Assumptions underlying the ANOVA

1. Assumption of independence
2. Assumption of normality
3. Assumption of homogeneity of variance

Independence



Normality



Homogeneity of variance

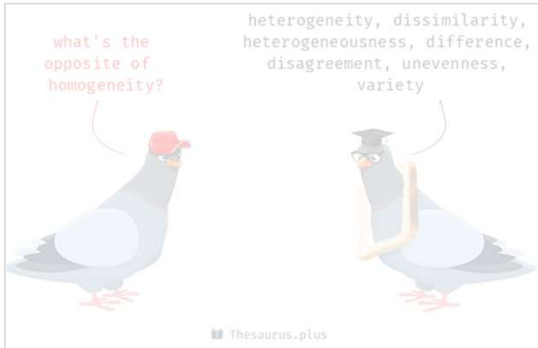
Assumptions underlying the ANOVA

Lancaster University

1. Assumption of independence
2. Assumption of normality
3. Assumption of homogeneity of variance

Independence

Normality



Homogeneity of variance

1. Assumption of independence

What is it?

- Participants should be randomly assigned to a group
- Participants should not cluster, sharing a classification variable
  - Gender
  - Skill level
- There should be no influence across one data point to another

1. Assumption of independence

Consequences of violation

- Becomes difficult to interpret results
- Did the manipulation have an effect, or was this driven by classification clustering or influence?

1. Assumption of independence

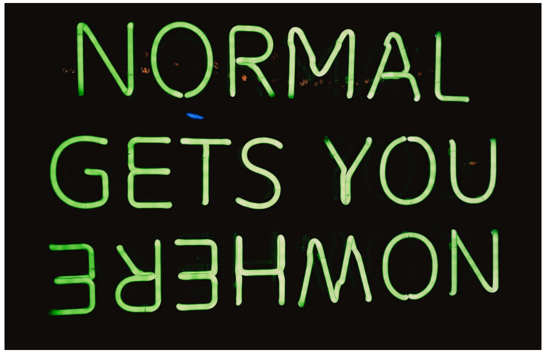
How to avoid it?

- Always randomly allocate participants to a condition
- Try to allocate equal numbers to each condition
- You can test to see whether you have significant differences on important classification variables

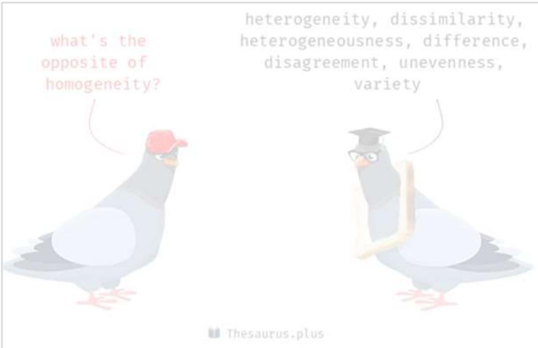
Assumptions underlying the ANOVA

1. Assumption of independence
2. Assumption of normality
3. Assumption of homogeneity of variance

Independence



Normality

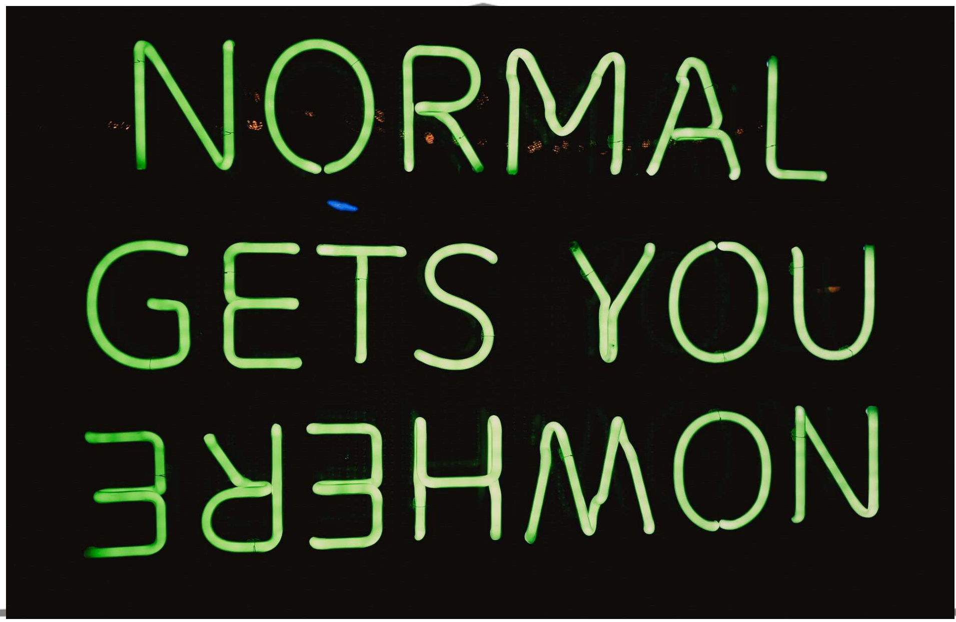


Homogeneity of variance

2. Assumption of normality

What is it?

- You want the overall data and the data for each subgroup to normally distributed
- This is because ANOVAs rely on the mean – and for skewed and bimodal data the mean is unlikely the best measure of central tendency



2. Assumption of normality

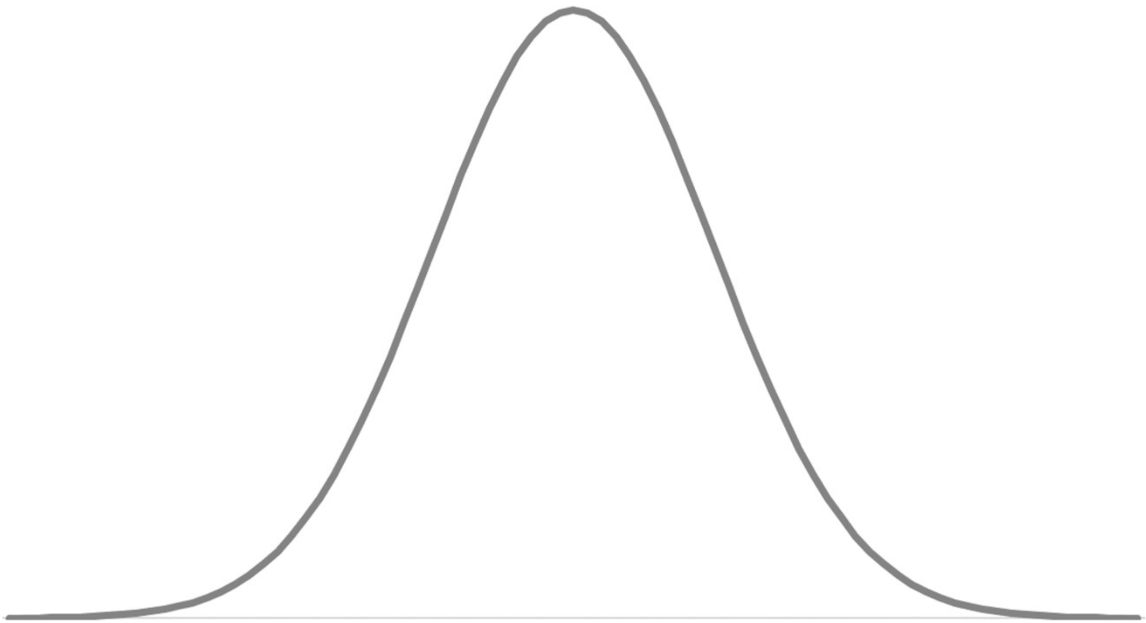
Consequences of violation

- If data are slightly skewed this is unlikely to cause problems
- If data are skewed by roughly the same degree in the same direction – unlikely a problem
- If skewed in different directions, this is a problem. Lead to type I and II errors!

2. Assumption of normality

How to avoid it?

- Avoid measures which often have ceiling or floor effects
- Transform data, changing every score in a systematic way
- Use a robust ANOVA (specialized test – more complex) or non-parametric alternatives

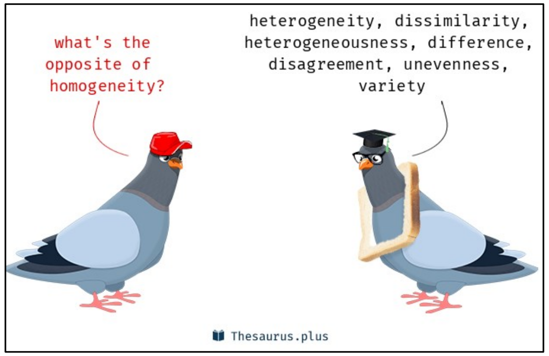


Assumptions underlying the ANOVA

1. Assumption of independence
2. Assumption of normality
3. Assumption of homogeneity of variance

Independence

Normality

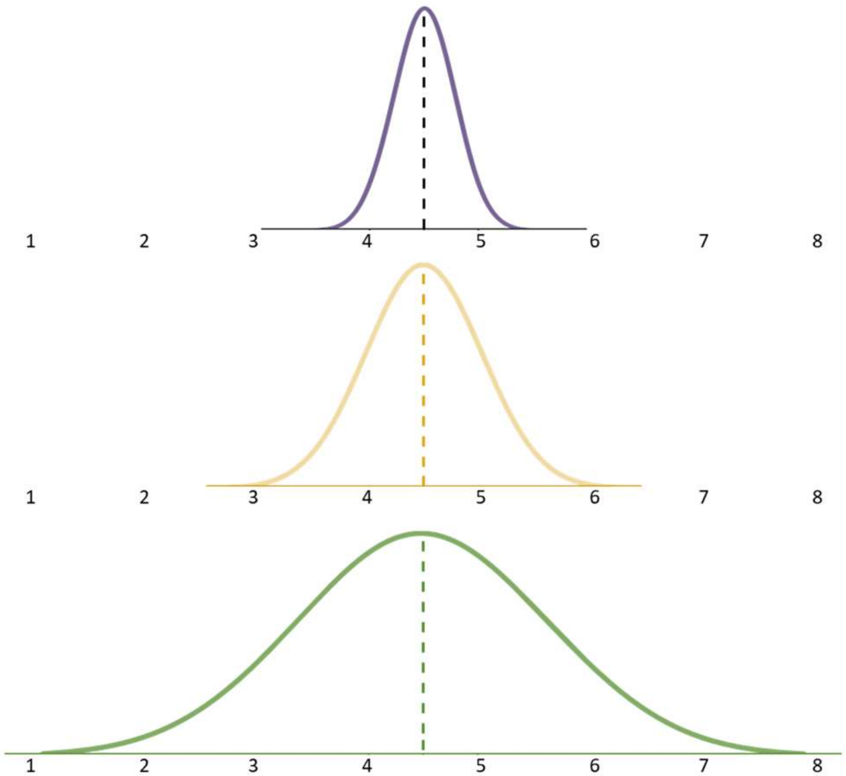


Homogeneity of variance

3. Homogeneity of variance

What is it?

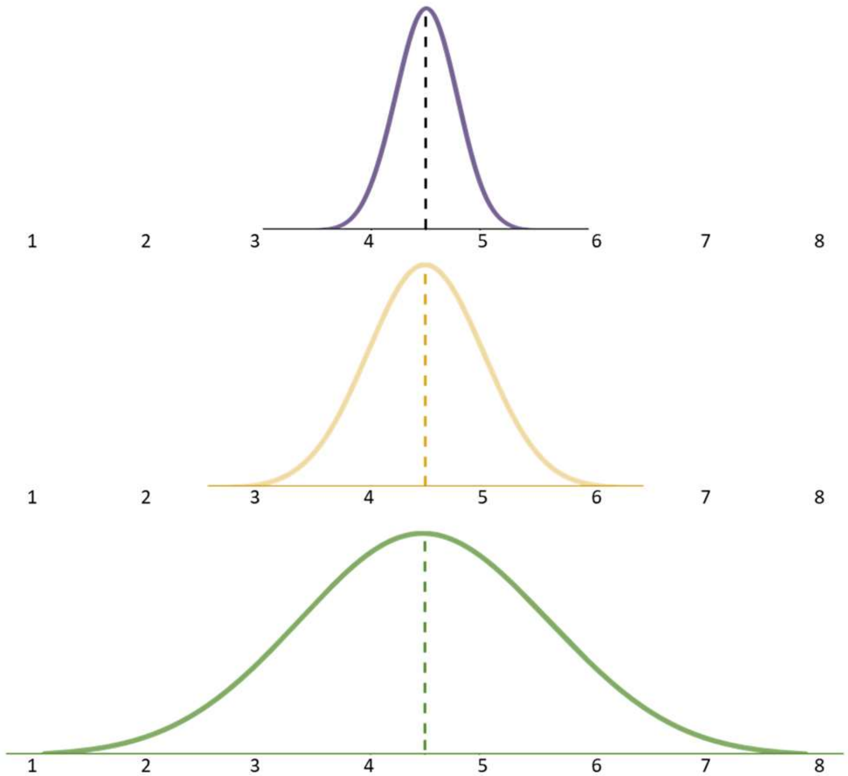
- Assumes that the variances of the distributions in the samples are equal
- Therefore the variances for each sample should not significantly vary from one another



3. Homogeneity of variance

Consequences of violation

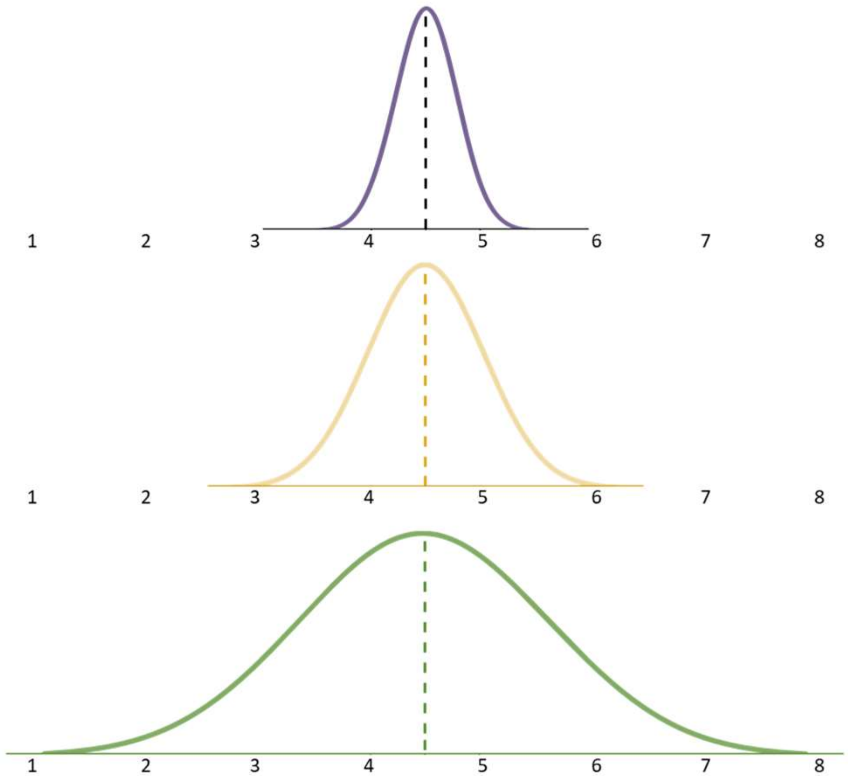
- The ANOVA tests the plausibility of the null hypothesis – i.e., all observations come from the same underlying population with the same degree of variability
- This is pointless to test when variance is already clearly different



3. Homogeneity of variance

How to avoid it?

- Difficult to avoid, but can be mitigated when testing
- As a rule of thumb, it is ok, as long as largest variance is no more than 4x the size of smallest
- Can also transform data or use non-parametric alternative

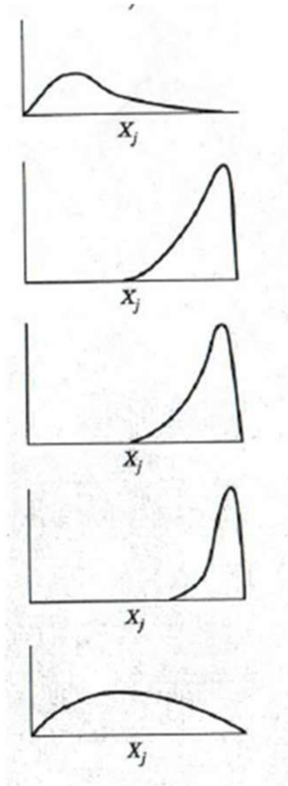


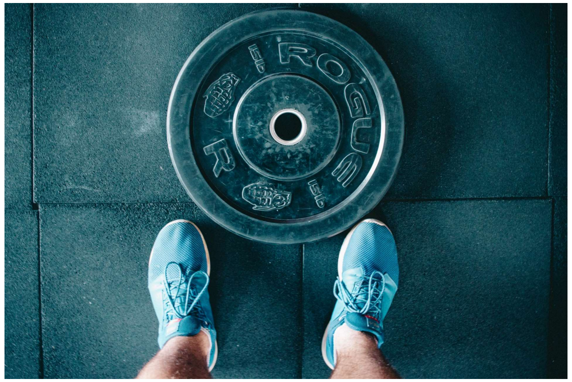
PSYC214: Statistics
Lecture 3 – Assumptions of ANOVA and follow-up procedures – Part II

Michaelmas Term
Dr Richard Philpot
r.philpot@lancaster.ac.uk

Dealing with 'rogue' data

- There are a number of strategies which may improve 'rouge' data
- None are panaceas and are unlikely to work in each situation
- If these aren't helpful, you can apply a non-parametric alternative
  - e.g., Kruskall-Wallace one-way Analysis of Variance by Ranks





Dealing with 'rogue' data

Transforming data

- This involves taking every score from each participant and applying a uniform mathematical function to each
- Report both the original data and the transformed data

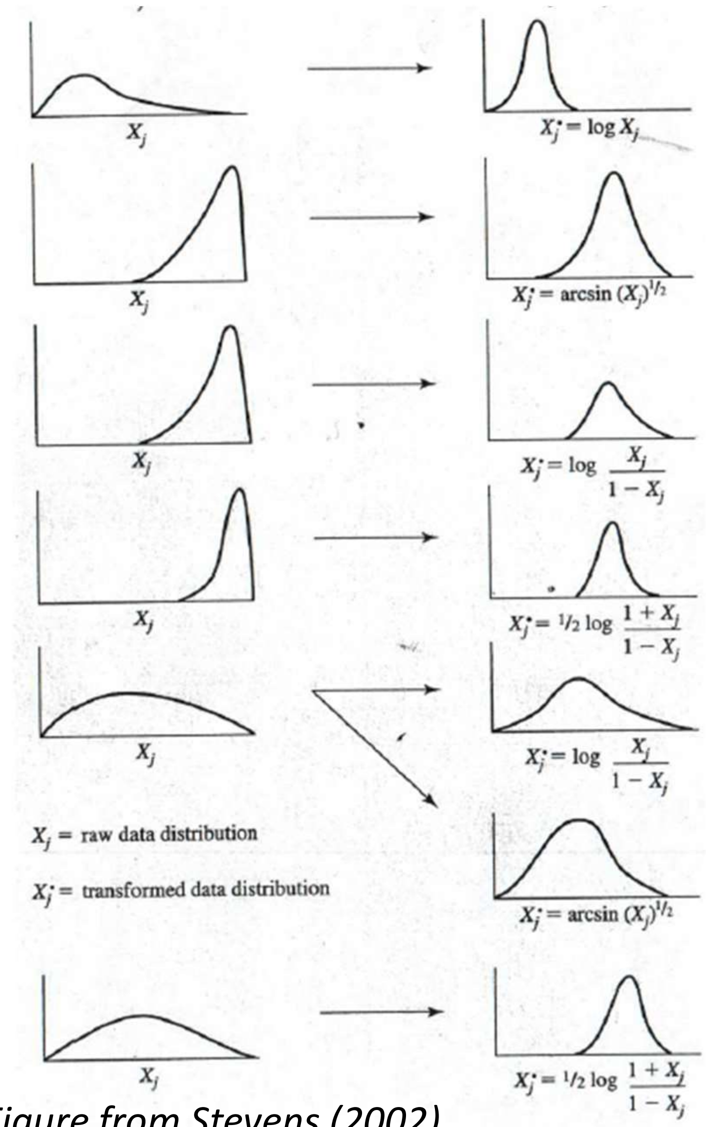


Figure from Stevens (2002)

Dealing with 'rogue' data

How to transform data

Untransformed	Square-root transformed	Log transformed
38	6.164	1.580
1	1.000	0.000
13	3.606	1.114
2	1.414	0.301
13	3.606	1.114
20	4.472	1.301
50	7.071	1.699
9	3.000	0.954
28	5.292	1.447
6	2.449	0.778
4	2.000	0.602
43	6.557	1.633

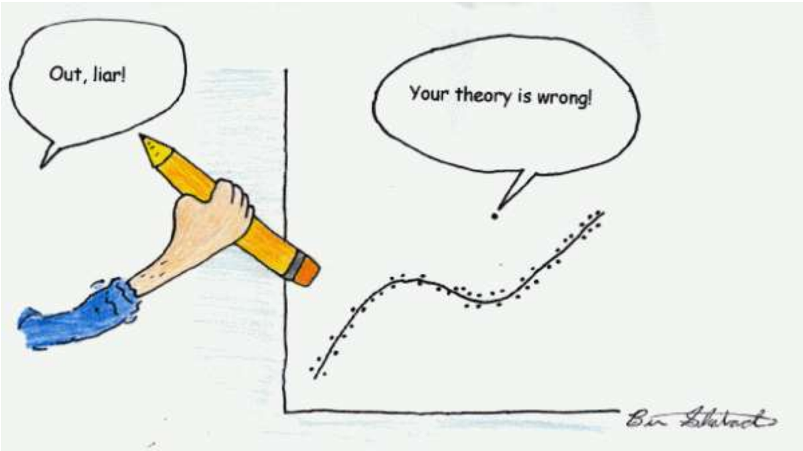
http://www.biostathandbook.com/transformation.html

Type of Data Transformation	Nature of Data
Log Transformation $(\log(X_i))$	Whole numbers and cover wide range of values, small values with decimal fractions.
Square-root Transformation $(\sqrt{X_i})$	Small whole number & Percentage data where the range is between 0 and 30 % or between 70 and 100 %

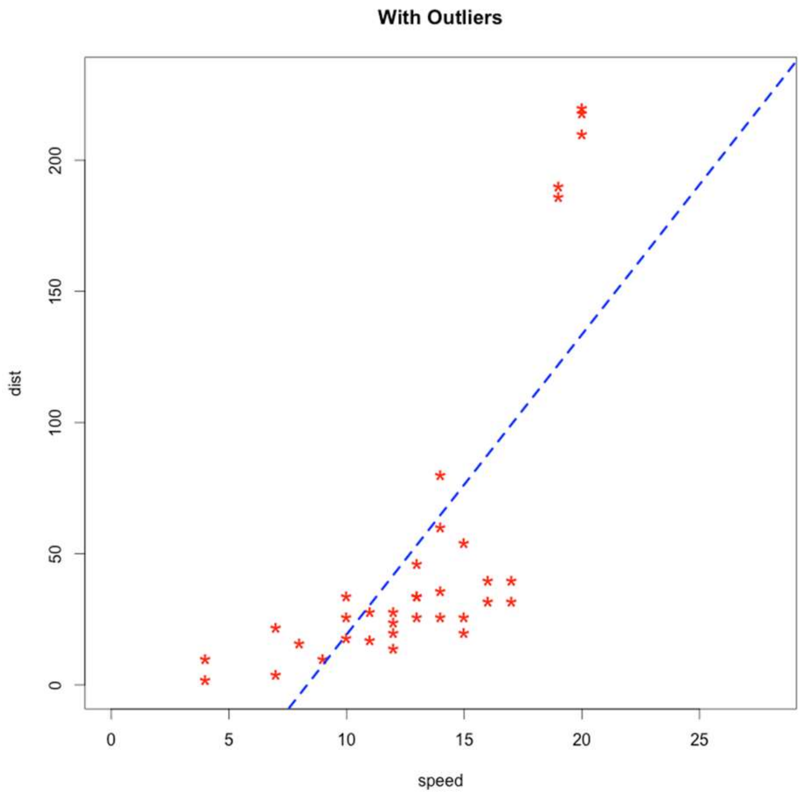
Maidapwad & Sananse (2014)

Outliers and their impact

- Outliers are data points which significantly differ from other observations
- Outliers can drastically bias/change predictive models
- Predictions can be exaggerated and present high error
- Outliers not only distort statistical analyses, they can violate assumptions

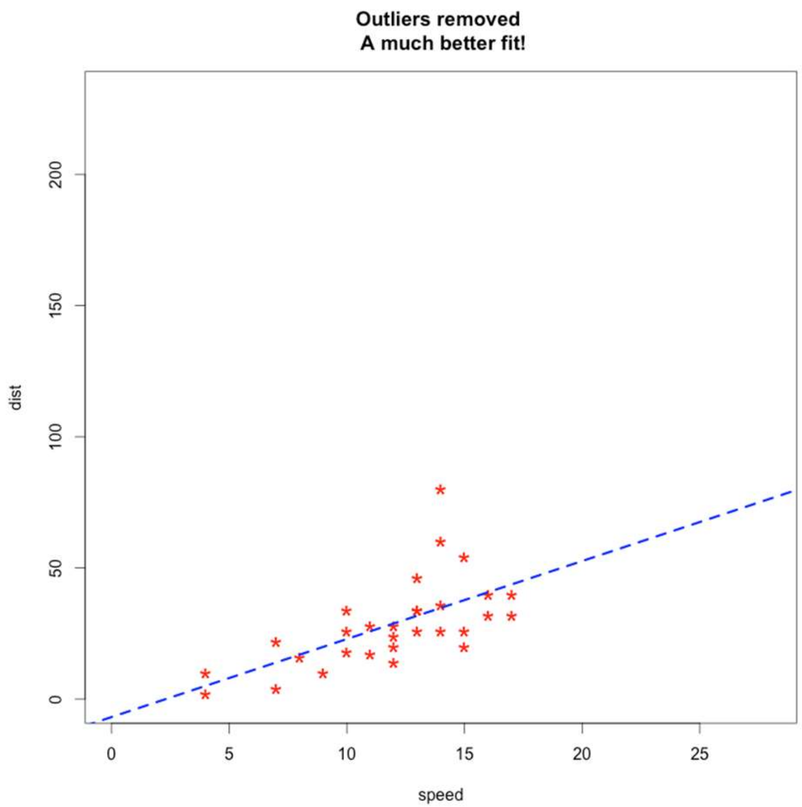


Outliers and their impact



- Given the problems outliers create, it may seem levelheaded to remove them
- However, it can be dishonest and misleading to do so if they are true scores
- It must be justifiable as to why it is necessary to remove data

Outliers and their impact



- Given the problems outliers create, it may seem levelheaded to remove them
- However, it can be dishonest and misleading to do so if they are true scores
- It must be justifiable as to why it is necessary to remove data

PSYC214: Statistics
Lecture 3 – Assumptions of ANOVA and follow-up procedures – Part III

Michaelmas Term

Dr Richard Philpot
r.philpot@lancaster.ac.uk

The meaning of an ANOVA output

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Group         2   1223   611.3   12.52 6.77e-06 ***
## Residuals   237  11571    48.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The meaning of an ANOVA output

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Group         2   1223   611.3   12.52 6.77e-06 ***
## Residuals   237  11571    48.8                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

$$F = \frac{611.3}{48.8}$$

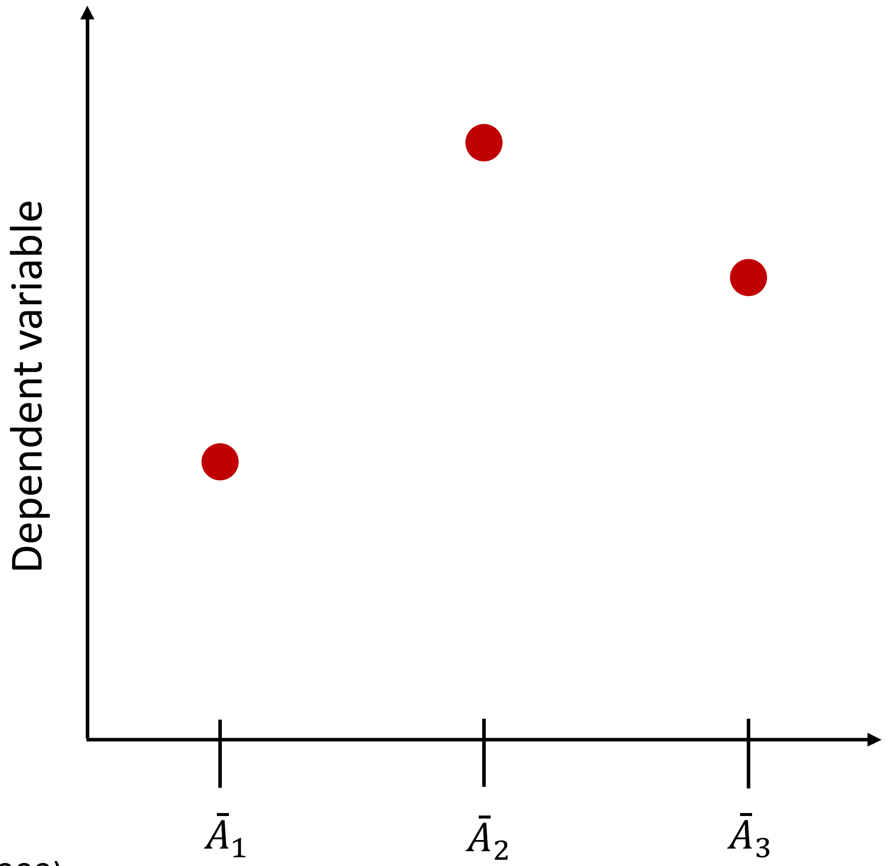
$\boxed{F = \ 12.52}$

$\boxed{p = 0.00000677}$

The meaning of an ANOVA output

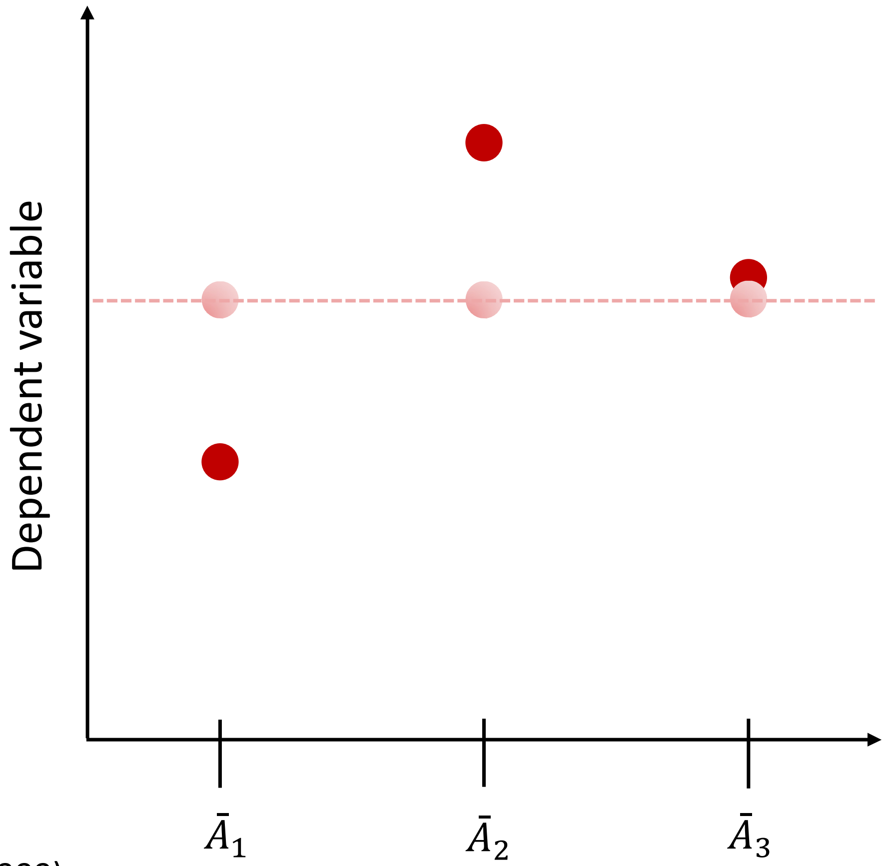
P-value	Definition
> .05	▪ We accept the null hypothesis (H_0) ▪ Under H_0, the samples come from the <u>same</u> population ▪ There is no statistical difference in the population means ($\mu_1 = \mu_2 = \mu_3$) ▪ Experimental effect = 0
≤ .05	▪ We reject the null hypothesis (H_1) ▪ Under H_1, the samples come from <u>different</u> populations ▪ Population means are statistically different ($\mu_1 \neq \mu_2 \neq \mu_3$) ▪ Experimental effect $\neq$ 0

Significant?



Adapted from Roberts and Russo (1999)

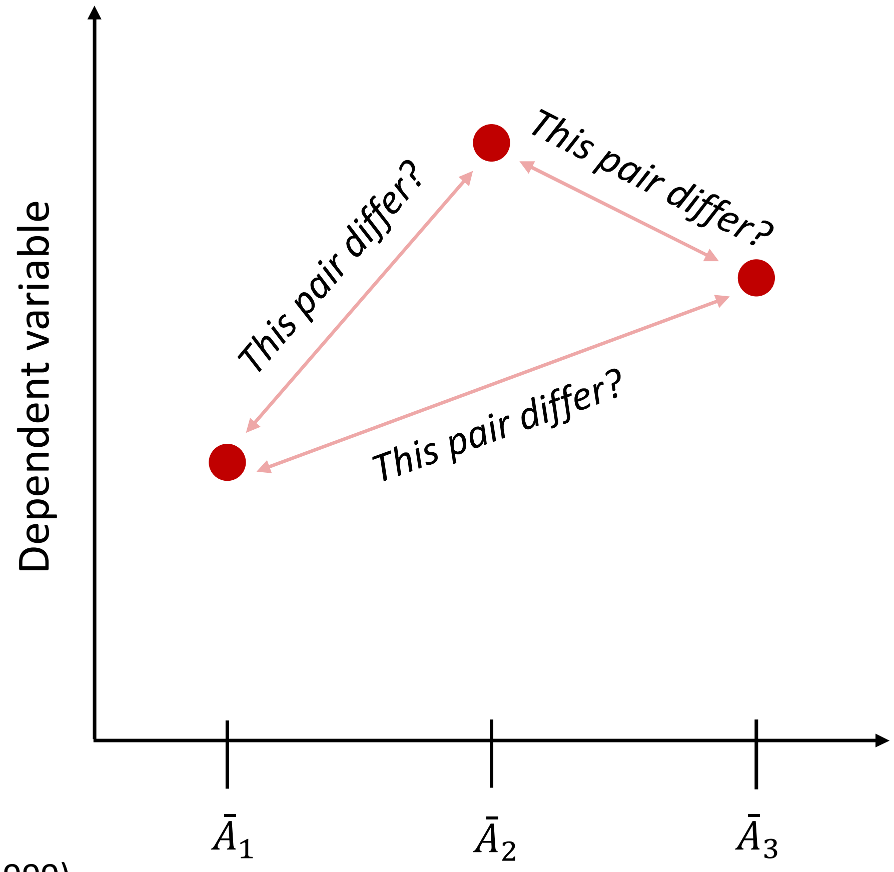
Non-significant



p > .05
There is insufficient evidence to conclude that any means significantly differs from any others

Adapted from Roberts and Russo (1999)

Significant



$p \leq .05$
At least one of the pairs of means is significantly different. The question is, which pairs?

Adapted from Roberts and Russo (1999)

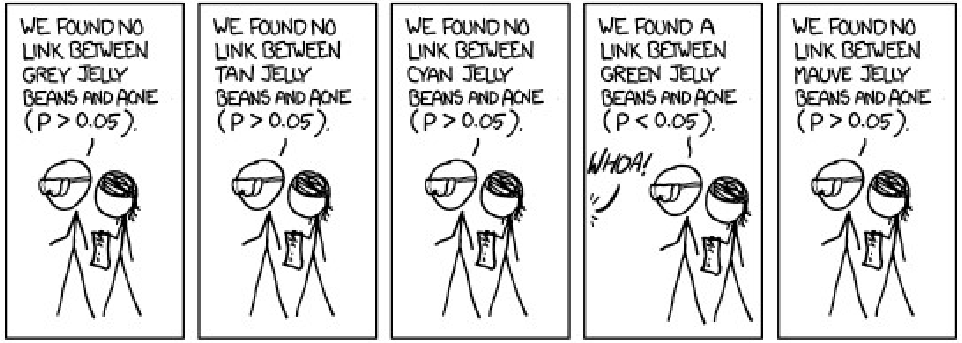
Pairwise comparisons

There are two strategies for following-up significant ANOVAs

- Planned comparisons
- Post-hoc comparisons

The problem of multiple comparisons

- Why not just run a bunch of t-tests?
- Multiple comparisons increase the probability of making a (familywise) type I error
- I.e., rejecting the null hypothesis when actually there was no effect



The problem of multiple comparisons

- Type 1 error - 1 test at $p \leq 0.05$ = 0.95 (i.e., 5% chance we get noise)
- Type 1 error - 2 tests = 0.95 * 0.95, = 0.903. (10% chance)
- Type 1 error - 3 tests = 0.95 * 0.95 * 0.95 = 0.857 (14% chance)
- Type 1 error – 4 tests = 0.95 * 0.95 * 0.95 * 0.95 = 0.815 (18.5% chance)
- Type 1 error – 5 tests = 0.95 * 0.95 * 0.95 * 0.95 * 0.95 = 0.774 (22.6% chance)

Pairwise comparisons

There are two strategies for following-up significant ANOVAs

- Planned comparisons
- Post-hoc comparisons

The problem of multiple comparisons

Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$	$\bar{A}_5$
$\bar{A}_1$	-	-	-	-	-
$\bar{A}_2$	●	-	-	-	-
$\bar{A}_3$	●	●	-	-	-
$\bar{A}_4$	●	●	●	-	-
$\bar{A}_5$	●	●	●	●	-

Planned comparisons

- Focussed approach to examine specific group differences
- Perfect when certain hypotheses can be tested without comparing all combinations of means
- Should be pre-specified
- Need to keep the number of planned comparisons as low as possible to negate Type I errors – (number of levels – 1)

Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$	$\bar{A}_5$
$\bar{A}_1$	-	-	-	-	-
$\bar{A}_2$	●	-	-	-	-
$\bar{A}_3$	●	●	-	-	-
$\bar{A}_4$	●	●	●	-	-
$\bar{A}_5$	●	●	●	●	-

Planned comparisons

Our options:

1. Run t-tests with a low number of pairs
2. Run t-tests with Bonferroni adjustment
3. ~~Specialized linear contrast~~

Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$	$\bar{A}_5$
$\bar{A}_1$	-	-	-	-	-
$\bar{A}_2$	●	-	-	-	-
$\bar{A}_3$	●	●	-	-	-
$\bar{A}_4$	●	●	●	-	-
$\bar{A}_5$	●	●	●	●	-

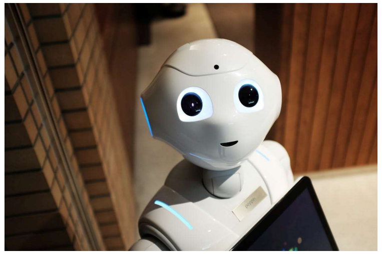
Planned comparisons – 1. Run t-tests

Lancaster University

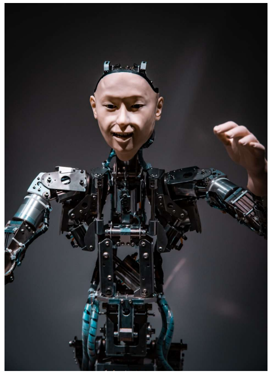
- Accept that we have inflated our risks
- Keep the number of planned comparisons as low as possible to negate Type I errors – (number of levels – 1)
- Even with two tests, however, our chance of a Type I error is 10%!

Planned comparisons – 1. Run t-tests

A_1 - Robot A(lpha)



A_2- Robot B(eta)



A_3 - Robot O(mega)

Planned comparisons

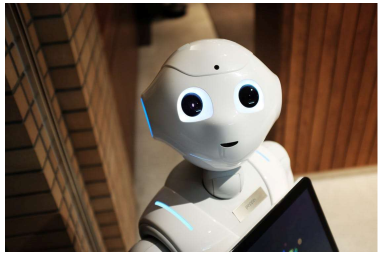
  

	Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$
	$\bar{A}_1$	-	-	-
	$\bar{A}_2$		-	-
	$\bar{A}_3$			-

Planned comparisons – 1. Run t-tests

A_1 - Robot A(lpha)

$$t = \frac{\bar{A}_1 - \bar{A}_2}{\sqrt{(Mean\ Square_{ERROR})(\frac{2}{NA})}}$$



A_2- Robot B(eta)

Planned comparisons – 1. Run t-tests

Mean differences between two levels

$$t = \frac{\boxed{\bar{A}_1} - \boxed{\bar{A}_2}}{\sqrt{\left(\boxed{Mean\ Square_{ERROR}}\right)\left(\frac{2}{\boxed{NA}}\right)}}$$

Within group variance from ANOVA output

Number of scores in each levels being compared

Planned comparisons – 1. Run t-tests

```
##   Group variable        n  mean    sd   min   max
##   <chr> <chr>       <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 A     Likeability    80  2.5  0.928     1     4
## 2 A     Score          80 58.1  6.45     44    72
## 3 B     Likeability    80  4.5  1.01      2     7
## 4 B     Score          80 60.4  7.27     40    74
## 5 O     Likeability    79  2.11 0.847     1     4
## 6 O     Score          80 63.6  7.22     47    79
```

$$t = \frac{\boxed{\bar{A}_1} - \boxed{\bar{A}_2}}{\sqrt{\left(\boxed{Mean\ Square_{ERROR}}\right)\left(\frac{2}{\boxed{NA}}\right)}}$$

80 + 80

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Group         2   1223   611.3   12.52 6.77e-06 ***
## Residuals   237  11571    48.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Planned comparisons – 1. Run t-tests

```
##   Group variable        n  mean    sd   min   max
##   <chr> <chr>       <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 A     Likeability    80   2.5 0.928     1     4
## 2 A     Score          80  58.1  6.45    44    72
## 3 B     Likeability    80   4.5  1.01     2     7
## 4 B     Score          80  60.4  7.27    40    74
## 5 O     Likeability    79  2.11 0.847     1     4
## 6 O     Score          80  63.6  7.22    47    79
```

$$t = \frac{\boxed{58.1} - \boxed{60.4}}{\sqrt{\boxed{(48.8)}\left(\frac{2}{160}\right)}}$$

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Group         2   1223   611.3   12.52 6.77e-06 ***
## Residuals   237  11571    48.8                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Planned comparisons – 1. Run t-tests

$$t = \frac{58.1 \ - 60.4}{\sqrt{(48.8)(0.0125)}}$$

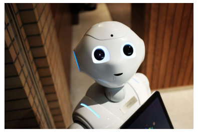
Planned comparisons – 1. Run t-tests

$$t = \frac{-2.3}{\sqrt{0.61}}$$

$$t = \frac{-2.3}{0.78}$$

$$t = -2.94$$

Planned comparisons – 1. Run t-tests



$t = $ -2.94, with 237 degrees of freedom

It's significant at $p = 0.05$ threshold

Degrees of Freedom	p=0.05	p=0.025	p=0.01	p=0.005
1	12.71	25.45	63.66	127.32
2	4.30	6.20	9.92	14.09
3	3.18	4.17	5.84	7.45
4	2.78	3.50	4.60	5.60
5	2.57	3.16	4.03	4.77
6	2.45	2.97	3.71	4.32
7	2.36	2.84	3.50	4.03
8	2.31	2.75	3.36	3.83
9	2.26	2.68	3.25	3.69
10	2.23	2.63	3.17	3.58
11	2.20	2.59	3.11	3.50
12	2.18	2.56	3.05	3.43
13	2.16	2.53	3.01	3.37
14	2.14	2.51	2.98	3.33
15	2.13	2.49	2.95	3.29
16	2.12	2.47	2.92	3.25
17	2.11	2.46	2.90	3.22
18	2.10	2.44	2.88	3.20
19	2.09	2.43	2.86	3.17
20	2.09	2.42	2.84	3.15
21	2.08	2.41	2.83	3.14
22	2.07	2.41	2.82	3.12
23	2.07	2.40	2.81	3.10
24	2.06	2.39	2.80	3.09
25	2.06	2.38	2.79	3.08
26	2.06	2.38	2.78	3.07
27	2.05	2.37	2.77	3.06
28	2.05	2.37	2.76	3.05
29	2.04	2.36	2.76	3.04
30	2.04	2.36	2.75	3.03
40	2.02	2.33	2.70	2.97
60	2.00	2.30	2.66	2.92
120	[illegible]	2.27	2.62	2.86
infinity	$\boxed{1.96}$	2.24	2.58	2.81

Planned comparisons – 2. Corrections

- Continue to run t-tests, but adjust the *p* value to make it more conservative
- Only accept significant if below this threshold
- Bonferroni Correction:
  - A new p-value is generated from the prior significance level divided by the number of tests

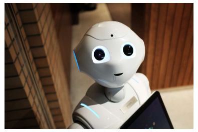
$$\boxed{0.05} \div \boxed{2} = \boxed{0.025}$$

P-value

Number of tests

Bonferroni adjusted P-value

Planned comparisons – 2. Corrections



$t =$ -2.94, with 237 degrees of freedom
It's significant at $p = 0.025$ threshold

$t =$ -2.14, with 237 degrees of freedom
It's significant at $p = 0.05$ threshold

Degrees of Freedom	p=0.05	p=0.025	p=0.01	p=0.005
1	12.71	25.45	63.66	127.32
2	4.30	6.20	9.92	14.09
3	3.18	4.17	5.84	7.45
4	2.78	3.50	4.60	5.60
5	2.57	3.16	4.03	4.77
6	2.45	2.97	3.71	4.32
7	2.36	2.84	3.50	4.03
8	2.31	2.75	3.36	3.83
9	2.26	2.68	3.25	3.69
10	2.23	2.63	3.17	3.58
11	2.20	2.59	3.11	3.50
12	2.18	2.56	3.05	3.43
13	2.16	2.53	3.01	3.37
14	2.14	2.51	2.98	3.33
15	2.13	2.49	2.95	3.29
16	2.12	2.47	2.92	3.25
17	2.11	2.46	2.90	3.22
18	2.10	2.44	2.88	3.20
19	2.09	2.43	2.86	3.17
20	2.09	2.42	2.84	3.15
21	2.08	2.41	2.83	3.14
22	2.07	2.41	2.82	3.12
23	2.07	2.40	2.81	3.10
24	2.06	2.39	2.80	3.09
25	2.06	2.38	2.79	3.08
26	2.06	2.38	2.78	3.07
27	2.05	2.37	2.77	3.06
28	2.05	2.37	2.76	3.05
29	2.04	2.36	2.76	3.04
30	2.04	2.36	2.75	3.03
40	2.02	2.33	2.70	2.97
60	2.00	2.30	2.66	2.92
120	1.98	2.27	2.62	2.86
infinity	1.96	$\boxed{2.24}$	2.58	2.81

PSYC214: Statistics
Lecture 3 – Assumptions of ANOVA and follow-up procedures – Part IV

Michaelmas Term

Dr Richard Philpot
r.philpot@lancaster.ac.uk

Pairwise comparisons

Lancaster University

There are two strategies for following-up significant ANOVAs

- Planned comparisons
  - T-tests
  - Bonferroni corrections
- Post-hoc comparisons

Post hoc tests

- Post hoc comes from Latin for "after the event"
- Post hoc tests assess all possible combinations of differences between groups by comparing each mean with the other
- Make adjustments to *p* value, but more conservative than Bonferroni correction

Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$	$\bar{A}_5$
$\bar{A}_1$	-	-	-	-	-
$\bar{A}_2$	●	-	-	-	-
$\bar{A}_3$	●	●	-	-	-
$\bar{A}_4$	●	●	●	-	-
$\bar{A}_5$	●	●	●	●	-

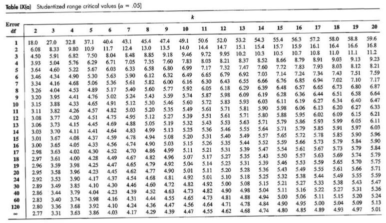
Post hoc tests

Method	Equal N F	Normality	Use	Error control	Protection
Fisher PLSD	Yes	Yes	Yes	All	Most sensitive to Type 1
Tukey-Kramer HSD	No	Yes	Yes	All	Less sensitive to Type 1 than Fisher PLSD
Spjotvoll-Stoline	No	Yes	Yes	All	As Tukey-Kramer
Student-Newman Keuls (SNK)	Yes	Yes	Yes	All	Sensitive to Type 2
Tukey-Compromise	No	Yes	Yes	All	Average of Tukey and SNK
Duncan's Multiple Range	No	Yes	Yes	All	More sensitive to Type 1 than SNK
Scheffé's S	Yes	No	No	All	Most conservative
Games/Howell	Yes	No	No	All	More conservative than majority
Dunnett's test	No	No	No	T/C	More conservative than majority
Bonferroni	No	Yes	Yes	All, TC	Conservative

https://www.researchgate.net/profile/Cyril-Iaconelli/post/The_choice_of_post-hoc_test/

Post hoc tests – Tukey-Kramer HSD

Table IX: Tukey α = 0.05



Studentized range statistic [num means, df]

$$W = \boxed{q_{(r, df_{ERROR})}} \sqrt{\frac{\boxed{Mean\ Square_{ERROR}}}{\boxed{N_A}}}$$

Within group variance from ANOVA output

Number of participants

	Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$
	$\bar{A}_1$	-	-	-
	$\bar{A}_2$	●	-	-
	$\bar{A}_3$	●	●	-

Table IX: Tukey α = 0.05

Lancaster University

Table IX(a) Studentized range critical values (α = .05)

Error df	k = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1	50.6	52.0	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.08	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	14.4	14.7	15.1	15.4	15.7	15.9	16.1	16.4	16.6	16.8
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.2	10.3	10.5	10.7	10.8	11.0	11.1	11.2
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

Post hoc tests – Tukey-Kramer HSD

	Group	$\bar{A}_1$ 	$\bar{A}_2$ 	$\bar{A}_3$ 
	$\bar{A}_1$	-	-	-
	$\bar{A}_2$	●	-	-
	$\bar{A}_3$	●	●	-

Studentized range statistic

$$W = \boxed{3.31}\,\frac{\sqrt{\boxed{48.8}}}{\boxed{239}}$$

Within group variance from ANOVA output

Number of participants

Post hoc tests – Tukey-Kramer HSD

				
	Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$
	$\bar{A}_1$	-	-	-
	$\bar{A}_2$	●	-	-
	$\bar{A}_3$	●	●	-

$$W = 3.31\sqrt{0.20}$$

$$W = 1.48$$

Means that differ over 1.48 will be statistically significant

Post hoc tests – Tukey-Kramer HSD

- Take home message
- As you add more and more mean comparisons, you require larger critical values (q) in the standardized table to find a statistical difference!
- As such, test what you need, not what you don't!

Group	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\bar{A}_4$	$\bar{A}_5$
$\bar{A}_1$	-	-	-	-	-
$\bar{A}_2$	●	-	-	-	-
$\bar{A}_3$	●	●	-	-	-
$\bar{A}_4$	●	●	●	-	-
$\bar{A}_5$	●	●	●	●	-

Lecture 3 – Assumptions of ANOVA and follow-up procedures

Lancaster University

Review of Lecture 3

- Assumptions of ANOVA
    - Assumption of independence
    - Assumption of normality
    - Assumption of homogeneity of variance
- Data transformations
- Pairwise between-level comparisons
    - Planned comparisons
    - Post-hoc tests

Thank you for attention! Questions?