UNIVERSITY OF CANBERRA

Survey Research and Design in Psychology

Lecture 4 - Correlations

Dr Ananthan Ambikairajah

University of Canberra

A quick recap

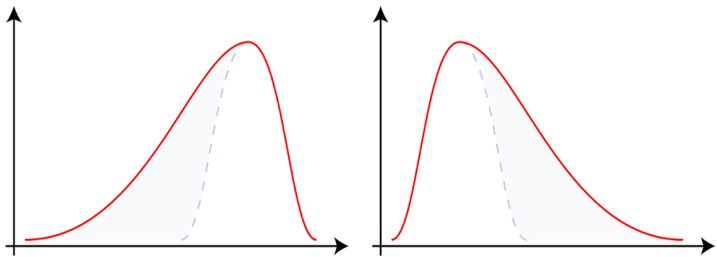
UNIVERSITY OF CANBERRA

- Getting to know your data
- Types of statistics to use for different levels of measurement
- Descriptive statistics
- Distributions of data
- The effect of skew on central tendency
- Principles of graphing
- Univariate graphical techniques

Some things to think about

UNIVERSITY OF CANBERRA

- If a survey question produces a floor effect, then where will the mean, median and mode lie in relation to one another?
- Would the mean number of cars owned in Australia exceed the median?
- Would the mean score on an easy test exceed the median performance?



Focus/Overview for today

UNIVERSITY OF CANBERRA

- Covariation
- Purposes of correlation
- Linear correlation
- Types of correlation
- Interpreting correlation
- Assumptions and limitations

The world is made of co-variations

UNIVERSITY OF CANBERRA

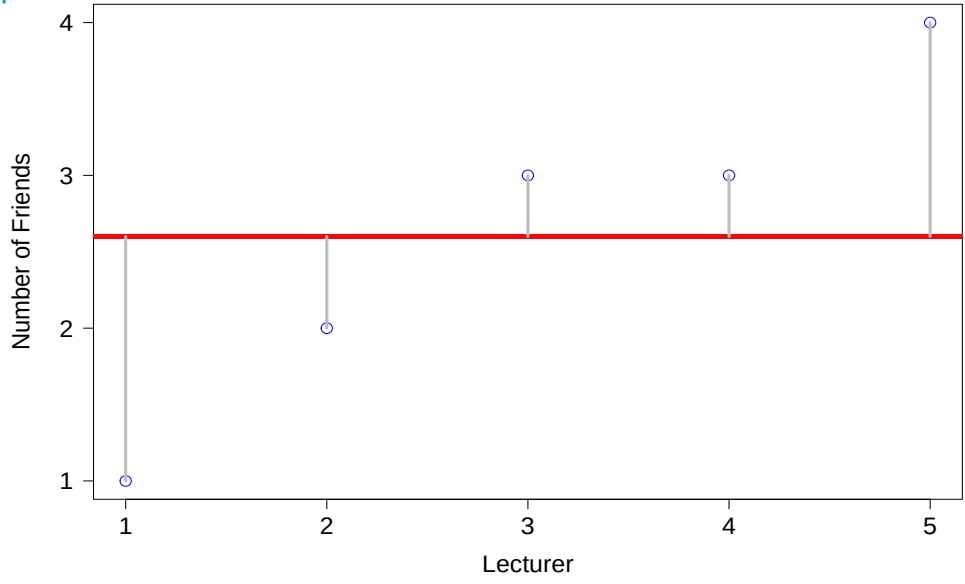
- News headline examples:
  - "Droughts mean fewer flowers for bees"
  - "Extra glass of wine a day 'will shorten your life by 30 minutes' "
  - "Shark attacks increase in El Nino years: Marine biologist"

This materials builds

UNIVERSITY OF CANBERRA

- Covariation is another key term to learn
- Correlation underlies many different statistical analyses we use in this unit e.g. internal consistency, regression, factor analysis.

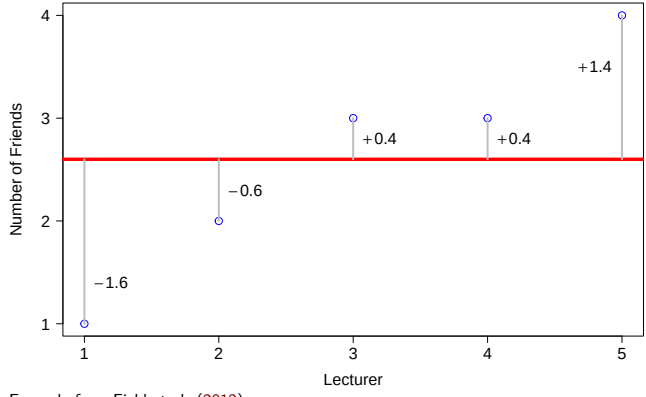
Concept Check - Variance



Example from Field et al. (2012)

Concept Check - Variance

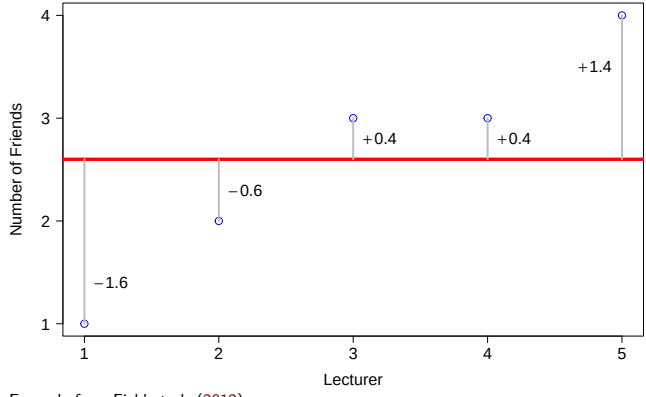
- total error = sum of deviances $= \sum(x_i - \bar{x})$



Example from Field et al. (2012)

Concept Check - Variance

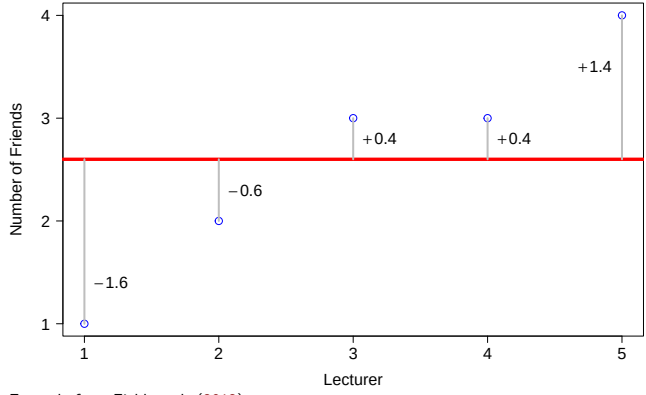
- total error = sum of deviances $= \sum(x_i - \bar{x})$
- sum of squared errors (SS) = $\sum(x_i - \bar{x})(x_i - \bar{x})$



Example from Field et al. (2012)

Concept Check - Variance

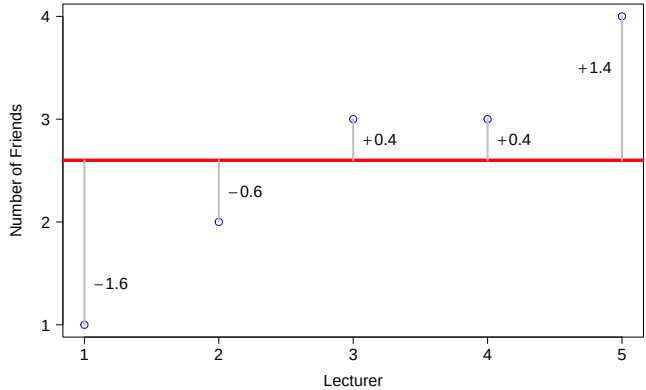
- total error = sum of deviances $= \sum(x_i - \bar{x})$
- sum of squared errors (SS) = $\sum(x_i - \bar{x})(x_i - \bar{x})$
- variance $(s^2) = \frac{SS}{N-1} = \frac{\sum(x_i - \bar{x})^2}{N-1}$



Example from Field et al. (2012)

Concept Check - Variance

- total error = sum of deviances $= \sum(x_i - \bar{x})$
- sum of squared errors (SS) = $\sum(x_i - \bar{x})(x_i - \bar{x})$
- variance $(s^2) = \frac{SS}{N-1} = \frac{\sum(x_i - \bar{x})^2}{N-1}$
- standard deviation $(s) = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}}$



Example from Field et al. (2012)

Why do we use correlation

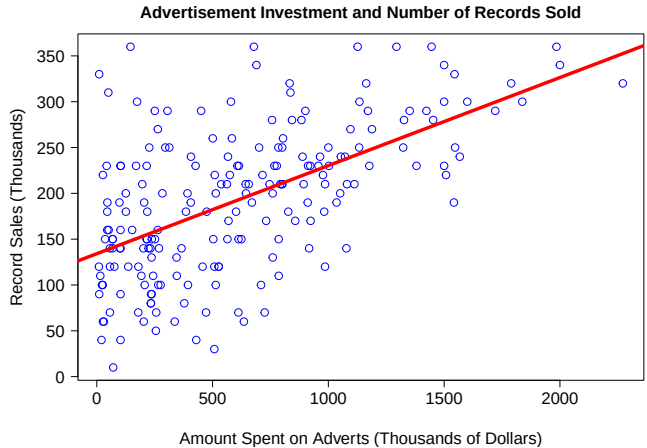
UNIVERSITY OF CANBERRA

Purpose of correlation

- What is the relationship/association/shared variance/co-relation between two variables?
- To what extent do two variables covary/depend on one another/explain one another?

Linear correlation

- To what extent do two variables have a linear relationship?
  - A simple straight line/constant relationship



Example and dataset from Field et al. (2012)

Linear correlation

UNIVERSITY OF CANBERRA

- To interpret a linear correlation, look at the:
  - Direction: is the relationship positive or negative? Check the sign (+/-)
  - Strength: how strong is the relationship? Check the size (-1 + 1)
  - Statistical significance: is there a significant relationship? Check the p-value ($<.05$ indicates a significant relationship)

Linear correlation

```
##
## Call:
## lm(formula = sales ~ 1 + adverts, data = album_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -152.949  -43.796   -0.393   37.040  211.866
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.341e+02  7.537e+00  17.799   <2e-16 ***
## adverts     9.612e-02  9.632e-03   9.979   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 65.99 on 198 degrees of freedom
## Multiple R-squared:  0.3346, Adjusted R-squared:  0.3313
## F-statistic: 99.59 on 1 and 198 DF,  p-value: < 2.2e-16
```

- Note - for simple linear regression $r = \sqrt{R^2}$

```
round(sqrt(0.3346), digits = 2)
```

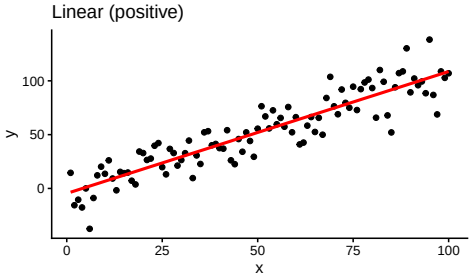
```
## [1] 0.58
```

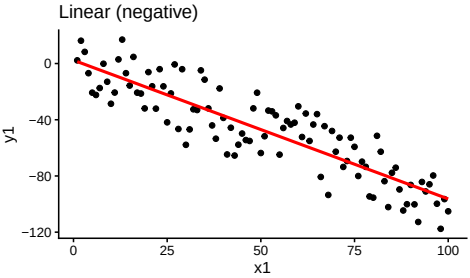
Types of correlation

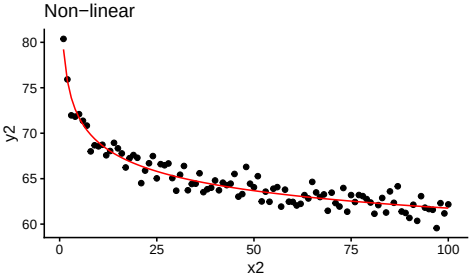
UNIVERSITY OF CANBERRA

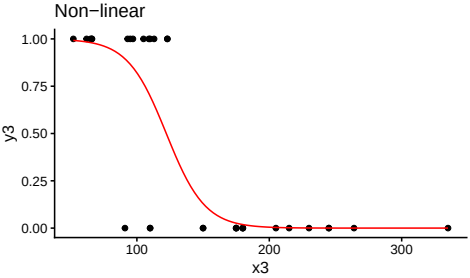
- No relationship (r is close to 0) -> your variables are independent
- Linear relationship -> your variables are dependent
  - As one increases, so does the other (positive correlation)
  - As one increases, the other decreases (negative correlation)
- Non-linear relationship

Types of correlation









Types of correlation

UNIVERSITY OF CANBERRA

- How do you choose which type of correlation to use?
- It depends on the levels of measurement of your variables

	Nominal	**Ordinal**	**Interval/Ratio**
Nominal	Chi-squared (X^2), Phi or Cramer's V, Clustered bar chart	← Treat as for	Point bi-serial correlation (r_{pb}), Scatterplot, bar chart, or error-bar chart
Ordinal		Spearman's Rho (r_s) or Kendall's Tau, Clustered bar chart OR scatterplot	Treat as for ↑ and ←
Interval/ Ratio			Pearson's Product-moment correlation (r), Scatterplot

Nominal x Nominal

UNIVERSITY OF CANBERRA

- Contingency (cross-tabs) table of:
  - Observed and expected frequencies
  - Row and/or column percentages
  - Marginal totals
- Clustered bar chart
- Chi-square
- Phi (ϕ) or Cramer's V

Example

UNIVERSITY OF CANBERRA

- A researcher took 200 cats and tried to train them to line-dance by giving them either food or affection as a reward for dance-like behaviour. At the end of the week he counted how many cats could line-dance and how many could not
- Categorical variables:
  - Training (the cat was trained with either food or affection, not both)
  - Dance (the cat either learnt to line dance or did not)

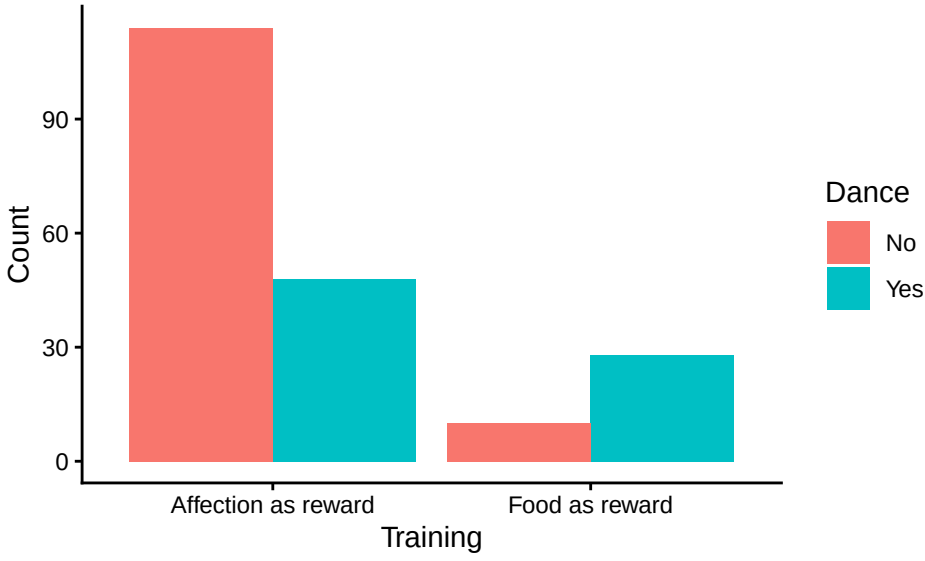
Example and dataset from Field et al. (2012)

Contingency table and chi-square test

UNIVERSITY OF CANBERRA

```
##
##    Cell Contents
## |-------------------------|
## |                   Count |
## |         Expected Values |
## |             Row Percent |
## |            Std Residual |
## |-------------------------|
##
## Total Observations in Table:  200
##
##                      | cats_data$Dance
##  cats_data$Training |        No  |       Yes  | Row Total |
## --------------------|------------|------------|-----------|
## Affection as Reward |       114  |        48  |       162  |
##                     |    100.44  |     61.56  |           |
##                     |     70.37% |    29.63%  |    81.00% |
##                     |      1.35  |     -1.73  |           |
## --------------------|------------|------------|-----------|
##      Food as Reward |        10  |        28  |        38  |
##                     |     23.56  |     14.44  |           |
##                     |     26.32% |    73.68%  |    19.00% |
##                     |     -2.79  |      3.57  |           |
## --------------------|------------|------------|-----------|
##        Column Total |       124  |        76  |       200  |
## --------------------|------------|------------|-----------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  25.35569     d.f. =  1     p =  4.767434e-07
##
## Pearson's Chi-squared test with Yates' continuity correction
## ------------------------------------------------------------
## Chi^2 =  23.52028     d.f. =  1     p =  1.236041e-06
##
##
## Fisher's Exact Test for Count Data
## ------------------------------------------------------------
## Sample estimate odds ratio:  6.579265
##
## Alternative hypothesis: true odds ratio is not equal to 1
## p =  1.311709e-06
## 95% confidence interval:  2.837773 16.42969
##
## Alternative hypothesis: true odds ratio is less than 1
## p =  0.9999999
## 95% confidence interval:  0 14.25436
##
## Alternative hypothesis: true odds ratio is greater than 1
## p =  7.7122e-07
## 95% confidence interval:  3.193221 Inf
##
##
##
##        Minimum expected frequency: 14.44
```

Clustered bar chart



Chi-square

UNIVERSITY OF CANBERRA

To calculate: $\chi^2 = \sum(\frac{(observed - expected)^2}{expected})$

- χ^2 = the test statistic that approaches a χ^2 distribution
- observed = frequencies observed
- expected = frequencies expected (by the null hypothesis) i.e. cell frequencies that should occur if the variables are not correlated
- Note: Chi-square is based on the squared differences between the actual and expected counts

Effect sizes for non-parametric measures of correlation

UNIVERSITY OF CANBERRA

- 2x2, 2x3 or 3x2 analyses -> use the Phi effect size
- 3x3 or greater analyses -> use Cramer's V

Write up

A Chi-squared analysis revealed a significant medium association between the type of training and whether or not cats would dance $\chi^2(1) = 25.36$, p $<.001$, $\phi = .36$. This seems to represent the fact that, based on the cross-tabulation results, cats were more likely to dance if they were trained with food, rather than trained with affection.

Nominal x Nominal

UNIVERSITY OF CANBERRA

- Determine the significance, strength and direction of the relationship
  - Significance = comes from the chi-square test
  - Direction = comes from the contingency table and clustered bar chart
  - Strength = comes from effect size measure, either Phi or Cramer's V

Ordinal x Ordinal

UNIVERSITY OF CANBERRA

	Nominal	**Ordinal**	**Interval/Ratio**
Nominal	Chi-squared (X^2), Phi or Cramer's V, Clustered bar chart	← Treat as for	Point bi-serial correlation (r_{pb}), Scatterplot, bar chart, or error-bar chart
Ordinal		Spearman's Rho (r_s) or Kendall's Tau, Clustered bar chart OR scatterplot	Treat as for ↑ and ←
Interval/ Ratio			Pearson's Product-moment correlation (r), Scatterplot

Correlational approaches

UNIVERSITY OF CANBERRA

- Spearman's rho (r_s)
- Kendall tau (τ)
- Can also use nominal by nominal techniques by first recoding the variables

Graphing

- Ordinal by ordinal data is difficult to visualise - it is non-parametric, but it can have many points
- You can use:
  - Non-parametric approaches (e.g. clustered bar chart)
  - Parametric approaches (e.g. scatterplot with line of best fit)

Spearman's rho

UNIVERSITY OF CANBERRA

- Spearman's rho is also called Spearman's rank order correlation
- Use this when your data is ranked (ordinal) level of measurement
- The formula is the same as the Pearson's product-moment correlation
- Be careful with your interpretation so you are taking into account the underlying ranked scales

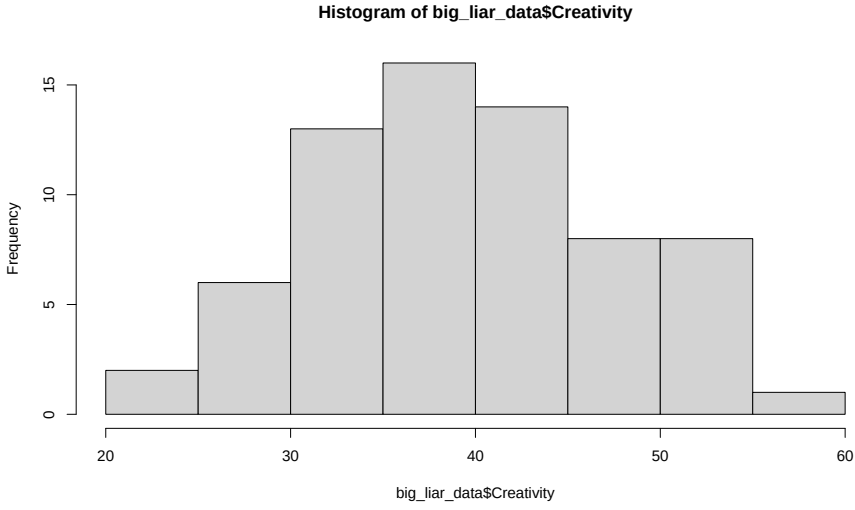
Question

- Are more creative people able to lie better? To test this, 68 participants who entered the World's Biggest Liar competition in England and told the best lie they could were asked where they were placed in the competition (i.e. first, second, third etc.). These participants also completed a creativity questionnaire (maximum score 60).
  - Since the participants placed in categories, but have meaningful order - Spearman's correlation should be used.

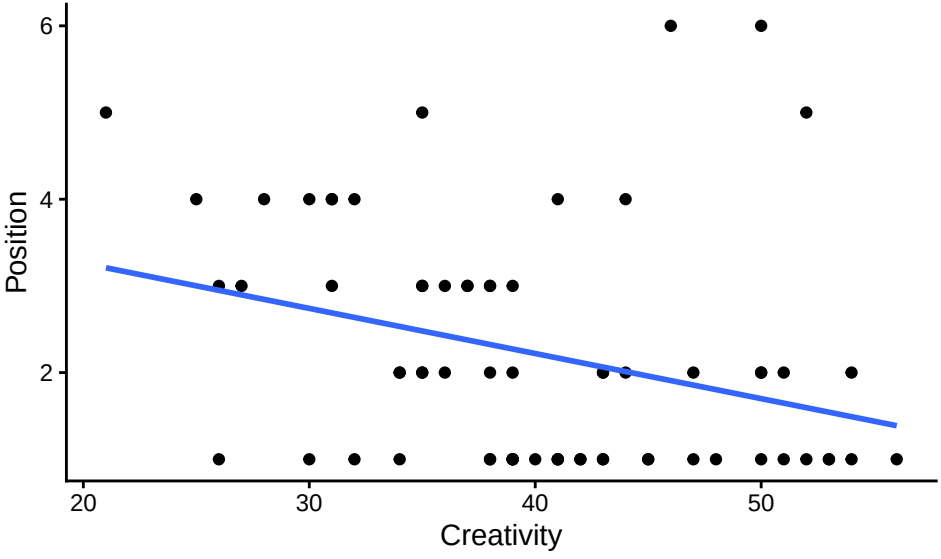
Example and dataset from Field et al. (2012)

Inspect data

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   21.00   35.00   39.00   39.99   45.25   56.00
```



Scatterplot



Correlation

UNIVERSITY OF CANBERRA

```
## Warning in cor.test.default(big_liar_data$Position, big_liar_data$Creativity, :
## Cannot compute exact p-value with ties

##
##  Spearman's rank correlation rho
##
## data:  big_liar_data$Position and big_liar_data$Creativity
## S = 71948, p-value = 0.00172
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## -0.3732184
```

Kendall's tau

UNIVERSITY OF CANBERRA

- Three versions:
  - Tau a - does not take joint ranks into account (e.g. ties for 2nd place)
  - Tau b - takes joint ranks into account, works for square tables (i.e. 2x2, 3x3, 4x4, etc.)
  - Tau c - takes joint ranks into account, works for rectangular tables (2x3, 4x2, etc.)

Correlation

UNIVERSITY OF CANBERRA

- Since our data set is small with a large number of tied ranks, kendall's tau is appropriate

```
##
##  Kendall's rank correlation tau
##
## data:  big_liar_data$Position and big_liar_data$Creativity
## z = -3.2252, p-value = 0.001259
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## -0.3002413
```

Interpretation

UNIVERSITY OF CANBERRA

There is a significant relationship between creativity scores and how well someone did in the World's Biggest Liar competition (τ = -.30, p = .001). Specifically, as creativity increased, position number/ranking decreased, indicating that higher levels of creativity were associated with improved competition performance.

Dichotomous x Interval/Ratio

UNIVERSITY OF CANBERRA

	Nominal	**Ordinal**	**Interval/Ratio**
Nominal	Chi-squared (X^2), Phi or Cramer's V, Clustered bar chart	$\leftarrow$ Treat as for	Point bi-serial correlation (r_{pb}), Scatterplot, bar chart, or error-bar chart
Ordinal		Spearman's Rho (r_s) or Kendall's Tau, Clustered bar chart OR scatterplot	Treat as for $\uparrow$ and $\leftarrow$
Interval/ Ratio			Pearson's Product-moment correlation (r), Scatterplot

Correlational approaches

UNIVERSITY OF CANBERRA

- You have one dichotomous and one interval/ratio variable
- The analysis is the point-biserial correlation (r_{pb})
- The calculation is the same as for Pearson's product-moment r
- Adjust your interpretation to consider the direction of the dichotomous scales

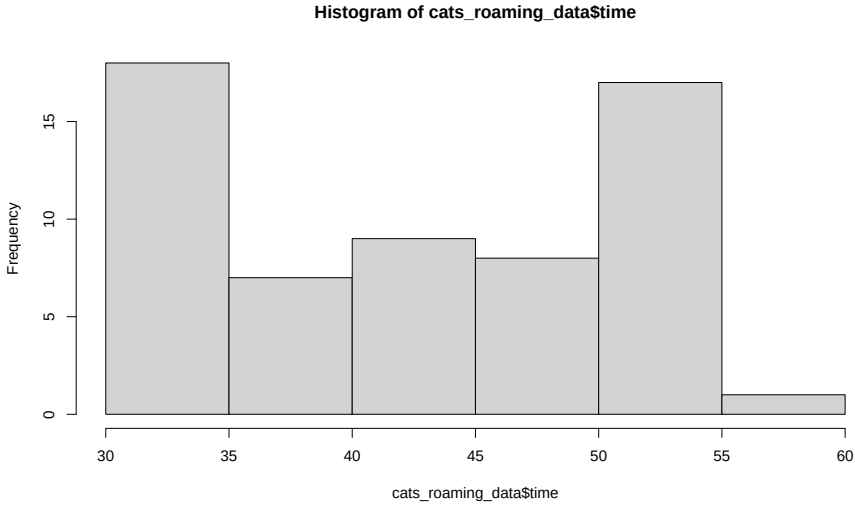
Question

- What is the association between the sex of a cat and how much time it spends away from home?
  - Note: Sex is coded 1 for male and 0 for female

Example and dataset from Field et al. (2012)

Inspect data

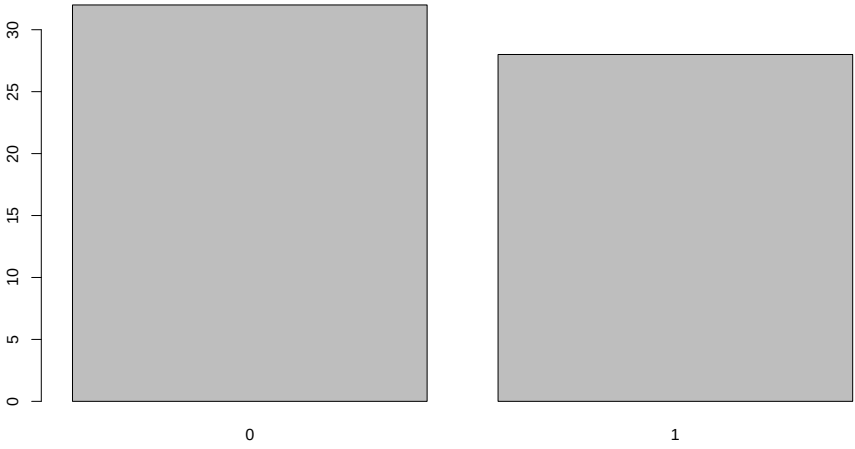
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    31.0    35.0    43.5    43.2    51.0    57.0
```



Inspect data

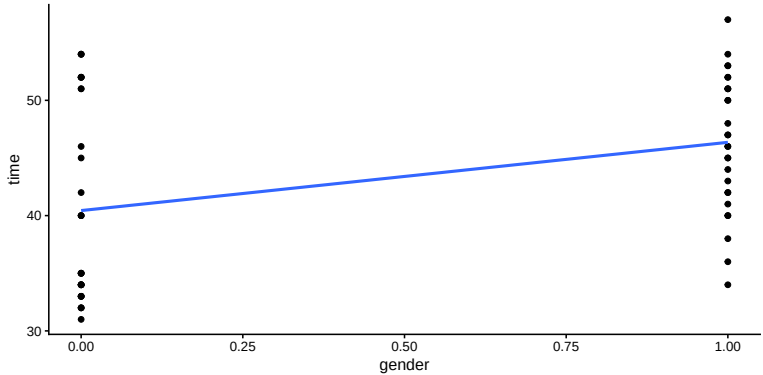
UNIVERSITY OF CANBERRA

```
##  0  1
## 32 28
```



Interpretation - Scatterplot

- Interpretation of effect size depends on coding of dichotomous variable
- Coded 0 = female and 1 = male



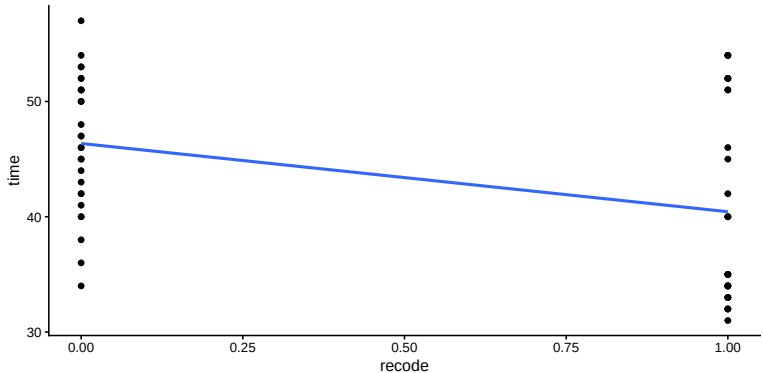
Correlation

UNIVERSITY OF CANBERRA

```
##
##  Pearson's product-moment correlation
##
## data:  cats_roaming_data$time and cats_roaming_data$gender
## t = 3.1138, df = 58, p-value = 0.002868
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.137769 0.576936
## sample estimates:
##       cor
## 0.3784542
```

Interpretation - Scatterplot

- Interpretation of effect size depends on coding of dichotomous variable
- Coded 1 = female and 0 = male



Correlation

UNIVERSITY OF CANBERRA

```
##
##  Pearson's product-moment correlation
##
## data:  cats_roaming_data$time and cats_roaming_data$recode
## t = -3.1138, df = 58, p-value = 0.002868
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.576936 -0.137769
## sample estimates:
##        cor
## -0.3784542
```

Interpretation

- Notice how the conclusion across both graphs and effect sizes are the same (i.e. male cats roam more than females), however, knowledge of how the dichotomous variable is coded is required to interpret the results and reach this conclusion.

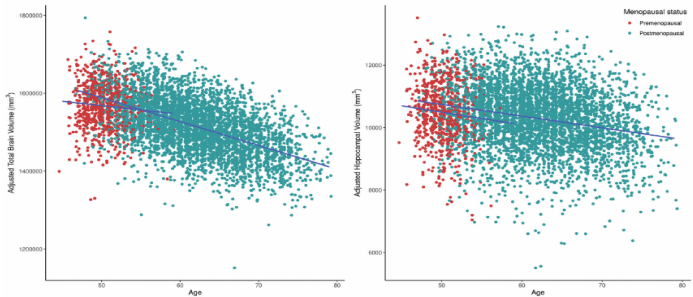
Interval/ratio x Interval/ratio

UNIVERSITY OF CANBERRA

- To graph an interval/ratio by interval/ratio relationship, use a scatterplot
- Plot each pair of observation (X and Y)
  - X = predictor variable (IV) on the x-axis
  - Y = criterion variable (DV) on the y-axis
- To interpret the direction of the relationship:
  - Positive relationship = trend is from bottom left to top right
  - Negative relationship = trend is from top left to bottom right

Example

- What is the relationship between age and brain volume in women?



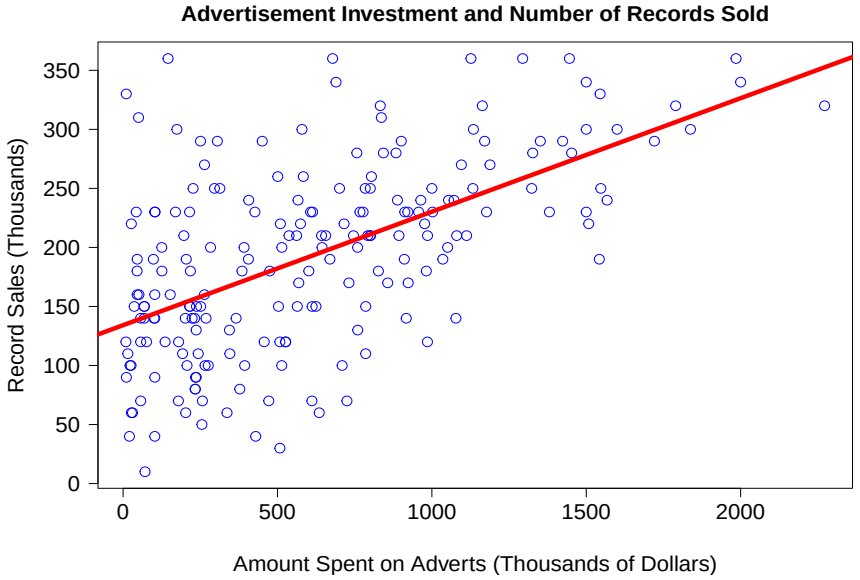
Ambikairajah et al. (2021)

Line of best fit

UNIVERSITY OF CANBERRA

- A correlation is a measure of the degree to which pairs of numbers (points) cluster together around a best-fitting straight line
- Line of best fit: $y = a + bx$
- Check for outliers and linearity - scatterplot will help with this

The method of least squares



Correlational approaches

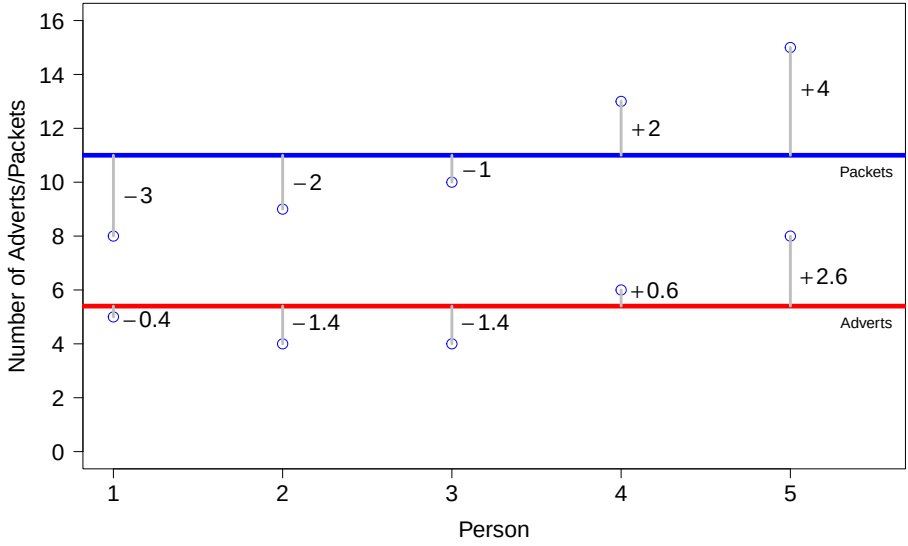
UNIVERSITY OF CANBERRA

- For interval/ratio by interval/ratio, use the Pearson product-moment correlation (r)

```
cor.test(album_data$sales, album_data$adverts, method = "pearson",
         conf.level = .95)
```

```
##
##  Pearson's product-moment correlation
##
## data:  album_data$sales and album_data$adverts
## t = 9.9793, df = 198, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4781207 0.6639409
## sample estimates:
##       cor
## 0.5784877
```

Covariance


Number of Adverts/Packets
Person
Packets
Adverts
−3
−2
−1
+2
+4
−0.4
−1.4
−1.4
+0.6
+2.6

Covariance

UNIVERSITY OF CANBERRA

- The covariance between two variables is the variance that they share.
- First, we multiply the deviation for one variable by the deviations from another variable to get the cross-product deviations: $(x_i - \bar{x})(y_i - \bar{y})$
- We then sum these cross-product deviations: $\sum(x_i - \bar{x})(y_i - \bar{y})$ and then divide by the degrees of freedom to get the covariance:
  - $cov(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1}$

Covariance

UNIVERSITY OF CANBERRA

- Covariance reflects the direction of the relationship (+/-)
- Covariance is unstandardised - depends on two variables having the same scale of measurement
- Problems with covariance:
  - It is not a standardised measure (i.e size depends on units). Therefore we cannot compare covariances in an objective way, nor can we say whether a covariance is particularly large or small relative to another data set (unless both data sets are measured in the same units).
- To overcome this problem, we use standardisation.

What is a correlation?

UNIVERSITY OF CANBERRA

- A correlation is a measure of the standardised covariance
- $r_{xy} = \frac{\text{cov(X,Y)}}{S_x S_y}$
- The relationship between X and Y is the covariance of X and Y divided by the product of the standard deviations of X and Y

What is the difference between covariance and correlation

UNIVERSITY OF CANBERRA

- The size of the covariance depends on the measurement scale used. If the variables use different scales, you can't compare them (e.g. height and weight)
- When you standardise the covariance (by dividing it by the cross-product of the standard deviations), you get the correlation
- Correlation is an effect size - a standardised measure of the strength of the relationship

Practice question

- The covariance between X and Y is 1.2. The standard deviation of X is 2 and the standard deviation of Y is 3. The correlation is:

a) 0.2
b) 0.3
c) 0.4
d) 1.2

$$r_{xy} = \frac{\text{cov(X,Y)}}{S_x S_y}$$

Practice question

- The covariance between X and Y is 1.2. The standard deviation of X is 2 and the standard deviation of Y is 3. The correlation is:

$$r_{xy} = \frac{1.2}{2 \times 3}$$

a) **0.2**
b) 0.3
c) 0.4
d) 1.2

Practice question

UNIVERSITY OF CANBERRA

- The covariance between X and Y is 5. The standard deviation of X is 2 and the standard deviation of Y is 4. The correlation is:

$$r_{xy} = \frac{\text{cov(X,Y)}}{S_x S_y}$$

a) 0.13
b) 0.28
c) 0.51
d) 0.63

Practice question

- The covariance between X and Y is 5. The standard deviation of X is 2 and the standard deviation of Y is 4. The correlation is:

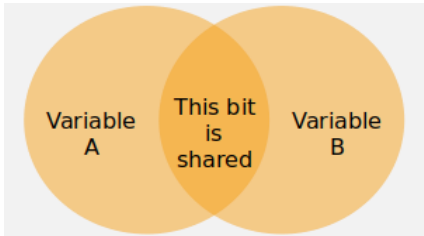
a) 0.13
b) 0.28
c) 0.51
d) **0.63**

Interpreting correlations

UNIVERSITY OF CANBERRA

Coefficient of Determination R^2

- R^2 = The proportion of variance in one variable that can be accounted for by another variable
- For example, if $r = .60$, $R^2 = .36$
  - This means 36% of the variance is shared between the variables



Interpreting correlations

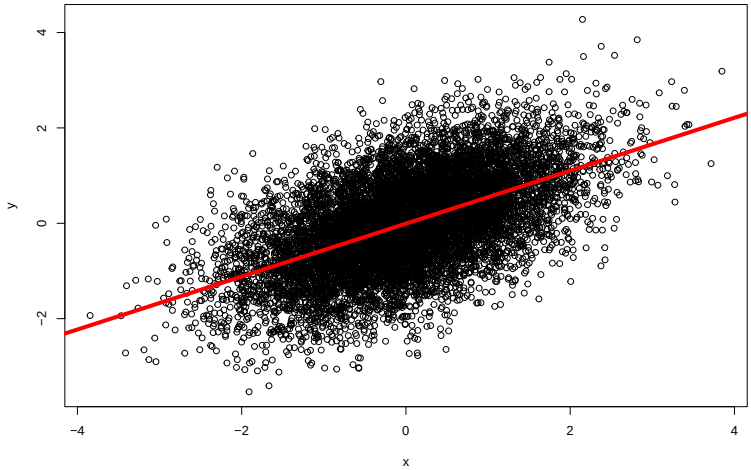
UNIVERSITY OF CANBERRA

- A correlation is an effect size
- We follow Cohen (1977) guidelines to determine the strength of a correlation:

Strength	r	r^2
Weak	.1 – .3	1 – 9%
Moderate	.3 – .5	10 – 25%
Strong	$>$ .5	$>$ 25%

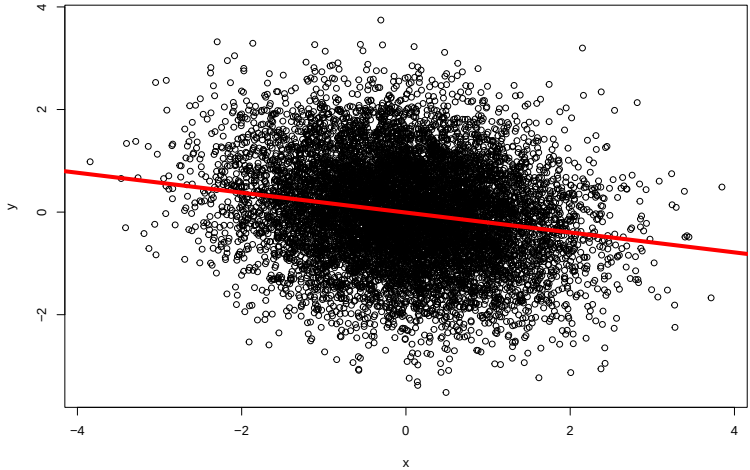
Size of correlation

- Strong correlation:
  - $r = .55$
    $p < .001$
- What is the percentage of shared variance?
  - 30%



Size of correlation

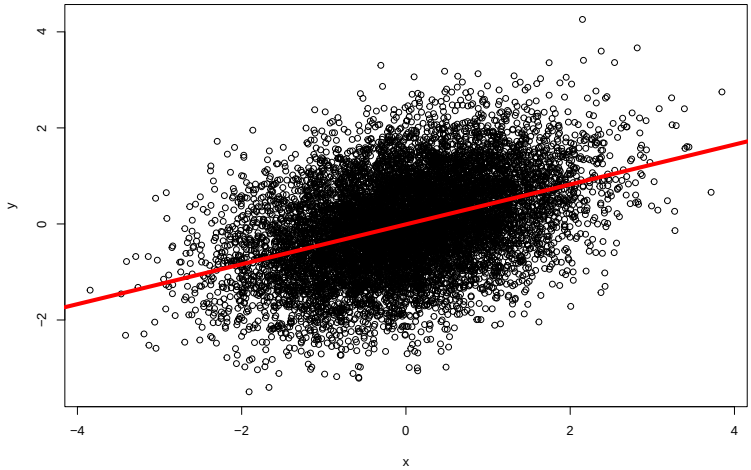
- Weak correlation:
  - r = -.2
    p $<$ .001
- What is the percentage of shared variance?
  - 4%



Size of correlation

UNIVERSITY OF CANBERRA

- Moderate correlation:
  - $r = .41$
    $p < .001$
- What is the percentage of shared variance?
  - 17%



Other interpretations

- Evans (1996) suggested the following for interpreting the strength of correlations:

Strength	r	r^2
Very weak	.00 – .19	0 – 4%
Weak	.20 – .39	4 – 16%
Moderate	.40 – .59	16 – 36%
Strong	.60 – .79	36 – 64%
Very strong	.80 – 1.00	64 – 100%

Assumptions and limitations

UNIVERSITY OF CANBERRA

- Assumptions and limitations of pearson product moment linear correlation
  - Levels of measurement
  - Normality
  - Linearity
  - Effects of outliers
  - Non-linearity
  - Homoscedasticity
  - No range restriction
  - Homogenous samples
  - Correlation is not causation
  - Dealing with multiple correlations

Normality

UNIVERSITY OF CANBERRA

- Assumptions that you have sampled from populations with normal distribution (of your residuals - but can start by looking at individual variables)
- Avoid relying on only one indicator of normality. Use histograms, skewness and kurtosis statistics

Effects of outliers

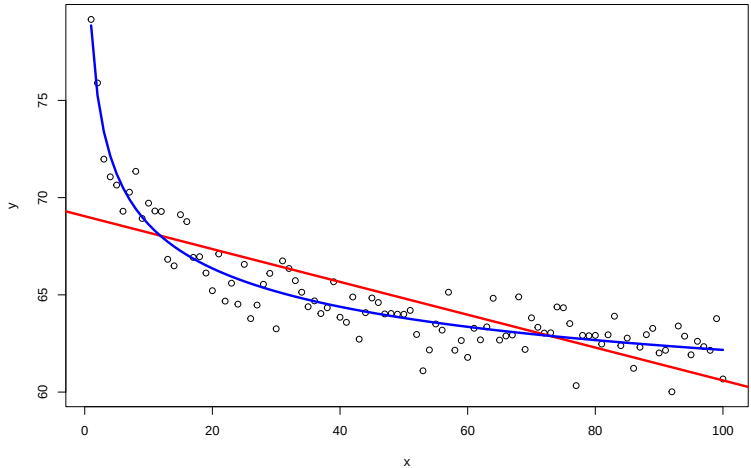
UNIVERSITY OF CANBERRA

- Outliers can disproportionately increase or decrease r
- How do you deal with this?
  - Compute r with and without outliers to see the effect
  - Get more data for outlying values
  - Recode outliers as having more conservative scores
  - Transformation
  - Recode variable into lower LOM and use a non-parametric approach

Non-linear relationship

UNIVERSITY OF CANBERRA

- Check scatterplot
- Does a linear relationship capture the lion's share of variance?



What do you do if it is non-linear

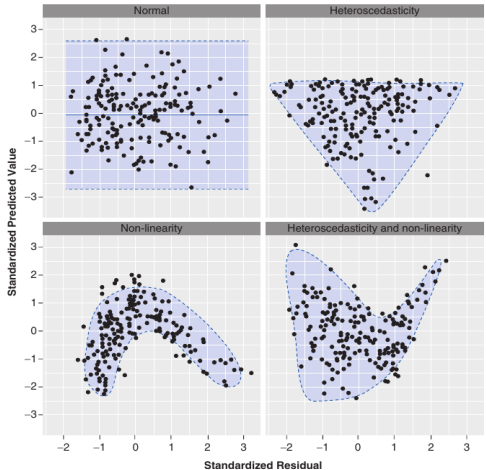
- Use a non-linear mathematical function to describe the relationship between the variables
- Transform variables to ‘create’ a linear relationship

Scedasticity

UNIVERSITY OF CANBERRA

- Homoscedasticity = even spread of observations around a line of best fit
- Heteroscedasticity = uneven spread of observations around a line of best fit
- Scedasticity can be assessed visually and by using Levene's test

Scedasticity



Example from Field et al. (2012)

Range restriction

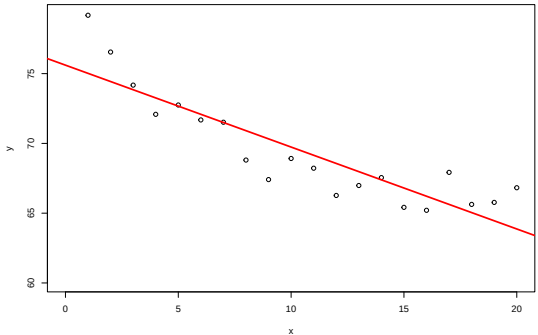
UNIVERSITY OF CANBERRA

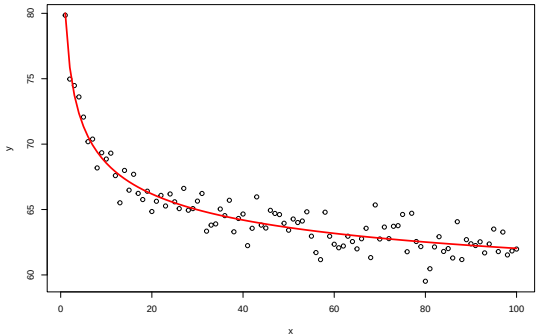
- Range restriction occurs when the sample has a restricted (or truncated) range of scores
  - E.g. age in a student sample - many participants are in the 18-20 age bracket
- If the range is restricted, be cautious about generalising beyond the range for which data is available
  - E.g. income may not continue to increase linearly with age

Range restriction

UNIVERSITY OF CANBERRA

Same data, different ranges





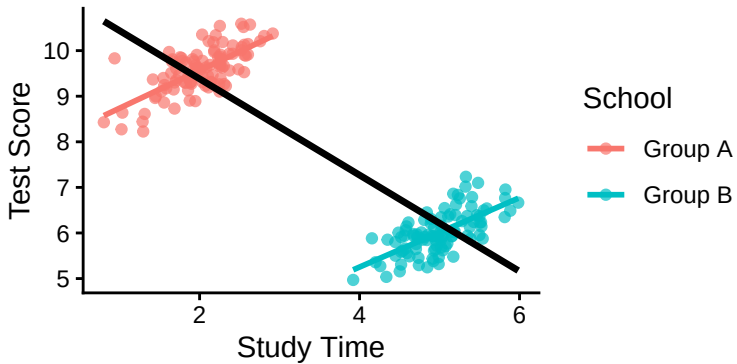
Heterogenous sample

UNIVERSITY OF CANBERRA

- If there are different subsamples in your data, this may artificially increase or decrease the overall r
- You can calculate r separately for subsamples and as well as overall, then look at the differences

Simpson's Paradox

- A relationship observed in aggregated data can reverse when a third variable is considered.



Correlation $\neq$ causation

UNIVERSITY OF CANBERRA

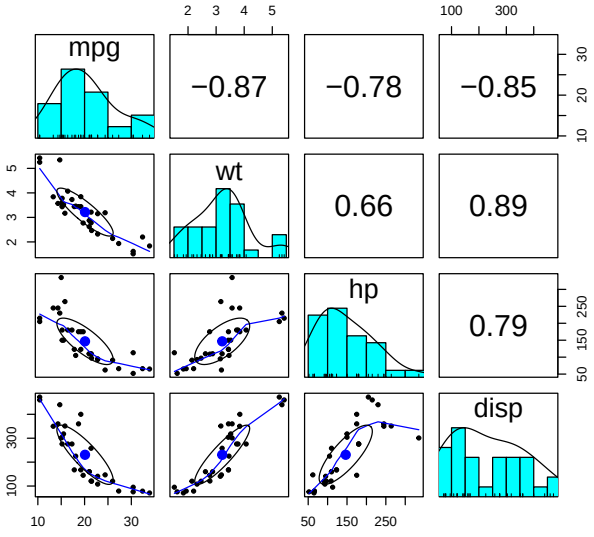
- Remember that just because two variables are related, this does not mean that one variable causes the other
- Classic example = shark attacks increase with more ice cream sales
- Causal interpretation = eating ice cream causes shark attacks
- Alternative explanation = people eat more ice cream when it's hotter and they also tend to swim more. With more people in the water, shark attacks become more likely.
- See more examples here

Dealing with several correlations

UNIVERSITY OF CANBERRA

- Scatterplot matrices organise scatterplots and correlations amongst several variables at once

Scatterplot matrix and correlation table




```
##          mpg    wt    hp  disp
## mpg    1.00 -0.87 -0.78 -0.85
## wt    -0.87  1.00  0.66  0.89
## hp    -0.78  0.66  1.00  0.79
## disp -0.85  0.89  0.79  1.00
```

Summary

- Covariation is everywhere!
- Correlation is a standardised measure of the covariance (extent to which two variables co-relate)
- It ranges from -1 to $+1$, with more extreme values indicating stronger relationships
- Correlation does not prove causation

Purpose of correlation

UNIVERSITY OF CANBERRA

- What is the relationship/association/shared variance/co-relation between two variables?

Correlations depend on the level of measurement

UNIVERSITY OF CANBERRA

	Nominal	**Ordinal**	**Interval/Ratio**
Nominal	Chi-squared (X^2), Phi or Cramer's V, Clustered bar chart	$\leftarrow$ Treat as for	Point bi-serial correlation (r_{pb}), Scatterplot, bar chart, or error-bar chart
Ordinal		Spearman's Rho (r_s) or Kendall's Tau, Clustered bar chart OR scatterplot	Treat as for $\uparrow$ and $\leftarrow$
Interval/ Ratio			Pearson's Product-moment correlation (r), Scatterplot

Steps to take

1) Identify the level of measurement of your variables
2) This will help guide the type of correlation and graph to use
3) Check your graphs for:

- Linearity
- Outliers
- Scedasticity
- Range restriction
- Subsamples to consider

Steps to take

4) Consider

- Effect size (e.g. ϕ, Cramer's V, r, R^2)
- Direction of the relationship
- Significance (inferential test/p-value)

5) Interpret and discuss your finding:

- Relate it back to your hypothesis
- What are the limitations? (e.g. based on hetereogeneity, range restriction, for making causal inference)

Interpreting correlations

UNIVERSITY OF CANBERRA

Strength	r	r^2
Weak	$.1 - .3$	$1 - 9\%$
Moderate	$.3 - .5$	$10 - 25\%$
Strong	$> .5$	$> 25\%$

Assumptions and limitations

UNIVERSITY OF CANBERRA

- Assumptions and limitations of pearson product moment linear correlation
  - Levels of measurement
  - Normality
  - Linearity
  - Effects of outliers
  - Non-linearity
  - Homoscedasticity
  - No range restriction
  - Homogenous samples
  - Correlation is not causation
  - Dealing with multiple correlations

Next week - exploratory factor analysis

UNIVERSITY OF CANBERRA

- Exploratory factor analysis, including:
  - An introduction to factor analysis
  - Exploratory factor analysis examples
  - Steps to run your own exploratory factor analysis

Contributions to this course

UNIVERSITY OF CANBERRA

Dr James Neill

Dr Samantha Stanley

Dr Jeroen van Boxtel

References I

Ambikairajah, A., Tabatabaei-Jafari, H., Hornberger, M., & Cherbuin, N. (2021). Age, menstruation history, and the brain. *Menopause*, *28*(2), 167–174. https://doi.org/10/ghtfz7

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences, Rev. ed* (pp. xv, 474). Lawrence Erlbaum Associates, Inc.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences* (pp. xxii, 600). Thomson Brooks/Cole Publishing Co.

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.