



RMHI PALS

Thursday 6th June 10AM-1PM

Disclaimer

We are **not** professionals.

We're all students here so don't just feel free to chime in every once in a while, we'd go as far as to encourage you to criticise us as much as you can!

You're all free to discuss with each other in groups if you'd prefer but please don't be too loud to disrupt those around you.

Concepts I will not be covering

(but you should really know them well!)

- Structure of a good research question
- Population vs sample
- Parameters vs statistics
- Types of distributions
- Constructs, measures and scores
- Properties of an estimator

Basics of R: Basic Operators

- `+`, addition
- `-`, subtraction
- `/`, division
- `*`, multiplication
- `^`, exponentiation

`2+2 #R is a fancy calculator`

```
## [1] 4
```

Basics of R: Logical Operators

- `==`, is equal to
- `!=`, is not equal to
- `>=`, is greater than or equal to
- `<=`, is less than or equal to
- `>`, is greater than
- `<`, is less than
- `|` is OR
- `&` is AND
- `!` is NOT

Basics of R: Functions

Functions are ways to make R do things!

```
ay2 <- function(x) {  
  x = x + 2  
} # "ay2" is a created function that adds the value inputed by 2  
  
x <- 5  
print(ay2(x))  
## [1] 7
```

```
ye <- function(feef,beep) {  
  beep = feef^2 + beep  
} #'ye' is another created function that squares the first value (the 'feef')  
and adds the second value (the 'beep')  
  
print(ye(2,3))  
## [1] 7  
  
#not equal to  
print(ye(3,2)) #the order of arguments are important!  
## [1] 11  
  
#alternatively  
print(ye(beep = 3, feef = 4)) #for sanity's sake you can set the arguments  
manually  
## [1] 19  
  
print(ye(feef = 4,beep=3))  
## [1] 19
```

Basics of R: Variable types

- a) Numeric = R considers numeric variables numbers.
- b) Characters = Also called 'strings', can contain letters or numbers.
- c) Logical = True/False, lot more powerful than you think.

Any variable with one or more values is called an (atomic) vector.

The values in a vector is called an element.

Basics of R: Packages

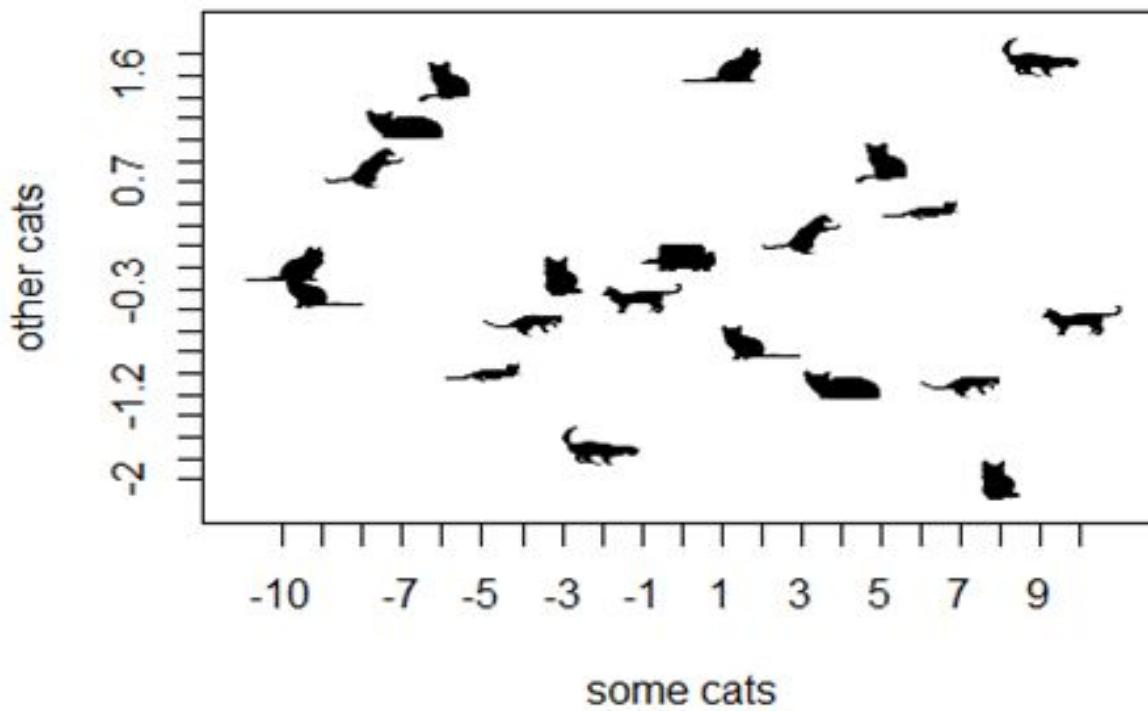
R allows for the installation and usage of 'packages'.

Packages contain functions and other fun-things that augment R in some way.

You can install packages and load them up to use functions that are relevant for what you're aiming to accomplish with your data wrangling, data visualisation, data analysis, etc.

Example: catterplots

Random Cats



Basics of R: Wrangling with Logical operators

Square brackets and logical expressions can be extremely powerful!

```
thingy <- c("ye", "na", "whep") #dont do this kind of variable naming
otherthing <- c(T,T,F) #seriously dont name things like this
thingy[otherthing] #viewing the contents of 'thingy' sorted by the
'otherthing' variable

## [1] "ye" "na"

thingy[1] #the first item in 'thingy'

## [1] "ye"

thingy[3] #the third item in 'thingy'

## [1] "whep"

#it's actually useful for data, trust me
bork <- c(5,10,15,24,30)
bork[bork<=20] #returning the contents of 'bork' that are Less than or equal
to 20

## [1] 5 10 15
```

```
bork[bork>=20]  
## [1] 24 30  
  
bork[bork/5 == round(bork/5,0)] #returning things that are divisible by 5  
(when divided by 5, is equal to a whole number)  
## [1] 5 10 15 30
```

Basics of R: Dataframes

Data-frames are variables combined together.

Data-frames can be conceptualised as a bunch of vectors mashed together into a dataset.

The \$ can be used as an operator to select particular variables.

```
cor(data$gattwto08, data$gdppc)  
## [1] 0.3615475
```

```
summary(reg5)$adj.r.squared  
## [1] 0.1827901
```

SidebaR: Documentation

Wise words from the amazing Danielle Navarro:

“Your worst enemy is you 6 months from now.”

Or in my case, probably 6 days from now.

Documentation to describe what portions of your script do (with the #) will
SAVE YOUR SANITY.

R ignores everything after the #, so use it away!

You could also use R markdown for more intensive documentation.

Quick R Exercise:

Consider this 5-entry data frame:

What will the output be when I input:

- a) `wheeble$agh[beh!="nay"]`
- b) `wheeble$agh[beh!="hah"]`
- c) `wheeble$beh[(wheeble$agh/2==round(wheeble$agh/2,0)) == T]`

	agh	beh
1	5	yay
2	7	nay
3	10	yay
4	15	aaah
5	20	aaah

Solution

```
> wheeble$agh[beh!="nay"]
[1] 5 10 15 20
> wheeble$agh[beh!="hah"]
[1] 5 7 10 15 20
> wheeble$beh[(wheeble$agh/2 == round(wheeble$agh/2,0))==T]
[1] yay aaah
```

These expressions combine logical operators and data frame variable notation.

I hope this example gives you an idea of the many ways we can manipulate a data frame.

Core Concepts: Manipulating scores

4 different types of scores:

- a) Raw/Observed scores: untouched metric from measure,
- b) Deviation score: scaled to the metric mean,
- c) Standardised score: scaled to a specified mean and a specified amount of standard deviations per unit,
- d) Z score: a type of standardised score with mean 0 and 1 standard deviation per unit.

Core Concepts: Random Variables

A variable consisting of the set of all possible outcomes defined by one or more population parameters.

I understand RVs to be the perpetually unknowable variable defined by the existing parameters of the natural world.

If it's easier to grasp as an infinite set ranging from infinity to negative infinity, that may be correct in some situations.

Core Concepts: Random Variables

What we do know is that they come in two flavours:

- a) Continuous (e.g. height, temperature, etc.) If naturally occurring decimals make sense- it's a continuous variable (13.4 cm, 34.2 celsius)
- b) Discrete (or Factor in R) (e.g. measure responses, no. of errors made, etc.)
Ones where naturally occurring decimals make no sense. Categorical variables are a type of discrete variables.

What matters is that we know these numbers behave differently; don't take continuous variables to be discrete and vice versa.

Sidebar: Standard Error

The standard deviation of a sampling distribution is called a standard error- why?

A way to think about this is that a sampling distribution talks of the set of different statistics you can get, so the spread in that distribution is the typical amount the statistic is different by, i.e. the typical amount of error from your computed statistic.

Sidebar: Observed Test Statistic

An application of the sampling distribution (I believe) is the calculation of the observed test statistic.

Observed Test Statistics (T_{obs}) are the statistics computed to encapsulate your data which are then compared to the corresponding theoretical probability distribution to construct a p-value.

This is your t-statistic, chi-square statistic, F-statistic, etc.

Corresponds with a p-value (more on that in a bit).

Core Concepts: Theoretical Probability Distribution

Sampling distributions are cool, but these are even cooler- because they exist by mathematical functions and don't need an unfeasible amount of resources to make.

A standardised sampling distribution is equivalent to a theoretical probability distribution.

Fitted normal curves are equivalent to sampling distributions- so we have the distributions that have the same properties as the sampling distributions without actually needing to construct them manually.

Core Concepts: NHST

Good ol' null hypothesis significance testing.

So what's a p-value? The p-value is a widely used statistical inference method to evaluate statistical hypotheses (e.g. null hypotheses).

Note that statistical inference is different from scientific inference, scientific inference may involve statistical inference, but you cannot say that a statistical inference implies a scientific inference. No. The tools we use are not infallible and reflects the researcher's intentions of inferring ideas from data.

Core Concepts: NHST

The typical method to reject the null hypothesis is:

Set null and alternative hypotheses,

Decide the alpha (typically .05 and consequently the critical test statistic(T_{crit})),

Calculate the observed test statistic (T_{obs}) from relevant sample stats,

Find the p-value corresponding to the T_{obs} of the data,

If $p < \alpha$, reject (consequently if $|T_{obs}| > |T_{crit}|$, reject) otherwise, retain.

NHST: What the p-value is NOT.

It's NOT:

The probability of the scientific hypothesis being true,

The probability of your alternative statistical hypothesis being false,

The probability of your null hypothesis being true,

The probability of your null hypothesis being false,

The probability of observing your sample statistic.

Core Concepts: Confidence Intervals (CIs)

CIs are a function of p-values (and of standard error in general)

The margin of error (interval area) is defined by what we don't consider to be statistically significantly different- and therefore 'plausible'.

Remember, the sampling distribution represents multiple samples- and values within the confidence interval are 'plausible' as a function of the preset alpha value and standard error of the sampling distribution to determine what values are 'plausible' in other samples.

In other words, it depends on the standard error value and our preset alpha criterion value.

CIs

The 'typical' way to formulate CIs are:

$$\text{Statistic} \pm T_{crit} * \text{Standard Error}$$

However, I think it might be easier to conceptualise CIs if you formulate them as:

$$\frac{\text{Statistic} \pm X}{\text{Standard Error}} = T_{crit}$$

With X representing the null hypothesised value and eventual confidence interval values.

CIs

$$\frac{\text{Statistic} \pm X}{\text{Standard Error}} = T_{crit}$$

Note that this expression is equivalent to a z-score and note the absence of absolute value ($|T_{crit}|$), we want both sides of the T_{crit} . Assuming alpha = .05, $T_{crits} = 1.96$ and -1.96 . We know three constants: the statistic, its standard error and then our a priori critical test statistics, and now we work backwards to find the two X values.

What I want you guys to take away is that CIs represent a range of null hypothesised values which cannot be considered statistically significantly different from your observed statistic.

Worked Example

```
#random vector
a <- c(90:110)
mean(a)

## [1] 100

sd(a)

## [1] 6.204837

#standard error:
plotrix::std.error(a)

## [1] 1.354006

sqrt((sd(a))^2/length(a))

## [1] 1.354006

#either way works.
```

```
#getting Tcrit, assuming normal distribution
#for alpha = 0.05
qnorm(1-(1-0.95)/2)

## [1] 1.959964

#and now we work backwards
#upper bound for alpha = 0.05
up <- (((qnorm(1-(1-0.95)/2)*plotrix::std.error(a))*(-1) - mean(a)))*(-1)
#note: the negatives in the equation reflect movement of certain portions of
the formula which can be quite confusing.

#Lower bound for alpha = 0.05
low <- (((qnorm(1-(1-0.95)/2)*plotrix::std.error(a)) - mean(a)))*(-1) #same
thing with the negatives. would not recommend you waste your time making sure
the negatives line up, it's tricky.

print(c(low, up)) #here's your interval from working backwards

## [1] 97.3462 102.6538
```

```
#compare it to the function representing the typical formula
ci <- function(x, conf){
  lower <- mean(x) - (qnorm(1-(1-conf)/2)*sqrt((sd(x))^2/length(x))) #Lower
  bound = mean - critical test value multiplied by the standard error
  upper <- mean(x) + (qnorm(1-(1-conf)/2)*sqrt((sd(x))^2/length(x))) #upper
  bound = mean + critical test value multiplied by the standard error
  out <- c(lower, upper) #creating a vector to place the lower and upper
  bounds
  print(out) #printing the confidence interval
  rm(out, lower, upper) #removing the variables created so the workspace
  doesn't get cluttered
}

ci(a,0.95) #pretty much the same.

## [1] 97.3462 102.6538
print(c(low, up)) #here's your interval from working backwards
## [1] 97.3462 102.6538
```

Sidebar: Alpha criterion

These serve to control type-1 error rates, and define boundaries of statistical significance.

$P(\text{Rejecting } H_0 | H_0 = T)$ is what the alpha represents.

In the metric of sampling distributions, the alpha criterion is directly equivalent to the critical test statistic (the value that the observed test statistic has to exceed to declare statistical significance).

Core Concepts: Effect Size

Last one. Hurray!

Kind of an ambiguous term to refer to a method of describing the strength of a relationship.

E.g., Pearson's r, Cramer's V, standardised regression coefficients or R-squared.

Effect size measures can also have their own conditions where they are robust or not, no matter whether or not your design has a statistical assumption that needs to be satisfied.

E.g., Bonett's delta, Hedges' g

Association

Systematic co-occurrence.

'These things occur together' is what association represents.

Association: Continuous Variables

We use pearson's correlation coefficient (I just call it Pearson's r)

But first, scatter-plots and covariance!

A neat game: <http://guessthecorrelation.com/>

Association: Covariance

The covariance of the plots directly refer to how two variables vary together- how much a sample varies on two variables together. In other words, it can tell you the strength (in its own metric) and direction of how much a sample systematically varies on two variables.

But here's a problem- it's in its own metric, so how do we change it to be meaningful?

We standardise it!

And that is what correlation is- standardised covariance. Or to be specific, covariance of z-scored variables.

Association: Pearson's r

r ranges from -1 and +1 and captures the strength and direction of association between two continuous variables.

'r' is the sample coefficient, the population coefficient is symbolised as rho (ρ).

Association: Categorical Variables

Contingency/Frequency tables!

The Cramer's V statistic is essentially a rescaled chi-square statistic to fit between 0 to 1.

The higher the V the stronger the association.

No direction since the chi-sq stat doesn't have a direction.

Association: Odds

Odds are the probability of something occurring relative to it not occurring.

i.e. 4:1 odds mean 80% to 20%, 9:1 odds mean 90:10.

20% chance it'll rain = 4:1 odds of not raining. (4 times more likely for it to not rain than it is to rain)

10% chance I'll bring my umbrella = 9:1 odds of not bringing my umbrella (9 times less likely for me to bring my umbrella than me not bringing my umbrella)

Association: Odds Ratios

Odds ratios (OR) take it a step further. The OR represents the ratio of two odds:

$$(9/1 / 4/1) = 2.25$$

It is 2.25x more likely for me to not bring my umbrella when it doesn't rain.

The reciprocal holds true too

$$4/9 = 0.44 \quad (4/1 / 9/1)$$

It is 0.44x more likely for me to bring my umbrella when it doesn't rain.

Association: Odds Ratios

Provided you don't mess with the raw numerical values (never mess with raw values!!), you can twist your odds ratio around to brain-numbing ways.

1/9 / 4/1 odds ratio of me bringing my umbrella when it doesn't rain

It is 1/36x more likely for me to bring my umbrella when it doesn't rain relative to not bringing my umbrella when it does.

9/1 / 1/4 odds ratio of me not bringing my umbrella when it rains

It is 36x more likely for me to not bring my umbrella when it rains relative to bringing my umbrella when it doesn't rain.

Couple o' Cheeky SAQs

- 1) Can a p-value of 0.21, equivalent to an observed test statistic of -1.30, be considered statistically significant if we've set the critical test statistic to be 1.18? (Arbitrary values, assuming a two-tailed null hypothesis)
- 2) What is the value of a 0% Confidence Interval?
- 3) Why are values in a confidence interval considered to be 'plausible'?
- 4) Why can covariance and correlation be negative, whereas variance can only be a positive value?

Answers to said Cheeky SAQs

- 1) Yes, it is considered statistically significant because the absolute observed test statistic is larger than the critical test statistic.
- 2) A 0% confidence interval would return the sample statistic used to compute the interval. The confidence level is determined by (1-alpha level), and a confidence level of 0 would mean an alpha level of 1- everything would be statistically significant, so there would be no range of plausible values. This is less a practical question and more on the understanding of constructions of CIs.

Answers to said Cheeky SAQs

- 3) Values within a confidence interval are considered 'plausible' as an effect of the level of significance and the computed standard error. The notion of 'plausibility' refers to how the range of values cannot be considered statistically significant from the computed statistic- and are therefore possible in other samples
- 4) Covariance and Correlation can have negative values since they are not squared values. Variance is squared, so it can never have a negative value.

Prediction

Prediction of a variable usually refers to use of regression, specifically for this course, we'll be considering two types of regression:

- a) Simple linear regression: 1DV, 1IV
- b) Multiple linear regression: 1 DV, more than 1 IVs

Prediction: Not Association

Correlation	Linear regression
Linear association between 2 variables	Prediction between dependent and independent variables
Bi-directional	Non bidirectional
Correlation coefficient indicates the strength of association between 2 variables (how 2 variables vary together)	Regression coefficient/Slope of regression line indicates the impact of one unit change in independent variables on dependent variables.
Pearson's r ranges from -1 to +1	R^2 measures strength of prediction, ranging from 0 to 1.

Prediction: Concepts

- **Variance**: The average amount of variability between observed scores and the mean. Can be referred to as 'sums of squares' (SS), and there are different types of sums of squares in regression models ($SStotal = SSreg + SSres$)
- **Standard deviation**: Square root of the variance
- **Correlation Coefficient**: Standardised covariance coefficient.
- **Covariance**: An unstandardised measure of calculating how much two variables vary together.

Standardised measures are typically more informative- since they are interpretable in terms of standard deviation opposed to the raw metric.

Prediction: Linear Regression

The Full simple linear regression line is expressed as:

$$Y = a + bX + e$$

Y is the dependent variable (predicted)

X is the independent variable (predictor)

a is the intercept of the line

b is the slope of the linear function- and indicates the direction of the relationship between variables.

And e is the 'error' term- also called residual.

Prediction: Linear Regression

The simplified linear regression line is written as

$$Y(\hat{}) = a + bX$$

There aren't any error terms in the simplified linear regression line because $Y(\hat{})$ refers to the predicted value of the dependent variable. In other words,
 $Y = Y(\hat{}) + \text{error}$

Prediction: Least Squares Regression

We will be using Ordinary Least Squares (OLS) regression for estimating our regression line.

What OLS regression does is minimise the value of the error terms- or to minimise the sums of squares of residuals (minimising the proportion of variance attributed to residuals)

Prediction: Multiple Linear Regression

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n + e$$

Similar to the simple regression line, but with multiple independent variables

$b_1, b_2, b_3, \dots, b_n$: partial regression coefficient: indicates the expected change in Y for a unit of change in an IV, holding other IVs constant.

Prediction: R-squared

R-squared (Coefficient of determination) is the proportion of variance explained by the regression model (SS_{reg}/SS_{total}).

It is a method to assess the strength of the complete regression model (all IVs are taken into consideration).

Another important aspect to consider is that SS_{total} is additively decomposed, $SS_{reg} + SS_{res}$ will always equal to SS_{total} .

Prediction: R-squared considerations

R-squared is not the same thing as adjusted R-squared.

The computed R-squared statistic may be biased if:

- a) There are a large amount of predictors,
- b) There is a small sample size.

The adjusted R-squared returns a less biased statistic of the R-squared, so pay attention if the model has a large amount of IVs and a small sample size.

Prediction: Strength of IVs

- a) Standardised regression coefficients: by using standardised data instead of data in its own metric, we can evaluate our model in terms of standard deviation and can allow you to compare strengths of multiple IVs- since they're now on the same metric: SD.
- b) Semi-partial correlation: the correlation between the chosen IV and the DV, holding all other IVs constant.
- c) Squared semi-partial correlation: The proportion of variance the chosen IV explains in the DV, holding all other IVs constant.

Prediction: Statistical Assumptions

1. Independence of observations: one participant's scores are observed and recorded independently of other participants.
2. Linearity. Linearity between the predictors and the predicted variable (multiple IVs and the one DV)
3. Homoscedasticity (Constant residual variance): errors vary independently (errors aren't associated to IVs).
4. Normality of residuals. We'd like our residuals to be normally distributed.

Homoscedasticity

Breusch-Pagan test assumes the constant residuals scores → so we would like to see a big value of p for this null hypothesis test.

Prediction MCQ

1. For the following regression equation, which option is incorrect?

$$Y = 2.6 - 0.41X_1 + 0.51X_2$$

- A. When X_1 increases by 1 and X_2 increases by 1, Y will increase by 0.10
- B. The value of Y is 2.6 when scores on all independent variables are zero
- C. Holding the scores on X_1 constant, 1 unit increase in X_2 predicts 0.51 unit increase in Y.
- D. Holding the scores on X_2 constant, 2 units decrease in X_1 predicts 0.82 unit increase in Y.

Prediction MCQ Answer

- For the following regression equation, which option is incorrect?

$$Y = 2.6 - 0.41X_1 + 0.51X_2$$

- A. When X_1 increases by 1 and X_2 increases by 1, Y will increase by 0.10
- B. The value of Y is 2.6 when scores on all independent variables are zero
- C. Holding the scores on X_1 constant, 1 unit increase in X_2 predicts 0.51 unit increase in Y.
- D. Holding the scores on X_2 constant, 2 units decrease in X_1 predicts 0.82 unit increase in Y.

2. Which of the following is not an assumption of a multiple linear regression model?
- A. All observed scores are recorded independently
 - B. Residuals scores are normally distributed
 - C. Residual scores are constant
 - D. Predictors are linearly correlated with each other.

2. Which of the following is not an assumption of a multiple linear regression model?
- A. All observed scores are recorded independently
 - B. Residuals scores are normally distributed
 - C. Residual scores are constant
 - D. Predictors are linearly correlated with each other.

3. Which of the following description is closest to the concept of heteroscedasticity?
- A. Constant residual variance on different predicted values of the dependent variable.
 - B. Non-constant residual variance on different values of independent variables.
 - C. Residuals are distributed equally above and below zero.
 - D. Residuals are not distributed equally above and below zero.

3. Which of the following description is closest to the concept of heteroscedasticity?
- A. Constant residual variance on different predicted values of the dependent variable.
 - B. Non-constant residual variance on different values of independent variables.
 - C. Residuals are distributed equally above and below zero.
 - D. Residuals are not distributed equally above and below zero.

4. In which of the following conditions should the adjusted R squared be used?

- A. Sample size and the number of IVs are large
- B. sample size and the number of IVs are small
- C. R-square statistic is computed repeatedly
- D. Sample size is small and number of IVs are large

4. In which of the following conditions should the adjusted R squared be used?

- A. Sample size and the number of IVs are large
- B. sample size and the number of IVs are small
- C. R-square statistic is computed repeatedly
- D. Sample size is small and number of IVs are large

5. Which of the following is the best interpretation if an IV has a 0.3 squared semi-partial correlation?
- a. The correlation coefficient between the IV and DV is 0.3
 - b. For one unit increase in IV, there will be 0.3 units increase in DV
 - c. The IV uniquely explains 30% of the variation in DV, holding other IVs constant
 - d. The IV uniquely explains 30% of the variation in DV

5. Which of the following is the best interpretation if an IV has a 0.3 squared semi-partial correlation?
- a. The correlation coefficient between the IV and DV is 0.3
 - b. For one unit increase in IV, there will be 0.3 units increase in DV
 - c. The IV uniquely explains 30% of the variation in DV, holding other IVs constant
 - d. The IV uniquely explains 30% of the variation in DV

6. Which of the following sentences is the most correct interpretation of the following analysis?

- A. Openness is a statistically significantly stronger predictor of "need for cognition" than agreeableness
- B. Agreeableness is a significant predictor of cognition
- C. The data is consistent with zero prediction by agreeableness
- D. Openness is a strong predictor of cognition

```
Call: lm(formula = cognition ~ agreeable + openness, data = dat.reg)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	16.46652	4.80669	3.426	0.000746
agreeable	0.08754	0.11260	0.777	0.437799
openness	1.09611	0.11333	9.672	< 2e-16

Residual standard deviation: 8.261 on 197 degrees of freedom

Multiple R-squared: 0.3872

F-statistic: 62.23 on 2 and 197 DF, p-value: < 2.2e-16

AIC	BIC
1417.17	1430.36

	Estimate	2.5 %	97.5 %
(Intercept)	16.46652060	6.9873398	25.9457014
agreeable	0.08754218	-0.1345043	0.3095887
openness	1.09610761	0.8726102	1.3196050

6. Which of the following sentences is the most correct interpretation of the following analysis?

- A. Openness is a statistically significantly stronger predictor of "need for cognition" than agreeableness
- B. Agreeableness is a significant predictor of cognition
- C. The data is consistent with zero prediction by agreeableness
- D. Openness is a strong predictor of cognition

```
Call: lm(formula = cognition ~ agreeable + openness, data = dat.reg)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	16.46652	4.80669	3.426	0.000746
agreeable	0.08754	0.11260	0.777	0.437799
openness	1.09611	0.11333	9.672	< 2e-16

Residual standard deviation: 8.261 on 197 degrees of freedom

Multiple R-squared: 0.3872

F-statistic: 62.23 on 2 and 197 DF, p-value: < 2.2e-16

AIC	BIC
1417.17	1430.36

	Estimate	2.5 %	97.5 %
(Intercept)	16.46652060	6.9873398	25.9457014
agreeable	0.08754218	-0.1345043	0.3095887
openness	1.09610761	0.8726102	1.3196050

GROUP DIFFERENCES

How can mutually-exclusive groups be distinguished from mutually-paired groups?

- a. Participants can belong to more than one group in mutually-exclusive design
- b. Participants can belong to only one group in mutually-exclusive groups
- c. Participants can belong to only one group in mutually-paired groups
- d. Participants can belong to more than one group in mutually-paired design

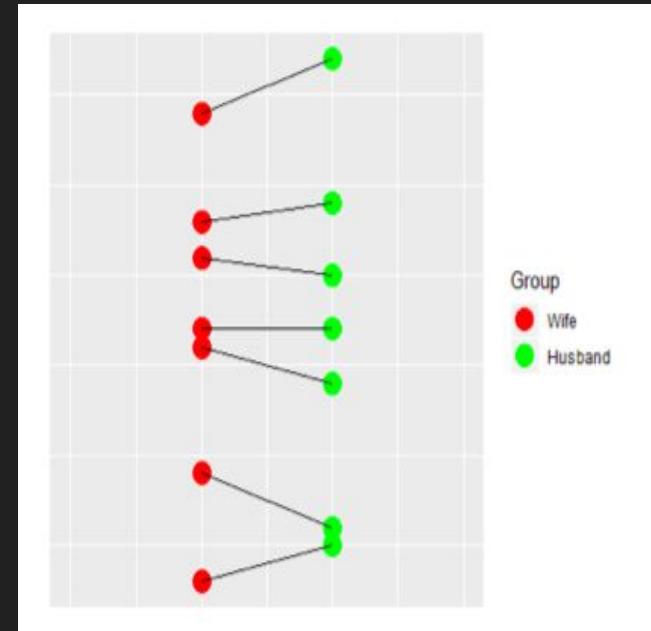
How can mutually-exclusive groups be distinguished from mutually-paired groups?

- a. Participants can belong to more than one group in mutually-exclusive design
- b. Participants can belong to only one group in mutually-exclusive groups
- c. Participants can belong to only one group in mutually-paired groups
- d. Participants can belong to more than one group in mutually-paired design



Two **mutually-exclusive** groups

- each score in one group is independent of all scores in the other group
- participants can only belong to one group
- the size of each group does not necessarily have to be the same



Two **mutually-paired** groups

- each score in one group is linked to one particular score in the other group
 - same person being measured twice
 - two people having a common dependency
- the size of each group must be the same by design

Which of the following indicates assumptions of the mean difference between two dependent groups?

- a. Normality of observed scores and homogeneity of variances
- b. Homogeneity of variances and independence of observations
- c. Independence of observations and normality of difference scores
- d. Normality of difference scores, homogeneity of variances and independence of observations

Which of the following indicates assumptions of the mean difference between two dependent groups?

- a. Normality of observed scores and homogeneity of variances
- b. Homogeneity of variances and independence of observations
- c. Independence of observations and normality of difference scores
- d. Normality of difference scores, homogeneity of variances and independence of observations

Which of the following is not an assumption of the mean difference between two independent groups?

- a. Normality of observed scores
- b. Independence of observations
- c. A balanced/ unbalanced design
- d. Group variances are the same

Which of the following is not an assumption of the mean difference between two independent groups?

- a. Normality of observed scores
- b. Independence of observations
- c. A balanced/ unbalanced design
- d. Group variances are the same

The balanced/ unbalanced design influences the robustness of confidence intervals when homogeneity of variances is violated.

Assumptions for mean differences between two INDEPENDENT GROUPS

1. Observations are independent
2. Observed scores on the construct measure are normally distributed
3. Variances in two groups are the same

Assumptions for mean differences between two DEPENDENT GROUPS

1. Observations are independent
2. Observed scores on the construct measure are normally distributed
 - Homogeneity of variance assumption is not relevant because the analysis is undertaken on the different scores

An *independent samples t-test* (n=250) was conducted, and result for homogeneity of variances was assessed. Is this result meaningful?

```
##  
##  Fligner-Killeen test of homogeneity of variances  
##  
##  data:  y by grp  
##  Fligner-Killeen:med chi-squared = 2.9496, df=1, p-value = 0.0859
```

- A. Yes. p-value is large ($p > .05$), so we fail to reject the null
- B. Yes, p-value is small ($p < .05$), so we fail to reject the null
- C. Yes, p-value is large ($p > .05$), so we reject the null
- D. No, there is not enough information

An *independent samples t-test* (n=250) was conducted, and result for homogeneity of variances was assessed. Is this result meaningful?

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data: y by grp  
## Fligner-Killeen:med chi-squared = 2.9496, df=1, p-value = 0.0859
```

- A. Yes. p-value is large ($p > .05$), so we fail to reject the null
- B. Yes, p-value is small ($p < .05$), so we fail to reject the null
- C. Yes, p-value is large ($p > .05$), so we reject the null
- D. No, there is not enough information

A dummy variable is used as an independent variable in a regression model when

- a. the variable involved is numerical
- b. the variable involved is categorical
- c. when two independent variables interact
- d. none of the above

A dummy variable is used as an independent variable in a regression model when

- a. the variable involved is numerical
- b. the variable involved is categorical
- c. when two independent variables interact
- d. none of the above

→ dummy variables are used to incorporate categorical variables into a meaningful set in a model

If a categorical independent variable contains two categories, then _____ dummy variable(s) will be needed to uniquely represent these categories

- a. 1
- b. 2
- c. 3
- d. 4

If a categorical independent variable contains two categories, then _____ dummy variable(s) will be needed to uniquely represent these categories

- a. 1
- b. 2
- c. 3
- d. 4

→ the dummy variable will contain 2 values, one for each category

OMNIBUS INVESTIGATIONS

- null hypothesis focuses on all group means are the same
- does not inform which groups differ

FOCUSED INVESTIGATIONS

- specific clear research questions
- provide identifiable differences
- include explanations for omnibus approach

One way within subject ANOVA

1. Differences between 3 or more dependent groups or one group with multiple levels. All the assumptions of two dependent group design also apply here.
2. Omnibus approach vs Focused research question

Assumption

Univariate approach: **Sphericity**

variances of **all possible difference scores between pairs of 3 or more within subject conditions being homogeneous at a population level.**

subject	T1	T2	T3	T1-T2	T1-T3	T2-T3
1	45	50	51	-5	-6	-1
2	42	42	45	0	-3	-3
3	36	41	43	-5	-7	-2
4	39	35	40	4	-1	-5
5	45	55	50	-10	-5	5
6	44	49	52	-5	-8	-3
		Variance	23.50	6.80	11.90	

Assumption

This can be checked by looking at the covariance matrix to see if it has the same variance in each diagonal and off diagonal element → Compound symmetry

Sufficient but not necessary condition

##		[,1]	[,2]	[,3]	[,4]	[,5]
##	[1,]	10	6	6	6	6
##	[2,]	6	10	6	6	6
##	[3,]	6	6	10	6	6
##	[4,]	6	6	6	10	6
##	[5,]	6	6	6	6	10

Focused research question

Contrast weights: weights of 2 comparison groups sum up to zero

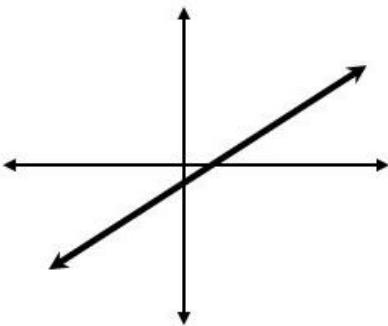
Orthogonal: difference in each contrast is unique and does not overlap.

- Contrast weights for each set sum up to zero
- Cross-products of corresponding weights from 2 linear contrasts sum up

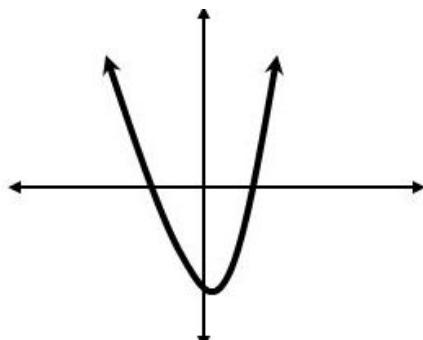
	T1	T2	T3	
Contrast 1		1	1	-2 $1 + 1 - 2 = 0$
Contrast 2		-1	1	0 $(-1) + 1 + 0 = 0$
Cross-product	$1 \times (-1) = -1$	$1 \times 1 = 1$	$(-2) \times 0 = 0$	$(-1) + 1 + 0 = 0$

The number of orthogonal contrast weights = the number of levels/cases - 1

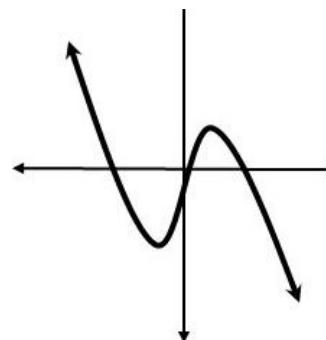
Using *polynomial orthogonal contrasts* to break down observed change over time into ***polynomial patterns of change*** over time: linear, quadratic, cubic, quartic, etc.



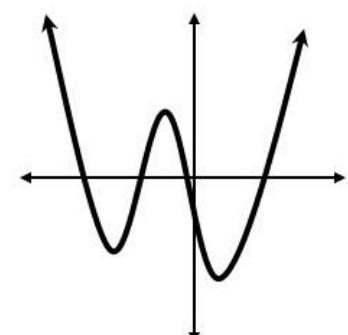
Linear



Quadratic



Cubic



Quartic

```
# Linear effect
poly.c1

##
## Response transformation matrix:
## trial.L
## baseline1 -0.6324555
## baseline2 -0.3162278
## week3      0.0000000
## week4      0.3162278
## week5      0.6324555
##
## Sum of squares and products for the hypothesis:
## trial.L
## trial.L 2016.4
##
## Sum of squares and products for error:
## trial.L
## trial.L 84.8
##
## Multivariate Tests:
## Df test stat approx F num Df den Df Pr(>F)
## Pillai      1  0.959642 190.2264      1     8 7.374e-07
## Wilks       1  0.040358 190.2264      1     8 7.374e-07
## Hotelling-Lawley 1 23.778302 190.2264      1     8 7.374e-07
## Roy         1 23.778302 190.2264      1     8 7.374e-07
```

```
# Quadratic effect
poly.c2

##
## Response transformation matrix:
## trial.Q
## baseline1  0.5345225
## baseline2 -0.2672612
## week3      -0.5345225
## week4      -0.2672612
## week5      0.5345225
##
## Sum of squares and products for the hypothesis:
## trial.Q
## trial.Q 89.1746
##
## Sum of squares and products for error:
## trial.Q
## trial.Q 59.39683
##
## Multivariate Tests:
##          Df test stat approx F num Df den Df   Pr(>F)
## Pillai      1  0.6002137 12.01069      1     8  0.008497
## Wilks       1  0.3997863 12.01069      1     8  0.008497
## Hotelling-Lawley 1  1.5013362 12.01069      1     8  0.008497
## Roy        1  1.5013362 12.01069      1     8  0.008497
```

```
# Cubic effect
poly.c3

##
## Response transformation matrix:
## trial.C
## baseline1 -3.162278e-01
## baseline2  6.324555e-01
## week3      -4.095972e-16
## week4      -6.324555e-01
## week5      3.162278e-01
##
## Sum of squares and products for the hypothesis:
## trial.C
## trial.C 256.7111
##
## Sum of squares and products for error:
## trial.C
## trial.C 38.08889
##
## Multivariate Tests:
## Df test stat approx F num Df den Df     Pr(>F)
## Pillai      1  0.870798 53.91832      1       8 8.0486e-05
```

```
# Quatric effect
poly.c4

##
## Response transformation matrix:
## trial^4
## baseline1  0.1195229
## baseline2 -0.4780914
## week3      0.7171372
## week4      -0.4780914
## week5      0.1195229
##
## Sum of squares and products for the hypothesis:
## trial^4
## trial^4 86.91429
##
## Sum of squares and products for error:
## trial^4
## trial^4 48.11429
##
## Multivariate Tests:
## Df test stat approx F num Df den Df Pr(>F)
## Pillai      1 0.6436733 14.45131      1     8 0.0052256
## Wilks       1 0.3563267 14.45131      1     8 0.0052256
## Hotelling-Lawley 1 1.8064133 14.45131      1     8 0.0052256
## Roy        1 1.8064133 14.45131      1     8 0.0052256
```

Confidence interval

The effects for quadratic, cubic and quartic is not readily understandable as the linear effects.

```
# Confidence interval for linear effect  
ci.poly.1 <- ci.lc.stdmean.ws(howell$baseline1,  
                                howell$baseline2,  
                                howell$week3,  
                                howell$week4,  
                                howell$week5,  
                                poly.c.mat[,2],  
                                level = .95 )
```

```
ci.poly.1
```

```
## $out
```

```
##
```

	MEAN CONTRASTS BTW W1-W5			
	Estimate	SE	LL	UL
## Observed Contrast Mean:	-15.7778	1.1440	-18.4158	-13.1398
## Sphericity-assumed:	-3.7263	0.2702	-4.3494	-3.1033
## Bonett delta:	-3.7263	0.8295	-5.3521	-2.1005

Two way ANOVA

Traditional ANOVA approach to 2 way design examines omnibus hypothesis.

1. *Is there a difference between schizophrenia and schizophreniform disorder in negative symptoms at remission in young people with a first psychotic disorder?*
2. *Is there a difference between non-affective psychotic disorder and bipolar disorders in negative symptoms at remission in young people with a first psychotic disorder?*
3. *Is there a difference between males and females in negative symptoms at remission in young people with a first psychotic disorder?*
4. *Do any differences between schizophrenia and schizophreniform disorder in negative symptoms at remission in young people with a first psychotic disorder vary according to sex?*
5. *Do any difference between non-affective psychotic disorder and bipolar disorders in negative symptoms at remission in young people with a first psychotic disorder vary according to sex?*

Factor A = Disorder (3 levels):

Sz: Schizophrenia

Sf: Schizopreniform

Bp: Bipolar Disorder with psychotic features

Factor B = Gender (2 levels):

Male

Female

DV = Negative symptoms

Main effect and Interaction effect

RQ 1, 2, 3 look at main effects of either sex or clinical diagnosis

RQ 4, 5 look at interaction between diagnosis and sex, above their main effects.



Cross-classified table

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97		
Females	27.17	11.17	10.33		
MarginalMeans					
MarginalEffects					

Means for each
level of *Sex*
irrespective of
Diagnosis

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97	29.30	
Females	27.17	11.17	10.33	16.22	
MarginalMeans					
MarginalEffects					

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97	29.30	
Females	27.17	11.17	10.33	16.22	
MarginalMeans	37.95	19.18	11.15		
MarginalEffects					

Means for
each level of
Diagnosis
irrespective
of Sex

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97	29.30	
Females	27.17	11.17	10.33	16.22	
MarginalMeans	37.95	19.18	11.15	22.76	
MarginalEffects					

The grand
mean for all
cases

Each marginal mean
minus the grand
mean

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97	29.30	+6.54
Females	27.17	11.17	10.33	16.22	-6.54
MarginalMeans	37.95	19.18	11.15	22.76	
MarginalEffects	+15.19	-3.58	-11.61		

Each marginal mean
minus the grand
mean

Cell mean = *grand mean*

+ *marginal effect* of Diagnosis

+ *marginal effect* of Sex

+ *interaction effect* of Sex and Diagnosis

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97	29.30	+6.54
Females	27.17	11.17	10.33	16.22	-6.54
MarginalMeans	37.95	19.18	11.15	22.76	
MarginalEffects	+15.19	-3.58	-11.61		

To obtain the *interaction effect* for each cell, we need to remove the *grand mean* and the *marginal effects of the two factors* (i.e., Diagnosis and Sex).

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	48.73	27.20	11.97	29.30	+6.54
Females	27.17	11.17	10.33	16.22	-6.54
MarginalMeans	37.95	19.18	11.15	22.76	
MarginalEffects	+15.19	-3.58	-11.61		

$$48.73 - 22.76 - 15.19 - 6.54 = +4.24$$

*Interaction effect
for the first cell*

	Sz	Sf	Bp	MarginalMeans	MarginalEffects
Males	+4.24	+1.48	-5.72	29.30	+6.54
Females	-4.24	-1.48	+5.72	16.22	-6.54
MarginalMeans	37.95	19.18	11.15	22.76	
MarginalEffects	+15.19	-3.58	-11.61		

Main Effects Focused Research Questions

RQ 1: Is there a difference between schizophrenia and schizophreniform disorder in negative symptoms?

If we only had the factor of *Diagnosis* to consider:

Sz	Sf	Bp

Main Effects Focused Research Questions

RQ 1: Is there a difference between schizophrenia and schizopreniform disorder in negative symptoms?

If we only had the factor of *Diagnosis* to consider:

Sz	Sf	Bp
+0.5	-0.5	0

But we also have the factor *Sex* to consider, so:

Sz	Sf	Bp	
+0.5	-0.5	0	Males
+0.5	-0.5	0	Females

We have 6 cells in our design. Each cell needs a weight.

```
##                                Estimate      SE       LL       UL
## Observed Mean Contrast:    18.7667 1.7794 15.2546 22.2787
## Hedges' g (Eq. Vars. Ass.): 1.9255 0.2097 1.5116 2.3394
## Bonett's d (Eq. Vars. Not Ass.): 1.9255 0.2244 1.4827 2.3683
##
## $coefs
##          Sz     Sf   Bp   Sz     Sf   Bp
## Contrast Coeff: 0.5 -0.5  0  0.5 -0.5  0
```

Main Effects Focused Research Questions

RQ 2: Is there a difference between the non-affective psychotic disorders and bipolar disorder?

If we only had the factor of *Diagnosis* to consider:

Sz	Sf	Bp
+0.25	+0.25	-0.5

Main Effects Focused Research Questions

RQ 2: Is there a difference between the non-affective psychotic disorders and bipolar disorder?

If we only had the factor of *Diagnosis* to consider:

Sz	Sf	Bp
+0.25	+0.25	-0.5

But we also have the factor *Sex* to consider, so:

Sz	Sf	Bp	
+0.25	+0.25	-0.5	Males
+0.25	+0.25	-0.5	Females

```
##                                     Estimate      SE       LL       UL
## Observed Mean Contrast:           17.4167 1.5410 14.3751 20.4582
## Hedges' g (Eq. Vars. Ass.):      1.7870 0.1849 1.4221 2.1519
## Bonett's d (Eq. Vars. Not Ass.): 1.7870 0.1804 1.4310 2.1430
##
## $coefs
##             Sz     Sf     Bp     Sz     Sf     Bp
## Contrast Coeff: 0.25 0.25 -0.5 0.25 0.25 -0.5
```

Main Effects Focused Research Questions

RQ 3: Is there a difference between males and females in negative symptoms?

If we only had the factor of Sex to consider:

+1	Males
-1	Females

Main Effects Focused Research Questions

RQ 3: Is there a difference between males and females in negative symptoms?

If we only had the factor of *Sex* to consider:

+1	Males
-1	Females

But we also have the factor *Diagnosis* to consider, so:

Sz	Sf	Bp	
+1	+1	+1	Males
-1	-1	-1	Females

```
##                               Estimate      SE       LL       UL
## Observed Mean Contrast:    13.0778 1.4529 10.2102 15.9454
## Hedges' g (Eq. Vars. Ass.): 1.3418 0.1655 1.0151 1.6685
## Bonett's d (Eq. Vars. Not Ass.): 1.3418 0.1691 1.0081 1.6755
##
## $coefs
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## Contrast Coeff: 0.3333333 0.3333333 0.3333333 -0.3333333 -0.3333333 -0.3333333
```

Interaction Focused Research Questions

RQ 4: Do differences between schizophrenia and schizophreniform disorder in negative symptoms vary according to sex?

Interaction Focused Research Questions

RQ 4: Do differences between schizophrenia and schizophreniform disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
				Males
				Females

Interaction Focused Research Questions

RQ 4: Do differences between schizophrenia and schizophreniform disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	-1	0		
				Males
				Females

Interaction Focused Research Questions

RQ 4: Do differences between schizophrenia and schizophreniform disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	-1	0		
			+1	Males
			-1	Females

Interaction Focused Research Questions

RQ 4: Do differences between schizophrenia and schizophreniform disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	-1	0		
+1 × +1	-1 × +1	0 × +1	+1	Males
+1 × -1	-1 × -1	0 × -1	-1	Females

Interaction Focused Research Questions

RQ 4: Do differences between schizophrenia and schizophreniform disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	-1	0		
+1	-1	0	+1	Males
-1	+1	0	-1	Females

```
#  
# Observed Mean Contrast:           Estimate      SE      LL      UL  
# Hedges' g (Eq. Vars. Ass.):       5.5333 3.5589 -1.4908 12.5575  
# Bonett's d (Eq. Vars. Not Ass.): 0.5677 0.3664 -0.1555 1.2909  
#  
## $coefs  
## [,1] [,2] [,3] [,4] [,5] [,6]  
# Contrast Coeff:    1   -1    0   -1    1    0
```

Interaction Focused Research Questions

RQ 5: Do differences between the non-affective psychotic disorders and bipolar disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
				Males
				Females

Interaction Focused Research Questions

RQ 5: Do differences between the non-affective psychotic disorders and bipolar disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	+1	-2		
				Males
				Females

Interaction Focused Research Questions

RQ 5: Do differences between the non-affective psychotic disorders and bipolar disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	+1	-2		
			+1	Males
			-1	Females

Interaction Focused Research Questions

RQ 5: Do differences between the non-affective psychotic disorders and bipolar disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	+1	-2		
+1 × +1	+1 × +1	-2 × +1	+1	Males
+1 × -1	+1 × -1	-2 × -1	-1	Females

Interaction Focused Research Questions

RQ 5: Do differences between the non-affective psychotic disorders and bipolar disorder in negative symptoms vary according to sex?

Sz	Sf	Bp		
+1	+1	-2		
+1	+1	-2	+1	Males
-1	-1	+2	-1	Females

```
##                                     Estimate      SE       LL       UL
## Observed Mean Contrast:           17.1667 3.0821 11.0836 23.2497
## Hedges' g (Eq. Vars. Ass.):      1.7613 0.3300 1.1100 2.4127
## Bonett's d (Eq. Vars. Not Ass.): 1.7613 0.3164 1.1369 2.3858
##
## $coefs
##          [,1] [,2] [,3] [,4] [,5] [,6]
## Contrast Coeff: 0.5  0.5   -1  -0.5 -0.5   1
##
## $test
##                               Sum of Sq  t Value     se      df p-
## value
## Contrast t Statistic: 2946.944 5.5698 3.0821 9.53158e-08
```

7. Which of the following is the most correct interpretation of the covariance matrix below?

- a. There is not much variation in the covariance matrix, so we do not have enough evidence to conclude that sphericity is not being met
- b. The variances are not the same in each diagonal and off diagonal element, so difference scores between pairs will be the heterogeneous at population level
- c. The fact that compound symmetry is not present shows that sphericity is not violated
- d. The variances are not the same in each diagonal and off diagonal element, so difference scores between pairs will be the heterogeneous at sample level

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   6.0  4.2  4.2  4.2  4.2
## [2,]   4.2  6.0  4.2  3.2  4.2
## [3,]   4.2  4.2  6.0  4.2  4.2
## [4,]   4.2  4.2  3.2  4.0  4.2
## [5,]   4.2  4.2  4.2  4.2  6.0
```

7. Which of the following is the most correct interpretation of the covariance matrix below?

- a. There is not much variation in the covariance matrix, so we do not have enough evidence to conclude that sphericity is not being met
- b. The variances are not the same in each diagonal and off diagonal element, so difference scores between pairs will be the heterogeneous at population level
- c. The fact that compound symmetry is not present shows that sphericity is not violated
- d. The variances are not the same in each diagonal and off diagonal element, so difference scores between pairs will be the heterogeneous at sample level

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   6.0  4.2  4.2  4.2  4.2
## [2,]   4.2  6.0  4.2  3.2  4.2
## [3,]   4.2  4.2  6.0  4.2  4.2
## [4,]   4.2  4.2  3.2  4.0  4.2
## [5,]   4.2  4.2  4.2  4.2  6.0
```

8. What is the best interpretation of the results below?

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai.L	1	0.959642	190.2264	1	8	7.374e-07
Pillai.Q	1	0.6002137	12.01069	1	8	0.008497
Pillai.C	1	0.870798	53.91832	1	8	8.0486e-05
Pillai^4	1	0.6436733	14.45131	1	8	0.0052256

- a. Because p is so small, the data is not consistent with non-zero linear, quadratic, cubic and quartic rates of change over time.
- b. The value of p is small enough to conclude that the data is consistent with non-zero quadratic and quartic rates of change over time.
- c. We can not make any conclusion from this result because confidence interval is not provided
- d. P values are very small, so there is enough evidence to suggest that our data is consistent with the non-zero linear, quadratic, cubic and quartic rates of change over time.

8. What is the best interpretation of the results below?

	Df	test	stat	approx F	num Df	den Df	Pr(>F)
Pillai.L	1	0.959642	190.2264		1	8	7.374e-07
Pillai.Q	1	0.6002137	12.01069		1	8	0.008497
Pillai.C	1	0.870798	53.91832		1	8	8.0486e-05
Pillai^4	1	0.6436733	14.45131		1	8	0.0052256

- a. Because p is so small, the data is not consistent with non-zero linear, quadratic, cubic and quartic rates of change over time.
- b. The value of p is small enough to conclude that the data is consistent with non-zero quadratic and quartic rates of change over time.
- c. We can not make any conclusion from this result because confidence interval is not provided
- d. P values are very small, so there is enough evidence to suggest that our data is consistent with the non-zero linear, quadratic, cubic and quartic rates of change over time.

9. What is true about main effects and interaction effects in between group subject design?

- a. Interaction effects are the combination of all main effects.
- b. Interaction effects are what's left over after all the marginal effects are accounted for
- c. While main effects are examined by using linear contrast weights, interaction effects are examined by using dummy coding.
- d. Interaction effects are what's left over after all the marginal effects and grand mean are accounted for

9. What is true about main effects and interaction effects in between group subject design?

- a. Interaction effects are the combination of all main effects.
- b. Interaction effects are what's left over after all the marginal effects are accounted for
- c. While main effects are examined by using linear contrast weights, interaction effects are examined by using dummy coding.
- d. Interaction effects are what's left over after all the marginal effects and grand mean are accounted for

10. Given that marginal means and grand means are 40 and 10 respectively, calculate marginal effect.

- a. 4
- b. -30
- c. 30
- d. 25

10. Given that marginal means and grand means are 40 and 10 respectively, calculate marginal effect.

- a. 4
- b. -30
- c. 30
- d. 25

11. Which of the following is most correct?

- a. Planned comparison weights for main effects and interaction effects should sum up to zero
- b. Planned comparison weights for main effects and interaction effects do necessarily sum up to zero
- c. Planned comparison weights for main effects and interaction effects should be orthogonal
- d. Planned comparison weights for main effects sum up to 1 while those for interaction effects should sum up to zero.

11. Which of the following is most correct?

- a. Planned comparison weights for main effects and interaction effects should sum up to zero
- b. Planned comparison weights for main effects and interaction effects do necessarily sum up to zero
- c. Planned comparison weights for main effects and interaction effects should be orthogonal
- d. Planned comparison weights for main effects sum up to 1 while those for interaction effects should sum up to zero.

12. If factor A has planned contrast (-2,+1,+1) and Factor B has planned contrast (2,-2), what are corresponding weights of contrast for interaction effect?

- a. (-4,2,-2,4,-2,-2)
- b. (4,-2,-2,4,-2,-2)
- c. (-4,2,2,4,-2,-2)
- d. (-4,2,-2,-4,2,2)

12. If factor A has planned contrast (-2,+1,+1) and Factor B has planned contrast (2,-2), what are corresponding weights of contrast for interaction effect?

- a. (-4,2,-2,4,-2,-2)
- b. (4,-2,-2,4,-2,-2)
- c. (-4,2,2,4,-2,-2)
- d. (-4,2,-2,-4,2,2)

PSYCHOLOGICAL ASSESSMENT

Which is not an assumption of Classical test theory?

- a. expected value of the error equals to 0
- b. $2Cov(\tau, \epsilon) = 0$
- c. errors associate with each other
- d. expected mean of observed scores = expected mean of true scores

Which is not an assumption of Classical test theory?

- a. expected value of the error equals to 0
- b. $2Cov(\tau, \epsilon) = 0$
 - i. errors do not correlate with true scores
- c. errors associate with each other
 - i. errors do not correlate with each other
- d. Expected mean of observed scores = expected mean of true scores

As the degree of reliability increases,

- a. error variance increases
- b. variance of observed scores increases
- c. variance of observed scores decreases
- d. none of the above

As the degree of reliability increases,

- a. error variance increases
- b. variance of observed scores increases
- c. variance of observed scores decreases
- d. none of the above

$$Reliability = \rho_{x\tau}^2 = \frac{\sigma_\tau^2}{\sigma_x^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\epsilon^2} = \frac{Signal}{Signal + Noise}$$

Where

- $\rho_{x\tau}^2$ is the theoretical reliability coefficient
- σ_x^2 is the variance of observed test results
- σ_τ^2 is true score variance
- σ_ϵ^2 is error variance
- In practice estimate $\rho_{x\tau}^2$ using the sample reliability coefficient r_{xx}

-> variance of observe test results decreases

-> error variance decreases -> increase reliability (we want to minimise the noise)

Which type of reliability is estimated by obtaining scores from two different tests on the same sample of people that measure same set of true scores and working out the correlation between them?

- a. Alternate-form reliability
- b. Split-half reliability
- c. Test-retest reliability
- d. None of the above

Which type of reliability is estimated by obtaining scores from two different tests on the same sample of people that measure same set of true scores and working out the correlation between them?

- a. Alternate-form reliability
- b. Split-half reliability
 - i. split your test into two parallel subtests & correlations
- c. Test-retest reliability
 - i. obtain scores from the same sample of people on two different administrations of the same set of test items

A source of carryover effects may take the form of

- a. participants lose interest and motivation to complete the test
- b. memory effects from previous tests on the retests
- c. test takers fail to return for the second test
- d. all of the above

A source of carryover effects may take the form of

- a. participants lose interest and motivation to complete the test
- b. memory effects from previous tests on the retests
- c. test takers fail to return for the second test
- d. all of the above

The standard error of measurement

- a. indicates the standard deviation of a distribution of observed test scores, obtained by one participant, under different administrations of the test
- b. indicates the variance of observed test scores, obtained by one participant, under different administrations of the test
- c. represents the mean differences of the sample
- d. none of the above

The standard error of measurement

- a. indicates the standard deviation of a distribution of observed test scores, obtained by one participant, under different administrations of the test
- b. indicates the variance of observed test scores, obtained by one participant, under different administrations of the test
- c. represents the mean differences of the sample
- d. none of the above

The standard error of measurement

- a. decreases when reliability decreases
- b. increases when reliability increases
- c. equals to 0 when reliability is perfect
- d. equals to 0 when the test has no reliability

The standard error of measurement

- a. decreases when reliability decreases
- b. increases when reliability increases
- c. equals to 0 when reliability is perfect
- d. equals to 0 when the test has no reliability

Formula: $SE_m = s_x \sqrt{1 - r_{xx}}$

(where s_x is the standard deviation of observed scores)

When:

- $r_{tt} = 1$, then $SE_{meas} = 0$
- $r_{tt} = .8$, then $SE_{meas} = \sigma_x \times .45$
- $r_{tt} = .6$, then $SE_{meas} = \sigma_x \times .63$
- $r_{tt} = .4$, then $SE_{meas} = \sigma_x \times .78$
- $r_{tt} = .2$, then $SE_{meas} = \sigma_x \times .89$
- $r_{tt} = 0$, then $SE_{meas} = \sigma_x \sqrt{(1 - 0)} = \sigma_x$

Reliability coefficient = 1
→ test has perfect reliability
→ indicates individual who take the test will always have the same score
→ SD of participants scores = 0
→ standard error of measurement = 0

Reliability coefficient = 0
→ test has completely no reliability
→ indicates SD of participants scores = SD of population

Given the population mean is 100 and the test's reliability is .5, what is the participant's predicted true score when the observed score was 150?

- a. 75
- b. 100
- c. 125
- d. 150

Given the population mean is 100 and the test's reliability is .5, what is the participant's predicted true score when the observed score was 150?

- a. 75
- b. 100
- c. 125
- d. 150

Predicted true scores

Participant's observed score

$$\hat{t} = r_{xx} \times x + (1 - r_{xx}) \times \mu_\tau$$

Population
mean scores

Predicted true scores

Reliability coefficient

$$PTS = (.5 \times 150) + (100 \times (1 - .5))$$

$$= 125$$

Given the standard error of measurement of IQ test is 25, one RMHI student obtains a predicted true score of 125.

Approximately, what is the 95% confidence interval for this test score?

- a. 100 - 150
- b. 75 - 150
- c. 75 - 175
- d. 100 - 175

Given the standard error of measurement of IQ test is 25, one RMHI student obtains a predicted true score of 125.

Approximately, what is the 95% confidence interval for this test score?

- a. 100 - 150
- b. 75 - 150
- c. 75 - 175
- d. 100 - 175

Standard error of estimation

Lower Bound: Predicted True Score – (1.96 × SE)

Upper Bound: Predicted True Score + (1.96 × SE)

$$\text{Lower Bound} = 125 - (1.96 \times 25) = 76$$

$$\text{Upper Bound} = 125 + (1.96 \times 25) = 174$$

95% CI [76, 174]

CRITERION VALIDITY

A test has criterion validity to the extent that it can predict scores on relevant criterion variables

CONTENT VALIDITY

A test has content validity to the extent its content reflects the full domain of the construct it is supposedly assessing

CONSTRUCT VALIDITY

A test has construct validity to the extent it reflects the construct it is meant to be reflecting.

CONCURRENT VALIDITY

PREDICTIVE VALIDITY

FACE VALIDITY

CONVERGENT VALIDITY

DISCRIMINANT VALIDITY

If the content of a test looks as though it is measuring what it is supposed to, it is said to have high

- a. construct validity
- b. face validity
- c. predictive validity
- d. discriminant validity

If the content of a test looks as though it is measuring what it is supposed to, it is said to have high

- a. construct validity
- b. face validity
- c. predictive validity
- d. discriminant validity

To assess the convergent validity of a new test of self-esteem, a researcher should establish that...

- a.it correlates highly with a different test of self-esteem
- b.it does not correlate too highly with a different test of self-esteem
- c.it correlates highly with a test of narcissism
- d.it does not correlate too highly with a test of narcissism

To assess the convergent validity of a new test of self-esteem, a researcher should establish that...

- a. it correlates highly with a different test of self-esteem
 - convergent validity indicates test scores should be correlated with tests of related constructs
- b. it does not correlate too highly with a different test of self-esteem
- c. it correlates highly with a test of narcissism
 - discriminant validity indicates test scores should not highly correlated with tests of unrelated constructs
- d. it does not correlate too highly with a test of narcissism

To test the predictive validity of a test of extraversion, a researcher could show that people's scores on the test correlate with ...

- a.their scores on another extraversion test
- b.their scores on the same extraversion test at a later date
- c.another person's ratings of their extraversion
- d.their frequency of attending parties over a two-month period

To test the predictive validity of a test of extraversion, a researcher could show that people's scores on the test correlate with ...

- a.their scores on another extraversion test
- b.their scores on the same extraversion test at a later date
- c.another person's ratings of their extraversion
- d.their frequency of attending parties over a two-month period

→ predictive validity indicates test scores evaluated against a criterion that would be measured later

→ concurrent validity indicates test scores evaluated against a criterion measured at the same time

- 1. When care has been taken to ensure that a measure is measuring all aspects of something fully, it has high..
 - a. construct validity
 - 2. When a measure correlates highly with other established measures of the same thing, it has high..
 - b. criterion validity
 - 3. When a measure is based on a full and close examination of the underlying concept, along with the related theoretical approaches, it has high..
 - c. content validity
 - d. face validity
-
- a. 1a - 2b - 3c
 - b. 1b - 2d - 3a
 - c. 1c - 2b- 3a
 - d. 1d - 2c - 3b

- 1. When care has been taken to ensure that a measure is measuring all aspects of something fully, it has high..
 - a. construct validity
 - 2. When a measure correlates highly with other established measures of the same thing, it has high..
 - b. criterion validity
 - 3. When a measure is based on a full and close examination of the underlying concept, along with the related theoretical approaches, it has high..
 - c. content validity
 - d. face validity
-
- a. 1a - 2b - 3c
 - b. 1b - 2d - 3a
 - c. 1c - 2b- 3a
 - d. 1d - 2c - 3b

- 1. When care has been taken to ensure that a measure is measuring all aspects of something fully, it has high..
 - a. construct validity
 - 2. When a measure correlates highly with other established measures of the same thing, it has high..
 - b. criterion validity
 - 3. When a measure is based on a full and close examination of the underlying concept, along with the related theoretical approaches, it has high..
 - c. content validity
 - d. face validity
-
- a. 1a - 2b - 3c
 - b. 1b - 2d - 3a
 - c. 1c - 2b- 3a
 - d. 1d - 2c - 3b

What's the sensitivity?

- A. healthy people testing positive/sick people testing positive
- B. sick people testing positive/people testing positive
- C. sick people testing positive/people with disease
- D. people testing positive/total

What's the sensitivity?

- A. healthy people testing positive/sick people testing positive
- B. sick people testing positive/people testing positive
- C. sick people testing positive/people with disease
- D. people testing positive/total

What's the specificity?

- A. healthy people testing positive/ people testing negative
- B. sick people testing positive/people testing positive
- C. sick people testing positive/sick people
- D. healthy people testing negative/healthy people

What's the specificity?

- A. healthy people testing positive/ people testing negative
- B. sick people testing positive/people testing positive
- C. sick people testing positive/sick people
- D. healthy people testing negative/healthy people

What's the Positive Predictive Power?

- A. healthy people testing positive/sick people testing positive
- B. sick people testing positive/people testing positive
- C. sick people testing positive/total
- D. people testing positive/total

What's the Positive Predictive Power?

- A. healthy people testing positive/sick people testing positive
- B. sick people testing positive/people testing positive
- C. sick people testing positive/total
- D. people testing positive/total

What's the Negative Predictive Power?

- A. sick people testing negative/ healthy people testing negative
- B. well people testing negative/ people testing negative
- C. well people testing negative/ total
- D. people testing negative/ total

What's the Negative Predictive Power?

- A. sick people testing negative/ healthy people testing negative
- B. well people testing negative/ people testing negative
- C. well people testing negative/ total
- D. people testing negative/ total

	Criterion Positive	Criterion negative	
Test Positive	a	b	$PPP = a/(a + b)$ Probability a positive test result indicates a positive case
Test Negative	c	d	$NPP = d/(c+d)$ Probability a negative test result indicates a negative case
	Sensitivity $Se = a/(a+c)$ Test's ability to correctly detect positive cases	Specificity $Sp = d/(b+d)$ Test's ability to correctly detect negative cases	
	Prevalence = $(a+c)/(a+b+c+d)$ Probability a random case from the study is criterion positive		

In attempt to estimate and evaluate validity evidence, which multitrait-multimethod correlation produces the highest reliability when research studies have several constructs and apply different methods?

- a. heterotrait-heteromethod correlations
- b. heterotrait-monomehtod correlations
- c. monotrait-heteromethod correlations
- d. monotrait-monomehtod correlations

In attempt to estimate and evaluate validity evidence, which multitrait-multimethod correlation produces the highest reliability when research studies have several constructs and apply different methods?

- a. heterotrait-heteromethod correlations
- b. heterotrait-monomehtod correlations
- c. monotrait-heteromethod correlations
- d. monotrait-monomehtod correlations

Multitrait-Multimethod Matrix

		Method 1			Method 2			Method 3			
		Traits	A ₁	B ₁	C ₁	A ₂	B ₂	C ₂	A ₃	B ₃	C ₃
Method 1	A ₁										
	B ₁										
	C ₁										
Method 2	A ₂	.57				.22			.09		
	B ₂		.22			.57			.10		
	C ₂			.11		.11			.46		
Method 3	A ₃	.56				.22			.11		
	B ₃		.23			.58			.12		
	C ₃			.11		.11			.45		

- one method measures the same trait → high correlations
→ convergent validity

different methods measure different constructs → weakest

Afterword

I'd also like to have you all think about these tools a bit in your own time.

Research tools are limited to the extent of the researcher's goals and ability to communicate them.

Believe me or not, I feel that understanding these tools are the easy part- communicating them and describing what they mean in practical terms take even more practice! And I'd argue that proper communication and understanding of these tools are immensely important in any piece of research.

Afterword

We've gone through a **LOT** in a short amount of time.

Don't let anyone tell you otherwise, these concepts take time and practice to understand.

Thank you everyone for listening and give yourselves a pat on the back!

Resources

You bet I'll list a bunch of things here that I find awesome:

- a) Learning Statistics with R. I find Danielle Navarro's communication style an absolute treat to read- 100% worth your time!
- b) Andy Field. He's got a bunch of titles for SPSS/SAS/R and again, he's got a wonderful communication style.
- c) Anything in R: <https://bookdown.org/>
- d) Thank Amy Perfors for this one-
<https://psychologicalsciences.unimelb.edu.au/research/hubs/chdh/ccs/sources-and-readings>
- e) I've got a bit more varied stuff, come have a chat with me!

We'd very much appreciate your feedback!

We'd really appreciate it if you could fill out this google form for feedback regarding the session we've just held.

<https://bit.ly/2IHVd4Z>

These slides will be available at

<http://github.com/jjanuar/rmhipals2019>