Week 10 & 11

These lectures really belong together. They provide a complete overview of when it is and is not applicable to use regression and ANOVA, and what to look out for, as well as discussing how to do research properly (and badly)

1) What can be used to check normality for linear regression?

A) PP-plot for the dependent variable
B) QQ-plot for the independent variable
C) Histogram of the residuals
D) Residual plot

2) What can be used to check linearity for linear regression?

A) QQ-plot of the residuals
B) Scatterplot of predicted values and residuals
C) Histogram of the dependent variable
D) Boxplot of the residuals

3) What can be used to check homoscedasticity for linear regression?

a) Residual plot
b) PP-plot of the dependent variable
c) Histogram of independent variable
d) Boxplot of independent variable

4) What can be used to check normality for ANOVA?

A) PP-plot of residuals
B) Boxplot of the dependent variable
C) QQ-plot of the dependent variable
D) Checking the skewness and kurtosis of the independent variable

5) What can be used to check linearity for ANOVA?

A) Residual plot
B) Separate boxplot for each group on the dependent variable
C) QQ-plot of the residuals
D) There is no assumption of linearity for ANOVA

6) What can be used to check homoscedasticity for ANOVA?

A) PP-plot of the dependent variable
B) Residual plot
C) Separate boxplot for each group on the dependent variable
D) There is no assumption of homoscedasticity for ANOVA

7) Which of the following is a bad research practice?

A) Determining required sample size beforehand
B) Failing to reject the null hypothesis due to a small sample size
C) Rejecting the null hypothesis due to a huge sample size
D) Continuously expanding the sample until the effect becomes statistically significant

8) Which of the following is a good research practice?

A) Continuously testing until you find a statistically significant result
B) Transforming data to meet assumptions of your chosen statistical model
C) Removing outliers, because they do not fit your hypothesis
D) Using incorrectly measured data to keep your sample size as high as possible

9) Which of the following statements is true?

A) All influential points are outliers
B) All outliers are influential points
C) If we find a value with a cook's distance of 3, this is not an influential point
D) If we find a value with a residual of 10, this is definitely an outlier

10) Which of the following statements is true?

A) Multicollinearity only happens in linear regression
B) Multicollinearity (R^2J) can never decrease when adding more predictors
C) Multicollinearity is never a problem, unless our number of predictors is larger than our sample size
D) Multicollinearity is not a problem if we find a VIF of 10

11) How can linearity be fixed?

A) Increasing sample size
B) Removing multicollinearity
C) Increasing homoscedasticity
D) Transforming independent variable

12) The VIF of a predictor is 2.8. What is the proportion explained variance that other predictors can explain (R^2J)?

13) You have computed a multiple linear regression with two predictors. The correlation between these predictors = 0.83. What are the Tolerance and VIF for these predictors?

Week 12 (Moderator & Mediator)

While the topic of interaction effect is broad and can be expanded upon a lot to threeway interactions and mixes between categorical and continuous variables, the focus is on interactions with two continuous variables, being able to calculate simple regression equations, and interpretation. To help gain more meaningful slopes, we may center the variables beforehand. Centering is not crucial for calculating simple regression equations.

1) Explain how centering helps with the interpretation in a model with interaction effect. Does it also help in a model without interaction effect?

Consider we find the following uncentered model:

$\hat{y} = -10 + 1X - 3Z + 2XZ$

2) What does the slope of X (1) tell us?

3) What does the slope of the interaction effect (2) tell us?

4) What is the simple regression equation for Z when we fill in X = 2?

5) The SE of the slope calculated in the previous question is 2, and the n = 30. Is this slope significant at a two-sided $\alpha = 0.05$?

6) What is the simple regression equation for X when we fill in Z = 4?

Consider we find the following centered model:

$\hat{y} = 2 - 0.5x + 2z - 1xz$

The standard deviation of x = 2, and the standard deviation of z = 3

7) What does the slope of z (2) tell us?

8) What is the simple regression equation of z for high levels of x?

9) What is the simple regression equation of x for low levels of z?

10) The SE of the slope calculated in question 9 is 1. N = 40. What is the 95% confidence interval for this slope?

11) Which of the following four slopes should <u>not</u> be significant if there is a mediator effect?

a) Slope of simple linear regression between IV and Mediator
b) Slope of simple linear regression between IV and DV
c) Slope of IV in the multiple linear regression that uses IV and Mediator to predict DV
d) Slope of Mediator in the multiple linear regression that uses IV and Mediator to predict DV

12) Which of the following statements is true?

a) The mediator variable has to always be continuous
b) A mediator effect can only be tested with regression
c) Both A and B are true
d) Both A and B are false

Week 13 (Code variables)

This lecture is all about the mastery of the following six steps when it comes to code variables

1) Determine the coding and model
2) Fill in the model for each group to determine group means
3) Change it from μ = … to ß = … through algebra
4) Calculate the regression coefficients
5) Compute a t-test or F-test to check for significance
6) Interpret regression coefficients

Let us go through this step-by-step for both a situation with one code variable and a situation with multiple code variables, using formal dummy coding. After that, a few more different coding schemes that you could practice the third step with.

1) There is one code variable coded as (Men = 0, Women = 1). Write down the model

2) Fill in the model you have created in the first step for both men and women, by filling in the appropriate code for all code variables. This step does not necessarily require numbers.

3) Find how you could calculate all ß's by only using a combination of μ's

4) μ of Men = 2, μ of Women = 5. Use the formulas you found in 3 to calculate the regression coefficients

5) The standard error of the slope is 0.75. There are 10 men and 10 women in the current sample. Calculate the t-value of the slope, and the F-value of the ANOVA table.

6) You should have found that the slope is 3, and the intercept is 2. What do these values mean in this context?

Next, rather than looking at gender, we could look at type of study. Students from Psychology, Business, Medicine, and Law are compared on general social intelligence.

7) The coding is such that Psychology has 1 on the first code variable (or dummy), Medicine has 1 on the second code variable (or dummy), and Business has 1 on the third code variable (or dummy). All other codes are 0. Write down the model

8) Fill in the model you have created in the first step for both men and women, by filling in the appropriate code for all code variables. This step does not necessarily require numbers.

9) Find how you could calculate all ß's by only using a combination of μ's

10) $\mu_{Psychology} = 25$, $\mu_{Medicine} = 23$, $\mu_{Business} = 22$, $\mu_{Law} = 22$. Use the formulas you found in the previous question to calculate the regression coefficients

11) The standard error of the slope that belongs with the first code variable (generally called "b1") is 2. The total sample is n = 40. Calculate the 95% confidence interval for this code variable

12) Interpret all four regression coefficients

Now, a few questions with more complicated coding schemes, which you can practice the algebraic step with. The goal is always the same. Codes are given, and the only goal is to define ß by one or more means.

13)

	Code 1	**Code 2**	**Code 3**
Group 1	1	0	0
Group 2	0	1	0
Group 3	0	0	1
Group 4	0	0	-1

14)

	Code 1	**Code 2**	**Code 3**
Group 1	1	0	0
Group 2	0	1	0
Group 3	0	0	1
Group 4	-1	-1	-1

15)

	Code 1	Code 2	Code 3
Group 1	1	0	0
Group 2	1	2	0
Group 3	1	2	3
Group 4	0	0	0

16)

	Code 1	Code 2	Code 3
Group 1	1	2	2
Group 2	2	1	2
Group 3	2	2	1
Group 4	2	2	2