## Week 10 & 11 - Answers

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?**
2. PP-plot for the dependent variable
3. QQ-plot for the independent variable
4. **Histogram of the residuals**
5. Residual plot
6. **What can be used to check linearity for linear regression?**
7. QQ-plot of the residuals
8. **Scatterplot of predicted values and residuals**
9. Histogram of the dependent variable
10. Boxplot of the residuals
11. **What can be used to check homoscedasticity for linear regression?**
12. **Residual plot**
13. PP-plot of the dependent variable
14. Histogram of independent variable
15. Boxplot of independent variable
16. **What can be used to check normality for ANOVA?**
17. **PP-plot of residuals**
18. Boxplot of the dependent variable
19. QQ-plot of the dependent variable
20. Checking the skewness and kurtosis of the independent variable
21. **What can be used to check linearity for ANOVA?**
22. Residual plot
23. Separate boxplot for each group on the dependent variable
24. QQ-plot of the residuals
25. **There is no assumption of linearity for ANOVA**
26. **What can be used to check homoscedasticity for ANOVA?**
27. PP-plot of the dependent variable
28. Residual plot
29. **Separate boxplot for each group on the dependent variable**
30. There is no assumption of homoscedasticity for ANOVA
31. **Which of the following is a bad research practice?**
32. Determining required sample size beforehand
33. Failing to reject the null hypothesis due to a small sample size
34. Rejecting the null hypothesis due to a huge sample size
35. **Continuously expanding the sample until the effect becomes statistically significant**
36. **Which of the following is a good research practice?**
37. Continuously testing until you find a statistically significant result
38. **Transforming data to meet assumptions of your chosen statistical model**
39. Removing outliers, because they do not fit your hypothesis
40. Using incorrectly measured data to keep your sample size as high as possible
41. **Which of the following statements is true?**
42. **All influential points are outliers**
43. All outliers are influential points
44. If we find a value with a cook’s distance of 3, this is not an influential point
45. If we find a value with a residual of 10, this is definitely an outlier
46. **Which of the following statements is true?**
47. Multicollinearity only happens in linear regression
48. **Multicollinearity (R²J) can never decrease when adding more predictors**
49. Multicollinearity is never a problem, unless our number of predictors is larger than our sample size
50. Multicollinearity is not a problem if we find a VIF of 10
51. **How can linearity be fixed?**
52. Increasing sample size
53. Removing multicollinearity
54. Increasing homoscedasticity
55. **Transforming independent variable**
56. The VIF of a predictor is 2.8. What is the proportion explained variance that other predictors can explain (R²J)?

R²J = 1 – (1 / VIF) = 1 – (1 / 2.8) = 0.643

1. 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?

R²J = 0.83² = 0.689  
Tolerance = 1 – 0.689 = 0.311  
VIF = 1 / Tolerance = 1 / 0.311 = 3.215

## Week 12 (Moderator & Mediator) - Answers

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?

In a model with interaction effect, the interpretation of the slopes of the main effect changes to specifically provide the slope at the level of “other predictor = 0”. Through centering, we make sure that this 0 is useful, so that the slope we get from SPSS is a useful slope to mention for our sample. In models without an interaction effect, it only changes the intercept, so it would generally not be considered useful.

Consider we find the following uncentered model:

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

When X increases by 1, Y increases by 1, given that Z = 0

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

When the part of XZ that is independent from X and Z increases by 1, Y increases by 2

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

-10 + 1 \* 2 – 3 \* Z + 2 \* 2 \* Z = -10 + 2 – 3Z + 4Z = -8 + 1Z

1. The SE of the slope calculated in question 4 is 2, and the n = 30. Is this slope significant at a two-sided α = 0.05?

t = 1 / 2 = 0.5  
DF = DFE = n – p – 1 = 30 – 3 – 1 = 26  
One-sided p > 0.25  
Two-sided p > 0.50, not significant

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

-10 + 1 \* X – 3 \* 4 + 2 \* X \* 4 = -10 – 12 + 1X + 8X = -22 + 9X

Consider we find the following centered model:

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

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

When Z increases by 1, Y increases by 2, given that X = 0 (or: given mean level of X)

1. What is the simple regression equation of z for high levels of x?

High level of x = µ + SD = 0 + 2 = 2, so we fill in x = 2

2 – 0.5 \* 2 + 2 \* z – 1 \* 2 \* z = 2 – 1 + 2z – 2z = 1

1. What is the simple regression equation of x for low levels of z?

Low level of z = µ - SD = 0 - 3 = -3, so we fill in z = -3

2 – 0.5 \* X + 2 \* (-3) – 1 \* X \* (-3) = 2 – 6 – 0.5X + 3X = –4 + 2.5X

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

DF = DFE = n – p – 1 = 40 – 3 – 1 = 36. Round down to first available DF = 30  
95% CI 🡪 t(30)\* = 2.042  
Margin of error = 2.042 \* 1 = 2.042  
95% CI: [2.5 – 2.042; 2.5 + 2.042] 🡪 [0.458; 4.542]

1. **Which of the following four slopes should not be significant if there is a mediator effect?**
2. Slope of simple linear regression between IV and Mediator
3. Slope of simple linear regression between IV and DV
4. **Slope of IV in the multiple linear regression that uses IV and Mediator to predict DV**
5. Slope of Mediator in the multiple linear regression that uses IV and Mediator to predict DV
6. **Which of the following statements is true?**
7. The mediator variable has to always be continuous
8. A mediator effect can only be tested with regression
9. Both A and B are true
10. **Both A and B are false**

## Week 13 (Code variables) - Answers

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.

t = 3 / 0.75 = 4  
DF = DFE = n – p – 1 = 20 – 1 – 1 = 18  
F = 4² = 16  
(It wasn’t asked, but it would be significant at two-sided α = 0.05)

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

Intercept = Mean of all men  
Slope = Difference between mean of men and women

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.

1. 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
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. µPsychology = 25, µMedicine = 23, µBusiness = 22, µLaw = 22. Use the formulas you found in the previous question to calculate the regression coefficients
5. 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

DF = DFE = n – p – 1 = 40 – 3 – 1 = 36. Round down to first available DF = 30  
95% CI 🡪 t(30)\* = 2.042  
Margin of error = 2.042 \* 2 = 4.084  
95% CI: [3 – 4.084; 3 + 4.084] 🡪 [-1.084; 7.084]

1. 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |