

## Chapter 12: How can I tell if one variable can predict another? Linear regression.

### Full answers to study questions

1. You could potentially argue that either variable should be the predictor or outcome, but these make the most sense to me:
  - 1.1. Parent's IQ is the predictor variable and child's IQ is the outcome.
  - 1.2. Number of siblings is the predictor variable and extraversion is the outcome.
  - 1.3. Months since acquiring brain injury is the predictor variable and performance on a language task is the outcome variable.
2. The unexplained variance is calculated as:  $1 - R^2$ . Importantly, when you look at these statistics it can look like there is a lot of unexplained variance. If you look at the significance of the original correlation  $r$  value, this will tell you whether the amount of variance explained by the model is "interesting" or not. If it is significant, then even if there is less explained variance than unexplained variance, the result is statistically interesting!
  - 2.1.  $R^2 = .029$ . There is 2.9% explained variance and 97.1% unexplained variance in the model.
  - 2.2.  $R^2 = .289$ . There is 28.9% explained variance and 71.1% unexplained variance in the model.
  - 2.3.  $R^2 = .552$ . There is 55.2% explained variance and 44.8% unexplained variance in the model.
3. The intercept value is negative, which makes interpretation a little trickier!
  - 3.1. The intercept value of -1.6 shows that if a participant gets a confidence score of 0 you would predict that their emotive rating would be -1.6. This is clearly impossible as the scale runs from 1 to 5! This has happened because the lowest score on the confidence questionnaire is not 0. It is always good to remember the minimum and maximum scores on a scale when calculating predicted values, as the analyses don't know what a possible number is and what is impossible!
  - 3.2. For each one point increase on the confidence questionnaire, positivity ratings increase by 0.23.
  - 3.3. Both of the predicted scores fall outside of the range of the scale for the positive emotion rating. This can sometimes happen in a regression analysis.
    - 3.3.1. Lowest possible score is 10: Predicted reaction time is 0.7
    - 3.3.2. Highest possible score is 50: Predicted reaction time is 9.9

### Full answers for additional dataset

1. Explain how you will analyse this dataset and why this method of analysis is appropriate?

Simpler linear regression will be used, with the number of Twitter posts per week being used to predict the outcome variable of narcissism.

2. Suggest a suitable hypothesis for this analysis.

There will be a significant positive relationship between the two variables whereby participants who post more frequently on Twitter will have higher scores on the narcissism measure.

3. Calculate the  $R^2$  (you may need to first look back at Chapter 11 to calculate Pearson's  $r$ ).

$$r = 0.969$$

$$R^2 = 0.969^2$$

$$R^2 = 0.93896$$

Conversion to percentage of variance explained:  $0.93896 * 100 = 93.8961$

4. Calculate the  $\beta_0$  and  $\beta_1$  values.

	Number of Twitter posts (x)	$x - \bar{x}$	Narcissism (y)	$y - \bar{y}$	$(x - \bar{x}) * (y - \bar{y})$	$(x - \bar{x})^2$
	15	4.9	10	3.0	14.7	24.0
	2	-8.1	3	-4.0	32.4	65.6
	24	13.9	12	5.0	69.5	193.2
	8	-2.1	6	-1.0	2.1	4.4
	10	-0.1	7	0.0	0.0	0.0
	16	5.9	9	2.0	11.8	34.8
	3	-7.1	5	-2.0	14.2	50.4
	7	-3.1	6	-1.0	3.1	9.6
	5	-5.1	4	-3.0	15.3	26.0
	11	0.9	8	1.0	0.9	0.8
Sum ( $\Sigma$ )	101		70		164.0	408.9
Mean	10.1		7			

Steps for calculating the slope ( $\beta_1$ ):

1: Calculate the Sum of Products (SP), which is  $\Sigma(x - \bar{x}) * (y - \bar{y})$  from the table above.

$$SP = 164.0$$

2: Calculate the Sums of Squares for the predictor (SSx), which is  $\Sigma(x - \bar{x})^2$  from the table.

$$SSx = 408.9$$

3: Calculate  $\beta_1 = SP/SSx$

$$\beta_1 = 164.0/408.9$$

$$\beta_1 = 0.4011$$

Steps for calculating the intercept ( $\beta_0$ ):

1: The values you need are the two mean values, and the slope.

2: Calculate  $\beta_0 = \bar{y} - (\beta_1 * \bar{x})$

$$\beta_0 = 7 - (0.4011 * 10.1)$$

$$\beta_0 = 7 - 4.05111$$

$$\beta_0 = 2.94889$$

5. Interpret and write up your findings using APA standards.

There is a significant positive correlation between the two variables ( $r(8) = .97, p < .001$ ) with the number of Twitter posts per week explaining about 93.9% of the variability in narcissism. The  $\beta_0$  intercept value is 2.95 and the  $\beta_1$  slope value is .40, showing that each extra Twitter post per week is associated with a 0.40 increase in narcissism.

6. Calculate the predicted narcissism scores for the following imaginary participants:

$$Y = 2.95 + (0.40 * x)$$

Number of Twitter posts: predictor value	Completed equation	Predicted job satisfaction: outcome value
6	$Y = 2.95 + (0.40 * 6)$ $Y = 2.95 + (2.4)$ $Y = 5.35$	5.35
0	$Y = 2.95 + (0.40 * 0)$ $Y = 2.95 + (0)$ $Y = 2.95$	2.95
14	$Y = 2.95 + (0.40 * 14)$ $Y = 2.95 + (5.6)$ $Y = 8.55$	8.55

### Are p values enough? An optional exercise

$$f^2 = \frac{0.93896}{1 - 0.93896}$$

$$f^2 = \frac{0.93896}{0.06104}$$

$$f^2 = 15.383$$

$f^2 = 15.383$ , indicating a large effect size.