Wednesday AM

- Presentation of yesterday's results
- Associations
- Correlation
- Linear regression
- Applications: reliability

Associations

- We're often interested in the association between two variables, especially two interval scales.
- Associations are measured by their:
  - direction (positive, negative, u-shaped, etc.)
  - magnitude (how well can you predict one variable by knowing the score on the other?)

Correlation

- The (Pearson) correlation (r) between two variables is the most common measure of association
  - Varies from -1 to 1
  - Sign represents direction
  - r^2 is the proportion of variance in common between the two variables (how much one can account for in the other)
  - Relationship is assumed to be *linear*

Correlation in SPSS

- Analyze…Correlate…Bivariate
- Enter variables to be correlated with one other.

		Q1	Q2	Q3
Q1	Pearson Correlation	1.000	.105	.109
	Sig. (2-tailed)	.	.111	.099
	N	233	233	231
Q2	Pearson Correlation	.105	1.000	.616
	Sig. (2-tailed)	.111	.	.000
	N	233	234	232
Q3	Pearson Correlation	.109	.616	1.000
	Sig. (2-tailed)	.099	.000	.
	N	231	232	232

- There was a significant positive correlation between Q2 and Q3 ($r = 0.62$, $p < .05$).

Linear regression

- Correlation is a measure of association based on a linear fit.
- Linear regression provides the equation for the line itself (e.g. $Y = b_1X + b_0$)
- That is, in addition to providing a correlation, it tells how much change in the independent variable is produced by a given change in the dependent variable...
- ... in both natural units and standardized units.

Linear regression in SPSS

- Analyze…Regression…Linear
- Enter dependent and independent variables
- Three parts to output:
  - Model summary: how well did the line fit?
  - ANOVA table: did the line fit better than a null model?
  - Regression equation: what is the line? How much change in the dependent variable do you get from a 1 unit (or 1 standard deviation) change in the independent variable

Linear regression output

- Predicting Q2 from Q3:

Model Summary

R	R Square	Adjusted R Square
.616	.380	.377

- R is the correlation
- R^2, the squared correlation, is proportion of variance in Q2 accounted for by variance in Q3
- Adjusted R^2 is a less optimistic estimate

Linear regression output

ANOVA

	Sum of Sq	df	Mean Square	F	Sig.
Regression	153.924	1	153.924	140.8	.000
Residual	251.455	230	1.093		
Total	405.379	231			

- Shows that the regression equation accounts for a significant amount of the variance in the dependent variable compared to a null model.
- (A null model is a model that says that the mean of Q2 is the predicted Q2 for all subjects).

Linear regression ouput

Coefficients

	Unstandardized		Standardized		
	B	Std. Error	Beta	t	Sig
(Constant)	.804	.315		2.554	.011
Q3	.693	.058	.616	11.866	.000

- Unstandardized coefficients (B) give the actual equation:

  Q2 = 0.693 * Q3 + 0.804

  - These are raw units. An increase of 1 point in Q3 increases Q2 by 0.693 points on average. People who have Q3 = 0 have Q2 = 0.804 on average, etc.
  - Because SE of B is estimated, we can perform t-tests to see if a B is significantly different than 0 (has a significant effect).
- Standardized coefficients (β) give the amount of change in Q2 caused by a change in Q3, measured in standard deviation units. They are useful in multiple regression (later)...

Measuring reliability of a scale

- Test-retest reliability is usually measured as the correlation between tests (ranks of subjects stay the same at each testing)
- Cronbach's α is another common internal reliability measure based on the average inter-item correlation of items in a scale.

Cronbach's α in SPSS

- Analyze…Scale...Reliability analysis
- Enter item variables that make up the scale
- Go to Statistics dialog box and ask for *scale* and *scale if item deleted* descriptives.

Cronbach's α in SPSS

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Q1	21.2913	9.2466	.3133	.6071
Q2	23.4000	6.0576	.4507	.5325
Q3	22.5826	6.4975	.4798	.5096
Q4	21.9043	8.5148	.3565	.5840
Q5	22.2130	7.4173	.3448	.5870

Reliability Coefficients

N of Cases = 230.0 **N of Items** = 5

Alpha = .6229 **Standardized item alpha** = .6367

Wednesday AM assignment

- Using the clerksp data set:
  - Examine the correlations between items 1-17 (self-ratings of different clerkship skills). What do you notice about the correlation matrix?
  - Select any one of those 17 items. Run a linear regression to determine if the pre-clerkship rating on that item predicts the post-clerkship rating.
  - Assume that we want to combine post-clerkship items 1-17 into a single scale of self-related clerk skill. What would the reliability of this scale be?