Week 11 lecture notes - PSYC 3435

April 3-7, 2017

This week, we'll talk about correlational designs

Definitions

- A *correlational study* is a type of research design that examines the relationships between multiple variables
  - Note: there is no manipulation involved, so this is a *nonexperimental* design
- Two types of questions involved:
  1. descriptive questions: is there a relationship between different behaviors?
  2. predictive questions: can one behavior be predicted from another behavior?

To motivate our discussion, let's take a look at a dataset (albumSales.csv) that shows several variables related to album sales for 200 different metal bands

Column definitions:

- Adverts: advertising budget (in $1000 units)
- Sales: total album sales (in $1000 units)
- Airplay: number of radio plays per week
- Attract: a 1-10 rating of the band's attractiveness

Testing descriptive hypotheses

Example (from Lab 1): is album sales related to advertising budget?

- the most common method for testing this relationship is a *Pearson correlation* test
- Null: the correlation between sales and advertising budget is 0 ($r = 0$)
- Alternative: the correlation between sales and advertising budget is not 0 ($r \neq 0$)
- Results from JASP: $r = 0.578$, $p < 0.001$
- since $p < 0.05$, we reject the null, from which we conclude that the correlation between sales and budget is nonzero.

Testing predictive hypotheses

Single predictor

Predictive hypotheses are composed of a hypothesized relationship between two classes of variables:

- outcome variables
- predictor variables

Typical model of predictive hypotheses is the *linear regression model*: $Y = b_0 + b_1 X$

- b_0: intercept
- b_1: slope (amount of change in Y that is due to a change in X)

Example: does advertising budget *predict* album sales?

- look at albumSales.csv in JASP
  - Sales = total album sales (in thousands of pounds)
  - Adverts = advertising budget (in thousands of pounds)
- build linear model: Sales $= b_0 + b_1 \cdot$ Adverts
- JASP output:

  - $R = 0.578$: this is the correlation between sales and advertising budget
  - $R^2 = 0.335$: proportion of variation in album sales that is explained by advertising budget
  - $b_0 = 167.68$
  - $b_1 = 0.096$
- this results in the following model:
  - $\text{Sales} = 167.68 + 0.096 \cdot \text{Adverts}$
- Meaning?
  1. we can predict sales from adverts: suppose advertising budget = 100,000 dollars. This means Adverts=100. Then *predicted* sales $= 167.68 + 0.096 \cdot 100 = 177.28$. Thus, we would predict total album sales of 177,280
  2. we can describe the *effect* of each predictor:
     - as advertising budget increases by one unit, album sales increases by 0.096 units.
     - since unit = 1000, each \$1000 increase in budget increases album sales by $0.096 \cdot 1000 = 96$ albums

Multiple predictors

There is no reason that we have to be limited to only one variable. We can have multiple predictor variables.

Example: can we predict album sales from advert budget, airplay, and attractiveness?

- Linear model: $\text{Sales} = b_0 + b_1 \cdot \text{Adverts} + b_2 \cdot \text{Airplay} + b_3 \cdot \text{Attract}$
- JASP output:
  - $R^2 = 0.665$ - these three predictors explain 2/3 of the variance in album sales
  - b_0 = -33.27
  - b_1 = 0.085 (as advertising budget increases by one unit, album sales increase by 0.085 units)
  - b_2 = 4.209 (as number of radio plays increases by one, album sales increases by 4.209 units)

- $b_3 = 13.86$ (as attractiveness increases by one unit, album sales increases by 13.86 units)