Bivariate analysis

vanderbi.lt/r

Steve Baskauf

Before we start…

- need to install MVN library

- test by:

```
library(MVN)
```

- install by:

```
install.packages("MVN")
```

- or use the package installer on the Packages tab of the lower right pane of RStudio

Categories of analysis for today:

- two continuous variables: correlation
- one continuous, one discontinuous variable: t-test of means
- two discontinuous variables: chi-squared contingency test (a.k.a. *test of association*). I think this is also known as "*cross tabulation*". NOT to be confused with *chi-squared goodness of fit test*.

Recap maternal_closeness

```
mutate(mc_H1PF1 = ifelse(!H1PF1==1 &
!is.na(H1PF1), 0, ifelse(H1PF1==1, 1,
NA)))
```

- **!H1PF1==1 & !is.na(H1PF1)** not a 1 and not a NA… : 0 if true, otherwise…
- H1PF1==1 if a 1: 1 if true, otherwise NA

- Final test requires all five responses to be 1, otherwise none to be NA

What is P ?

- What is P?
  - P is the probability that we would see results like this if nothing interesting were going on (variation is random).
  - P = 0.6 (could be like this 60% of the time if random; likely to be random)
  - P = 0.001 (could be like this 0.1% of the time if random; not likely to be random)
- If it's really unlikely that our results would occur when only random things are happening, we think something interesting is going on.

Why do we like it when P<0.05 ?

- Hypotheses:
  - things are different (alternate hypothesis)
  - things are the same (null hypothesis)
- Strategy:
  - show that the null hypothesis is wrong
- If P < 0.05, then we assume the null hypothesis is wrong because it's so unlikely.
- If P > 0.05, then either the null hypothesis is wrong or our experiment SUCKS !
- We probably know what's going on if P < 0.05 but not if P > 0.05

Statistical power

- Power is the ability to show that different things are different ($P < 0.05$)
- We get more statistical power if there's less variation in the data or a larger sample size.
- We may be able to control variation by experimental conditions
- We should be able to increase the sample size (if we have time and money).
- If not different, increasing power won't reduce P.
- If different, increasing power will make P get smaller.

Power tradeoff

- Too little statistical power:
  - can't show that different things are different
  - Unable to get P > 0.05 when there are differences.

- It seems like more statistical power would always be a good thing, but…
- Too much statistical power
  - tiny unimportant things are shown to be different
  - P< 0.05 for factors with a very small effect.