

Inference for Categorical Variables

General Tips

if you can, pick a continuous outcome over a binary outcome

Why? For a binary outcome, you'll need a much larger sample size. Continuous outcomes also allow more precision.

logistic regression = linear model for the log-odds of the outcome

analyzing relationship between categorical outcome and a continuous covariate

effect modification vs confounding

if we don't take effect modification into account, we get an over-generalized estimate of the relationship between the outcome and the exposure for the entire co-hort

- Breslow-Day Test examines if evidence of a differential association between two variables across the level of a third variable
 - similar limitations to Cochran-Mantel-Haenszel test

Cochran-Mantel_haenszel test

- limitations
 - can only adjust for one variable at a time

looks at two binary categorical variables while adjusting for the value of a third categorical variable

Parametric One-Sample Inference of Categorical Variables

- one-sample proportion test
 - do NOT use Yate's continuity, so specify:
 - `prop.test(..., correct = FALSE)`
- χ^2 goodness of fit test
 - to ensure sufficient sample size: $n \cdot p_{0j} > 5$
 - don't use continuity corrections!
 - `chisq.test(..., correct = FALSE)`

NOTE: *one-sample single proportion test* gives a 95% CI – χ^2 does not!

Types of Probabilities

Joint, Marginal and Conditional Probabilities

- QI

- [SQUIRE 2.0 for QI Reporting](#)
- ☐ Stepped-wedge trial
 - [Link](#)
- Linear regression
 - [Q-Q plot](#)
 - plot of residuals
 - [Cook's Distance](#)
 - these are different!
 - correlative
 - descriptive
 - predictive
 - associative
- confounding vs. effect modification

to assess a paired difference

- create histogram
- plot as box plot
- make [Q-Q plot](#)

From:

<https://ewrobbins.com/> - ewrobbins.com

Permanent link:

https://ewrobbins.com/doku.php?id=duke_notes&rev=1763415978

Last update: **2025/11/17 21:46**

