Hypothesis tests

STA9750

Spring 2019

Conceptual setup

Basic setup

- Data come from a probability model which has some unknown characteristics (parameters)
- We usually make some assumptions about the data-generating mechanism (DGM)
  - Example: The data are normal with unknown mean and variance
- Our goal is learning about an unknown feature of the DGM (a parameter), given the data

A way of doing hypothesis tests

Start with a null hypothesis H_0 for the DGM, which you don't want to reject *unless you have enough evidence to reject it,* and an alternative hypothesis H_1

Desired properties

- If H_0 is true, we want to falsely reject it at most 100 α % of the time (you decide what α is before you do the analysis)
- If H_1 is true, we want the probability of rejecting H_0 to be as high as possible

Implementation: *p*-values

- If the *p*-value is less than α, reject H_0; otherwise, don't reject H_0
- If the *p*-value is less than α, we say that the result is "statistically significant" (there is significant evidence against H_0)

Some terminology

Truth

	H_0 is true	H_1 is true
Do not reject H_0	Correct (True -)	Type 2 error (False -)
Reject H_0	Type 1 error (False +)	Correct (True +)

Decision

Tests we saw last time

Tests for one group

- Testing proportions (*z*-test)
  - A pharmaceutical wants to market a new drug. They'd like to argue that their drug has a recovery rate of at least, say, 50%
  - Null hypothesis: recovery rate less than or equal to 50%
  - Alternative hypothesis: recovery rate greater than 50%
- Testing means (*t*-test)
  - You want to argue that the highest speed that people drive at is, "on average," greater than the highest speed limit in the country (85 mph)
  - Null hypothesis: average maximum driving speed is less than or equal to 85mph
  - Alternative: average maximum driving speed is greater than 85mph

Assumptions / conditions one group

- For testing proportions (z-test)
  1. Assume data come independently
  2. Check that sample size is "big enough" (some people would say than more than 30 observations is fine)
- Assumptions for testing means (t-test)
  1. Assume data come independently
  2. Assume DGM with finite variance
  3. Check that either
     - Sample size is "big enough"
     - Sample is is small, but data look bell-shaped (normal)

Two groups: Independent means *t*-test

- Example:
  - Want to know if standardized scores in math are the same "on average" for men and women
  - Null hypothesis: scores don't depend on gender
  - Alternative: they do
- Assumptions / conditions
  - Assume the data within groups are independent, groups are independent
  - Check that either / or
    - Sample size is big enough
    - Data within each of the groups look normal
  - Some versions of the test require that the variance of the groups be equal, some don't

One- and two-sided alternative hypotheses

- An alternative hypothesis is said to be one-sided if it's of the type "greater than" or "smaller than"
  - Example: The recovery rate of a new drug is greater than 50%
- An alternative hypothesis is two-sided if it's of the type "not equal to"
  - Example: Average math scores are not equal for men and women

“New” tests

Paired means testing

- Two measurements on the same individual, under different circumstances
- For example, we measure some biomarker before and after treatment, and we want to know if there is a *significant* change
- The two measures (before, after) are correlated
- Paired means testing
  1. Take the difference "after – before"
  2. Do a test for one group
- Assumptions: individuals are independent, sample size is "big enough" or difference looks bell-shaped (normal)

z-test for 2 proportions

- Compare probabilities of success in 2 independent groups
  - Are they the same? Is one greater than the other?
- Example: Want to test if two treatments have the same recovery rate
- Assumptions
  - Data within groups are independent, groups are independent
  - Either / or
    - Sample size is big enough
    - Data within each of the groups look normal

Tests of independence: Categorical variables

- Suppose we have 2 categorical variables
- Null hypothesis: the variables are independent
- Alternative hypothesis: the variables are dependent
- Example:
  - Variables: X = socioeconomic status, Y = Type of high-school attended (public or private)
  - Null hypothesis: the type of high-school you attended does not depend on your socioeconomic status
  - Alternative: the type of high-school you attended depends on your socioeconomic status [e.g. rich people go to private schools more than working-class people]
- Assumptions:
  - Data come independently
  - Expected counts under independence are "big enough" for most cells

Confidence intervals

Confidence intervals are random intervals that come with a long-run guarantee:

- *If you report 95% confidence intervals all your life, 95% of them will capture the true value*
- You can't say anything about a particular interval; it either contains the truth or it doesn't

Visualization: http://rpsychologist.com/d3/CI/

- In SAS, you can find CIs for means and proportions (one and two groups) using the same PROCs we used for testing