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

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

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

One proportion with SAS

- Example: drug.csv (0 = no recovery; 1 = recovery)
  - H_0: recovery rate less than or equal to 0.5
  - H_a: recovery rate greater than 0.5

```
PROC FREQ data = drug;
      TABLES recovery / binomial (p = 0.5);
RUN;
```

"boundary" case

Sample recovery rate < 0.5…

Definitely not rejecting the null! No need to look at any p-values

recovery	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	34	56.67	34	56.67
1	26	43.33	60	100.00

Test of H0: Proportion = 0.5	
ASE under H0	0.0645
Z	1.0328
One-sided Pr > Z	0.1508
Two-sided Pr > |Z|	0.3017

Smallest p-value out of one-sided tests

Test 1

H_{01}: recovery rate ≤ 0.5

H_{a1}: recovery rate > 0.5

Test 2

H_{02}: recovery rate ≥ 0.5

H_{a2}: recovery rate < 0.5

p-value(test 2) ≤ p-value(test 1) if sample recovery rate < 0.5

p-value(test 2) = 1 – p-value(test 1)

p-value for

H_0: recovery rate = 0.5

H_a: recovery rate ≠ 0.5

One mean with SAS

- Example: speed data
  - H_0: max speed less than or equal to 85mph
  - H_a: max speed more than 85mph

"boundary" case

```
PROC TTEST data = speed sides = U H0 = 85 ;
       VAR speed;
RUN;
```

- Alternative "greater than" `sides = U`
- Alternative "less than" `sides = L`
- Alternative "not equal to", don't type `sides`

DF	t Value	Pr > t
1306	9.25	<.0001

← p-value

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

2 independent means with SAS

Example: speed data

H_0: max speed in men = max speed in women

H_a: max speed in men ≠ max speed in women

```
PROC TTEST data = speed;
       VAR speed;
       CLASS gender;
RUN;
```

Can use options "H0" and "sides" as in one-sided tests (order of difference is alphabetical; here it's "female – male")

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1305	-8.48	<.0001
Satterthwaite	Unequal	840.93	-8.33	<.0001

← p-value equal variances

← p-value unequal variances

“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" (in a DATA step)
  2. Do a test for one group (use PROC TTEST)
- 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

2 proportions with SAS

- Example: 2drugs
  - H_0: recovery rate drug A = recovery rate drug B
  - H_a: recovery rate drug A ≠ recovery rate drug B

```
PROC FREQ data = twodrugs;
      TABLES drug*recovery / chisq;
RUN;
```

Statistic	DF	Value	Prob
Chi-Square	1	2.1825	0.1396
Likelihood Ratio Chi-Square	1	2.1967	0.1383
Continuity Adj. Chi-Square	1	1.4556	0.2276
Mantel-Haenszel Chi-Square	1	2.1429	0.1432
Phi Coefficient		-0.1992	
Contingency Coefficient		0.1954	
Cramer's V		-0.1992	

← p-value

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

Tests of independence with SAS

- Using the hsb2 dataset:
  - H_0: school type independent of soc/econ status
  - H_a: school type dependent of soc/econ status

```
PROC FREQ data = hsb2;
      TABLES schtyp*ses / chisq;
RUN;
```

Statistic	DF	Value	Prob
Chi-Square	2	6.3342	0.0421
Likelihood Ratio Chi-Square	2	7.9060	0.0192
Mantel-Haenszel Chi-Square	1	0.2191	0.6397
Phi Coefficient		0.1780	
Contingency Coefficient		0.1752	
Cramer's V		0.1780	

← p-value

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/](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