Inference 1: Hypothesis testing (cont)

Last updated: September 14, 2017

Recap: hypothesis testing

Hypothesis testing = statistical proof by contradiction

- State: null hypothesis H_0, alternative hypothesis H_A.
- Assume H_0 is right.
- Observe s from data.
- Compute p-value.
- Smaller p-value: H_A better than H_0. Larger p-value: H_0 better than H_A.

H_A: "extremely large". Test: $\mathbb{P}(S \geq ...|H_0)$
H_A: "extremely small". Test: $\mathbb{P}(-S \leq ...|H_0)$
H_A: "extreme" (either way). Test: $\mathbb{P}(|S| \geq ...|H_0)$.

The VERY IMPORTANT questions

- **How should I choose H_0? How should I choose H_A?**
- **How small should the p-value be so that we can reject H_0?**
- **When should I set the significance level α?**
- **Should I keep testing for alternative hypotheses until I can reject H_0?**
- What is the 'best' S for a given (H_0, H_A) pair?

The VERY IMPORTANT questions

- **How should I choose H_0? How should I choose H_A?**
  - $H_0 =$ common sense BEFORE seeing data
  - $H_A =$ next common sense if reject H_0 BEFORE seeing data
- **How small should the p-value be so that we can reject H_0?**
- **When should I set the significance level α?**
- **Should I keep testing for alternative hypotheses until I can reject H_0?**
- What is the 'best' S for a given (H_0, H_A) pair?

The VERY IMPORTANT questions

- **How should I choose H_0? How should I choose H_A?**
  - $H_0 =$ common sense BEFORE seeing data
  - $H_A =$ next common sense if reject H_0 BEFORE seeing data
- **How small should the p-value be so that we can reject H_0?**
  - p-value $< \alpha \Rightarrow$ reject H_0 in favor of H_A
  - Significance level $\alpha =$ probability that you wrongly reject H_0!
- **When should I set the significance level α?**
- **Should I keep testing for alternative hypotheses until I can reject H_0?**
- What is the 'best' S for a given (H_0, H_A) pair?

The VERY IMPORTANT questions

- **How should I choose H_0? How should I choose H_A?**
  - H_0 = common sense BEFORE seeing data
  - H_A = next common sense if reject H_0 BEFORE seeing data
- **How small should the p-value be so that we can reject H_0?**
  - p-value $< \alpha \Rightarrow$ reject H_0 in favor of H_A
  - Significance level α = probability that you wrongly reject H_0!
- **When should I set the significance level α?**
  - BEFORE seeing the data!
- **Should I keep testing for alternative hypotheses until I can reject H_0?**
- What is the 'best' S for a given (H_0, H_A) pair?

The VERY IMPORTANT questions

- **How should I choose H_0? How should I choose H_A?**
  - $H_0 =$ common sense BEFORE seeing data
  - $H_A =$ next common sense if reject H_0 BEFORE seeing data
- **How small should the p-value be so that we can reject H_0?**
  - p-value $< \alpha \Rightarrow$ reject H_0 in favor of H_A
  - Significance level $\alpha =$ probability that you wrongly reject H_0!
- **When should I set the significance level α?**
  - BEFORE seeing the data!
- **Should I keep testing for alternative hypotheses until I can reject H_0?**
  - No! If you test enough, you will most likely do reject H_0 even if it were true
  - The goal of inference $\neq$ reject H_0.
- What is the 'best' S for a given (H_0, H_A) pair?

What test is best? Power, significance and errors

The power β of a test with significance level α is

$$\beta := \mathbb{P}(\text{ reject } H_0 | H_A) = \mathbb{P}(p - \mathit{value} < \alpha | H_A).$$

- $\alpha =$ probability that the test wrongly reject H_0
- $\beta =$ probability that the test does reject H_0 if the alternative were correct.

Higher $\beta =$ better. Smaller $\alpha =$ better.

Is there a test with $\beta = 1$?

What test is best? Power, significance and errors

The power β of a test with significance level α is

$$\beta := \mathbb{P}(\text{ reject } H_0 | H_A) = \mathbb{P}(p - value < \alpha | H_A).$$

- $\alpha =$ probability that the test wrongly reject H_0
- $\beta =$ probability that the test does reject H_0 if the alternative were correct.

Higher $\beta =$ better. Smaller $\alpha =$ better.

Yes: the test that always reject H_0 has $\beta = 1$!

What test is best? Power, significance and errors

The power β of a test with significance level α is

$$\beta := \mathbb{P}(\text{ reject } H_0 | H_A) = \mathbb{P}(p - \mathit{value} < \alpha | H_A).$$

- $\alpha =$ probability that the test wrongly reject H_0
- $\beta =$ probability that the test does reject H_0 if the alternative were correct.

Higher $\beta =$ better. Smaller $\alpha =$ better.

Yes: the test that always reject H_0 has $\beta = 1$!

Is there a test with $\alpha = 0$?

What test is best? Power, significance and errors

The power β of a test with significance level α is

$$\beta := \mathbb{P}(\text{ reject } H_0|H_A) = \mathbb{P}(p - value < \alpha|H_A).$$

- $\alpha =$ probability that the test wrongly reject H_0
- $\beta =$ probability that the test does reject H_0 if the alternative were correct.

Higher $\beta =$ better. Smaller $\alpha =$ better.

Yes: the test that always reject H_0 has $\beta = 1$!

Yes: The test that always accept H_0 has $\alpha = 0$!

What test is best? Power, significance and errors

The power β of a test with significance level α is

$$\beta := \mathbb{P}(\text{ reject } H_0 | H_A) = \mathbb{P}(p - value < \alpha | H_A).$$

- $\alpha =$ probability that the test wrongly reject H_0
- $\beta =$ probability that the test does reject H_0 if the alternative were correct.

Higher $\beta =$ better. Smaller $\alpha =$ better.

Yes: the test that always reject H_0 has $\beta = 1$!

Yes: The test that always accept H_0 has $\alpha = 0$!

There is a trade-off between higher power and smaller significance.
However, for fixed α, the likelihood ratio test is one with maximal β. Most tests you learn are equivalent to the likelihood ratio test (for that situation).

Which test to use

- Depends on data
- S (test stat) depends on H_A (alternative)
- justifying the assumptions = CRUCIAL.

Tests we learn	num	cat	cat. vs cat.	num. vs cat.
Fisher's exact			✓	
permutation (randomization)	✓	✓	✓	✓
chi-square			✓	
z-test for proportions		✓		
t-test	✓			
two-sample t-test				✓
ANOVA				✓

Modern Fisher's exact test for contingency tables

Data: $m \times n$ contingency table (cat. vs cat.)
Example: `sex` vs `survived`, `true` vs `guessed`, `class` vs `survived`

- H_0: row and column variables are independent
- H_A: they are not independent
- Test statistic S: the contingency table, ordered by their likelihood under H_0 (ie: $s \geq t$ means $\mathbb{P}(S = s|H_0) \geq \mathbb{P}(S = t|H_0)$).
- p-value: $\mathbb{P}(S \leq s|H_0)$
- Key: by combinatorics, one can compute exactly. (Hence the name 'exact' test)
- One-sided tests only available for 2×2 table.

Data example: Titanic