STAT 135, Concepts of Statistics

Helmut Pitters

Hypothesis testing 1

Department of Statistics
University of California, Berkeley

February 28, 2017

Hypothesis testing.

So far concerned ourselves with estimating population parameters.

Another important pillar of classical statistical inference:
testing whether or not an hypothesis about a population parameter has to be rejected when checked against data.

Contexts in which hypothesis testing might appear in applications:

- Does specific drug/medical treatment have a positive effect on patients' health?
- Does specific ad increase sales of a product?
- Establish the authorship of documents.
- Is specific die fair?

Hypothesis testing.

Example

Political candidate T claims to gain 50% of votes in city election. Conservative hypothesis: assume T is right, i.e.

$$\text{“null hypothesis”} \qquad H_0\colon p = \frac{1}{2},$$

where

$$p = \text{proportion of supporters of T in the electorate.}$$

Might be skeptical of T's claim and seek to support

$$\text{“alternative hypothesis”} \qquad H_A\colon p < \frac{1}{2}.$$

Among $n = 15$ randomly selected eligible voters, 8 favor T. Does this data support T's claim? Could high percentage $8/15 > 0.5$ of supporters be explained just by the chance of sampling?

Hypothesis testing.

Example

To decide whether or not to reject H_0, define *test statistic*

$$S_n := \text{\# supporters of T in sample of size } n.$$

Provided H_0 is true (i.e. $p = 0.5$),

$$S_n \sim \text{binomial}(n, 0.5),$$

Distribution of S_n under H_0 is called *null distribution*.

Hypothesis testing.

Example

Small values of S_n contradict H_0, large values of S_n support H_0.

If S_n is "small enough," want to reject H_0.
How should we decide what a "small enough" value for S_n is?

Hypothesis testing.

Example

k	0	1	2	3	4	5		
$\mathbb{P}\{S_n \leq k\}$	0.0	0.0	0.004	0.018	0.059	0.151		
k	6	7	8	9	10	11	12	13
$\mathbb{P}\{S_n \leq k\}$	0.304	0.5	0.696	0.849	0.941	0.982	0.996	1.0

Table: Cumulative distribution function of Binomial(15, 0.5) (rounded to 3 decimals).

Decision rule. Want to find critical value k such that

$$\begin{cases} H_0 \text{ is rejected if } S_n \leq k \\ H_0 \text{ is not rejected if } S_n > k. \end{cases} \tag{1}$$

Hypothesis testing.

Remark (Types of errors)

Notice: irrespective of decision rule, there are two types of errors we can make:

	H_0 true	H_0 false
reject H_0	type I error	correct decision
do not reject H_0	correct decision	type II error

Table: Two types of errors in hypothesis testing.

Usually, H_0 taken to be a conservative hypothesis,
$\rightarrow$ type I error considered "more serious" than type II error.

Example (Murder trial)

Null hypothesis should be: “accused is innocent” (‘in dubio pro reo’), since type I error: “innocent person is convicted” considered more serious than type II error: “murderer is acquitted.”

Hypothesis testing.

Remark

	H_0 true	H_0 false
reject H_0	type I error	correct decision
do not reject H_0	correct decision	type II error

Table: Two types of errors in hypothesis testing.

Probability

$$\alpha := \mathbb{P}\{\text{reject } H_0 | H_0\} = \mathbb{P}\{\text{type I error}\}$$

is called the *significance level* of a test.

Hypothesis testing.

In Neyman-Pearson paradigm of hypothesis testing controlling these two types of errors is central.
One starts by specifying significance level α. Commonly used values are $\alpha = 0.05$ or $\alpha = 0.01$.

Hypothesis testing.

Example

Let us fix

$$\text{significance level} \qquad \alpha = 0.018.$$

Let

$$k = 3$$

i.e. reject H_0 if $S_n \leq 3$. Then[1]

$$\mathbb{P}\{\text{reject } H_0 | H_0\} = \mathbb{P}\{S_n \leq 3 | H_0\} = \sum_{i=0}^{3} \mathrm{binomial}(15, 0.5)(i) = 0.018,$$

Our risk to conclude that T will lose, if, in fact, he wins, is 0.018, i.e. we'd make this error in 18 out of 1.000 cases.

[1] For large values of n and k find these values in table of binomial probabilites or via R.

Hypothesis testing.

Example (Speed limit)

Consider certain stretch of a highway.
Original speed limit: 65mph

original avg. speed: 63mph.

On this stretch a new speed limit is set:

55mph.

In sample of $n = 100$ cars:

new avg. speed: $\bar{X}_n = 61.4$mph with standard deviation SD$=4.6$mph

Suppose speeds $X_1, \ldots, X_{100}$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$.
Was avg. speed genuinely reduced, or is reduced speed in sample due to chance in sampling?
(Find a test with significance level $\alpha = 0.01$.)

Hypothesis testing.

Example (Speed limit)

Find a test with significance level $\alpha = 0.01$.

- $H_0 : \mu = 63.0$mph (avg. speed unchanged)
- $H_A : \mu < 63.0$mph
- Under H_0, distribution of test statistic

$$Z := \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim \mathcal{N}(0,1) \quad \text{is approximately standard normal.}^2$$

- Small values of Z contradict H_0, that is we're looking for r s.t.

$$\alpha = \mathbb{P}\left\{\text{type I error}\right\} = \mathbb{P}\left\{Z \leq (-\infty, r] | H_0\right\},$$

  Look up α-percentile normal distribution to find $r \approx -2.32$.
- Decision rule > next slide

---

[2] Would use t-distribution for small sample size n.

Hypothesis testing.

Example (Speed limit)

- Decision rule for test with significance level $\alpha = 0.01$ is

$$\begin{cases} \text{reject } H_0 & \text{if } S \leq -2.32 \\ \text{do not reject } H_0 & \text{if } S > -2.32. \end{cases}$$

Hypothesis testing.

Example (Speed limit)

- Decision rule for test with significance level $\alpha = 0.01$ is

$$\begin{cases} \text{reject } H_0 & \text{if } S \leq -2.33 \\ \text{do not reject } H_0 & \text{if } S > -2.33. \end{cases}$$

- From data we find:

$$Z = \frac{61.4 - 63.0}{4.6/10} = -3.48 \leq -2.33$$

and hence reject H_0 based on the data.

Notice that rejecting H_0 is not a statement about whether or not the new speed limit *caused* the reduction of avg. speed.
(There could be numerous other reasons causing the reduction that are unknown to us.)

Hypothesis testing.

Terminology of Neyman-Pearson paradigm. Consider data $x_1, \dots, x_n$ modeled as random samples from r.v.s $X_1, \dots, X_n$ with common distribution $\mathbb{P}_\theta$ that depends on some parameter $\theta \in \Theta$.[3]
E.g.

- $\mathbb{P}_\theta = \mathcal{N}(\theta, 1)$, $\Theta = (-\infty, \infty)$.
- $\mathbb{P}_\theta = \mathrm{Exponential}(\theta)$, $\Theta = (0, \infty)$.
- $\mathbb{P}_\theta = \mathrm{binomial}(N, \theta)$, $\Theta = [0, 1]$.

[3] Tests dealing with such parameterized models are referred to as parametric tests.

Hypothesis testing.

Terminology of Neyman-Pearson paradigm. Two hypotheses to be examined on the basis of data:

$$\boxed{\text{null hypothesis} \qquad H_0 \colon \theta \in \Theta_0}$$

This is usually a conservative hypothesis that is not to be rejected unless there is clear evidence to do so.

$$\boxed{\text{alternative hypothesis} \qquad H_A \colon \theta \in \Theta_A}$$

H_A specifies the kind of departure from H_0 one is interested in.
One has $\Theta_0 \cup \Theta_A = \Theta$ and $\Theta_0 \cap \Theta_A = \emptyset$.

If hypothesis specifies precisely one distribution, e.g.

$$H_A \colon \ \text{Poisson}(2.7)$$

it is called *simple*, otherwise it is called *composite*.

Hypothesis testing.

Terminology of Neyman-Pearson paradigm. Based on a *test statistic* $T = T(X_1, \ldots, X_n)$ a test is defined by the *rejection region*, C say, and the decision rule

$$\begin{cases} \text{reject } H_0 & \text{if } T(X) \in C \\ \text{do not reject } H_0 & \text{if } T(X) \notin C \end{cases} \tag{2}$$

The complement $\bar{C}$ of C is referred to as the *acceptance region*. As before

$$\alpha := \mathbb{P}\left\{\text{type I error}\right\},$$

and

$$\beta := \mathbb{P}\left\{\text{type II error}\right\} = \mathbb{P}\left\{\text{accept } H_0 | H_0 \text{ is false}\right\}.$$

The probability

$$1 - \beta = \mathbb{P}\left\{\text{reject } H_0 | H_0 \text{ is false}\right\}$$

to reject H_0 when it is false is called the *power* of the test.