Statistical Methods: Lecture 6

Lecture Overview

Testing a Claim

Aim: test a claim, i.e., "hypothesis". 2 sorts: null hypothesis / alternative hypothesis.

E.g., coin tossing. Test: "unbiased" vs. "biased". Toss 100 times.

Outcome not "fitting" claim (null hypothesis)? Then reject it.

- 47×H, 53×1\$ — $\hat{p}_{100} = 0.47$ — fair?
- 5×H, 95×1\$ — $\hat{p}_{100} = 0.05$ — biased
- 37×H, 63×1\$ — $\hat{p}_{100} = 0.37$ —fair/biased??

7.2 Basics of Hypothesis Testing

To test hypotheses we need data (from experiments).

Claim: coin is fair (unbiased)
(vs. claim: coin favors tails)
Experiment: toss coin 100 times
Outcome: 37 $\times$ heads, 63 $\times$ tails

Probability of 37 heads or less assuming claim is true?

If coin is fair, this probability is < 0.01.

$\Rightarrow$ Result "37 heads" unlikely with a fair coin.
$\Rightarrow$ Deviation statistically significant at level 1.

Conclusion: based on outcome it is unlikely that coin is fair.

7.2 Basics of Hypothesis Testing

Now we are considering test for p, later we will consider tests for μ and other tests.

2 hypotheses:

Null hypothesis

$$H_0 : \text{population parameter} \;=\; \text{hypothetical value}$$

Alternative hypothesis

Depending on the claim we choose one out of three:

$$H_a : \begin{cases} \text{population parameter} \;<\; \text{hypothetical value} & \text{left-tailed test} \\ \text{population parameter} \;>\; \text{hypothetical value} & \text{right-tailed test} \\ \text{population parameter} \;\neq\; \text{hypothetical value} & \text{two-tailed test} \end{cases}$$

Left- and right-tailed tests: one-sided tests, two-tailed test: two-sided test.

7.2 Basics of Hypothesis Testing

Measure "fitting"

Aim: find probability of
"if H_0 were true, the experiment outcome is at least as extreme as observation."

- Is $\hat{p}_{100} = 0.37$ likely if we assume that $p = 0.5$?
- Is $\bar{x}_n = 145.85$ likely if we assume that $\mu = 155$?

Problem?

Sample statistic's (proportion/mean) distribution depends on parameter to be tested (see CLT).

7.2 Basics of Hypothesis Testing

Solution?

– Test statistic ... is a random variable used for deciding about H_0.
Convert sample statistic by pretending H_0 is true ((null) hypothetical value).

Aim: distribution of test statistic under H_0 (null hypothesis).
Must not depend on true p (next time: μ) but can depend on n.

Example: Test statistic for testing about p

$H_0 : p = 0.4$, i.e. hypothetical value $p_0 = 0.4$.

$$\text{sample statistic } \hat{P}_n \qquad \Rightarrow \qquad \text{test statistic } \frac{\hat{P}_n - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{n}}}$$

(i.e. insert hypothetical value $p_0 = 0.4$ into $\frac{\hat{P}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$)

What do we know about this test statistic?

7.2 Basics of Hypothesis Testing

Five steps of hypothesis testing

0. identify population parameter
1. formulate H_0 and H_a and choose significance level α
2. collect data

3a. choose test statistic and identify its distribution under H_0

3b. calculate the value of test statistic based on data

3c. find P-value (depending on which alternative):
probability under H_0 that test statistic has more extreme values than observed value

4. formulate, based on P-value and α, conclusion regarding H_0:
reject (small P-value) or not (large P-value)

Significance level

Denoted by α, small, usually 0.05, but also 0.01 or 0.1.

Be careful

Hypothesis testing is asymmetric: if H_0 not rejected, we still do not "accept it"!

7.3 Testing a Claim About a Proportion

Example: Fair coin?

37 out of 100 coin tosses were heads. Claim: the coin is biased towards tails.

Step 0: Identify population parameter: population proportion of heads p

Step 1: Formulate H_0 and H_1 and choose significance level α:
$H_0 : p = 0.5$ vs. $H_1 : p < 0.5$ with significance level $\alpha = 0.05$.

Other examples

More than 60% of students have a Snapchat account

$$H_0 : p = 0.6 \qquad H_1 : p > 0.6$$

20% of Dutch citizens support VVD

$$H_0 : p = 0.2 \qquad H_1 : p \neq 0.2$$

Less than 35% of women smoke

$$H_0 : p = 0.35 \qquad H_1 : p < 0.35$$

7.3 Testing a Claim About a Proportion

Step 2: Collect data: $37 \times H, \;\; 63 \times T$

Step 3a: Choose test statistic and identify its distribution under H_0:

Use that the sampling distribution of the sample proportion approximately has a $N(p, \frac{p(1-p)}{n})$ distribution. A good test statistic is

$$Z = \frac{\hat{P}_n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where $p = 0.5$ is the claimed value from H_0.

Under H_0, $\;Z \sim N(0,1)$ approximately (if n is large enough).

Be careful

We use the claimed value (under H_0), and not $\hat{p}_n$.

Step 3b: Calculate the value of the test statistic based on data:

It follows from the data that $\hat{p}_n = 0.37$, so the observed value of the test statistic is

$$z = \frac{0.37 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{100}}} = -2.6$$

7.3 Testing a Claim About a Proportion

Step 3c: Find P-value:
We observed $z = -2.6$.
What is the probability of "test statistic takes even more extreme values"?

Our test is left-tailed: $H_1 : p < 0.5$
That means we have to compute

$$P(Z \leq -2.6) = 0.0047.$$

Step 4: Conclude:
The P-value is smaller than the significance level $\alpha = 0.05$, so H_0 is rejected.
We have enough evidence to confirm the coin is baised towards tails.

7.3 Testing a Claim About a Proportion

Calculation of P-value

If the test is

- left-tailed: P-value = area to the left of z,
- right-tailed: P-value = area to the right of z,
- two-tailed & $z < 0$: P-value = $2\times$ area to the left of z,
- two-tailed & $z > 0$: P-value = $2\times$ area to the right of z.

P-value and conclusion

- If P-value $\leq \alpha$: reject H_0.
- If P-value $> \alpha$: do not reject H_0.

7.3 Testing a Claim About a Proportion

Another example

Claim(s): (Not) 10% of population is left-handed
Data: $n = 750$ people surveyed, 92 are left-handed

0. identify population parameter
1. formulate H_0 and H_a and choose significance level α
2. collect data

3a. choose test statistic and identify its distribution under H_0

3b. calculate value of test statistic based on data

3c. find P-value (depending on H_a)

4. formulate conclusion regarding H_0: reject or do not reject H_0

7.3 Testing a Claim About a Proportion

Another example

Null hypothesis: 10% of population is left handed
Data: $n = 750$ people surveyed, 92 are left-handed

0: population parameter: population proportion p

1: $H_0 : p = 0.1$ $\quad H_a : p \neq 0.1$ $\quad$ significance level $\alpha = 0.01$

2: data: check; (but H_0, H_a, and α need to be chosen before collecting data.)

3: Calculate score, where claimed value $p_0 = 0.1$

$$z = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{92}{750} - 0.1}{\sqrt{\frac{0.1\cdot 0.9}{750}}} \approx 2.07$$

Under H_0: test statistic has approx. $N(0,1)$-distribution.
Test: two-tailed, and $z > 0$. Hence,

$$P\text{-value} = 2 \cdot P(Z \geq 2.07) = 2 \cdot (1 - P(Z \leq 2.07)) = 2 \cdot 0.0192 = 0.0384.$$

4: conclusion?

7.2 Basics of Hypothesis Testing

Recall:
Claim: 20% of Dutch citizens support VVD — $H_0 : p = 0.2$ $H_a : p \neq 0.2$
Claim: Less than 35% of women smoke — $H_0 : p = 0.35$ $H_a : p < 0.35$

Claims and hypotheses

More examples of translating claims into hypotheses in the book.
Careful: H_0 AND H_a both consist of competing claims.

No golden rule for choosing hypotheses; ... some guidelines:

- present/typical situation is in H_0,
- hypothesis you wish to reject is in H_0,
- hypothesis you wish to confirm is in H_a.

7.2 Basics of Hypothesis Testing

Three approaches to hypothesis testing

- confidence interval method (not used here)
- P-value method
- critical value method

Methods: P-value vs critical value

Same conclusions. Critical value sometimes simpler.

7.2 Basics of Hypothesis Testing

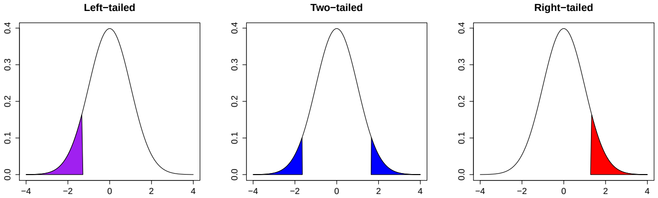
Critical region: all extreme scores for which H_0 is rejected (at significance level α).

Critical region: depends on α, distribution of test statistic and H_a.

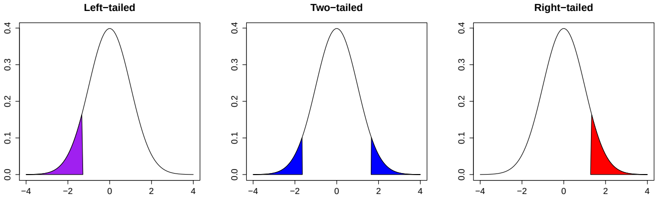
Critical regions

Fix α and look at density of test statistic's distribution. If test is

- left-tailed: critical region = extreme left tail
- two-tailed: critical region = both extreme left AND right tails
- right-tailed: critical region = extreme right tail



7.2 Basics of Hypothesis Testing



- area of purple region $= \alpha$
- total area of blue regions $= \alpha$, hence each part has area $= \alpha/2$
- area of red region $= \alpha$

Critical regions: test about proportion

Test statistic's distribution approx. $N(0, 1)$ under H_0.
For instance, for $\alpha = 0.05$, we reject H_0 if

- $z \leq -1.645$ (left-tailed test),
- $z \leq -1.96$ OR $z \geq 1.96$ (two-tailed test),
- $z \geq 1.645$ (right-tailed test).

7.3 Testing a Claim About a Proportion

Example: Fair coin?

Only step 3 is different when the critical region is used instead of the P-value.
We test $H_0 : p = 0.5$ vs. $H_1 : p < 0.5$ with significance level $\alpha = 0.05$.
The test statistic is

$$Z = \frac{\hat{P}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1) \text{ under } H_0$$

The test is left-tailed, so the critical region is $z \leq -1.645$.

We compute

$$z = \frac{0.37 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{100}}} = -2.6.$$

The observed value lies in the critical region, therefore H_0 is rejected.
We have enough evidence to confirm the coin is biased towards tails.

7.3 Testing a Claim About a Proportion

Example: Left-handed people

Only step 3 is different when the critical region is used instead of the P-value.
We test $H_0 : p = 0.1$ vs. $H_1 : p \neq 0.1$ with significance level $\alpha = 0.01$.
The test statistic is

$$Z = \frac{\hat{P}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1) \text{ under } H_0$$

The test is two-tailed, so the critical region is $z \leq -2.575$ or $z \geq 2.575$. (See Table 2; now $\alpha = 0.01$.)

We compute

$$z = \frac{92/750 - 0.1}{\sqrt{\frac{0.1 \cdot 0.9}{750}}} \approx 2.07.$$

The observed value does not lie in the critical region, therefore H_0 is not rejected.
We do not have enough evidence to confirm that the percentage of left-handed people is unequal to 10%.

7.2 Basics of Hypothesis Testing

Since hypothesis testing is based on data, randomness can cause errors.

	H_0 true	H_0 false
Reject H_0	Type I error	correct
Do not reject H_0	correct	Type II error

- Type I error: mistake of rejecting H_0 when it is true. Probability of a type I error $\leq \alpha$ (α is called the significance level).
- Type II error: mistake of not rejecting H_0 when it is false. Its probability is denoted by β.

Properties of two errors

- Probability of type I error is fixed.
- Prob. of type II error depends on (α,) n and actual value of population parameter. Usually decreases with n.