Introductory Statistics Lectures

Introduction to Hypothesis Testing

Testing a claim about a population proportion

ANTHONY TANBAKUCHI
DEPARTMENT OF MATHEMATICS
PIMA COMMUNITY COLLEGE

(Compile date: Tue May 19 14:50:31 2009)

Contents

1 Introduction to Hypothesis Testing

1.1 Introduction

Example 1. A 2001 study estimated 56% of people in the US wear corrective lenses.[1]

However, you believe the proportion of people in the US who wear corrective lenses is less than 56 percent.

[1] Source: Walker, T.C. and Miller, R.K. 2001 Health Care Business Market Research Handbook, fifth edition, Norcross (GA): Richard K. Miller & Associates, Inc., 2001. Study estimated about 160 million people in US wear glasses. 2001 population was estimated to be 286 million.

Question 1. How could you support your claim?

Question 2. You conduct a study of our class and find the proportion of students who wear corrective lenses is 55.6%. Does this support our hypothesis that the proportion of people in the US who wear corrective lenses is less than 56 percent? Why?

Question 3. What would we need to know to support our hypothesis that the proportion of people in the US who wear corrective lenses is less than 56 percent?

Goal

- Find probability of observing a sample proportion **at least as extreme** as $\hat{p} = 0.556$.
- If we can determine that it is unlikely to observe $\hat{p} = 0.556$ assuming $p_0 = 0.56$ then the rare event rule would make us question our assumption that $p_0 = 0.56$ and allow us to support our claim that $p < 0.56$.

Sampling distribution of $\hat{p}$

If np and $nq \geq 5$ then p will have a normal distribution[2] and the CLT tells us that $\hat{p}$ **is approximately normally distribution** where:

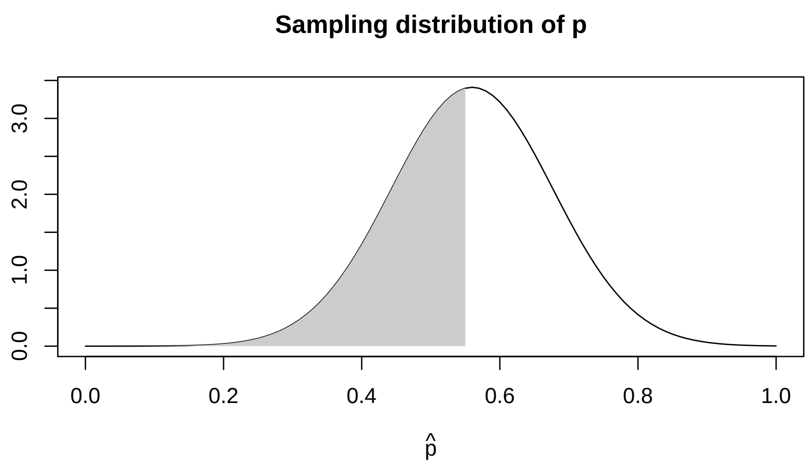
$$\mu_{\hat{p}} = p \tag{1}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \tag{2}$$

Probability of observing our sample data.

In our case $p = 0.56$, $n = 18$. We want to find the probability of observing a sample proportion **at least as extreme** as 0.556: $P(\hat{p} < 0.556)$.

[2]Normal approximation of binomial.



The above plot is the sampling distribution for $\hat{p}$ assuming $\mu_{\hat{p}} = p = 0.56$ and the **shaded area** 0.5**.**

Since p-value $= 0.5$:

Question 4. Using the rare event rule, would it be unusual to observe a sample proportion at least as extreme as 0.556 if the true population value is 0.56?

Question 5. Can we support our claim that the proportion of people in the US who wear corrective lenses is less than 56 percent?

Question 6. If we decided to support our claim that the proportion of people in the US who wear corrective lenses is less than 56 percent, what is the probability that we made the wrong decision? In other words, what is the probability that we would observe $\hat{p} = 0.556$ from a random sample drawn from a population with $p = 0.56$

Question 7. **Under what conditions can we support our claim** via the rare event rule?

Question 8. **Under what conditions can't we support our claim** via the rare event rule?

1.2 Hypothesis testing

Goal of hypothesis testing
The basic conceptual steps for hypothesis testing are:

1. Assume the status quo — the **null hypothesis** — is true.
2. Calculate the probability of observing the sample data assuming the the null hypothesis is true — the ***p*-value**.
3. If the *p*-value is small it is unlikely that we would have observed our sample data if the null hypothesis is true. Thus, we can reject the null hypothesis and we have evidence to support our claim — the **alternative hypothesis**.
4. If the *p*-value is not small it is not surprising to observe our sample data under the assumption that the null hypothesis is true. We cannot support our alternative hypothesis.

Two key concepts in hypothesis testing.

1. A hypothesis test is designed to **disprove the null hypothesis.**
   We don't prove anything. We simply show that the null hypothesis is statistically unlikely in light of sample data and the data **supports** our **alternative hypothesis**.
2. The null hypothesis always involves **equality**.
   We never support claims with equality![3]

STEPS

Eight simple steps

0. Write down what is **known**.
1. Determine which type of hypothesis **test** to use.
2. Check the test's **requirements**.
3. Formulate the **hypothesis**: H_0, H_a
4. Determine the **significance level** α.
5. Find the ***p*-value**.
6. Make the **decision**.
7. State the final **conclusion**.

You must know by heart and write down all eight steps when working problems![4]

K-T-R-H-S-P-D-C: "Know The Right Hypothesis So People Don't Complain"[5]

Step 1: Determine which type of hypothesis test to use.

[3]If we wish to do so, we must go beyond the content of this course and calculate the probability of a Type II error β.

[4]Note: I am showing you the *p*-value method for hypothesis testing. The book discusses it as well as the critical-value and confidence interval methods. The *p*-value method provides more information, is more precise (using R), and is more meaningful as compared to the critical-value method. Use the *p*-value method.

[5]Thanks to Maria Starzk for the nemonic.

Some common tests:

Single sample tests : Test for

1. Population proportion ($H_0 : p = p_0$)
2. Population mean ($H_0 : \mu = \mu_0$)
3. Population std. dev. ($H_0 : \sigma = \sigma_0$)
4. No correlation ($H_0 : \rho = 0$)
5. Normality (H_0 : pop. is normally dist.)

Where p_0, μ_0, σ_0 are all constants, (the status quo).

Two sample tests : Test for

1. Equality of two proportions ($H_0 : \Delta p = 0$)
2. Equality of two mean ($H_0 : \Delta\mu = 0$)
3. Equality of two std. devs. ($H_0 : \Delta\sigma = 0$)

Step 2: Check the test's requirements.

Each test has specific requirements. If you can't satisfy the requirements then the results will be meaningless.

HYPOTHESES H_0, H_A

Step 3: Formulate the hypothesis: H_0, H_a

Formulate the problem in terms of the null hypothesis that we want to disprove H_0 and the **alternative hypothesis** H_a[6] that we want to support.

Null hypothesis H_0. Definition 1.1

Represents the status quo which we hope to disprove. It always involves **equality**.

ex: $H_0 : p = 0.5$, $H_0 : \mu = 70\text{in}$, $H_0 : \Delta p = 0$

We never support H_0. It's like the control in an experiment.

Alternative hypothesis H_a. Definition 1.2

Represents the hypotheses that we want to support. If H_a involves $\neq$ it is a **two tailed test**, otherwise it is a **one tailed test**.

ex: $H_a : p \neq 0.5$, $H_a : p < 0.5$, $H_a : p > 0.5$.

Given the following statement:

"The proportion of people who think the sun revolves around the earth is 1/5."

Question 9. What would the null hypothesis be?

Question 10. What would the alternative hypothesis be?

Given the following statement:

"The proportion of people who think the sun revolves around the earth is more than 1/5."

Question 11. What would the null hypothesis be?

[6]Another notation for the alternative hypothesis is H_1.

Question 12. What would the alternative hypothesis be?

SIGNIFICANCE LEVEL

Step 4: Determine the significance level.

DEFINITION 1.3 SIGNIFICANCE LEVEL α.

Determine the maximum allowable type I error α you can live with. (Often 0.05 but you must decide what is right)

The type I error is the probability you made the wrong decision if you rejected the null hypothesis.

Confidence level $= 1 - \alpha$

P-VALUE

Step 5: Calculate p-value.

DEFINITION 1.4 p-VALUE.

The p-value is the probability of observing a **test statistic** at least as extreme as the one observed **assuming the null hypothesis** H_0 **is true.**

Common form of a test statistic

$$\text{test statistic} = \frac{(\text{sample statistic}) - (\text{null hypothesis of parameter})}{(\text{standard deviation of sample statistic})} \tag{3}$$

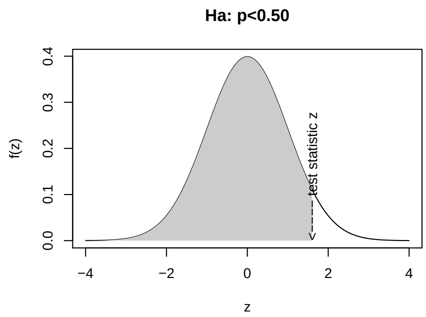
Finding the p-value "manually"

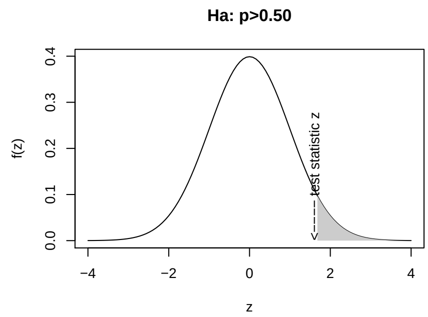
To find a p-value for any hypothesis test:

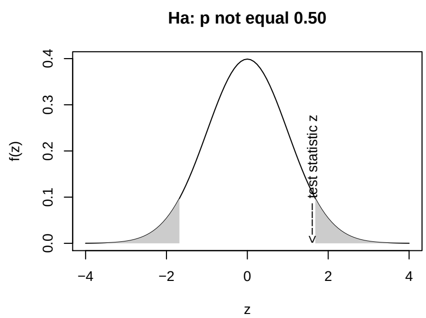
1. Calculate the test statistic.
2. Using the distribution of the test statistic:
   - If H_a contains $<$: p-value is the **area in the lower tail** bounded by the test statistic.
   - If H_a contains $>$: p-value is the **area in the upper tail** bounded by the test statistic.
   - If H_a contains $\neq$: p-value is **double** the tail area bounded by the test statistic. If the test statistic is negative use lower tail area. If the test statistic is positive use upper tail area.

Visualizing p-values for various H_a

Shaded region represents the p-value. Sample $\hat{p} = 0.67$ with $H_0 : p = 0.5$.







FORMAL DECISION

Step 6: Make the formal decision

- **Reject** H_0 if p-value $\leq \alpha$.
- **Fail to reject** H_0 if p-value $> \alpha$.

FINAL CONCLUSION

Step 7: State final conclusion

- If rejecting H_0:
  "The sample data supports the claim that (state H_a in words)."
- If failing to reject H_0:
  "The sample data does not contradict the claim that (state H_0 in words)."

TYPE I & II ERRORS

Type I Error α / p-value. Definition 1.5

If you **reject the null hypothesis** H_0, the p-value is the probability that you made the wrong decision, that is, H_0 is true and should not have been rejected.

The maximum Type I error you are willing to accept is α.

Type II Error β. Definition 1.6

If you **fail to reject the null hypothesis** H_0, β is the probability that you made the wrong decision, that is, H_0 is false and should have been rejected.

Decisions and errors

Decision	H_0 is true	H_0 is false
Reject H_0	**Type I Error,** α / p**-value**	Correct!
Fail to reject H_0	Correct!	**Type II Error,** β

You are on trial for murder in the US judicial system. If convicted you will be sentenced to death.

Question 13. What is H_0?

Question 14. What is H_a?

Question 15. What α should a jury use to convict you?

Question 16. If a jury commits a Type I error, what have they done?

Question 17. If a jury commits a Type II error, what have they done?

Tests in the medical field

For a medical test that should detect breast cancer:

sensitivity $= 1 - \beta$, proportion of **patients with breast cancer** that the **test marks correctly** as having breast cancer. (True positive — Power)

specificity $= 1 - \alpha$, proportion of **patients without breast cancer** and the **test marks correctly** as not having breast cancer. (True negative)

Question 18. What would we ideally like sensitivity and specificity to be?

Question 19. Is one more important than the other?

POWER

Definition 1.7

Power of a test.

Probability of correctly rejecting a **false** null hypothesis.[7]

$$\text{power} = 1 - \beta \tag{4}$$

[7] This is generally our goal. If as we suspect H_0 is false, we should know how likely we are to support H_a.

- Tells us the likelihood of supporting a true alternative hypothesis (making the correct decision). Good tests have powers of at least 0.8-0.9.
- Generally not simple to calculate, β depends on (1) α level, (2) sample size, (3) effect size, (4) specific test being used, (5) variance in population, ...
- **Before doing a study to support a hypothesis:** you should determine the minimum effect size you wish to detect and the desired power, then calculate the required n.

If we fail to reject the null hypothesis there are 3 possibilities:

1. The null hypothesis is true, there is no true effect. (Must calculate β to support this conclusion.)
2. The study design is too weak to detect a true alternative hypothesis (true effect).
3. The study had a good power, but random chance (sampling error) prevented us from rejecting a true alternative hypothesis.

Importance of Power Analysis:
Before conducting a study, we need to ensure that the sample size is large enough so it will be likely that we can actually detect a true alternative hypothesis. If the study design is too weak, the alternative hypothesis may be true but it is unlikely that we would be able to detect this. To determine the sample size n we must decide what is the minimum effect size — difference from the null hypothesis — that we are interested in detecting.

1.3 Single sample proportion test

USE

Often used to help answer:

1. Is the proportion of a population equal to p_0? *Is the proportion of people who smoke 20%?*
2. Is the proportion of a population different than p_0? *Is the proportion of people who smoke more than 20%? Is the proportion of contaminants in the water below the EPA standard?*

COMPUTATION

SINGLE SAMPLE PROPORTION TEST. DEFINITION 1.8

To support an alternative hypothesis concerning a population proportion:

requirements (1) simple random sample, (2) binomial distribution, (3) normal approximation of binomial np, $nq \geq 5$.

null hypothesis $H_0 : p = p_0$

alternative hypothesis (1) $H_a : p \neq p_0$, (2) $H_a : p < p_0$, (3) $H_a : p > p_0$

p_0 is a constant representing the status quo.

test statistic : described by the z distribution

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \tag{5}$$

Question 20. What distribution do you use to calculate the p-value?

SIMPLE EXAMPLE USING TEST STATISTIC

Question 21. If $H_0 : p = 0.25$, $H_a : p > 0.25$ and a study finds $x = 30$ and $n = 100$, find the test statistic and p-value.

Question 22. What would the p-value have been if $H_a : p \neq 0.25$?

R COMMAND

SINGLE SAMPLE PROPORTION TEST:

```
prop.test(x, n, p=0.5,
alternative="two.sided", conf.level=0.95)
```

x study number of successes
n study sample size
p null hypothesis value of p_0.
alternative H_a $\neq$:"two.sided", $<$:"less", $>$:"greater"
conf.level 1-α
The `conf.level` optional argument has no bearing on the p-value, it is used to make a confidence interval for p using the sample data.

Note on `prop.test()` :

- It uses the continuity correction (which our book does not), so the results are more accurate.
- It does give the test statistic, $z = \sqrt{\chi^2}$
- For an exact test try: `binom.test(...)`

A COMPLETE EXAMPLE

Example 2. A 2001 study estimated 56% of people in the US wear corrective lenses.[8]

However, you believe the proportion of people in the US who wear corrective lenses is less than 56 percent. You conduct a study of our class and find the proportion of students who wear corrective lenses is 55.6%.

K-T-R-H-S-P-D-C: "Know The Right Hypothesis So People Don't Complain"

Step 0: Gather the known information

```
R: p0
[1] 0.56
R: p.hat
[1] 0.55556
R: n
[1] 18
R: x = n * p.hat
R: x
[1] 10
```

Step 1: Determine test. Single sample proportion test. $(H_0 : p = p_0)$

Step 2: Requirements. (1) simple random sample, (2) binomial distribution, (3) normal approximation

```
R: n * p0
[1] 10.08
R: n * (1 - p0)
[1] 7.92
```

Question 23. Have we satisfied the requirements?

Step 3: Determine hypothesis. $H_0 : p = 0.56$, $H_a : p < 0.56$

Question 24. Is this a two tailed, lower tailed, or upper tailed test?

[8]Source: Walker, T.C. and Miller, R.K. 2001 Health Care Business Market Research Handbook, fifth edition, Norcross (GA): Richard K. Miller & Associates, Inc., 2001. Study estimated about 160 million people in US wear glasses. 2001 population was estimated to be 286 million.

Step 4: Determine significance level α. Not a life or death situation, we will use standard significance level of 0.05. (Thus our confidence level is 0.95).

Step 5: Find the p-value.

Question 25. Write what you would type to do this in R?

```
R: alt
[1] "less"
R: prop.test(x, n, p = p0, alternative = alt)
        1-sample proportions test with continuity correction

data:  x out of n, null probability p0
X-squared = 0, df = 1, p-value = 0.5
alternative hypothesis: true p is less than 0.56
95 percent confidence interval:
 0.00000 0.73176
sample estimates:
      p
0.55556
```

Question 26. What is the p-value?

Question 27. What is the test statistic z?

Question 28. What is the probability of a Type I error?

Step 6: Decision. Fail to reject H_0 since p-value is NOT less than or equal to 0.05

Step 7: Conclusion. "The sample evidence does not contradict the claim that the proportion of people in the US who wear corrective lenses is 56 percent."

Two tailed H_a

If we had a different alternative hypothesis: $H_a : p \neq 0.56$

Find the p-value:

```
R: prop.test(x, n, p = p0, alternative = "two.sided")
        1-sample proportions test with continuity correction

data:  x out of n, null probability p0
X-squared = 0, df = 1, p-value = 1
alternative hypothesis: true p is not equal to 0.56
95 percent confidence interval:
 0.33334 0.75789
sample estimates:
      p
0.55556
```

Note how the p-value has doubled.

1.4 Discussion

What hypothesis testing is
Assuming our null hypothesis H_0 is true, we calculate the probability (the p-value) that sampling error could cause our observed sample statistic to differ from H_0's claim about the population parameter using the sampling distribution. If the probability is very small — unusual — then, as stated by the Rare Event Rule, we it is unlikely that H_0 is true. If p-value $\leq \alpha$, we reject H_0 and have evidence to support H_a.

Important points

- In many situations, researchers define the beginning of reasonable doubt as $\alpha = 0.05$ or less.
- If we reject H_0, we have evidence to support H_a. The probability that we made the wrong decision (the Type I error) is the p-value.
- If we fail to reject H_0, we don't know if H_0 is true. The p-value **does not** represent the probability that H_0 is true. Only β — which we don't calculate in this class — tells us the probability that we made the wrong decision and H_0 is false.
- If you fail to meet the test's requirements, the results are meaningless.
- If you fail to sample properly, the results are meaningless.

1.5 Summary

Hypothesis testing steps

0. Write down what is **known**.
1. Determine which type of hypothesis **test** to use.
2. Check the test's **requirements**.
3. Formulate the **hypothesis**: H_0, H_a
4. Determine the **significance level** α.
5. Find the p**-value**.
6. Make the **decision**.
7. State the final **conclusion**.

K-T-R-H-S-P-D-C: "Know The Right Hypothesis So People Don't Complain"

Single sample proportion test
requirements (1) simple random sample, (2) binomial distribution, (3) normal approximation of binomial np, $nq \geq 5$.
null hypothesis $H_0 : p = p_0$
alternative hypothesis (1) $H_a : p \neq p_0$, (2) $H_a : p < p_0$, (3) $H_a : p > p_0$
Test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \tag{6}$$

Finding p-value manually See page 6

test in R : `prop.test(x, n, p=0.5, alternative="two.sided")`
`alternative="two.sided", "less", "greater"`

1.6 Additional Examples

Try this yourself. Do all 8 steps.

Example 3. Among 734 randomly selected Internet users, it was found that 360 of them use the Internet for making travel plans. Use a 0.01 significance level to test the claim that among Internet users, less than 50% use it for making travel plans.