Chapter 20 Testing Hypotheses about Proportions

Problem*:* ***Suppose we tossed a coin 100 times and we have obtained 38 Heads and 62 Tails. Is the coin biased toward tails****?*

There is no way to say yes or no with 100% certainty. But we may evaluate the strength of support to the hypothesis that *"the coin is biased".*

In statistics, a **hypothesis is a claim or statement about a parameter (a property of a population)**.
A **hypothesis test** (or "test of significance") **is a standard procedure for testing a claim.**
If, under a given assumption, we observe an event with likelihood exceptionally small, we conclude that the assumption is probably not correct.

We start by making two statements called the **Hypotheses:**

Null hypothesis (denoted by H_0) is a statement about an established fact, no change of known value of a population parameter. Expressed as Math equation i**t must contain a condition of equality:** =, ≥, or ≤. We replace all of above with a simple "="
Example: H_0: the coin is fair, and 50% of tosses end with H.

Alternative hypothesis (denoted by H_1 or H_a) is the statement that the parameter has a value that somehow differs from the null hypothesis. Needs a strong support from data to change our thinking and contradicts Ho. Expressed as Math statement it contains ≠, <, or >

Example: We contradict the statement that the coin is fair. Three ways are possible: the coin is biased toward heads (proportion of heads is bigger than tails). Or – it is less. Or – simply – not equal to 50%

In practice, there are three 3 ways to set up the hypotheses:

1. H_0: the parameter= given number, H_1: the parameter ≠ given number (2 tails)
2. H_0: the parameter= given number, H_1: the parameter < given number (left tail)
3. H_0: the parameter= given number, H_1: the parameter > given number (right tail

Example: Set up the hypotheses

Summarizing "*Testing a coin": If p is the probability that the coin turns "Heads" state both hypotheses*

Back to **Problem*: Suppose we tossed a coin 100 times and we have obtained 38 Heads and 62 Tails. Is the coin biased toward tails***?

H_O: coin is fair, ***p = 0.5*** *(population proportion of heads is the same as tails)*

H_1: there are three ways to <u>*disagree with Ho*</u>. We can say:

- *coin is biased toward heads,* ***p > 0.5*** *(more heads than tails were observed), or*
- *coin is biased toward tails* ***p<0.5****(less heads than tails), or*
- *coin is biased* ***p≠0.5*** *(the numbers of heads and tails are not nearly equal)*

Exercises:

For each of the following claims, determine the null and alternative hypotheses. State whether the test is two-tailed, left-tailed or right-tailed.

a) In 2008, 62% of American adults regularly volunteered their time for charity work. A researcher believes that this percentage is different today.

b) According to a study published in March, 2006 the mean length of a phone call on a cellular telephone was 3.25 minutes. A researcher believes that the mean length of a call has increased since then.

Attitude: Assume that the null hypothesis H_O is true and uphold it, unless data strongly speaks against it.

Test the Null Hypothesis directly.
In conclusion: Reject H_0 or fail to reject H_0
<u>NEVER</u> reject or fail to reject the <u>alternative,</u> H_1. NEVER state that any hypothesis is "proven".

Assumptions: We assume that all conditions for CLT are met: large enough random sample (more than 10 successes and failures), but at the same time, "small enough" sample (less than 10% of the population).
Method: By CLT the statistic p-hat has approximately normal distribution with the center at population proportion p and standard deviation $SD(\hat{p}) = \sqrt{\frac{p_o q_o}{n}}$

(In this formula, p, often denoted as p_o, is the population proportion of interest stated in H_o, and q, or q_o=1-po, and n=sample size)

Test mechanics: From data compute the value of a proper test statistics. *In our example **test statistic is the z-score*** computed for your observed statistic $\hat{p}$*:*

$$z = \frac{(\hat{p} - p_O)}{SD(p_0)}$$

where p_0 is the H_O value of the parameter (in our example, p_0 =0.5).
If H_O is correct then our z-score should be close to 0, the center of z-distribution.
If it is far from what is expected under the null model H_O assumption, then we reject H_O.

H_0: p=0.5 H_1: p<0.5

$$\hat{p} = 0.38, \quad p_O = 0.50$$

if H_0 is true, if the coin is fair, then

$$SD(p_0) = \sqrt{\frac{p_O q_O}{n}} = \sqrt{\frac{0.5 \times 0.5}{100}} = 0.05$$

$$z = \frac{(\hat{p} - p_O)}{SD(\hat{p})} = \frac{(.38 - .50)}{0.05} = -2.4$$

Our observed proportion 0.38 has been translated into a z-score z=-2.4
How far down is z=-2.4 from 0? How likely is to see z=-2.4 or less assuming H_O were true, that is, that the coin is not biased? To answer this we'll find P(z<-2.4)

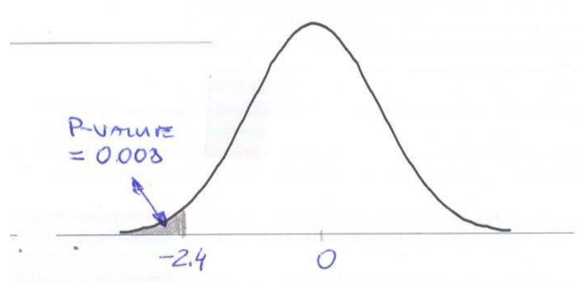
Level of significance α: Should be selected before we attempt to solve the problem.
It separates "likely" from "unlikely" events.

P-value: The probability of obtaining a test statistic at least as extreme as the one actually obtained, assuming null hypothesis is true.

P-value is the smallest level of significance at which we can reject null hypothesis. It measures the strength of evidence against null hypothesis. **The smaller p-value the stronger evidence against H_O**

In this problem let's agree to α=1%: if P-value is larger than 1%, then our observed statistic does not give sufficient evidence that the coin is unfair. If the P-value is less than the level of significance α, then we got sufficient evidence to reject H_o.

In our example P-value = P(z<-2.4) = 0.0082, less than 1%.
Meaning: if the coin is fair, then the probability of observing 38 or fewer heads in 100 tosses is less than 1%.



Conclusion: Two statements
1. Decide if you reject Ho or not (only one of the two options is possible
- We reject H_O
- We fail to reject H_O

2. Answer the original question (ex. support the claim that… or, do not support the claim that…)

In our example the conclusion is:
Reject the null hypothesis that the coin is fair (at significance level 0.01) and support H_1 stating that the coin is biased toward tails.

SUMMARY - One-proportion z-test

Assumptions
1. Random sample
2. Independent observations
3. If sampling without replacement, the sample size *n* should be no more than 10% of the population.
4. "Large" sample size *n* (np >10 and nq >10)

Hypotheses:

- Null hypothesis H_O: $p = p_O$
- Alternative hypothesis H_A: $p > p_O$ or H_A: $p < p_O$ or H_A: $p \neq p_O$

Attitude: Assume that the null hypothesis H_O is true and uphold it, unless data strongly speaks against it.

Level of significance α (more about it in the next chapter): it is marked alpha (α); we treat is as a threshold between “likely” and “unlikely” value of our test statistic; helps to make a decision about H_o.
Common significance levels: α=0.10, **α=0.05**, α=0.01 (but can be another)

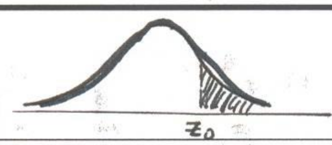
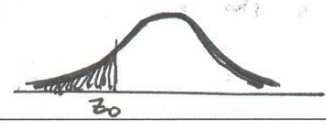
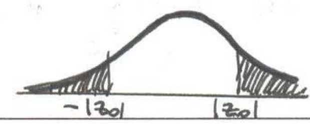
Test statistic:

$$z = \frac{\hat{p} - p_O}{SD(p_O)}$$

where $\hat{p}$ *is a sample proportion, and* $SD(\hat{p}) = \sqrt{\frac{p_O q_O}{n}}$

Distribution: If H_O is true, then test statistic z is approximately <u>standard normal</u> (and should be close to 0).

Let z_o be the observed value of the test statistic. The way we compute the P-value depends on H_A

H_A	P-value	
H_A: $p > p_O$	$P(z > z_o)$	
H_A: $p < p_O$	$P(z < z_0)$	
H_A: $p \neq p_O$	$P(z > \|z_o\|) + P(z < -\|z_o\|)$ $=2\,P(z < -\|z_o\|)$	

Decision:

- if the P-value is smaller than or equal α, we reject H_O *at the significance level* α,

- if the P-value is bigger than α, we fail to reject H_O *at the significance level* α

Note: we do not EVER "accept" or "prove" null hypothesis!

<u>Classwork</u>

2. **More hypotheses.** Write the null and alternative hypotheses you would use to test each of the following situations.
   a) In the 1950s only about 40% of high school graduates went on to college. Has the percentage changed?
   b) 20% of cars of a certain model have needed costly transmission work after being driven between 50,000 and 100,000 miles. The manufacturer hopes that a redesign of a transmission component has solved this problem.
   c) We field-test a new-flavor soft drink, planning to market it only if we are sure that over 60% of the people like the flavor.

4. **Dice.** The seller of a loaded die claims that it will favor the outcome 6. We don't believe that claim, and roll the die 200 times to test an appropriate hypothesis. Our P-value turns out to be 0.03. Which conclusion is appropriate? Explain.

a) There's a 3% chance that the die is fair.
b) There's a 97% chance that the die is fair.
c) There's a 3% chance that a loaded die could randomly produce the results we observed, so it's reasonable to conclude that the die is fair.
d) There's a 3% chance that a fair die could randomly produce the results we observed, so it's reasonable to conclude that the die is loaded.

6. **Origins.** In a 1993 Gallup poll, 47% of the respondents agreed with the statement "*God created human beings pretty much in their present form at one time within the last 10,000 years or so.*" When Gallup asked the same question in 2001, only 45% of those respondents agreed. Is it reasonable to conclude that there was a change in public opinion given that the P-value is 0.37? Explain.

16. **Educated mothers.** The National Center for Education Statistics monitors many aspects of elementary and secondary education nationwide. Their 1996 numbers are often used as a baseline to assess changes. In 1996, 31% of students reported that their mothers had graduated from college. In 2000, responses from 8368 students found that this figure had grown to 32%. Is this evidence of a change in education level among mothers?

a) Write appropriate hypotheses.
b) Check the assumptions and conditions.
c) Perform the test and find the P-value.
d) State your conclusion.
e) Do you think this difference is meaningful? Explain.

Test an appropriate hypothesis and state your conclusion. Perform the test at significance level=5%.

****

For a possible bonus: “**critical region”**, or
“classical” method.

“To do” list for Hypotheses Testing

a. **What is being tested?** The population mean, or population proportion? _______

b. **Hypotheses.** H_0: ________________ vs. H_1: ________________.

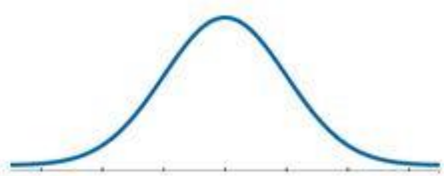
c. **Type of the test :** Right/Left Tail or Two-Tail Test? __________

Significance level: α=…….. (if not given, 5%)

d. Calculate test statistic:

e. Choose the method or use both

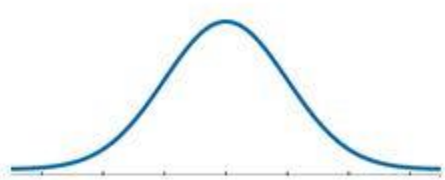
I Rejection region: Find the critical value and mark clearly the rejection region and critical value on the graph.



If α = ____ then z_α =__________,

Test statistic **is / is not** in the rejection region.

II: P-value method:



P-value=________ (Mark clearly P-value and the test statistic)

Compare with α: **P-value< α or P-value > α?**

f. The conclusion (Two statements):

a) Reject/fail to reject H_0

b) Support / do not support the alternative, that is, the claim that …..

Example:
A researcher obtains a random sample of 1000 people and finds that 534 are in favor of the banning cell phone use while driving, so p-hat = 534/1000. Does this suggest that the majority, that is, more than 50% of people favor the policy? In other words, would it be unusual to obtain a sample proportion of 0.534 or higher from a population whose proportion is 0.5? What is convincing, or statistically significant, evidence?

When observed results are unlikely under the assumption that the null hypothesis is true, we say the result is **statistically significant**. When results are found to be statistically significant, we reject the null hypothesis.

Hypothesis testing procedure;
Step 1 – check the assumptions (above) and determine the hypotheses
There are three ways to set up a hypothesis testing problem:

Two-Tailed	Left-Tailed	Right-Tailed
$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
$H_1: p \neq p_0$	$H_1: p < p_0$	$H_1: p > p_0$

Note: p_0 is the assumed value of the population proportion.

Our choice of the hypotheses:
Ho: no difference. The same proportion favors as does not favor new policy.
H1: the majority favors new policy. Write mathematical statements:

Ho: p=.50 H1: p>0.5

Step 2 - Select a level of significance, α, based on the seriousness of making a Type I error (the more serious consequences, the smaller alpha). Typical error is alpha = 0.05 or 5%

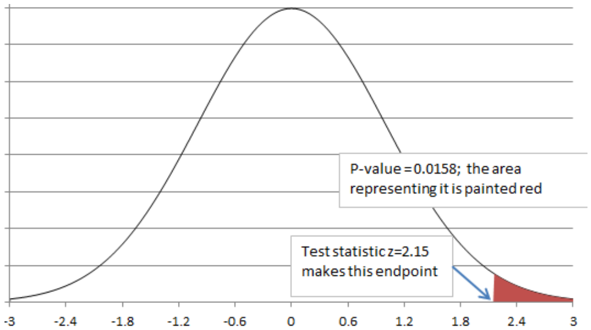
Step 3 – Compute test statistic **using p_0**, **not p-hat**, to compute standard error

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad Z_0 = \frac{.534 - .5}{\sqrt{.5 \cdot 0.5 / 1000}} = 2.15$$

Step 4 – either Classical or Modern approach

Modern:

P-value (**the probability that your observation is AT LAST as extreme as you found it if the null hypothesis is true)**
Its value equals the area of the corner(s) cut by the test statistic(s):
P(p-hat>0.534)=normalcdf(.534, 1, 0.5, $\sqrt{0.5*0.5/1000}$ =0.0158

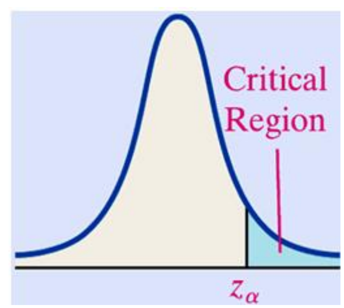


Compare P-value against alpha:
P-value<α=5%. Our observed proportion is unusual (unusually small, or unusually large).
We conclude that null hypothesis is not right.

Conclusion: Reject null hypothesis. Support alternative hypothesis which said that the majority of the population favors banning cell phones while driving.

Classical method:

Find $z_{\alpha/2}$ and Critical Region:
Alpha = 5%.
The problem is a right-tail problem.
z-score (a number of standard deviations) separating 5% in a right corner of normal distribution is 1.645



Our test statistic is $Z_0 = \dfrac{.534-.5}{\sqrt{.5 \cdot 0.5/1000}} = 2.15$

Our z-statistic is in critical (Rejection) region. Reject Ho!

Step 5: Conclusion (both methods give the same result)

Basing on our data, at 5% significance level we reject null hypothesis that there is a fifty-fifty support. We have enough evidence to support the claim that the majority does support new policy.