Chapter 16: Testing Hypotheses About Proportions

Example: Process of a court trial.

Suppose that someone is on trial for allegedly committing a crime:

- The accused person is assumed to be innocent. assume H_0 true
- The prosecutors collect and present evidence to contradict the innocence hypothesis. data
- The jury considers how likely it is that the evidence occurred, given that the person is innocent. P-value
- If there is enough evidence against the presumption of innocence, the court rejects the innocence hypothesis and declares the person guilty.
- If the evidence is not strong enough to prove that the defendant is guilty (that is, to reject the innocence hypothesis), then the court finds the person not guilty.

ompare -value with α

Note: In the case of insufficient evidence, the court does **not** declare the person as **innocent**. It finds the person **not guilty**. It was **not proven** that the person is innocent, but merely that there was not enough evidence to conclude that the person is guilty.

- fail to reject H_0
- did not prove H_0 true

A **hypothesis** is a claim or statement about the value of a population parameter (or parameters). A hypothesis usually states that a parameter takes a specific value or lies in a specific range of values.

A **hypothesis test** is a method of using sample data to decide whether or not to reject a hypothesis (in favour of another hypothesis).

A hypothesis test for a population parameter has five steps:

1. **Assumptions/Conditions:** Identify the variable/parameter of interest. State any assumptions and check any conditions needed to carry out the test.
   - independence assumptions
   - randomization conditions
   - sample size requirements
   - shape of distributions (population, sampling)

     ↳ normal

2. **Hypotheses:** A hypothesis test involves two types of hypotheses about parameters:
   - **Null Hypothesis:** denoted H_0, is a statement that a parameter takes a specific value. The null hypothesis is assumed to be true.
   - H_a, H_1

     negation of null

     Alternative Hypothesis: denoted H_A, is a statement that the parameter lies in some alternative set of values (values considered plausible if we reject the null hypothesis). Either a two-sided alternative or one-sided alternative.

   one-sided test

   choose one before looking at data

   In Step 2, we state H_0 and H_A.

   $H_0: \mu = 4$
   $H_A: \mu \neq 4$
   two-sided test or

   $H_0: p = 0.2$
   $H_A: p > 0.2$
   upper-tailed test or

   $H_0: \mu = 29$
   $H_A: \mu < 29$
   lower-tailed test

3. **Test Statistic:** a quantity calculated from sample data that measures the distance (usually the number of standard deviations) between the point estimate of the parameter and the null hypothesis value.

   In Step 3, we compute the test statistic. Z - score probability about data

4. **P-value:** the probability of observing a test statistic value at least as far away from the null hypothesis value as the test statistic computed in Step 3 (assuming the null hypothesis is true).

   - Measures the strength of the evidence against the null hypothesis.
   - If the P-value is small, it means it's very unlikely we would see data such as these if the null hypothesis were true. An outcome that would rarely occur assuming that the null hypothesis is true provides evidence against the null hypothesis, in favour of the alternative hypothesis. Probability

   not ~~P(Ha true)~~

   In Step 4, we compute the P-value.

   Conditional probability

   P(observed value or more extreme | H_0 true)

5. **Conclusion:** Report the P-value and interpret it in context. Make a decisions as to whether or not to reject H_0. One way to do this is to compare the P-value to a pre-determined value α, called a **significance level**. If P-value $\leq \alpha$, we reject H_0 and if P-value $> \alpha$, we do not reject H_0 (fail to reject H_0).

→ only two conclusions: reject H_0
do not reject H_0

Note: If we reject H_0 using a significance level, the result is said to be **statistically significant**. → means p-value $\leq \alpha$

→ does not mean of practical importance

Note: In the event that we reject H_0, following up with a confidence interval gives us a range of plausible values for the parameter

↳ use correct values in formula

Hypothesis Test for a Population Proportion: One-Proportion z-Test

A hypothesis test for a population proportion p has five steps:

1. **Assumptions/Conditions:** ↳ binary categorical variable.

   - Individuals in sample must be independent of each other.
   - Data collected using randomization.
   - The sample size n is large enough: $np_0 \geq 10$ and $n(1-p_0) \geq 10$, where p_0 is the assumed value of p in the null hypothesis.
   - Sample size should be less than 10% of population size.

   CLT $\hat{p} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$

2. **Hypotheses:**

   The null hypothesis is $H_0 : p = p_0$, where $0 \leq p_0 \leq 1$, and the alternative hypothesis is either

   choose one

$H_A : p \neq p_0$	or	$H_A : p < p_0$	or	$H_A : p > p_0$
(two-tailed test)		(lower-tailed test)		(upper-tailed test)

3. **Test Statistic:** observed sample proportion $\hat{p} = \frac{\# \text{ successes}}{n}$

   Measures how far the sample proportion $\hat{p}$ falls from p_0.

   z-score

   $$z_0 = \frac{\hat{p} - p_0}{\underbrace{\sqrt{\frac{p_0(1-p_0)}{n}}}_{\sigma_{\hat{p}} = SD(\hat{p})}}$$

   ← assumed population proportion. (H_0)

   $\mu_{\hat{p}} = p_0$

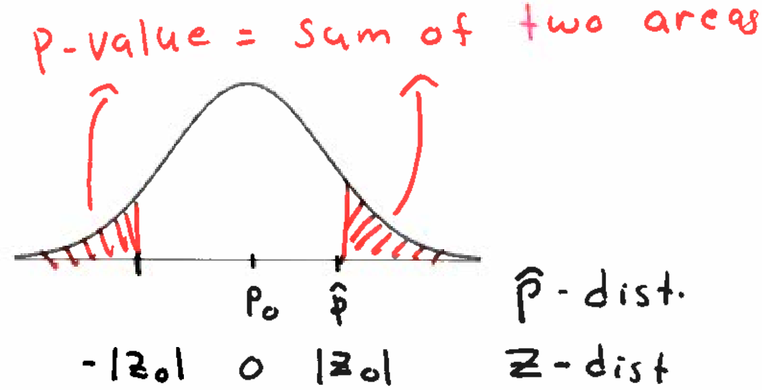
   $\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$

4. ***P*-value:** Calculate the probability of observing a test statistic value at least as far away from $p = p_0$ as the $\hat{p}$ we actually observed:

two-tailed test

$H_0 : p = p_0$
$H_A : p \neq p_0$

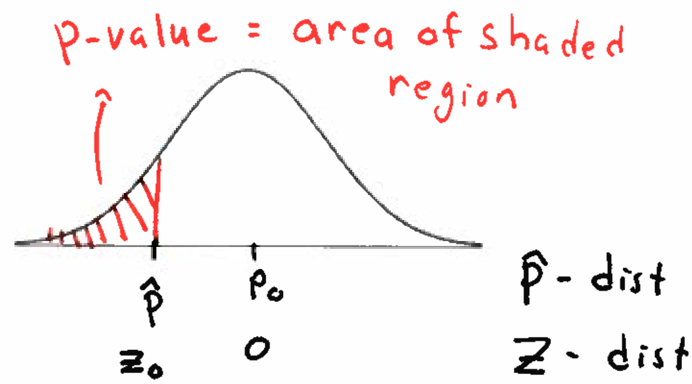
$2P(Z > |z_0|)$
$= 2P(Z < -|z_0|)$



lower-tailed test

$H_0 : p = p_0$
$H_A : p < p_0$

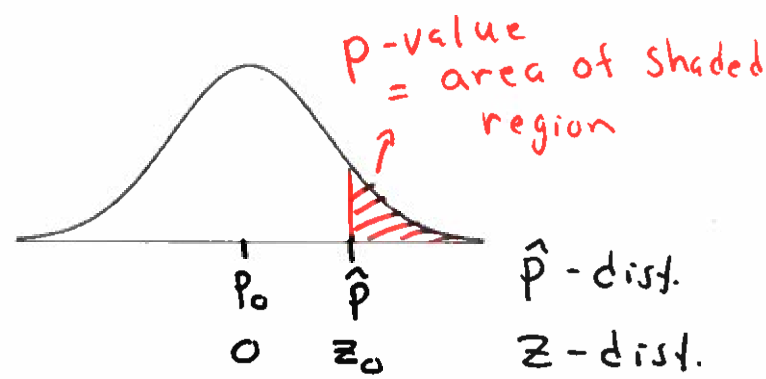
$P(Z < z_0)$



upper-tailed test

$H_0 : p = p_0$
$H_A : p > p_0$

$P(Z > z_0)$



5. **Conclusion:** Report and interpret the P-value in context. Given a significance level α, if

$$P\text{ - value} \leq \alpha,$$

then we reject H_0 at level α and if

$$P\text{ - value} > \alpha,$$

then we do not reject H_0 at level α.

Note:

Null hypothesis:

- Assumed to be true until it is declared false. → not the claim
- Represents the status quo.
- Represents no change from traditional value, no difference, or no effect.
- Either rejected or not rejected.
- You can **never** accept or prove H_0. → can disprove.
- Not rejecting H_0 does **not** prove it is true. It means there isn't enough evidence to show it is false. → data consistent with model.
- When we do not reject H_0, we say, "the data have failed to provide sufficient evidence to reject the null hypothesis".

Alternative hypothesis: (negation of H_0)

- Claim someone wishes to establish.
- Try to find evidence for.

Example:

State the null hypothesis and alternative hypothesis in each of the following situations:

a) A cable company claims that, due to improved procedures, the proportion of cable subscribers that have complaints against them is now less than 0.13.

$H_0: p = 0.13$

$H_A: p < 0.13$

b) According to a large study, a drug company's old pain reliever formula provided relief to 89% of the people who used it. The company wants to test a new formula to determine if it is better.

$H_0: p = 0.89$

$H_A: p > 0.89$

c) You meet a man on the street who claims to be psychic. Being skeptical of his claim, you take a standard deck of 52 cards and ask him to identify the suit of a randomly selected card without seeing it. You repeat the experiment 100 times, placing the card back in the deck after each trial.

$H_0: p = 0.25$ (guessing)

$H_A: p > 0.25$

d) In 1984, the LA Times reported that 15% of all California motorists had tampered with their vehicle emission-control devices so that they could save money by using leaded gas instead of unleaded gas. Suppose that in a random sample of 200 cars from one county in California, 21 had been modified. Does this suggest that, at the time, the proportion of cars in this county with modified devices differed from the statewide proportion?

$H_0: p = 0.15$

$H_A: p \neq 0.15$

Example: According to a US national statistics agency, 16% of all elementary school teachers in the US are male. A researcher randomly selected 1000 elementary school teachers in California and found that 142 were male. Does this sample provide sufficient evidence that the percentage of male elementary school teachers in California is different from the national percentage? (Use $\alpha = 0.05$.)

p = proportion of male elementary school teachers in California.

$\hat{p} = \frac{142}{1000} = 0.142$

1. **Assumptions/Conditions:**

- teachers randomly selected
- $1000(0.16) = 160 \geq 10$, $1000(0.84) = 840 \geq 10$ } Sample size large enough.
- independence
- < 10% population sampled

$\hat{p} \sim N\left(0.16, \sqrt{\frac{(0.16)(0.84)}{1000}}\right)$

2. **Hypotheses:**

$H_0: p = 0.16$

$H_A: p \neq 0.16$

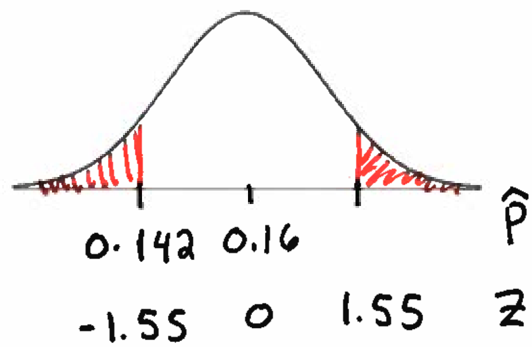
} two-tailed test

3. **Test Statistic:**

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{142}{1000} - 0.16}{\sqrt{\frac{(0.16)(0.84)}{1000}}} = \frac{-0.018}{0.01159} \simeq -1.55$$

4. ***P*-value:**

$2P(z < -1.55)$
$= 2(0.0606)$
$= 0.1212 > 0.05$



5. **Conclusion:** Since P-value $> \alpha = 0.05$,
We do not reject H_0 at the 0.05 significance level.
No, the sample does not provide sufficient evidence to conclude the percentage is different.

Example: A battery manufacturer wants to be reasonably certain that fewer than 6% of the batteries in a very large shipment are defective. Suppose that 300 of the batteries are randomly selected from this shipment. Each one is tested and 10 are found to be defective.

p = proportion of defective batteries in shipment.

a) Does this provide sufficient evidence, at the 0.05 level of significance, for the manufacturer to conclude that the fraction of defectives in the entire shipment is less than 0.06?

1. **Assumptions/Conditions:**
- random selection of batteries
- $300(0.06) = 18 \geq 10$
  $300(0.94) = 282 \geq 10$
- independence
- $<10\%$ population sampled.

$\hat{p} \sim N\left(0.06, \sqrt{\frac{(0.06)(0.94)}{300}}\right)$

2. **Hypotheses:** $H_0: p = 0.06$, $H_A: p < 0.06$ } lower tailed test

3. **Test Statistic:**

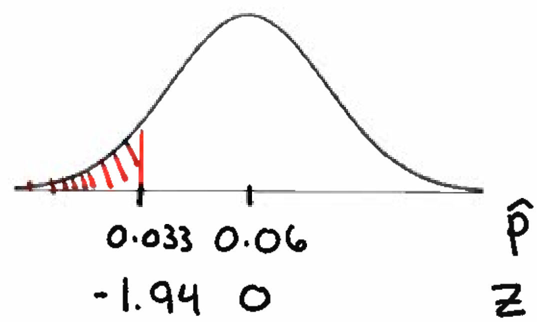
$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{10}{300} - 0.06}{\sqrt{\frac{(0.06)(0.94)}{300}}} = \frac{-0.0266667}{0.013711} = -1.94$$

4. ***P*-value:**

$P(\hat{p} < 0.033)$

$= P(z < -1.94)$

$= 0.0262 \leq 0.05$



5. **Conclusion:** Since p-value $\leq \alpha = 0.05$, we reject H_0 at the 0.05 significance level, that is, there is enough statistical evidence to conclude that less than 6% of the batteries in the shipment are defective.

b) at the 0.01 level of significance?

Since P-value $= 0.0262 > \alpha = 0.01$, we do not reject H_0 at the 0.01 significance level.