Chapter 17 - Testing Hypotheses About Proportions

In Statistics, a **hypothesis** proposes a model for the world we are interested in. This will be some statement about the parameters in a population.

Significance Test (or Hypothesis Testing) is a formal way to compare observed data (sample) with a hypothesis whose truth we want to assess.

1. If the data is consistent with the model, then we have no reason to disbelieve the hypothesis. The data supports the hypothesis, but does not prove the hypothesis.
2. If the data is inconsistent with the model, then we need to decide if the data is inconsistent enough to disbelieve the hypothesis. If inconsistent enough the data causes us to reject the hypothesis.

There are Four basic parts to Hypothesis Testing

1. State the Hypothesis
2. Determine the Model
3. Perform the Mechanics
   - Calculate the Test Statistic
   - Find P-Value
4. State Conclusion

17.3 The Reasoning of Hypothesis Testing

1. State the Hypothesis (17.1)

Determine the Hypothesis or Original Claim that is being tested. Write in symbolic form along with the opposite claim.

Null Hypothesis (H_0) is the statement of no effect or no difference.
Usually something of the form:
H_0: parameter = hypothesized value. $H_0: p = \#$

Alternative Hypothesis (H_A) is the hypothesis we accept if the null is rejected.
Usually something in one of the following forms:
H_A: parameter > hypothesized value (**Right-Tailed Alternative**),
H_A: parameter < hypothesized value (**Left-Tailed Alternative**), or
H_A: parameter ≠ hypothesized value (**Two-Tailed Alternative**).

2. Determine the Model

Each Model has some name that should be reported. Also, any conditions for a test must be satisfied.
The test for proportions is the **one-proportion *z*-test**. Normal Distribution

Option 5 in TESTS Menu of STAT Button on calculator.

Conditions: Plausible Independence, Random, 10% condition, Success/Fail Condition ($np_0 \geq 10$ & $nq_0 \geq 10$)

3. Perform the Mechanics

Usually this will be done by computer software or in our case the calculator, but we will learn the calculations involved to better understand the process.

a) Calculate the Test Statistic

The formula for the test statistic will vary depending on what test is being used. The formula for the one-proportion *z*-test is presented here.

To test the Null Hypothesis: $H_O : p = p_o$

p_o is hypothesized Value (a number)

We use the test statistic (t.s.): $z = \dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o q_o}{n}}}$

$q_o = 1 - p_o$

z score formula.

Why do we use $SD_{\hat{p}} = \sqrt{\dfrac{p_o q_o}{n}}$?

Sampling Distribution from Chapter 15 and z-score formula.

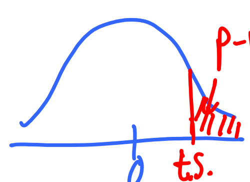
b) Find the P-Value

The **P-Value** is the probability that the test statistic will take a value at least as extreme as the observed value given that H_0 is true. **(17.2)**

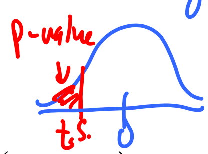
Conditional Probability.

The smaller the p-value the stronger the evidence is against H_0. (The smaller the p-value the less likely it will occur.)

$H_A: p > p_0$ **Right-Tail:** $\text{p-value} = P(z > \text{test statistic})$

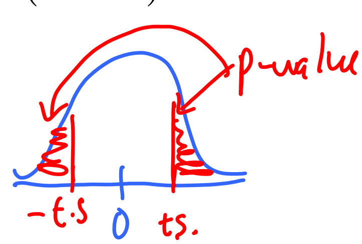


$H_A: p < p_0$ **Left-Tail:** $\text{p-value} = P(z < \text{t.s.})$



$H_A: p \neq p_0$ **Two-Tail:** $\text{p-value} = 2 \cdot P(z > \text{t.s.})$ or $2 \cdot P(z < \text{t.s.})$ **(17.4)**

We will calculate P-value using 1-Prop z test



4. **State the Conclusion** **(17.5)**

The conclusion is a statement about the hypothesis.
The conclusion must state either that we reject the null hypothesis or fail to reject the null hypothesis.
The conclusion should be stated in context of the original problem.

Significance Level (α) - The fixed value, determined in advance, that will be used to decide if the p-value is too extreme or not. (If not given use $\alpha = 0.05$) $[\alpha = 1 - C]$

- If $\text{p-value} \le \alpha$ then data is statistically significant at the α level and we reject the H_0 and accept the H_A.
- If $\text{p-value} > \alpha$ then data does not provide sufficient evidence to reject the null hypothesis. We fail to reject the H_0. This does not prove H_0 is true, but data supports H_0.

Statistically Significant does not mean importance. In statistics significant is used as in identifying (signifying) a level or standard.

If $\text{p-value} = 0.03$, this is significant at $\alpha = 0.05$ but not at $\alpha = 0.01$.

---

p.469

12 a) Original Claim: Percentage of high school students going to college has changed from 40%.

$H_0: p = .4$ — $p_0 = .4$ $q_0 = 1 - .4 = .6$

$H_A: p \neq .4$ — 2 tailed Alternative test.

b) Original Claim: Redesign of Transmission has solved problem.

$H_0: p = .2$

$H_A: p < .2$ — left tailed test

c) Original Claim: Over 60% like the new soft Drink flavor.

$H_0: p = .6$

$H_A: p > .6$ — Right tailed test.

p.470 (22) Original Claim: Abnormalities has increased from 5% since 1980's.

* a) $H_0: p = .05$ $\qquad$ $p_0 = .05 \quad q_0 = 1 - .05 = .95$

$H_A: p > .05$ Right Tailed test.

* b) Should be independent. Don't know if random, Assume it is.

10% condition: $n = 384$ should be less than 10% of all children.

S/F condition: $np_0 = 384 \times .05 = 19.2 \geq 10$

$nq_0 = 384 \times .95 = 364.8 \geq 10$

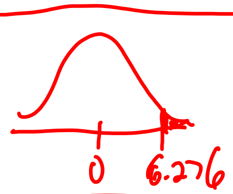
Conditions are met. One proportion z (Normal) test can be used.

* c) $\hat{p} = \frac{x}{n} = \frac{46}{384} = .1198$

$$t.s.\ z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{(.1198 - .05)}{\sqrt{\frac{.05 \times .95}{384}}} \approx 6.276$$

use calculator (1-prop z test) to find p-value

$P\text{-Value} = 1.75 \times 10^{-10} = .000000000175 \approx 0$



p-value $= P(z > 6.276)$

Normalcdf$(6.276, 100000)$

d) If the abnormalities rate has not increased then the probability of observing 46 out of 384 children with the abnormality is 1.75×10^{-10} or approximately 0.

* e) p-value $\quad \alpha$ since not given

$1.75 \times 10^{-10} < .05$ $\quad$ Reject H_0 & Accept H_A.

The data shows the abnormality rate has increased.

f) Even though rate has increased we cannot say if chemicals caused the increase.

p.472 32

Original Claim: There is a home field advantage ($p > 50\%$ or $p > .5$)

step 1

$H_0: p = .5$ $\quad p_0 = .5 \quad q_0 = 1 - .5 = .5$

$H_A: p > .5$ ← Right tailed test.

step 2

Random is not specified. Games should be independent.

10% cond: $n = 246$ is less than 10% of ALL NFL games.

S/F cond: $np_0 = 246 \times .5 = 123 \geq 10$

$nq_0 = 246 \times .5 = 123 \geq 10$

Conditions are met so One proportion z-test can be used.

step 3

$$\hat{p} = \frac{x}{n} = \frac{143}{246} = .5813$$

$$\text{t.s. } z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{(.5813 - .5)}{\sqrt{\frac{.5 \times .5}{246}}} \approx 2.550$$

P-value $= .0054$

step 4

P-value $\quad \alpha$

$.0054 < .05$ $\quad$ Reject H_0 & Accept H_A

* The data shows strong evidence that the home team does have an advantage.