STA 1013 : Statistics through Examples

Lecture 29: Statistical Hypothesis 3

Hwiyoung Lee

November 20, 2019

Department of Statistics, Florida State University

Review of the Z - test : Hypothesis Testing for Population Mean (when σ known)

Rejection Region approach

- **Test Statistic under H_0 (when σ known):**

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- We will reject H_0 if our test statistic (z_0) is **too "extreme."**
- "Too extreme" means the test statistic is located in the **rejection region**

Rejection Region approach

Rejection Regions

1. Left-Tailed test :

$$\textbf{Rejection Region} < -z_\alpha$$

2. Right-Tailed test :

$$\textbf{Rejection Region} > z_\alpha$$

3. Two-Tailed test :

$$\textbf{Rejection Region} > z_{\alpha/2} \text{ or } \textbf{Rejection Region} < -z_{\alpha/2}$$

P-value approach

1. Left-Tailed test

$$\textbf{P-value} \ = \ P(Z \leqslant z_0)$$

2. Right-Tailed test

$$\textbf{P-value} \ = \ P(Z \geqslant z_0)$$

3. Two-Tailed test

$$\textbf{P-value} = 2 * P(Z \geqslant |z_0|)$$

- Reject H_0 when : P-value $\leqslant \alpha$
- Can't reject H_0 when : P-value $> \alpha$

Calculator (Z-Test)

Z-Test from Data

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 1: **Z-Test**
3. Inpt : Highlight **Data**
4. Enter values for
   - μ_0 : Claimed value in the null hypothesis (H_0)
   - σ : Population standard deviation
   - List : Data (ex : L_1)
   - μ : Select the test type ($\underbrace{\neq \mu_0}_{Two-Tailed}$, $\underbrace{< \mu_0}_{Left-Tail}$, $\underbrace{> \mu_0}_{Right-Tail}$)

Example

Coke

Randomly selected cans of Coke are measured for the amount of cola, in ounces. The sample values listed below.

12.3	12.1	12.2	12.3	12.2	12.3	12.0	12.1	12.2
12.1	12.3	12.3	11.8	12.3	12.1	12.1	12.0	12.2
12.2	12.2	12.2	12.2	12.2	12.4	12.2	12.2	12.3
12.2	12.2	12.3	12.2	12.2	12.1	12.4	12.2	12.2

Assume that we want to use a 0.05 significance level to test the **claim that cans of Coke have a mean amount of cola greater than 12 ounces.** Assume that the population has a standard deviation of $\sigma = 0.115$ ounce.

1. **State the Hypothesis**
2. **Perform Z-test**

T - test : Hypothesis Testing for Population Mean (when σ unknown)

Population Test σ unknown

When we don't know the population standard deviation σ,

- Use t distirbution
- **Test statistic for t-test :**

$$t_0 = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

- Note : Compare the test statistic for z-test :

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Rejection Region Approach

1. Left-Tail Test

$$\textbf{Rejection Region} < -t_{\alpha,df}$$

2. Right-Tail Test

$$\textbf{Rejection Region} > t_{\alpha,df}$$

3. Two-Tailed Test

$$\textbf{Rejection Region} < -t_{\alpha/2,df} \text{ or } \textbf{Rejection Region} > t_{\alpha/2,df}$$

- Rejection region approach depends on t distribution
- <u>We will not use the rejection region approach</u>

P-value approach

1. Left-Tailed test

$$\textbf{P-value} \ = \ P(T_{df} \leqslant t_0)$$

2. Right-Tailed test

$$\textbf{P-value} \ = \ P(T_{df} \geqslant t_0)$$

3. Two-Tailed test

$$\textbf{P-value} = 2 * P(T_{df} \geqslant |t_0|)$$

- Reject H_0 when : P-value $\leqslant \alpha$
- Can't reject H_0 when : P-value $> \alpha$

Calculator (T-Test)

T-Test from Statistics

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 2: **T-Test**
3. Inpt : Highlight **Stats**
4. Enter values for
    - μ_0 : Claimed value in the null hypothesis (H_0)
    - $\bar{X}$: Sample mean
    - S_x : Sample standard deviation
    - n : sample size
    - μ : Select the test type ($\underbrace{\neq \mu_0}_{Two-Tailed}$, $\underbrace{< \mu_0}_{Left-Tail}$, $\underbrace{> \mu_0}_{Right-Tail}$)

Example

Fuel consumption

According to the Energy Information Administration (Federal Highway Administration data), the average gas mileage of all automobiles is 21.4 miles per gallon. For a random sample of 40 sport utility vehicles (SUVs), the mean gas mileage is 19.8 miles per gallon with a standard deviation of 3.5 miles per gallon. Test the claim that the mean mileage of all SUVs is less than 21.4 miles per gallon. ($\alpha = 0.05$)

1. **State the Hypothesis**

2. **Find the P-value**

3. **Conclusion**

Example

Baseballs

A random sample of 40 new baseballs is obtained. Each ball is dropped onto a concrete surface, and the bounce heights have a mean of 92.67 inches and a standard deviation of 1.79 inches (based on data from USA Today). Test the claim that the new baseballs have a mean bounce height that is less than the mean bounce height of 92.84 inches found for older baseballs. ($\alpha = 0.05$)

1. **State the Hypothesis**

2. **Find the P-value**

3. **Conclusion**

Example

coin Weight

According to the U.S. Department of the Treasury, the mean weight of a quarter is 5.670 grams. A random sample of 50 quarters has a mean weight of 5.622 grams with a standard deviation of 0.068 gram. Test the claim that the mean weight of quarters in circulation is 5.670 grams. ($\alpha = 0.05$)

1. **State the Hypothesis**

2. **Find the P-value**

3. **Conclusion**

Hypothesis Tests For Population Proportions

Hypothesis Tests For Population Proportions

- We now consider hypothesis testing with proportions

$$
\begin{aligned}
&H_0 : p = p_0 \\
&H_a : p \neq p_0 \ (Two - Tailedtest) \\
&H_a : p < p_0 \ (Left - Tailedtest) \\
&H_a : p > p_0 \ (Right - Tailedtest)
\end{aligned}
$$

- All the ideas from previous tests apply
- p (or π) : denotes the population proportion
- $\hat{p}$: denotes the sample proportion

Motivation

Poll

Suppose a political candidate commissions a poll in advance of a close election. Using a random sample of n = 400 likely voters, the poll finds that 204 people support the candidate. Should the candidate be confident of winning?

- We now cast the question as a hypothesis test.
- For a hypothesis test, we ask whether the poll results (which are the sample statistics) support the hypothesis that the candidate has more than 50% of the vote.
- Hypothesis :

  $H_0 : p = 0.5$ (50% of voters favor the candidate)

  $H_a : p > 0.5$ (more than 50% of voters favor the candidate)

Test Statistic

We will use z test statistic (Derivation)

- First let X_i's are independent Bernoulli dist (p)

$$E(X_i) = p, Var(X_i) = p(1-p)$$

- Second let $X = \sum_{i=1}^{n} X_i$, then X is Binomial dist (n, p)

$$E(X) = np, Var(X) = np(1-p)$$

- Then sample proportion can be viewed as $\hat{p} = \frac{X}{n}$

$$E\left(\frac{X}{n}\right) = \frac{np}{n} = p,$$
$$Var\left(\frac{X}{n}\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Test Statistic

Moreover,

- The sample proportion is an average form of independent Bernoulli dist (p)

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- We know that $\mu = p, \sigma = \sqrt{p(1-p)}$ are mean and standard deviation of X_i
- Thus apply CLT then

$$\hat{p} \sim N\left(p, \quad \underbrace{\sqrt{\frac{p(1-p)}{n}}}_{\text{Standard deviation of } \hat{p}}\right)$$

Test Statistic

By the z-transformation

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

Under the H_0, our z test statistic is given by,

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$$

Example : Poll

Poll

Suppose a political candidate commissions a poll in advance of a close election. Using a random sample of n $=$ 400 likely voters, the poll finds that 204 people support the candidate. Should the candidate be confident of winning? ($\alpha = 0.05$)

1. State the hypothesis :

   $H_0 : p = 0.5$ (50% of voters favor the candidate)

   $H_a : p > 0.5$ (more than 50% of voters favor the candidate)

2. Calculate the sample proportion :

$$\hat{p} = \frac{204}{400} = 0.51$$

Example : Poll

3. Calculate the test statistic :

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.51 - 0.5}{0.025} = 0.4$$

4. Rejection Region :

$$\textbf{Rejection Region} \geqslant z_\alpha = 1.645$$

5. Conclusion :

   Since $z_0 < 1.645$, which means test statistic in **not** located in the rejection region, We can conclude that there is no strong evidence to support H_a. Therefore, the candidate is not confident of winning.

Example

Local Unemployment Rate

Suppose the national unemployment rate is 10%. In a survey of n = 500 people in a rural Wisconsin county, 75 people are found to be unemployed. County officials apply for state aid based on the claim that the local unemployment rate is higher than the national average. Test this claim at a 0.05 significance level.

1. State the hypothesis :
2. Calculate the sample proportion :
3. Find the rejection region :
4. Conclustion :

Solution

1. State the hypothesis :

$$H_0 : p = 0.1$$
$$H_a : p > 0.1$$

2. Calculate the sample proportion :

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.15 - 0.1}{\sqrt{0.1 \cdot 0.9/500}} = 3.727$$

3. Find the rejection region :

$$\textbf{Rejection Region} \geqslant z_\alpha = 1.645$$

4. Conclusion : (z_0 is located in the rejection region)
   We can reject the H_0. So there is strong evidence to aid the rural Wisconsin county.

Calculator (1-PropZTest)

Z-Test for proportion

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 5: **1-PropZTest**
3. Enter values for
   - p_0 : Claimed proportion in the null hypothesis (H_0)
   - x : Number of success
   - n : Sample size
   - prop : Select the test type

     $\neq p_0$: Two-Tailed
     $< p_0$: Left-Tail
     $> p_0$: Right-Tail

Example

Left-Handed Population

A random sample of n $=$ 750 people is selected, of whom 92 are left-handed. Use these sample data to test the claim that 10% of the population is left-handed. ($\alpha = 0.05$)

1. State the hypothesis :

2. Perform Z-Test for proportion :

Example

smoking and college Education

smoking and college Education. A survey showed that among 785 randomly selected subjects who completed four years of college, 144 smoke and the others do not smoke (based on data from the American Medical Association). Test the claim that the rate of smoking among those with four years of college is less than the 27% rate for the general population. ($\alpha = 0.05$)

1. State the hypothesis :

2. Perform Z-Test for proportion :

Example

grade Pressure

A study commissioned by the U.S. Department of Education, based on responses from 2,000 randomly selected teenagers, concluded that 44% of teenagers cite grades as their greatest source of pressure. Test the claim that fewer than half of all teenagers in the population feel that grades are their greatest source of pressure. ($\alpha = 0.05$)

1. State the hypothesis :

2. Perform Z-Test for proportion :