Chapter 9-10

Confidence Intervals and Hypothesis Testing

HT for Single Proportion

Statistics

Mehmet Güray Güler, PhD

Last updated 23.07.2020

Inferences on Population Proportions

- The relevant test statistics
- Single Proportion (single populations)
  - Hypothesis testing
  - Confidence Intervals
- Two Proportions (two populations)
  - Hypothesis testing
  - Confidence Intervals

The Test Statistic

- sample mean $\bar{X}$ ⇔the population mean μ,
- sample proportion $\hat{P}$ ⇔ the population proportion *p*.

- population proportion: *p* (A parameter = unknown constant)
- sample proportion: $\hat{P} = \frac{X}{n}$ (A statistic = random variable)

The Test Statistic

- The formula for the sample proportion $\hat{P} = \frac{X}{n}$.
- Here X represents the number of “*success*”es in n trials.
- ***X_1, X_2, …, X_n* :** the observations in the random sample.
- We define
  - $X_i = 1$, if it is a “success”, and $X_i = 0$ otherwise.
  - X (=the total number of “success”es in the sample)

$$X = \sum_{i=1}^{n} X_i$$

What is the distribution of X?

The Test Statistic

- X is binomial random variable
- **Binomial Distribution (Let’s remember)**
  - Binomial RV X, counts the number of successes in n experiments.
  - p is the success probability
  - $\mu = np \quad \sigma^2 = np(1-p)$
- Then from $\hat{P} = X/n$ we have
  - $\mu_{\hat{P}} = \mu_X/n$ = p
  - $\sigma_{\hat{P}}^2 = \frac{\sigma_X^2}{n^2} = p(1-p)/n$

The Test Statistic

- How about its distribution?
- Recall $\hat{P} = \frac{X}{n}$. Then
- $\hat{P} = \sum_{i=1}^{n} \frac{X_i}{n}$ is a **sample mean**
- Then, by the CLT, for large *n*,
  - the distribution of $\hat{P} = X/n$ is approximately normal
  - Hence $\mathrm{Z} = \frac{\hat{P}-p}{\sqrt{p(1-p)/n}}$ is approx. standard normal

Hypothesis Testing for Proportions

Single Proportion

Two Proportions

One Sample: Test on a Single Proportion

The standardized test statistic: $Z = \dfrac{(\hat{p} - p_0)}{\sqrt{p_0(1 - p_0)/n}}$

Where $\hat{p} = \dfrac{X}{n}$ is the sample proportion of successes.

DECISION RULES for $H_0 : p = p_0$

(large sample, standardized test statistic)

$A.$	$H_1 : p > p_0$	Reject H_0 if $Z > z_\alpha$
$B.$	$H_1 : p < p_0$	Reject H_0 if $Z < -z_\alpha$
$C.$	$H_1 : p \neq p_0$	Reject H_0 if $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$

(small sample, Binomial distribution)

A, B or C: Reject H_0 if P-value $< \alpha$. **(Later)**

Test On a Single Proportion – Large Sample

- Example: A commonly prescribed [used] drug for relieving nervous tension is believed to be only 60% effective.
- Experimental results with a new drug administered to a random sample of 100 adults who were suffering from nervous tension show that 70 received relief.
- Is this sufficient evidence to conclude that the new drug is better than the current standard relief 60%?, i.e., is it better than the one commonly prescribed. Use $\alpha = 0.05$)

Test On a Single Proportion – Large Sample

Example 3. Hypotheses: H_0: $p = 0.6$ vs H_1: $p > 0.6$.

Since $n = 100$, we can use the normal approximation.

Decision Rule: Reject H_0 if $Z > 1.645$

Computations: $x = 70$, $n = 100$ $\hat{p} = \frac{x}{n} = 0.70$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.7 - 0.6}{\sqrt{(0.6)(0.4)/100}} = 2.04.$$

Decision: Reject H_0 since $z_{obs} = 2.04 > 1.645$

Conclusion: The new drug is superior to the one commonly used.

P-value = P(Z>2.04)<0.0207 Very small!

Test On a Single Proportion – Small Sample

Using Binomial Distribution for HT in Small Samples

Test On a Single Proportion – Small Sample

- Steps for testing a proportion for small samples:

1. H_0: $p = p_0$ vs.

- A. H_1: $p > p_0$, B. H_1: $p < p_0$, or C. H_1: $p \neq p_0$.

2. Choose α, the level of significance = P(Type I error),

3. Test statistic: X is Binomial with parameters n, and p (= p_0),

- X = number of "success"es in the sample,

4. Compute the P-value, based on the observed value of X.

5. Make your decision and draw appropriate conclusions based on the rejection rule (large sample, normal approximation) or the P-value (small sample, Binomial distribution).

Test On a Single Proportion – Small Sample

- Example 2. A builder claims that heat pumps are installed in at least 70% of all homes constructed after 2000 in İstanbul.
- A real estate agent disagrees and claims that the actual percentage of such homes with heat pumps is much less.
- Assume a random sample of 60 newly built houses are selected at random from İstanbul, and inspected for heat pumps. 35 of these houses has an installed heat pump.
- Devise a hypothesis testing problem to settle this argument. Using 0.10 level of significance.

Test On a Single Proportion – Small Sample

- SOLUTION: First, we will write the hypotheses:
- H_0: $p = 0.7$ vs. H_1: $p < 0.7$.

- **Test Stats:** X:The number of built homes with a heat pump

  X: a Binomial variable with $p = 0.7$ and n.

- **Sample Proportion**: $\hat{p} = \frac{35}{60} = 0.583 < 0.7 \Longrightarrow$ Evidence for H1.

Test On a Single Proportion – Small Sample

- H_0: p = 0.7   vs. H_1: p < 0.7.
- **Test statistic:**
  - Binomial variable X with p = 0.7 and n = 60.
  - X_{obs} = 35
- Compute P-value = P(X ≤ 35 when p = 0.7)

$$\sum_{x=0}^{35} b(x;60,0.7) = 0.0362.$$

- Hence P-value ≈ 0.04.

Test On a Single Proportion – Small Sample

- H_0: $p = 0.7$ vs. H_1: $p < 0.7$.
- Test statistic: Binomial variable X with $p = 0.7$ and $n = 60$.
- **P-value** ≈ 0.04.
- **Decision:** We reject H_0 since P-value < $\alpha = 0.10$.
- **Conclusion:** The data support the real estate agent's claim that the true proportion of homes with installed heat pumps is below 70%.