PROPORTION TEST

HYPOTHESIS TESTING

prepared by:

Gyro A. Madrona
Electronics Engineer

TOPIC OUTLINE

1-Proportion Test

2-Proportion Test

ANOM

1-PROPORTION TEST

1-PROPORTION TEST

1-Proportion test is a statistical method used to determine whether a **sample proportion** ($\hat{p}_1$) differs significantly from a hypothesized population proportion (P_o).

Test Statistic

$$z = \frac{\hat{p}_1 - P_o}{\sqrt{\frac{P_o(1 - P_o)}{n}}}$$

where:

z = z-statistic

$\hat{p}_1$ = sample proportion $\left(\frac{x}{n}\right)$

P_o = hypothesized population proportion

x = number of success in the sample

n = sample size

BINOMIAL TEST

syntax

```
from scipy import stats
result = stats.binomtest(
    k = number of success,
    n = sample size,
    p = population proportion,
    alternative = 'two-sided'
)
p_value = result.pvalue
```

Null Hypothesis

$$H_o: P_1 = P_o$$

Alternative Hypothesis

$$H_a: P_1 \neq P_o \text{ (p-value} \leq \alpha)$$

Assumption

- Discrete data

Z-TEST

syntax

```
from statsmodels.stats.proportion
import proportions_ztest

z_stat, p_value = proportions_ztest(
    count = number of success,
    nobs = number of observations,
    value = population proportion,
    alternative = 'two-sided'
)
```

Null Hypothesis

$$H_o: P_1 = P_o$$

Alternative Hypothesis

$$H_a: P_1 \neq P_o \text{ (p-value} \leq \alpha)$$

Assumption

- Discrete data

EXERCISE

In a survey of **1250** people, **600** preferred product A. Test if this is significantly **different** from the expected **50%** preference.

Solution

Let $\alpha = 0.05$

Null Hypothesis

$$\boldsymbol{H_o: P_1 = 0.5}$$

Alternative Hypothesis

$$\boldsymbol{H_a: P_1 \neq 0.5} \text{ (p-value} \leq 0.05)$$

2-PROPORTION TEST

2-PROPORTION TEST

2-Proportion test is a statistical method used to determine whether the proportions of **two independent groups** $(\hat{p}_1, \hat{p}_2)$ are significantly different from each other.

Test Statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where:

z = z-statistic

$\hat{p}_1$ = sample proportion of group 1 $\left(\frac{x_1}{n_1}\right)$

$\hat{p}_2$ = sample proportion of group 2 $\left(\frac{x_2}{n_2}\right)$

p = pooled proportion $\left(\frac{x_1+x_2}{n_1+n_2}\right)$

x_1, x_2 = number of success in each group

n_1, n_2 = sample size in each group

Z-TEST

syntax

```
from statsmodels.stats.proportion
import proportions_ztest

z_stat, p_value = proportions_ztest(
    count = [success_1, success_2],
    nobs = [size_1, size_2],
    alternative = 'two-sided'
)
```

Null Hypothesis

$$H_o: P_1 = P_2$$

Alternative Hypothesis

$$H_a: P_1 \neq P_2 \text{ (p-value} \leq \alpha)$$

Assumption

- Discrete data

EXERCISE

A company produces two types of circuit boards, Board A and Board B. In a quality test:

- **35** out of **150** Board A samples were defective
- **25** out of **120** Board B samples were defective

Is there a significant **difference** in the defect rates between Board A and Board B at a **5%** significance level?

Solution

Let $\alpha = 0.05$

Null Hypothesis

$\boldsymbol{H_o}$: **Board A = Board B**

Alternative Hypothesis

$\boldsymbol{H_a}$: **Board A ≠ Board B** (p-value ≤ 0.05)

ANOM

ANOM

ANOM (Analysis of Mean) is a statistical method used to test whether the **means** of **several groups** $(\hat{p}_1, \hat{p}_2, \hat{p}_3, \dots \hat{p}_n)$ differ significantly from the overall mean.

Test Statistic

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where:

x^2 = chi-square test statistic

O_i = number of observation in each group

E_i = expected number of observation in each group

CHI-SQUARE TEST FOR PROPORTION

syntax

```
from statsmodels.stats.proportion
import proportions_chisquare
chi_stat, p_value, table =
proportions_chisquare(
    counts = [success_1,success_2,
        success_3,...success_n],
    nobs = [observation_1,observation_2,
        observation_3,...observation_n],
)
```

Null Hypothesis

$$H_o: P_1 = P_2 = P_3 = \cdots P_n$$

Alternative Hypothesis

H_a: at least 1 ≠ (p-value ≤ α)

Assumption

- Discrete data

EXERCISE

A company produces two types of circuit boards, Board A and Board B. In a quality test:

- **35** out of **150** Board A samples were defective
- **25** out of **120** Board B samples were defective
- **30** out **85** Board C samples were defective

Is there a significant **difference** in the defect rates between the boards at a **5%** significance level?

Solution

Let $\alpha = 0.05$

Null Hypothesis

$\boldsymbol{H_o}$: **Board A = Board B = Board C**

Alternative Hypothesis

$\boldsymbol{H_a}$: **at least 1 board is different** (p-value ≤ 0.05)

LABORATORY