SESSIONS 4

Inferential Statistics: Hypothesis Testing Part 3

Data Science Program

Purwadhika
Startup and Coding School

Hypothesis test for proportion

- Z test single population proportion
- Z test double population proportion
- Chi-Squared test more than two population proportion

Purwadhika
Startup and Coding School

Use Case: A/B Testing categorical measurements

1. An newly launched online store want to test whether their conversion rate is greater than 50% or not. So they **gathered 1000** visitor **data randomly** and observed whether they are buying or not. As a Data Scientist you are asked to perform a hypothesis testing whether their conversion rate is greater than 50%?
2. A few months later the online store launched a new web design. They **randomly assign** visitor to layout A or layout B and 1000 for each design. You are asked again to perform hypothesis testing whether the conversion rate is increasing or not ?

Similar Ex. Amazon, Ben, Ubisoft

Z-Test Single Population Proportion

Assumption:

1. Dichotomy categorical variable with population proportion ***po***
2. Data collected using randomization
3. sample size (n) large enough:

   n x ***po*** ≥ 15 and n x (1 - ***po***) ≥ 15

Hypothesis:

Ho : ***p*** = ***po*** (ex. ***p*** = 0.5)

Ho : ***p*** ≠ ***po*** (two sided) or

p < ***po*** or ***p*** > ***po*** (one sided)

Test Statistics z:

$$z = \frac{\hat{p} - p_0}{se_0} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Rejection Criteria (spicy) :

P-value ≤ α (two-sided)

P-value/2 ≤ α (one-sided)

Purwadhika
Startup and Coding School

Z-Test Double Population Proportion

Assumption:

1. Dichotomy categorical variable observed in two group
2. Data collected using randomization
3. sample size (n1 and n2) large enough that there are five successive and five failure in each group.

Hypothesis:

Ho : ***p1*** = ***p2*** (ex.***p1*** - ***p2*** = 0)

Ho : ***p*** ≠ ***po*** (two sided) or

p1 < ***p2*** or ***p1*** > ***p2*** (one sided)

Test Statistics z:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} \text{ with } se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $\hat{p}$ is the pooled estimate.

Rejection Criteria (spicy) :

P-value ≤ α (two-sided)

P-value/2 ≤ α (one-sided)

Purwadhika
Startup and Coding School

Use Case: A/B Testing categorical measurements more than two groups

- A few months later the online store launched a new web design. They **randomly assign** visitor to layout A (current design), layout B1, or layout B2 and 1000 for each design.
- You are asked again to perform hypothesis testing whether the conversion rate is increasing or not and which design perform best?

Similar Ex. Zalora Case

Chi-Squared Test for Independence between Two Categorical Variables

Assumption:

1. Two categorical variable :

   check out (yes or no) vs web design (design A, B1, B2)

2. Randomization

Hypothesis:

Ho : The two variables are independent

Ho : The two variables are dependent.

Test Statistics z:

$$X^2 = \sum \frac{(observed\ count - expected\ count)^2}{expected\ count}$$

Rejection Criteria: P-value ≤ α

Purwadhika
Startup and Coding School

A/B Testing part 2

- Logistics Regression should be used in A/B Testing for Zalora use case.
- A/B Testing with more advanced method (ex using logistics regression) should be discuss in module 3 after learning regression and logistics regression.

Goodness of Fit Test

- We need normality assumption for some test
- We can check the distribution but that is very vulnerable to subjectivity
- To Test if sample data follow certain distribution (ex. Normal)
- Commonly used goodness of fit test:
  - Chi - square
  - Anderson - darling
  - Shapiro - wilk
  - Kolmogorov – smirnov

Hypothesis

Ho : The sample follow certain distribution

H1 : The sample do not follow the distribution

Rejection Region : P-Value < α

Non-Parametric

Non-parametric Statistics usually used when

- We don't have enough data (small sample size)
- We can't assume that our data population follow certain distribution. ex. result of our normality test is reject Ho

	Parametric	Non-parametric
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers
Tests		
Correlation test	Pearson	Spearman
Independent measures, 2 groups	Independent-measures t-test	Mann-Whitney test
Independent measures, >2 groups	One-way, independent-measures ANOVA	Kruskal-Wallis test
Repeated measures, 2 conditions	Matched-pair t-test	Wilcoxon test
Repeated measures, >2 conditions	One-way, repeated measures ANOVA	Friedman's test

Hypothesis testing for mean: Non-Parametrics

- **Mann Whitney** for independence double population mean
- **Wilcoxon** for paired mean test
- **Kruskal Walis** for more than two population mean

- ✓ These method are usually used when normality assumption doesn't meet
- ✓ For example: Income data or money related data usually have right skew distribution

Wilcoxon

Assumption:

1. numerical value with population mean **Mo**
2. Data collected using randomization
3. Test for median of population
4. sample size too small

Test Statistics : Sumrank

Rejection Criteria (spicy) :

P-value ≤ α (two-sided)

P-value/2 ≤ α (one-sided)

Hypothesis:

Ho : **M** = **Mo** (ex. **μ** = 40)

Ho : **M** ≠ **Mo** (two sided) or

M < **Mo** or **μ** > **μo** (one sided)

Purwadhika
Startup and Coding School

Mann Whitney

Assumption:

1. numerical variable (ordinal, interval, ratio) in each group
2. Data collected using randomization
3. Test for median of population
4. sample size too small

Test Statistics : U-Statistics

Rejection Criteria (spicy) :

P-value ≤ α (two-sided)

P-value/2 ≤ α (one-sided)

Hypothesis:

Ho : **M1** = **M2** (ex. **M1** - **M2** = 0)

Ho : **M1** ≠ **M2** (two sided) or

M1 < **M2** or **M1** > **M2** (one sided)

Purwadhika

Startup and Coding School

Kruskal-Wallis

Assumption:

1. The population distribution of the numerical variable is assumed to be normal
2. Data collected using randomization
3. Test for median of population
4. sample size too small

Hypothesis:

Ho : **M1** = **M2** = … = **Mk** (there are k groups)

Ho : at least one pair of the population are not equal (and we don't know which)

Test Statistics: V-Statistics

Rejection Region: p-value < α

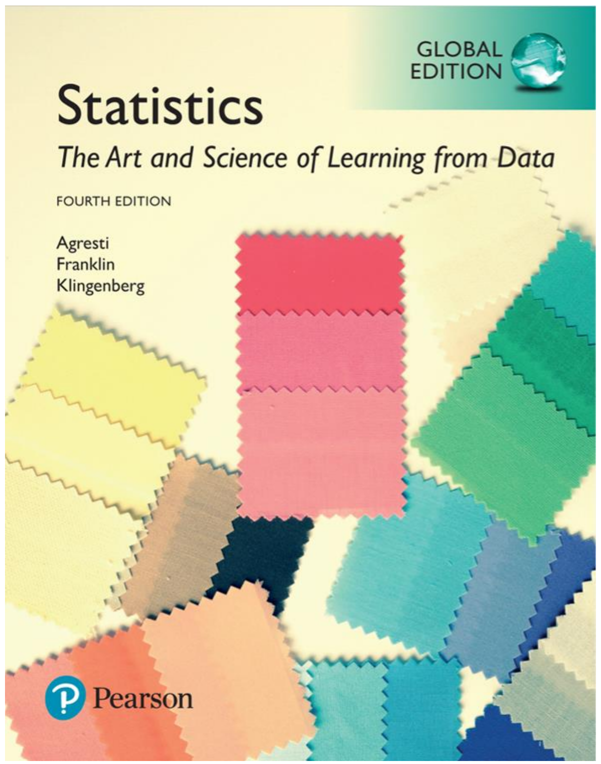
Purwadhika
Startup and Coding School

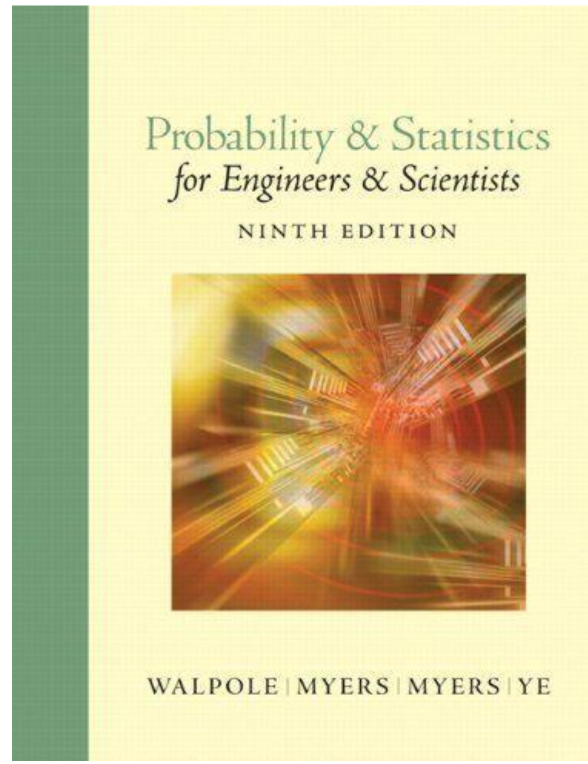
Another example of Hypothesis testing

Methods that you have learned can also be applied to another cases:

- Medicine, to understand if a drug works or not
- Economics, to understand human behavior
- Foreign aid and charitable work (the reputable ones at least), to understand which interventions are most effective at alleviating problems (health, poverty, etc)

Reference





Purwadhika
Startup and Coding School

Reference

https://towardsdatascience.com/data-science-you-need-to-know-a-b-testing-f2f12aff619a

https://towardsdatascience.com/data-science-fundamentals-a-b-testing-cb371ceecc27

https://www.niagahoster.co.id/blog/ab-testing-adalah/

https://vwo.com/blog/ab-testing-examples/

https://www.scribbr.com/methodology/sampling-methods/

Purwadhika
Startup and Coding School