Statistical Inference

Parametric Tests II

Contents

Paired samples t-test

- The paired sample t-test is used to determine whether the mean difference between two sets of observations is zero ,where each subject or entity is measured twice resulting in pair of observations.

- Commonly used when observations are recorded ‘before’ and ‘after’ the treatment / training and objective is to test whether the treatment/training is effective.

Case Study

To execute Parametric test in Python, we shall consider the below case as an example.

Background

The company organized a training program to improve efficiency. Time taken to complete MIS report before and after training are recorded for 15 employees.

Objective

To test whether the average time taken to complete MIS before and after training is not different.

Sample Size

Sample size: 15
Variables: time_before, time_after

Data Snapshot

PAIRED t TEST

Variables

time_before	time_after
85	74
95	91
92	80
102	91

servations

Columns	Description	Type	Measurement	Possible values
time_before	Time to complete MIS report before training	Numeric	Hours	Positive values
time_after	Time to complete MIS report after training	Numeric	Hours	Positive values

Assumptions for paired sample t-test

- The assumptions of the paired-sample t-test are listed below:

  - Random sampling from a defined population
    (employees are selected at random from the company)

  - Population of the testing variable is normally distributed
    (Difference time taken to complete MIS report should be normally distributed).

- Normality test can be performed by any of the methods explained earlier.

- The validity of the test is not seriously affected by moderate deviations from 'Normality' assumption.

Paired sample t-test

Testing whether means of two dependent groups are equal.

Objective	To test the average time taken to complete MIS before and after training is not different.

Null Hypothesis (H_0): There is no difference in average time before and after the training. i.e. D=0

Alternate Hypothesis (H_1):Average time is less after the training. (Training is effective.) D>0

D= μBefore – μAfter

Test Statistic	$t = \frac{\bar{d}}{s_d / \sqrt{n}}$ Where $\bar{d}$ is the sample mean of the difference before-after , s_d is the sample standard deviation of the difference, n is the sample size of difference. The quantity t follows a distribution called as 't distribution' with n-1 degrees of freedom.
Decision	Reject the null hypothesis if p-value < 0.05

Computation

	Notation	Value
Sample Size	n	12
Mean difference (before-after)	$\bar{d}$	8.3333
Standard Deviation	s_d	3.9219
t	$t = \frac{\bar{d}}{s_d / \sqrt{n}}$	8.2295

Paired sample t-test in Python

```
# Import data
data=pd.read_csv('PAIRED t TEST.csv')

# t-test for paired samples
stats.ttest_rel(data['time_before'],data['time_after']
,alternative='greater')
```

- **data[‘time_before’]** and **data[‘time_after’]** are the variables under study.
- **ttest_rel()** from scipy, returns t & pvalue

* Before performing t test, normality test is done to ensure difference variable is normally distributed.

Paired sample t-test in Python

```
# Output:

Ttest_relResult(statistic=8.22948711672449, pvalue=4.918935850301797e-07)
```

Interpretation :

- Since p-value is <0.05, reject H0. Average time taken to

Ttest_relResult(statistic=8.22948711672449, pvalue=4.918935850301797e-07)

  training is effective.
- 95% C.I does not contain value of D=0 (under H0),reject H0.

t-test for Correlation

- Correlation coefficient summarizes the strength of a linear relationship between two variables.
- t-test is used to check if there is significant correlation between two variables.
- Sample correlation coefficient (r) is calculated using bivariate data.
- Null hypothesis of this test is
  H0: there is no correlation between 2 variables under study ($\rho=0$)

Case Study

To execute Parametric test in Python, we shall consider the below case as an example.

Background

A company with 25 employees has calculated job proficiency score & aptitude test score for its employees

Objective

To test if there is significant correlation between job proficiency and aptitude test score.

Sample Size

Sample size: 25
Variables: Empcode, Aptitude, Job_prof

Data Snapshot

Correlation test

Variables

Empcode	aptitude	job_prof
E101	86	88
E102	62	80
E103	110	96
E104	101	76
E105	100	80
E106	78	73

bservations

Columns	Description	Type	Measurement	Possible values
Empcode	Employee code	Numeric	-	
Aptitude	Score of aptitude test	Numeric	-	Positive values
Job_prof	Job proficiency score	Numeric	-	Positive values

Correlation t-test

Testing for correlation coefficient value.

Objective	To test whether there exists significant correlation between job proficiency and aptitude score.

Null Hypothesis (H_0): There is no correlation between Job proficiency and Aptitude test

Alternate Hypothesis (H_1):There is correlation between Job proficiency and Aptitude test.

Test Statistic	$t = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}}$ where r is the sample correlation coefficient, the sample size. The quantity t follows a distribution called as 't distribution' with n-2 degrees of freedom.
Decision Criteria	Reject the null hypothesis if p-value < 0.05

Computation

	Notation	Value
Sample Size	n	25
Sample correlation coefficient	r	0.514411
t	$t = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}}$	2.8769

Correlation t-test in Python

```
# Import data
data=pd.read_csv('Correlation test.csv')

# t-test for correlation
stats.pearsonr(data['aptitude'], data['job_prof'])
```

- **data['aptitude']** and **data['job_prof ']**are the variables under study.
- **pearsonr()** from scipy, returns t & pvalue

Correlation t-test in Python

```
# Output:

(0.5144106946654772, 0.00851721615248713 7)
```

Interpretation :

- Since p-value is <0.05, reject H0. There is correlation between aptitude test and job proficiency.
- 95% C.I does not contain value $\rho=0$ (under H0), reject H0.

Quick Recap

In this session, we continued to learn various parametric tests . Here is a quick recap :

Paired sample t test	• Used to determine whether the mean difference between two sets of observations is zero ,where each subject or entity is measured twice resulting in pair of observations. • H0: μ1-μ2=d=0
t test for correlation	• Used to check if there is significant correlation between two variables. • H0: ρ=0