

Hypothesis Testing

By

Dr Shaik A Qadeer

Professor MJCET

Hypothesis Testing..

- Types of Hypothesis testing: Parametric and non-parametric
 1. Parametric tests: They make use population parameters such as mean, standard deviations etc. Example: Z-test, T-test, ANOVA etc.
 2. Non- Parametric tests: They make use data distribution to comment on the claim. Example Chi-Square etc.
- Few examples of the null hypothesis are as follows:
 1. Children who drink the health drink Complan are likely to grow taller
 2. Women use camera phone more than men (Freier, 2016)
 3. Vegetarians miss few flights (Siegel, 2016)
 4. Smokers are better sales people

ANOVA(Analysis of Variance)

- **One-way ANOVA** – used to study the impact of a single treatment at different levels on a continuous response variable.
- The null and alternative hypothesis for comparing 3 groups are given by.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_A: \text{Not all } \mu \text{ values are equal}$$

ANOVA(Analysis of Variance)...:Example

- Question: Ms Rachael Khanna the brand manager of ENZO detergent powder at the “one-stop” retail was interested in understanding whether the price discounts have any impact on the sales quantity of ENZO. To test whether the price discounts had any impact, price discounts of 0%, 10%, and 20% were given on randomly selected days. The quantity of ENZO sold in a day under different discount levels is shown in Table 3.1. Conduct a one way ANOVA to check whether discount had any significant impact on the average sales quantity at $\alpha = 0.05$

ANOVA(Analysis of Variance)...:Example..

TABLE 3.1 Sales for different discount levels

No Discount (0% discount)									
39	32	25	25	37	28	26	26	40	29
37	34	28	36	38	38	34	31	39	36
34	25	33	26	33	26	26	27	32	40
10% Discount									
34	41	45	39	38	33	35	41	47	34
47	44	46	38	42	33	37	45	38	44
38	35	34	34	37	39	34	34	36	41
20% Discount									
42	43	44	46	41	52	43	42	50	41
41	47	55	55	47	48	41	42	45	48
40	50	52	43	47	55	49	46	55	42

ANOVA(Analysis of Variance)...:Solution

- Read the records from the file and print the first few records

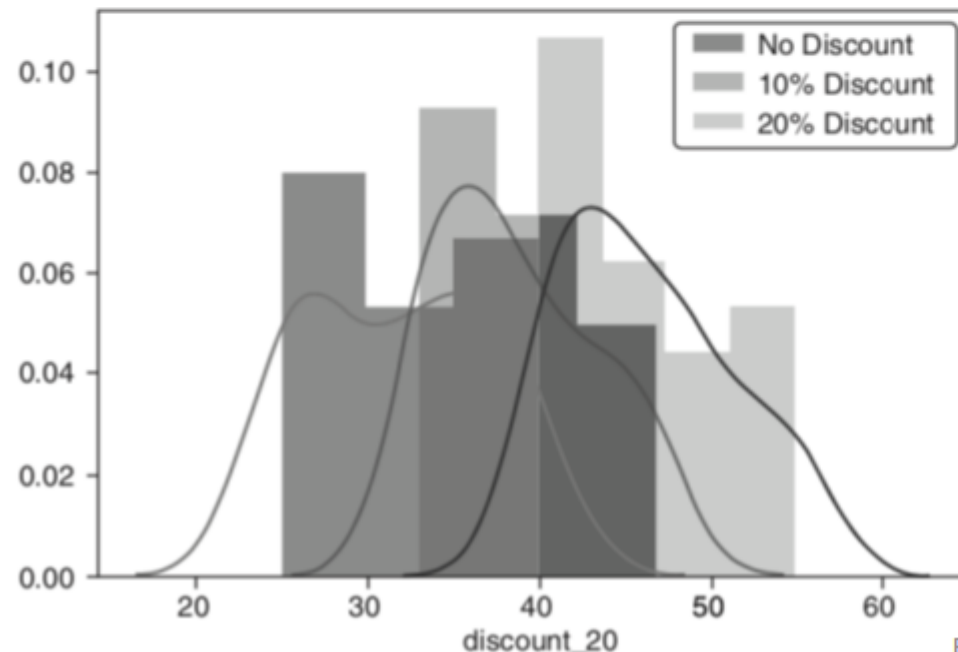
```
onestop_df = pd.read_csv('onestop.csv')  
onestop_df.head(5)
```

	discount_0	discount_10	discount_20
0	39	34	42
1	32	41	43
2	25	45	44
3	25	39	46
4	37	38	41

ANOVA(Analysis of Variance)...:Solution

- Distribution plot of the groups

```
sn.distplot(onestop_df['discount_0'], label = 'No Discount')  
sn.distplot(onestop_df['discount_10'], label = '10% Discount')  
sn.distplot(onestop_df['discount_20'], label = '20% Discount')  
plt.legend();
```



By Dr Shaik Qadeer

FIGURE 3.11 Comparing distributions of sales for different discount levels

ANOVA(Analysis of Variance)...:Solution

- Conducting the test

```
from scipy.stats import f_oneway  
  
f_oneway(onestop_df['discount_0'],  
         onestop_df['discount_10'],  
         onestop_df['discount_20'])
```

```
F_onewayResult(statistic=65.8698, pvalue=0.00)
```

- P-value is less than 0.05 value.
- We reject the null hypothesis.
- i.e., the mean sales quantity values under different discounts are different.

Non-Parametric Test(Chi-Square Goodness for fit test)

- It is a non-parametric test use for comparing the observed distribution of data with the expected distribution of the data
- Chi-square statistics is given by

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

- Where O_i is the observed frequency
- E_i is the expected frequency of the i th category.

Chi-Square Goodness for fit test.. Example

- Question: Hanuman Airlines (HA) operated daily flights to several Indian cities. One of the problems HA faces is the food preferences by the passengers. Captain Cook, the operations manager of HA, believes that 35% of their passengers prefer vegetarian food, 40% prefer non-vegetarian food, 20% low calorie food, and 5% request for diabetic food. A sample of 500 passengers was chosen to analyze the food preferences and the observed frequencies are as follows:

1. Vegetarian: 190
2. Non-vegetarian: 185
3. Low-calorie: 90
4. Diabetic: 35

Conduct a chi-square test to check whether Captain Cook's belief is true at $\alpha=0.05$.

Chi-Square Goodness for fit test.. Solution

- Parameters used for chi-square test
 1. `f_obs`: array_like – Observed frequencies in each category
 2. `f_exp`: array_like – Expected frequencies in each category
- From the data we can create the following arrays:

```
## Observed frequencies  
f_obs = [190, 185, 90, 35]  
## Expected frequencies from the percentages expected  
f_exp = [500*0.35, 500*0.4, 500*.2, 500*0.05]  
print(f_exp)
```

```
[175.0, 200.0, 100.0, 25.0]
```

Chi-Square Goodness for fit test.. Solution

- Conducting the test

```
stats.chisquare(f_obs, f_exp)
```

```
Power_divergenceResult(statistic=7.4107, pvalue=0.0598)
```

- P-value is more than 0.05 value.
- We retain the null hypothesis.
- That is, Captain Cook's belief is true.