IBM Developer
SKILLS NETWORK

Hypothesis Testing

Estimated time needed: **30** minutes

The goal of hypothesis testing is to answer the question, “Given a sample and an apparent effect, what is the probability of seeing such an effect by chance?” The first step is to quantify the size of the apparent effect by choosing a test statistic (t-test, ANOVA, etc). The next step is to define a null hypothesis, which is a model of the system based on the assumption that the apparent effect is not real. Then compute the p-value, which is the probability of the null hypothesis being true, and finally interpret the result of the p-value, if the value is low, the effect is said to be statistically significant, which means that the null hypothesis may not be accurate.

Objectives

- Import Libraries
- Lab exercises
  - Stating the hypothesis
  - Levene's Test for equality
  - Preparing your data for hypothesis testing
- Quiz

Import Libraries

All Libraries required for this lab are listed below. The libraries pre-installed on Skills Network Labs are commented. If you run this notebook in a different environment, e.g. your desktop, you may need to uncomment and install certain libraries.

In [1]:

```
#install specific version of libraries used in lab
#! mamba install pandas==1.3.3
#! mamba install numpy=1.21.2
#! mamba install scipy=1.7.1-y
#!  mamba install seaborn=0.9.0-y
#!  mamba install matplotlib=3.4.3-y
#!  mamba install statsmodels=0.12.0-y
```

Import the libraries we need for the lab

In [2]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.stats
```

Read in the csv file from the URL using the request library

In [3]:

```
ratings_url = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ST0151EN-SkillsNetwork/labs/teachingratings.csv'
ratings_df = pd.read_csv(ratings_url)
```

Lab Exercises

T-Test: Using the teachers' rating data set, does gender affect teaching evaluation rates?

We will be using the t-test for independent samples. For the independent t-test, the following assumptions must be met.

- One independent, categorical variable with two levels or group
- One dependent continuous variable
- Independence of the observations. Each subject should belong to only one group. There is no relationship between the observations in each group.
- The dependent variable must follow a normal distribution
- Assumption of homogeneity of variance

State the hypothesis

- $H\_0: µ\_1 = µ\_2$ ("there is no difference in evaluation scores between male and females")
- $H\_1: µ\_1 ≠ µ\_2$ ("there is a difference in evaluation scores between male and females")

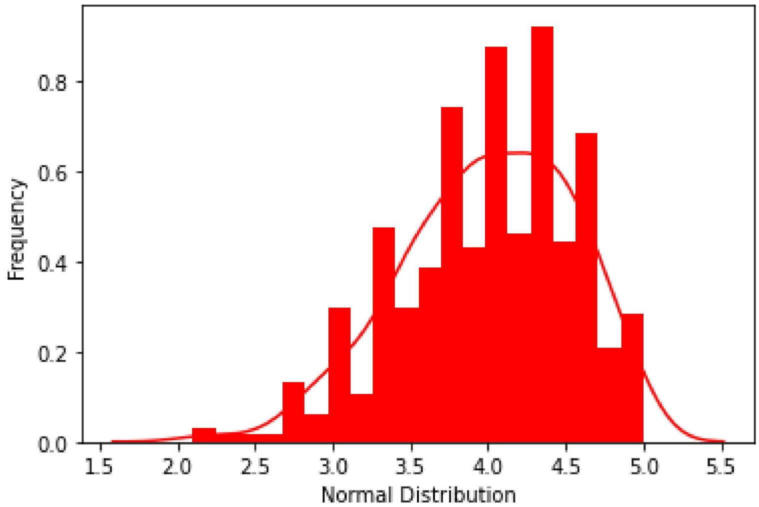
We can plot the dependent variable with a historgram

In [4]:

```
ax = sns.distplot(ratings_df['eval'],
                  bins=20,
                  kde=True,
                  color='red',
                  hist_kws={"linewidth": 15,'alpha':1})
ax.set(xlabel='Normal Distribution', ylabel='Frequency')
## we can assume it is normal
```

Out[4]:

```
[Text(0.5, 0, 'Normal Distribution'), Text(0, 0.5, 'Frequency')]
```



We can use the Levene's Test in Python to check test significance

In [5]:

```
scipy.stats.levene(ratings_df[ratings_df['gender'] == 'female']['eval'],
                   ratings_df[ratings_df['gender'] == 'male']['eval'], center='mean')

# since the p-value is greater than 0.05 we can assume equality of variance
```

Out[5]:

```
LeveneResult(statistic=0.19032922435292574, pvalue=0.6628469836244741)
```

Use the `ttest_ind` from the `scipy_stats` library

In [6]:

```
scipy.stats.ttest_ind(ratings_df[ratings_df['gender'] == 'female']['eval'],
                   ratings_df[ratings_df['gender'] == 'male']['eval'], equal_var = True
)
```

Out[6]:

```
Ttest_indResult(statistic=-3.249937943510772, pvalue=0.001238760944952221
7)
```

Conclusion: Since the p-value is less than alpha value 0.05, we reject the null hypothesis as there is enough proof that there is a statistical difference in teaching evaluations based on gender

ANOVA: Using the teachers' rating data set, does beauty score for instructors differ by age?

First, we group the data into cateries as the one-way ANOVA can't work with continuous variable - using the example from the video, we will create a new column for this newly assigned group our categories will be teachers that are:

- 40 years and younger
- between 40 and 57 years
- 57 years and older

In [7]:

```
ratings_df.loc[(ratings_df['age'] <= 40), 'age_group'] = '40 years and younger'
ratings_df.loc[(ratings_df['age'] > 40)&(ratings_df['age'] < 57), 'age_group'] = 'betwe
en 40 and 57 years'
ratings_df.loc[(ratings_df['age'] >= 57), 'age_group'] = '57 years and older'
```

State the hypothesis

- $H_0: \mu_1 = \mu_2 = \mu_3$ (the three population means are equal)
- $H_1:$ At least one of the means differ

Test for equality of variance

In [8]:

```
scipy.stats.levene(ratings_df[ratings_df['age_group'] == '40 years and younger']['beaut
y'],
                   ratings_df[ratings_df['age_group'] == 'between 40 and 57 years']['be
auty'],
                   ratings_df[ratings_df['age_group'] == '57 years and older']['beauty'
],
                   center='mean')
# since the p-value is less than 0.05, the variance are not equal, for the purposes of
 this exercise, we will move along
```

Out[8]:

```
LeveneResult(statistic=8.60005668392584, pvalue=0.000215366180993476)
```

First, separate the three samples (one for each job category) into a variable each.

In [9]:

```
forty_lower = ratings_df[ratings_df['age_group'] == '40 years and younger']['beauty']
forty_fiftyseven = ratings_df[ratings_df['age_group'] == 'between 40 and 57 years']['be
auty']
fiftyseven_older = ratings_df[ratings_df['age_group'] == '57 years and older']['beauty'
]
```

Now, run a one-way ANOVA.

In [10]:

```
f_statistic, p_value = scipy.stats.f_oneway(forty_lower, forty_fiftyseven, fiftyseven_o
lder)
print("F_Statistic: {0}, P-Value: {1}".format(f_statistic,p_value))
```

```
F_Statistic: 17.597558611010122, P-Value: 4.3225489816137975e-08
```

Conclusion: Since the p-value is less than 0.05, we will reject the null hypothesis as there is significant evidence that at least one of the means differ.

ANOVA: Using the teachers' rating data set, does teaching evaluation score for instructors differ by age?

Test for equality of variance

In [11]:

```
scipy.stats.levene(ratings_df[ratings_df['age_group'] == '40 years and younger']['eval'
],
                   ratings_df[ratings_df['age_group'] == 'between 40 and 57 years']['ev
al'],
                   ratings_df[ratings_df['age_group'] == '57 years and older']['eval'],
                   center='mean')
```

Out[11]:

```
LeveneResult(statistic=3.820237661494229, pvalue=0.022621418520219 39)
```

In [12]:

```
forty_lower_eval = ratings_df[ratings_df['age_group'] == '40 years and younger']['eval'
]
forty_fiftyseven_eval = ratings_df[ratings_df['age_group'] == 'between 40 and 57 years'
]['eval']
fiftyseven_older_eval = ratings_df[ratings_df['age_group'] == '57 years and older']['ev
al']
```

In [13]:

```
f_statistic, p_value = scipy.stats.f_oneway(forty_lower_eval, forty_fiftyseven_eval, fi
ftyseven_older_eval)
print("F_Statistic: {0}, P-Value: {1}".format(f_statistic,p_value))
```

```
F_Statistic: 1.2226327996572206, P-Value: 0.29540894225417536
```

Conclusion: Since the p-value is greater than 0.05, we will fail to reject the null hypothesis as there is no significant evidence that at least one of the means differ.

Chi-square: Using the teachers' rating data set, is there an association between tenure and gender?

State the hypothesis:

- $H\_0:$ The proportion of teachers who are tenured is independent of gender
- $H\_1:$ The proportion of teachers who are tenured is associated with gender

Create a Cross-tab table

In [14]:

```
cont_table  = pd.crosstab(ratings_df['tenure'], ratings_df['gender'])
cont_table
```

Out[14]:

gender	female	male
tenure		
no	50	52
yes	145	216

Use the `scipy.stats` library and set correction equals False as that will be the same answer when done by hand, it returns: χ2 value, p-value, degree of freedom, and expected values.

In [15]:

```
scipy.stats.chi2_contingency(cont_table, correction = True)
```

Out[15]:

```
(2.20678166999886,
 0.1374050603563787,
 1,
 array([[ 42.95896328,  59.04103672],
        [152.04103672, 208.95896328]]))
```

Conclusion: Since the p-value is greater than 0.05, we fail to reject the null hypothesis. As there is no sufficient evidence that teachers are tenured as a result of gender.

Correlation: Using the teachers rating dataset, Is teaching evaluation score correlated with beauty score?

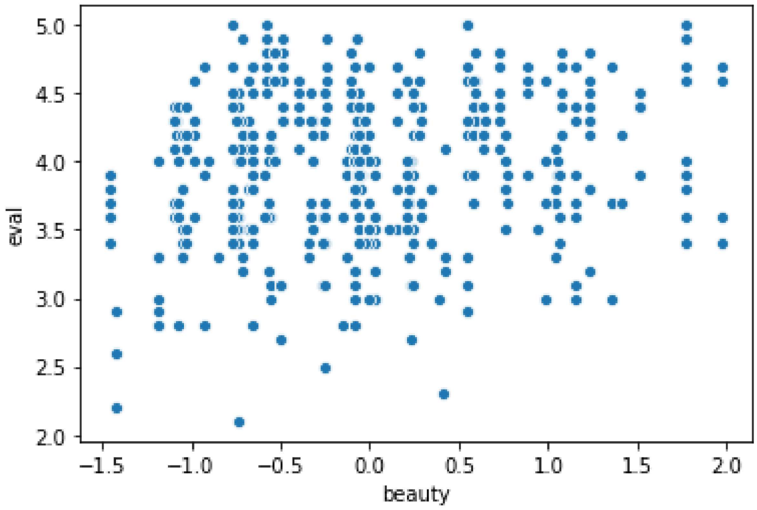
State the hypothesis:

- $H\_0:$ Teaching evaluation score is not correlated with beauty score
- $H\_1:$ Teaching evaluation score is correlated with beauty score

Since they are both continuous variables we can use a pearson correlation test and draw a scatter plot

In [16]:

```
ax = sns.scatterplot(x="beauty", y="eval", data=ratings_df)
```



In [17]:

```
scipy.stats.pearsonr(ratings_df['beauty'], ratings_df['eval'])
```

Out[17]:

```
(0.1890390908404521, 4.247115419812614e-05)
```

Conclusion: Since the p-value (Sig. (2-tailed) < 0.05, we reject the Null hypothesis and conclude that there exists a relationship between beauty and teaching evaluation score.

Practice Questions

Question 1: Using the teachers rating data set, does tenure affect teaching evaluation scores?

- Use α = 0.05

In [18]:

```
## insert code here
scipy.stats.ttest_ind(ratings_df[ratings_df['tenure'] == 'yes']['eval'],
                      ratings_df[ratings_df['tenure'] == 'no']['eval'], equal_var = True)
```

Out[18]:

```
Ttest_indResult(statistic=-2.8046798258451777, pvalue=0.005249471210198792)
```

Double-click **here** for the solution.

Question 2: Using the teachers rating data set, is there an association between age and tenure?

- Discretize the age into three groups 40 years and youngers, between 40 and 57 years, 57 years and older (This has already been done for you above.)
- What is your conclusion at α = 0.01 and α = 0.05?

In [25]:

```
## insert code here
cont_table  = pd.crosstab(ratings_df['tenure'], ratings_df['age_group'])
print(cont_table)
## state your hypothesis
##Null Hypothesis: There is no association between age and tenure
##Alternative Hypothesis: There is an association between age and tenure


scipy.stats.chi2_contingency(cont_table, correction = True)


#At the α = 0.01, p-value is greater, we fail to reject null hypothesis as
#there is no evidence of an association between age and tenure


#At the α = 0.05, p-value is less, we reject null hypoothesis as
#there is evidence of an association between age and tenure
```

```
age_group  40 years and younger  57 years and older  between 40 and 57 years
tenure
no                           15                  25                       62
yes                          98                  97                      166
```

Out[25]:

```
(8.749576239010711,
 0.012590809706820843,
 2,
 array([[ 24.89416847,  26.87688985,  50.22894168],
        [ 88.10583153,  95.12311015, 177.77105832]]))
```

Double-click **here** for a hint.

Double-click **here** for the solution.

Question 3: Test for equality of variance for beauty scores between tenured and non-tenured instructors

- Use α = 0.05

In [31]:

```
scipy.stats.levene(ratings_df[ratings_df['tenure'] == 'yes']['beauty'],
                   ratings_df[ratings_df['tenure'] == 'no']['beauty'],
                   center='mean')
```

Out[31]:

```
LeveneResult(statistic=0.4884241652750426, pvalue=0.4849835158609811)
```

Double-click **here** for the solution.

Question 4: Using the teachers rating data set, is there an association between visible minorities and tenure?

- Use α = 0.05

In [34]:

```
## insert code here
#cont_table  = pd.crosstab(ratings_df['tenure'], ratings_df['minority'])
#print(cont_table)
#scipy.stats.chi2_contingency(cont_table, correction = True)


cont_table  = pd.crosstab(ratings_df['vismin'], ratings_df['tenure'])
print(cont_table)
scipy.stats.chi2_contingency(cont_table, correction = True)
```

```
tenure  no  yes
vismin
0       92  307
1       10   54
```

Out[34]:

```
(1.3675127484429763,
 0.24223968800237178,
 1,
 array([[ 87.90064795, 311.09935205],
        [ 14.09935205,  49.90064795]]))
```

Double-click **here** for a hint.

Double-click **here** for the solution.

Authors

Aije Egwaikhide (https://www.linkedin.com/in/aije-egwaikhide/?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=SkillsNetwork-Channel-SkillsNetworkCoursesIBMDeveloperSkillsNetworkST0151ENSkillsNetwork20531532-2021-01-01) is a Data Scientist at IBM who holds a degree in Economics and Statistics from the University of Manitoba and a Post-grad in Business Analytics from St. Lawrence College, Kingston. She is a current employee of IBM where she started as a Junior Data Scientist at the Global Business Services (GBS) in 2018. Her main role was making meaning out of data for their Oil and Gas clients through basic statistics and advanced Machine Learning algorithms. The highlight of her time in GBS was creating a customized end-to-end Machine learning and Statistics solution on optimizing operations in the Oil and Gas wells. She moved to the Cognitive Systems Group as a Senior Data Scientist where she will be providing the team with actionable insights using Data Science techniques and further improve processes through building machine learning solutions. She recently joined the IBM Developer Skills Network group where she brings her real-world experience to the courses she creates.

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-08-14	0.1	Aije Egwaikhide	Created the initial version of the lab