



Department of Computer Science and Engineering (Data Science)
B.Tech. Sem: III Subject: Statistics for Data Science
Experiment 7

Name: Rishabh Singhvi

SAP ID:60009210206

Date:	Experiment Title:
Aim	To study test of significance for small samples.
Software	Google Colab
Theory	<p>Question 1: Given 11 ages, check whether average age is 30 or not.</p> <pre>[3]: def pvalue(test, prob): if test == 'two-tailed': p_value = prob else: p_value = prob/2 return p_value</pre> <p>Q1.</p> <pre>[6]: age = [32, 34, 22, 29, 22, 39, 38, 37, 38, 36, 30] print("Null Hypothesis is that average age is 30") print("Alternate Hypothesis is that average age is not 30") alpha=0.05 t_statistics, prob=stats.ttest_1samp(age,30) test='two-tailed' p_value=pvalue(test,prob) print("t-statistics is:",t_statistics, "and p value is:",p_value) if p_value>alpha: print("Failed to reject null hypothesis for level of significance = ",alpha) else: print(f"Null hypothesis is rejected for level of significance ={alpha}")</pre> <p>Null Hypothesis is that average age is 30</p> <p>Alternate Hypothesis is that average age is not 30</p> <p>t-statistics is: 1.3263370175675118 and p value is: 0.21422150477795807</p> <p>Failed to reject null hypothesis for level of significance = 0.05</p> <pre>[5]: from scipy.stats import t t_statistics=(st.mean(age)-30)/(st.stdev(age)/math.sqrt(len(age))) print(t_statistics) print(2*(1-t.cdf(t_statistics,10)))</pre> <p>1.326337017567512</p> <p>0.2142215047779581</p> <p>Q2.</p> <p>Question 2: Test the hypothesis that average score in math is less than 50 from the scores given below.</p> <p>10, 20, 35, 50, 28, 40, 55, 18, 16, 55, 30, 25, 65, 18, 43, 23, 21, 20, 19, 70</p>

```
[0]: marks = [10,20,35,50,20,40,55,10,16,55,30,25,65,18,43,23,21,20,19,70]
print("Null Hypothesis is that average score in math is 50")
print("Alternate Hypothesis is that average score in math is not 50")
alpha=0.05
t_statistics, prob=pvalue(stats.ttest_ind(marks,50))
test='left-tailed'
p_value=pvalue(test,prob)
print("t-statistics is:",t_statistics, "and p value is:",p_value )
if p_value>alpha:
    print("Failed to reject null hypothesis for level of significance = ",alpha)
else:
    print(f"Null hypothesis is rejected for level of significance = {alpha}")
```

Null Hypothesis is that average score in math is 50

Alternate Hypothesis is that average score in math is not 50

t-statistics is: -4.284640191237414 and p value is: 0.00020006970464996746

Null hypothesis is rejected for level of significance = 0.05

Question 3: Consider two populations following Poisson distribution with mean age 35 years and 10 years respectively. The size of population 1 is 150000 and that of population 2 is 100000. Combine both of them to make one population. Two samples are taken from Gujarat state with size 30 and 20 and mean value 30 and 10 respectively. Combine both samples to make one sample. Test the hypothesis that mean age of Gujarat is equal to mean age of the country.

```
: from scipy.stats import norm

pop1_mean=35
pop1_var=pop1_mean
pop1_size=150000
pop2_mean=10
pop2_var=pop2_mean
pop2_size=100000
pop_comb_mean=(pop1_size*pop1_mean + pop2_size*pop2_mean)/(pop1_size + pop2_size)
pop_comb_var=(pop1_size*pop1_var + pop2_size*pop2_var)/(pop1_size + pop2_size) + (pop1_size*pop2_size)*(pop1_mean-pop2_mean)**2/(pop1_size + pop2_size)**2
pop_comb_std = math.sqrt(pop_comb_var)

samp1_mean = 30
samp2_mean = 10
samp1_size = 30
samp2_size = 20

samp_comb_mean=(samp1_size*samp1_mean + samp2_size*samp2_mean)/(samp1_size + samp2_size)
print("Null Hypothesis is that mean age of Gujarat is equal to mean age of the country.")
print("Alternate Hypothesis is that mean age of Gujarat is not equal to mean age of the country.")
alpha=0.05
test='two-tailed'
z_statistics=(samp_comb_mean - pop_comb_mean)/(pop_comb_std/math.sqrt(samp1_size + samp2_size))
prob=2*(norm.sf(abs(z_statistics)))
p_value=pvalue(test,prob)
print("z-statistics is:",z_statistics, "and p value is:",p_value )
if p_value>alpha:
    print("Failed to reject null hypothesis for level of significance = ",alpha)
else:
    print(f"Null hypothesis is rejected for level of significance = {alpha}")
```

Null Hypothesis is that mean age of Gujarat is equal to mean age of the country.

Alternate Hypothesis is that mean age of Gujarat is not equal to mean age of the country.

z-statistics is: -1.6035674514745464 and p value is: 0.10800943004054568

Failed to reject null hypothesis for level of significance = 0.05

Question 4: Check the hypothesis that mean values of temperature for week 1 and week 2 are the same.

```
: temp_week1 = [25, 22, 22, 29, 22, 29, 18, 25, 28, 20, 24, 16, 25, 24, 21, 28, 29, 25, 27, 33, 25, 28, 28, 26,
temp_week2 = [18, 31, 34, 21, 28, 39, 35, 27, 33, 25, 28, 30, 26, 26, 26, 26, 19, 32, 24, 28, 20, 22, 32, 35, 29,
print("Null Hypothesis is that the mean values of temperature for week 1 and week 2 are same.")
print("Alternate Hypothesis is that the mean values of temperature for week 1 and week 2 are not same.")
alpha=0.05
test='two-tailed'
t_statistics, prob=stats.ttest_ind(temp_week1,temp_week2)
p_value=pvalue(test,prob)
print("t-statistics is:",t_statistics, "and p value is:",p_value )
if p_value>alpha:
    print("Failed to reject null hypothesis for level of significance = ",alpha)
else:
    print(f"Null hypothesis is rejected for level of significance = {alpha}")
```

Null Hypothesis is that the mean values of temperature for week 1 and week 2 are same.

Alternate Hypothesis is that the mean values of temperature for week 1 and week 2 are not same.

t-statistics is: -3.1146814975012775 and p value is: 0.002415039212632959

Null hypothesis is rejected for level of significance = 0.05

Question 5: Consider patient's data with blood pressure values before and after treatment. Check the effectiveness of treatment.

```
[13]: from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[14]: path = "/content/drive/MyDrive/SDS dataset/blood_pressure.csv"
data = pd.read_csv(path)
data

[14]: ,
/
/
.....

   patient  sex  agegrp  bp_before  bp_after
0         1  Male   30-45        143        153
1         2  Male   30-45        163        170
2         3  Male   30-45        153        168
3         4  Male   30-45        153        142
4         5  Male   30-45        146        141
...     ...   ...     ...         ...         ...
115      116  Female  60+         152         152
116      117  Female  60+         161         152
117      118  Female  60+         165         174
118      119  Female  60+         149         151
119      120  Female  60+         185         163

/
120 rows x 5 columns

/
.....
/
```

```
[16]: #incorrect
print("Null Hypothesis is that there is no change in BP before and after treatment")
print("Alternate Hypothesis is that BP before treatment is more than the BP after treatment")
alpha=0.05
test='right-tailed'
t_statistics, prob=stats.ttest_ind(data['bp_before'],data['bp_after'])
p_value=pvalue(test,prob)
print("t-statistics is:",t_statistics, "and p value is:",p_value )
if p_value>alpha:
    print("Failed to reject null hypothesis for level of significance = ",alpha)
else:
    print(f"Null hypothesis is rejected for level of significance = {alpha}")
```

Null Hypothesis is that there is no change in BP before and after treatment

Alternate Hypothesis is that BP before treatment is more than the BP after treatment

t-statistics is: 3.0669836819036274 and p value is: 0.0012061387390394456

Null hypothesis is rejected for level of significance = 0.05

```
[18]: print("Null Hypothesis is that there is no change in BP before and after treatment")
print("Alternate Hypothesis is that BP before treatment is more than the BP after treatment")
Diff=data['bp_before']-data['bp_after']
alpha=0.05
test='right-tailed'
t_value,prob=stats.ttest_1samp(Diff,0)
p_value=pvalue(test,prob)
print("t-statistics is:",t_value, "and p value is:",p_value )
if p_value>alpha:
    print("Failed to reject null hypothesis for level of significance = ",alpha)
else:
    print(f"Null hypothesis is rejected for level of significance = {alpha}")
```

Null Hypothesis is that there is no change in BP before and after treatment

Alternate Hypothesis is that BP before treatment is more than the BP after treatment

t-statistics is: 3.3371870510833657 and p value is: 0.0005648957322420411

Null hypothesis is rejected for level of significance = 0.05

Question 6. Construct a 95% Confidence Interval for population variance based on the following data:

40.2, 32.8, 38.2, 43.5, 47.6, 36.6, 38.4, 45.5, 44.4, 40.3, 34.6, 55.6, 50.9, 38.9, 37.8, 46.8, 43.6, 39.5

```
from scipy.stats import chi2
data= [40.2, 32.8, 38.2, 43.5, 47.6, 36.6, 38.4, 45.5, 44.4, 40.3, 34.6, 55.6, 50.9, 38.9, 37.8, 46.8, 43.6, 39.5]
s_var=st.variance(data)
n=len(data)
dof=n-1
alpha=0.05
chi_L=chi2.ppf(0.975,dof)
chi_U=chi2.ppf(0.025,dof)
L=dof*s_var/chi_L
U=dof*s_var/chi_U
conf_interval=(L,U)
print(f"95% Confidence interval for a given population variance is {conf_interval}")
```

95% Confidence interval for a given population variance is (21.313726207189386, 78.61721262359482)

Question 7: Two samples having 21 and 9 observations have variances given as 16 and 8 respectively. Test the hypothesis that the first population variance is greater than the second one at 0.05 and 0.01 level of significance.

	<pre>[26]: from scipy.stats import f print("Null Hypothesis is that the first population variance is same as that of the second one.") print("Alternate Hypothesis is that the first population variance is greater than that of the second one.") n1,n2=21,9 samp_var1,samp_var2=16,8 pop_est_var1=n1*samp_var1/(n1-1) pop_est_var2=n2*samp_var2/(n2-1) if pop_est_var1 > pop_est_var2: f_statistics=pop_est_var1/pop_est_var2 dof1=n1-1 dof2=n2-1 else: f_statistics=pop_est_var2/pop_est_var1 dof1=n2-1 dof2=n1-1 p_value=f.sf(f_statistics,dof1,dof2) alphas=[0.05,0.01] print("f-statistics is:",f_statistics, "and p value is:",p_value) for alpha in alphas: if p_value>alpha: print("Failed to reject null hypothesis for level of significance = ",alpha) else: print(f"Null hypothesis is rejected for level of significance = {alpha}")</pre> <p>Null Hypothesis is that the first population variance is same as that of the second one.</p> <p>Alternate Hypothesis is that the first population variance is greater than that of the second one.</p> <p>f-statistics is: 1.8666666666666667 and p value is: 0.18409481717733314</p> <p>Failed to reject null hypothesis for level of significance = 0.05</p> <p>Failed to reject null hypothesis for level of significance = 0.01</p>
Conclusion	

Signature of Faculty