

-*- coding: utf-8 -*-

''''

Created on Wed Jun 22 22:12:32 2022

@author: Jeevi

Download the dataset by clicking this link

<https://visualize.data.gov.in/?inst=45a38dad-9dc8-4b99-860f-f1e6c3eb92c5>

Execute Hypothesis Test (One tail Test) X-axis is Year and Y-axis is Cyclonic Disturbance

Total. Your random samples should be greater than 30 following the Z statistics method

and the significance value is $\alpha=5\%$ its Z value is $Z=1.645$

(Graphical Visualization is Mandatory)

''''

```
from sklearn.metrics import mean_squared_error #to calculate mean square algorithm
```

```
import pandas as pd
```

```
import csv
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
from scipy import stats #mode calcula
```

```
from scipy.stats import norm
```

```
import statistics
```

```
import math #importing mat library
```

```
data= pd.read_csv('C:/Users/Public/acc/hypo.csv')
```

```
print(data.info())
```

```
# Creating a dataset
```

```
X = np.array(data['Year'])
```

```
print("X values in the dataset are:\n", X)
```

```

Y = np.array((data['Cyclonic Disturbances - TOTAL']))
print("Y valuesCyclonic Disturbances in the dataset are:\n", Y)

# Calculation of Mean Squared Error (MSE)
MSE = mean_squared_error(X,Y)
print("Mean Square Error dataset are:\n", MSE)


Med= np.median(Y)
print("Median dataset are:\n", Med)
Mea=np.mean(Y)
print("Mean dataset are:\n", Mea)
Mod=stats.mode(Y)
print("Mode dataset are:\n", Mea)


t_value,p_value=stats.ttest_1samp(Y,Mea) # tvaue calculation based on mean

one_tailed_p_value=float("{:.6f}".format(p_value/2)) # Since alternative hypothesis is one tailed, We
need to divide the p value by 2.

print('Test statistic is %f'%float("{:.6f}".format(t_value)))

print('p-value for one tailed test is Since alternative hypothesis is one tailed, We need to divide the p
value by 2.final we get %f'%one_tailed_p_value)

alpha = 0.05 # ALPHA 5%

if one_tailed_p_value<=alpha:

    print('Conclusion','n','Since p-value(=%f)%p_value,<','alpha(=%f)%alpha',"We reject the null
hypothesis H0.

```

So we conclude that there is no significant mean difference in Cyclonic Disturbances - TOTAL.

i.e., $\mu = 12.31007751$ at %.2f level of significance""%alpha)

else:

```
print('Conclusion','n','Since p-value(=%f)'%one_tailed_p_value,'>',alpha(=%.2f)'%alpha,""We do
not reject the null hypothesis H0."")
```

```
from statsmodels.stats import weightstats as stests
```

```
ztest ,pval = stests.ztest(data['Cyclonic Disturbances - TOTAL'], x2=None, value=Mea)
```

```
print("pvalues of data set for ztest :",float(pval))
```

```
if pval<0.05:
```

```
    print("reject null hypothesis")
```

```
else:
```

```
    print("accept null hypothesis")
```

```
#calculation through matplotlib to find skew left/right or normal
```

```
dat=np.array((data['Cyclonic Disturbances - TOTAL']))
```

```
# plt.hist(dat, bins=7, align='right',alpha=0.5, color='purple', edgecolor='black')
```

```
plt.xlabel('values Cyclonic Disturbances in the dataset are:\n")
```

```
meanDATA= statistics.mean(dat)
```

```
sdDATA = statistics.stdev(dat)
```

```
plt.plot(Y, norm.pdf(Y, meanDATA, sdDATA))
```

```
plt.show()
```

```
# plotting histogram
```

```
#plt.hist(norm.pdf(Y, meanDATA, sdDATA), 10)
```

```
# plotting mean line
```

```
#plt.axvline(norm.pdf(Y, meanDATA, sdDATA).mean(), color='k', linestyle='dashed', linewidth=2)
```

```
# showing the plot
```

```
plt.show()
```

```
# Calculating mean and standard deviation
```

```
Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\Jeevi\spyder-py3\assignment5.py

37
38 Med= np.median(Y)
39 print("Median dataset are:\n", Med)
40 Mea=np.mean(Y)
41 print("Mean dataset are:\n", Mea)
42 Mod=stats.mode(Y)
43 print("Mode dataset are:\n", Mea)
44
45
46 t_value,p_value=stats.ttest_1samp(Y,Mea) # tvaue calculation based on mean
47
48 one_tailed_p_value=float("{:.6f}".format(p_value/2)) # Since alternative hypothesis is o
49
50 print('Test statistic is %f'%float("{:.6f}".format(t_value)))
51
52 print('p-value for one tailed test is Since alternative hypothesis is one tailed, We nee
53
54 alpha = 0.05 # ALPHA 5%
55
56 if one_tailed_p_value<alpha:
57
58     print('Conclusion', 'n', 'Since p-value(=%f)' %p_value, '<', 'alpha(=%f)' %alpha, 'We r
59
60 So we conclude that there is no significant mean difference in Cyclonic Disturbances - n
61 i.e.,  $\mu = 12.31007751$  at  $\alpha$  level of significance'' %alpha)
62
63
64 else:
65
66     print('Conclusion', 'n', 'Since p-value(=%f)' %one_tailed_p_value, '>', 'alpha(=%f)' %al
67
68
69
70
71
72 from statsmodels.stats import weightstats as stests
73 ztest, pval = stests.ztest(data['Cyclonic Disturbances - TOTAL'], x2=None, value=Mea)
74 print("pvalues of data set for ztest : ",float(pval))
75 if pval<0.05:
76     print("reject null hypothesis")
77 else:
78     print("accept null hypothesis")
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```

memory usage: 13.2 KB

None

X values in the dataset are:

[1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019]

Y valuesCyclonic Disturbances in the dataset are:

[13 12 12 12 11 10 12 13 7 10 6 13 14 9 10 11 15 9 8 6 7 9 10 8 9 14 10 11 11 9 10 13 16 13 20 13 18 13 15 14 11 14 16 16 15 17 19 10 19 16 19 14 14 19 15 17 18 18 12 16 15 17 10 14 13 14 7 12 16 15 18 13 17 16 14 18 15 13 14 15 15 18 16 12 20 14 18 14 11 14 12 20 8 7 15 8 9 9 10 11 8 12 5 8 8 9 9 13 9 7 6 6 7 10 12 12 10 8 8 10 5 10 8 12 10 10 14 12]

Mean Square Error dataset are:

3775484.596899225

Median dataset are:

12.0

Mean dataset are:

12.310077519379846

Mode dataset are:

12.310077519379846

Test statistic is 0.000000

p-value for one tailed test is Since alternative hypothesis is one tailed, We need to divide the p value by 2.Final we get 0.500000

Conclusion n Since p-value(=0.500000) > alpha(=0.05) We do not reject the null hypothesis H0.

pvalues of data set for ztest : 1.0

```
Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\Jeevi\spyder-py3\assignment5.py

60 So we conclude that there is no significant mean difference in Cyclonic Disturbances - TO
61 i.e.,  $\mu = 12.31007751$  at  $\alpha$  level of significance'' %alpha)
62
63
64 else:
65
66     print('Conclusion', 'n', 'Since p-value(=%f)' %one_tailed_p_value, '>', 'alpha(=%f)' %al
67
68
69
70
71
72 from statsmodels.stats import weightstats as stests
73 ztest, pval = stests.ztest(data['Cyclonic Disturbances - TOTAL'], x2=None, value=Mea)
74 print("pvalues of data set for ztest : ",float(pval))
75 if pval<0.05:
76     print("reject null hypothesis")
77 else:
78     print("accept null hypothesis")
79
80 #calculation through matplotlib to find skew left/right or normal
81
82
83
84 dat=np.array([data['Cyclonic Disturbances - TOTAL']])
85 # plt.hist(dat, bins=7, align='right',alpha=0.5, color='purple', edgecolor='black')
86 plt.xlabel('values Cyclonic Disturbances in the dataset are:\n")
87
88 meanDATA= statistics.mean(dat)
89 sdDATA = statistics.stdev(dat)
90
91 plt.plot(Y, norm.pdf(Y, meanDATA, sdDATA))
92 plt.show()
93
94
95
96
97
98
99
100
```

1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019]

Y valuesCyclonic Disturbances in the dataset are:

[13 12 12 12 11 10 12 13 7 10 6 13 14 9 10 11 15 9 8 6 7 9 10 8 9 14 10 11 11 9 10 13 16 13 20 13 18 13 15 14 11 14 16 16 15 17 19 10 19 16 19 14 14 19 15 17 18 18 12 16 15 17 10 14 13 14 7 12 16 15 18 13 17 16 14 18 15 13 14 15 15 18 16 12 20 14 18 14 11 14 12 20 8 7 15 8 9 9 10 11 8 12 5 8 8 9 9 13 9 7 6 6 7 10 12 12 10 8 8 10 5 10 8 12 10 10 14 12]

Mean Square Error dataset are:

3775484.596899225

Median dataset are:

12.0

Mean dataset are:

12.310077519379846

Mode dataset are:

12.310077519379846

Test statistic is 0.000000

p-value for one tailed test is Since alternative hypothesis is one tailed, We need to divide the p value by 2.Final we get 0.500000

Conclusion n Since p-value(=0.500000) > alpha(=0.05) We do not reject the null hypothesis H0.

pvalues of data set for ztest : 1.0

accept null hypothesis

In [151]:

