Hypothesis Testing

By G M Nithin Sai | Guidance From Kanav Bansal | Updated on 22^{nd} May, 2021.

- Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis.

  KEY TAKEAWAYS

  - Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.
  - The test provides evidence concerning the plausibility of the hypothesis, given the data.
  - Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

  ## How Hypothesis Testing Works

  - In hypothesis testing, an analyst tests a statistical sample, with the goal of providing evidence on the plausibility of the null hypothesis.

- Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis.

- The null hypothesis is usually a hypothesis of equality between population parameters; e.g., a null hypothesis may state that the population mean return is equal to zero.

- The alternative hypothesis is effectively the opposite of a null hypothesis (e.g., the population mean return is not equal to zero). Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.

In [2]:

```
import matplotlib.pyplot as plt
import numpy as np

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

In [9]:

```
def tscore(sample_size,sample_mean,pop_mean,sample_std):
    nume=sample_mean-pop_mean
    denom=sample_std/(sample_size)**0.5
    return nume/denom
```

Understanding the Hypothesis Testing

Step - 1:

Alternate Hypothesis(Bold Claim): $ H_1 $ $\implies $ $\lt$, $\gt$, $\neq$
Null Hypothesis (status Quo): H_0 $\implies$ $\leq$, $\geq$, $=$

Step - 2:

- Collect a sample of size n
- Calculate the Mean of the sample ($\bar{x}$)

Step - 3: Compute test statistic:

- If standard deviation is known : $$ z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt[2]{n}}} $$
- If standard deviation is not known:
  $$ t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt[2]{n}}}$$

step - 4: Decide Significance level α. If the alpha is lower remember we need stronger evidence to reject Null hypothesis.

step - 5.1: Apply Decision Rule:

- If test statistic is Z - score:
  - Two Tailed Z-test:
    $$ |z| > z_\frac{\alpha}{2}\, \implies \, Accept H_1 or Reject H_0 $$ -Right Tailed Z-Test:
    $$ z > z_\alpha\, \implies \, Accept H_1 or Reject H_0 $$
  - Left Tailed Z-Test:
    $$ z < -z_\frac{\alpha}{2}\, \implies \, Accept H_1 or Reject H_0 $$
- If test statistic is t-score:
  - Two tailed test:
    $$ |t| > t_{n-1,\alpha/2} \implies Accept H_1 or Reject H_0 $$
  - Right Tailed t-test:
    $$ t > t_{n-1,\alpha} \implies Accept H_1 or Reject H_0 $$
  - Left tailed t-test:
    $$ t < t_{n-1,\alpha} \implies Accept H_1 or Reject H_0 $$

Step - 5.2 Compute P-Value $$ P(test statistics|H_0) $$

- For Two tailed test:
  $$ pvalue=2*(1.0-cdf(test statistic)) $$
- For one tailed test:
  $$ pvalue=(1.0-cdf(test statistic))$$ Now,
  $$ if(pvalue<\alpha)\implies Accept H_1 or Reject H_0 $$

Practical Example-1

Q-1: Pista House selling Hyderabadi Chicken Dum biryani claims that each parcel packet has 500 grams of biryani (also mentioned on the label of packet). You are sceptic of their claims and believe that on average each packet does not contain 500 grams of biryani. How do you prove your claim?

Step 1:

Alternate Hypothesis (Bold Claim) :
$$ H_1: On \,an\, Average\, each\, Chicken\, dum\, biryani\, packet \,does\, not\, weigh\, 500 \,grams$$ $$\mu \neq 500$$ Null Hypothesis (status quo):
$$ H_0: Pista\, House\, claims\, that\, each\, chickendum\, biryani \,packet \,weighs\, 500 grams $$ $$\mu = 500$$ step - 2:

- Collect a sample of size n=12

$$ [500,510,505,498,476,520,521,480,510,504,550,469]$$

- Compute sample mean
  $$\bar{x}=497.0 $$

Step - 3: compute Test Statistic:

since standard deviation is not provide we compute t score where,
$$ t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt[2]{n}}}=-0.5725114387470406$$
Step - 4: Decide α or significance level
$$\alpha=0.95$$ step - 5.1: Two Tailed t-test:
$$ Reject\, H_0\, if \, tscore\, >\, t\, critical $$ $$From\, the \,above\,condition \,we \,failed\,to\,reject\,Null\,Hypothesis$$
step - 5.2:Calculate P-value
$$Calculated \,P- value \,is\, :\, 0.20641847943818092 $$ $$if \,P -value > alpha\,: \,reject \,null\, hypothesis$$ $$ Hence,\,we\, failed\, to\, reject\, null\, hypohthesis$$

In [3]:

```
samples=[490,220,470,500,495,496,496,498,508,480]
```

```
sample_mean=np.array(samples).mean()
print("sample mean:",sample_mean)
def sample_std_(samples,sample_mean):
    summation=0
    for i in range(len(samples)):
        summation=summation+(samples[i]-sample_mean)**2

    return (summation/9)**0.5
sample_std=sample_std_(samples,sample_mean)
print("sample standard deviation:",sample_std)
```

```
sample mean: 465.3
sample standard deviation: 86.84859110991827
```

In [4]:

```
confidence_level=0.95
alpha=1-confidence_level
t_critical=t.ppf(1-alpha/2,df=9)
print(t_critical)
```

```
2.2621571627409915
```

In [5]:

```
sample_size=10
pop_mean=500
```

In [6]:

```
t_=tscore(sample_size,sample_mean,pop_mean,sample_std)
print(t_)
```

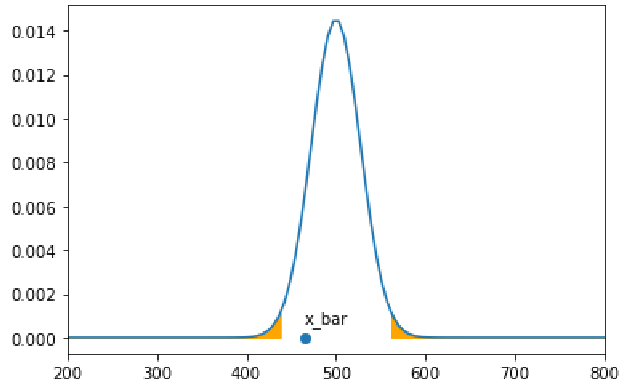
```
-1.2634751284446715
```

In [7]:

```
x_min = 200
x_max = 800


# Defining the sampling distribution mean and sampling distribution std
mean = pop_mean
std = sample_std / sample_size**0.5


# Ploting the graph and setting the x limits
x = np.linspace(x_min, x_max, 100)
y = norm.pdf(x,mean,std)
plt.xlim(x_min, x_max)
plt.plot(x, y)



# Computing the left and right critical values (Two tailed Test)
t_critical_left = pop_mean + (-t_critical * std)
t_critical_right = pop_mean + (t_critical * std)


# Shading the left rejection region
x1 = np.linspace(x_min, t_critical_left, 100)
y1 = norm.pdf(x1,mean, std)
plt.fill_between(x1, y1, color='orange')


# Shading the right rejection region
x2 = np.linspace(t_critical_right, x_max, 100)
y2 = norm.pdf(x2,mean,std)
plt.fill_between(x2, y2, color='orange')


# Ploting the sample mean and concluding the results
plt.scatter(sample_mean, 0)
plt.annotate("x_bar", (sample_mean, 0.0007))
```

Out[7]:

```
Text(465.3, 0.0007, 'x_bar')
```



In [8]:

```python
if(abs(t_) > t_critical):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
Fail to reject Null Hypothesis
```

In [9]:

```python
p_value = 2 * (1.0 - norm.cdf(np.abs(t_)))
print("p_value = ", p_value)
if(p_value < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
p_value =  0.20641847943818092
Fail to reject Null Hypothesis
```

Practical Example - 2

Q-2: You have developed a new Natural Language Processing Algorithms and done a user study. You claim that the average rating given by the users is greater than 4 on a scale of 1 to 5. How do you prove this to your client?

Step 1:

Alternate Hypothesis (Bold Claim) :
$$ H_1: Average\, rating\, given\, by\, the \,users\, is\, greater\, than \,4 $$ $$\mu>4$$ Null Hypothesis (status quo):
$$ H_0: Average\, rating\, given\, by \,the \,users\, i, less thanorequalto\, 4 $$ $$\mu\leq4$$ step - 2:

- Collect a sample of size n=12

$$ [4,3,5,4,5,3,5,5,4,2,4,5,5,4,4,5,4,5,4,5] $$

- Compute sample mean
  $$\bar{x}=4.25 $$

Step - 3: compute Test Statistic:

since standard deviation is not provide we compute t score where,
$$ t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt[2]{n}}}=1.3483997249264843 $$
Step - 4: Decide α or significance level
$$\alpha=0.95$$ step - 5.1: one Tailed t-test:

- Right Tailed t-test:
  $$ t > t_{n-1,\alpha} \implies Accept H_1 or Reject H_0 $$

$$From\, the \,above\,condition \,we \,failed\,to\,reject\,Null\,Hypothesis$$

step - 5.2:Calculate P-value
$$Calculated \,P- value \,is\, :\, 0.17752985241215358 $$ $$if \,P -value > alpha\,: \,reject \,null\, hypothesis$$ $$ Hence,\,we\, failed\, to\, reject\, null\, hypohthesis$$

In [10]:

```python
# One Tail - Calculating the z-critical value

confidence_level = 0.95
```

```
alpha = 1 - confidence_level

t_critical = t.ppf(1 - alpha,df=19)

print(t_critical)
```

```
1.729132811521367
```

In [11]:

```
sample_size = 20
sample_mean = 4.25
pop_mean=4
```

In [12]:

```
samples=[4,3,5,4,5,3,5,5,4,2,4,5,5,4,4,5,4,5,4,5]
sample_mean=np.array(samples).mean()
print("sample mean:",sample_mean)
def sample_std_(samples,sample_mean):
    summation=0
    for i in range(len(samples)):
        summation=summation+(samples[i]-sample_mean)**2

    return (summation/len(samples))**0.5
sample_std=sample_std_(samples,sample_mean)
print("sample standard deviation:",sample_std)
```

```
sample mean: 4.25
sample standard deviation: 0.82915619758885
```

In [13]:

```
t_=tscore(sample_size,sample_mean,pop_mean,sample_std)
print(t_)
```

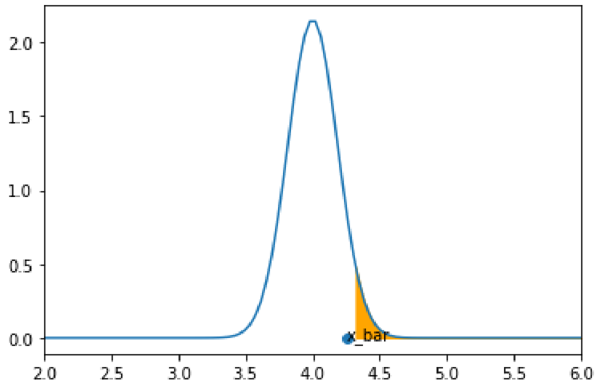
```
1.3483997249264843
```

In [14]:

```
#plotting
x_min = 2
x_max = 6


# Defining the sampling distribution mean and sampling distribution std
mean = pop_mean
std = sample_std / sample_size**0.5


# Ploting the graph and setting the x limits
x = np.linspace(x_min, x_max, 100)
y = norm.pdf(x,mean,std)
plt.xlim(x_min, x_max)
plt.plot(x, y)



# Computing the left and right critical values (Two tailed Test)
t_critical_left = pop_mean + (-t_critical * std)
t_critical_right = pop_mean + (t_critical * std)




# Shading the right rejection region
x2 = np.linspace(t_critical_right, x_max, 100)
y2 = norm.pdf(x2,mean,std)
plt.fill_between(x2, y2, color='orange')


# Ploting the sample mean and concluding the results
plt.scatter(sample_mean, 0)
plt.annotate("x_bar", (sample_mean, 0.0007))
```

Out[14]:

```
Text(4.25, 0.0007, 'x_bar')
```



In [15]:

```
if(abs(t_) > t_critical):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
Fail to reject Null Hypothesis
```

In [16]:

```
p_value = 2 * (1.0 - norm.cdf(np.abs(t_)))
print("p_value = ", p_value)
if(p_value < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
p_value =  0.17752985241215358
Fail to reject Null Hypothesis
```

Practical Example - 3

Q-3: TATA has developed a better fuel management system for the SUV segment. They claim that with this system, on average the SUV's mileage is at least 15 km/litre?

Step 1:

Alternate Hypothesis (Bold Claim) :
$$ H_1: Average SUV's mileage is less than 15km/hr $$ $$\mu<15km/hr$$ Null Hypothesis (status quo):
$$ H_0: Average SUV's mileage is greater than or equal to 15km/hr $$ $$\mu\geq15km/hr$$ step - 2:

- Collect a sample of size n=10

$$ [14.08, 14.13, 15.65, 13.78, 16.26, 14.97, 15.36, 15.81, 14.53, 16.79, 15.78, 16.98, 13.23, 15.43, 15.46, 13.88, 14.31, 14.41, 15.76, 15.38] $$

- Compute sample mean
  $$\bar{x}=15.1 $$

Step - 3: compute Test Statistic:

since standard deviation is not provide we compute t score where,
$$ t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt[2]{n}}}=0.44748014931892083 $$
Step - 4: Decide α or significance level
$$\alpha=0.95$$ step - 5.1: one Tailed t-test:

- Right Tailed t-test:
  $$ t > t_{n-1,\alpha} \implies Accept H_1 or Reject H_0 $$

$$From\, the \,above\,condition \,we \,failed\,to\,reject\,Null\,Hypothesis$$

step - 5.2:Calculate P-value
$$Calculated \,P- value \,is\, :\, 0.6545284174132713 $$ $$if \,P -value > alpha\,: \,reject \,null\, hypothesis$$ $$ Hence,\,we\, failed\, to\, reject\, null\, hypohthesis$$

In [17]:

```
# One Tail - Calculating the z-critical value
```

```
confidence_level = 0.95

alpha = 1 - confidence_level

t_critical = t.ppf(1 - alpha,df=19)

print(t_critical)
```

```
1.729132811521367
```

In [18]:

```
samples=[14.08, 14.13, 15.65, 13.78, 16.26, 14.97, 15.36, 15.81, 14.53, 16.79, 15.78, 16.98, 13.23, 15.43,
sample_mean=np.array(samples).mean()
print("sample mean:",sample_mean)
def sample_std_(samples,sample_mean):
    summation=0
    for i in range(len(samples)):
        summation=summation+(samples[i]-sample_mean)**2

    return (summation/len(samples))**0.5
sample_std=sample_std_(samples,sample_mean)
print("sample standard deviation:",sample_std)
```

```
sample mean: 15.099
sample standard deviation: 0.9994043225842081
```

In [19]:

```
sample_size = 20
sample_mean = 15.1
pop_mean = 15
pop_std = 1
```

In [20]:

```
t_=tscore(sample_size,sample_mean,pop_mean,sample_std)
print(t_)
```

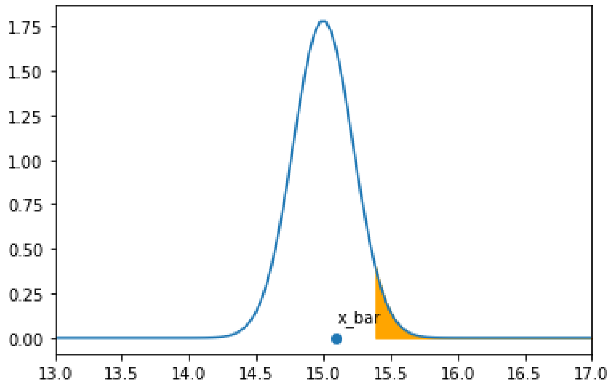
```
0.44748014931892083
```

In [21]:

```
x_min = 13
x_max = 17

mean = pop_mean
std = pop_std / (sample_size**0.5)

x = np.linspace(x_min, x_max, 100)
y = norm.pdf(x, mean, std)

plt.xlim(x_min, x_max)
# plt.ylim(0, 0.03)

plt.plot(x, y)

t_critical_right = pop_mean + (t_critical * std)

x1 = np.linspace(t_critical_right, x_max, 100)
y1 = norm.pdf(x1, mean, std)
plt.fill_between(x1, y1, color='orange')

plt.scatter(sample_mean, 0)
plt.annotate("x_bar", (sample_mean, 0.1))
```

Out[21]:

```
Text(15.1, 0.1, 'x_bar')
```



In [22]:

```
if(abs(t_) > t_critical):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
Fail to reject Null Hypothesis
```

In [23]:

```
p_value = 2 * (1.0 - norm.cdf(abs(t_)))
print("p_value = ", p_value)
if(p_value < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
p_value =  0.6545284174132713
Fail to reject Null Hypothesis
```

In [ ]:

Practical Example - 4

Q-4: You have developed a new Machine Learning Application and claim that on average it takes less than 100 ms to predict for any future datapoint. How do you convince your client about this claim?

Step 1:

Alternate Hypothesis (Bold Claim) :
$$ H_1: Average\,MachineLearning\, Model\,takes \,lessthan\,100ms\,to\,predict \,any\,future \,data point 4 $$
$$\mu < 100$$ Null Hypothesis (status quo):
$$ H_0: Average\,MachineLearning\, Model\,takes \,greater than or equal to\,100ms\,to\,predict \,any\,future \,data point $$
$$\mu\geq100$$ step - 2:

- Collect a sample of size n=10

  $$[120,100,98,88,115,88,78,99,121,102] $$

- Compute sample mean
  $$\bar{x}=100.9 $$

Step - 3: compute Test Statistic:

since standard deviation is not provide we compute t score where,
$$ t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt[2]{n}}}= 0.20987632600435635$$
Step - 4: Decide α or significance level
$$\alpha=0.99$$ step - 5.1: one Tailed t-test:

- Left tailed t-test:
  $$ t < t_{n-1,\alpha} \implies Accept H_1 or Reject H_0 $$ $$From\, the \,above\,condition \,we \,failed\,to\,reject\,Null\,Hypothesis$$

Step - 5.2:Calculate P-value
$$Calculated \,P- value \,is\, :\, 0.8337641998328718 $$ $$if \,P -value > alpha\,: \,reject \,null\, hypothesis$$ $$ Hence,\,we\, failed\, to\, reject\, null\, hypohthesis$$

In [4]:

```
alpha = 1 - 0.99

t_critical = t.ppf(1 - alpha,df=9)

print(t_critical)
```

```
2.8214379233005493
```

In [5]:

```
samples=[120,100,98,88,115,88,78,99,121,102]
sample_mean=np.array(samples).mean()
print("sample mean:",sample_mean)
def sample_std_(samples,sample_mean):
    summation=0
    for i in range(len(samples)):
        summation=summation+(samples[i]-sample_mean)**2

    return (summation/len(samples))**0.5
sample_std=sample_std_(samples,sample_mean)
print("sample standard deviation:",sample_std)
```

```
sample mean: 100.9
sample standard deviation: 13.560604706280618
```

In [12]:

```
sample_size =10
sample_mean=100.9
pop_mean=100
pop_std=15
```

In [13]:

```
t_=tscore(sample_size,sample_mean,pop_mean,sample_std)
print(t_)
```

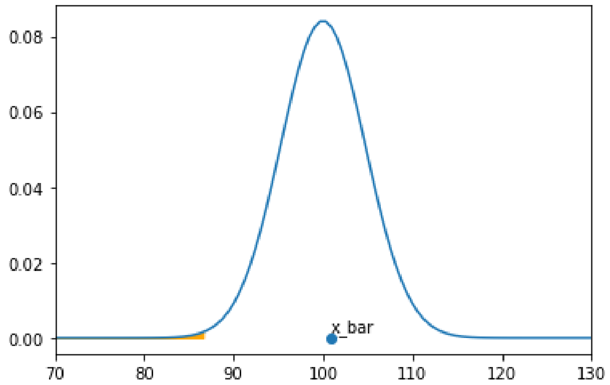
```
0.20987632600435635
```

In [17]:

```
x_min = 70
x_max = 130

mean = pop_mean
std = pop_std / (sample_size**0.5)

x = np.linspace(x_min, x_max, 100)
y = norm.pdf(x, mean, std)

plt.xlim(x_min, x_max)
# plt.ylim(0, 0.03)

plt.plot(x, y)

t_critical_left = pop_mean + (-t_critical * std)

x1 = np.linspace(x_min, t_critical_left, 100)
y1 = norm.pdf(x1, mean, std)
plt.fill_between(x1, y1, color='orange')

plt.scatter(sample_mean, 0)
plt.annotate("x_bar", (sample_mean, 0.002))
```

Out[17]:

```
Text(100.9, 0.002, 'x_bar')
```



In [15]:

```
if(abs(t_) > t_critical):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
Fail to reject Null Hypothesis
```

In [16]:

```
p_value = 2 * (1.0 - norm.cdf(abs(t_)))
print("p_value = ", p_value)
if(p_value < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

```
p_value =  0.8337641998328718
Fail to reject Null Hypothesis
```

In [ ]: