Z-Test

Z-Test

- A statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution.
- Because of the central limit theorem, many test statistics are approximately normally distributed for large samples
  - *(Usually when sample size ≥ 30)*
- Z-tests are closely related to t-tests, but t-tests are best performed when an experiment has a small sample size
- T-tests assume the standard deviation is unknown, while z-tests assume it is known.

Lets generate a Population

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

```
pop_mean=25
pop_std=30
pop_size=1_00_000
```

```
np.random.seed(123)
pop=np.random.normal(loc=pop_mean,
                     scale=pop_std,size=pop_size)
print(pop[:10])
```

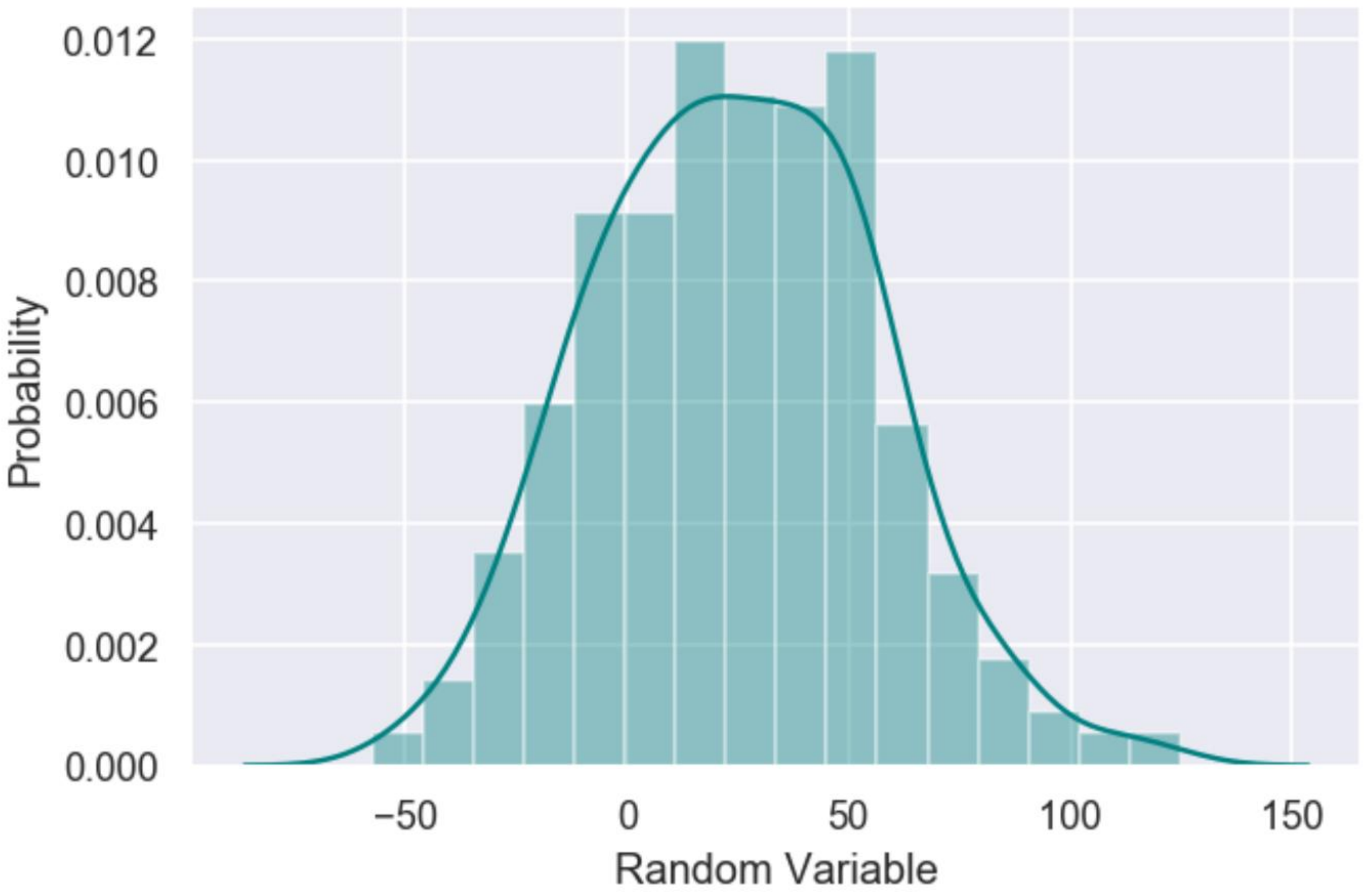
```
[ -7.5689181   54.9203634   33.48935494 -20.18884142   7.64199244
  74.54309611 -47.8003773   12.13262113  62.97808776  -1.00221207]
```

Extract a Sample from the Population

```
sample_size=500
np.random.seed(123)
sample=np.random.choice(pop,size=sample_size)
print("samples",sample[:10])
sample_mean=np.mean(sample)
print("Mean of the samples ",sample_mean)
```

```
samples [ 31.0199469  109.59719132  29.49145446 -15.75025585  19.19351326
  62.30673661  -7.81839642  39.48893064  16.41863051 -14.84027253]
Mean of the samples  24.547698264765508
```

Probability Distribution of the sample



Q. Does the sample have the same Characterstics of the Population

Q. Is the mean of the Sample equal to the mean of the Population

Lets Find our using Z-Test

Define the Hypothesis

Null Hypothesis

$$H_0 : \mu = 25$$

Alternative Hypothesis

$$H_1 : \mu \neq 25$$

Calculation of Standard Error for Mean

$$S.E. = \frac{\sigma}{\sqrt{(n)}}$$

s - Population SD

n - Number of samples

```
SE = pop_std/np.sqrt(sample_size)
print("Standard Error SE ",SE)
```

```
Standard Error SE  1.3416407864998738
```

Calculation of Z-Statistics

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

or

$$z = \frac{\bar{x} - \mu}{S.E}$$

```
zstatistics=(sample_mean-pop_mean)/SE
print("Zstatistics is ",zstatistics)
```

```
Zstatistics is  -0.3371258087751455
```

$\bar{x}$ - Sample Mean

μ - Population Mean

s - Standard deviation of population

n - Sample size

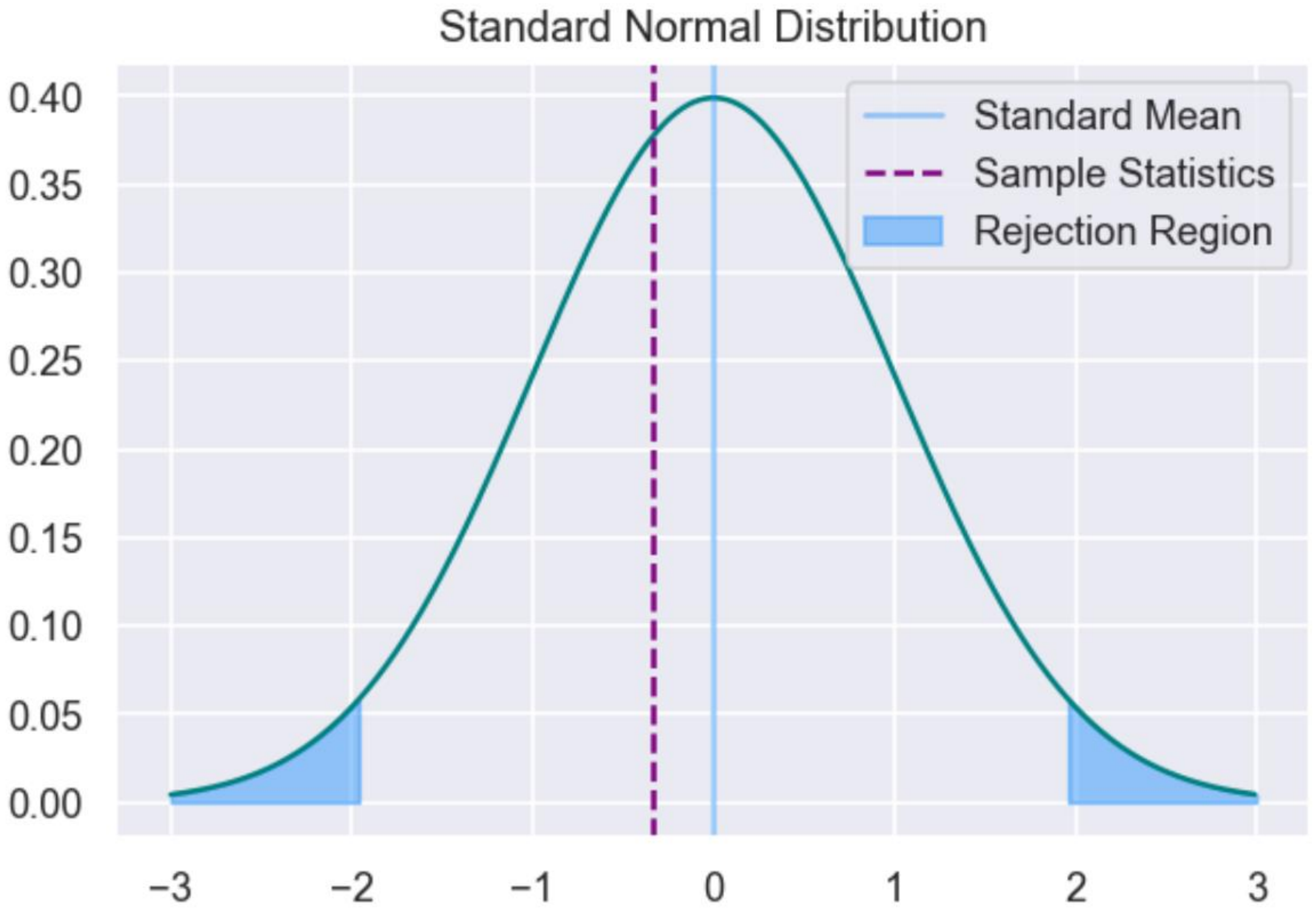
$S.E.$ - Standard Error for Mean

Finding the Z Critical Value

```
significant=0.05
zcritical_l=norm.ppf(q=significant/2)
zcritical_u=-zcritical_l
print("Critical Values are ",zcritical_l,zcritical_u)
```

```
Critical Values are  -1.9599639845400545 1.9599639845400545
```

Plot in Standard Normal Distribution



Confidence Interval

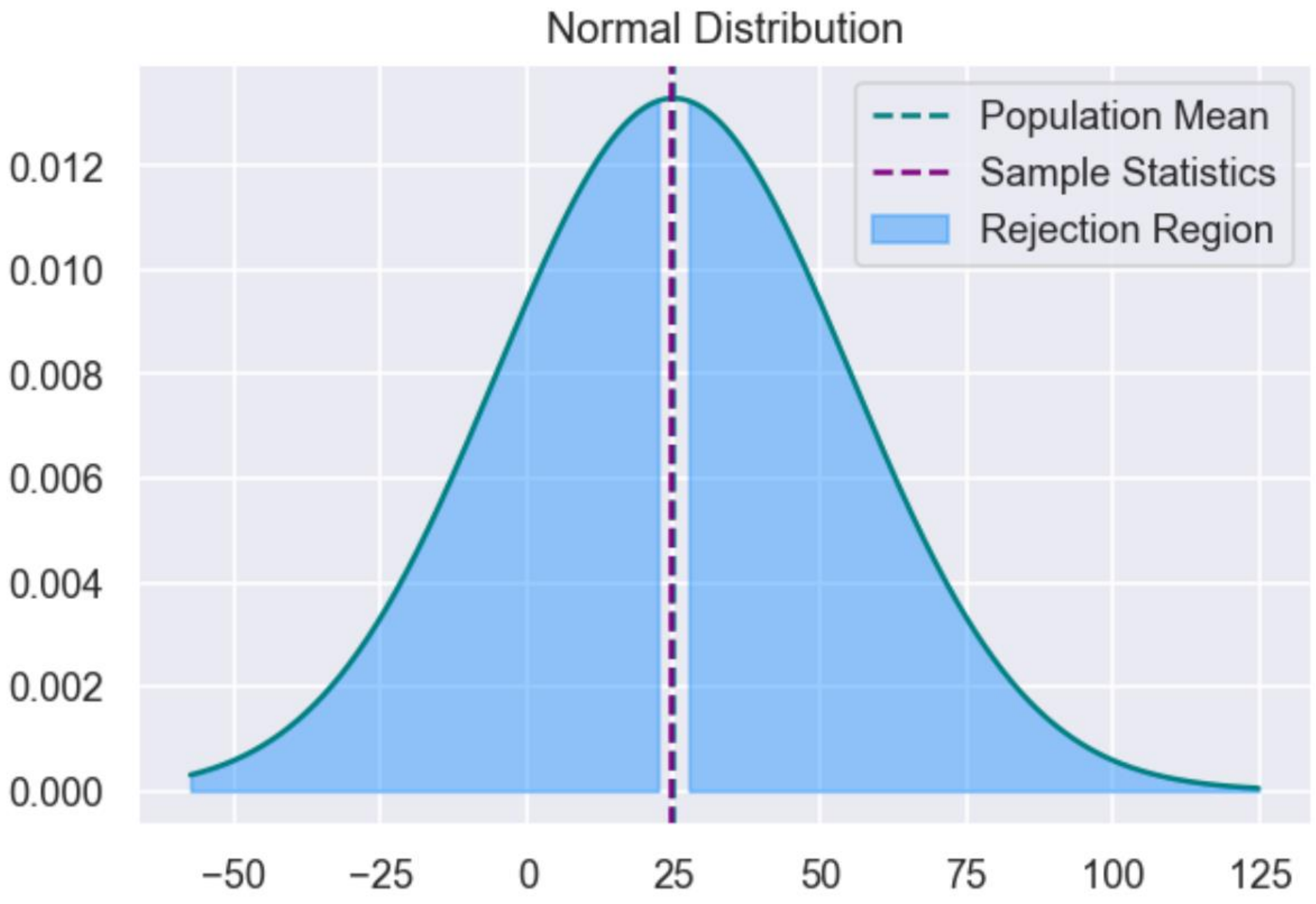
$$C.I. = \bar{x} \pm SE * critical_{\frac{\alpha}{2}}$$

```
CI_l=pop_mean+(zcritical_l*SE)
CI_u=pop_mean+(zcritical_u*SE)

print("Confidence Interval\nLower limit",CI_l,
      "\nUpper limit",CI_u)
```

```
Confidence Interval
Lower limit 22.370432378270255
Upper limit 27.629567621729745
```

Sample Distribution



Sample Distribution (Zoomed)