Department of Computer Science and Engineering (Data Science)
S.Y. B.Tech. Sem: IV Subject: Statistics for Data Science
Experiment 5

Name: Smayan Kulkarni **SAP ID: 60009230142**

Date:	**Experiment Title: Confidence Interval**
Aim	To implement Confidence Interval using Python.
Software	Google Colab
Theory	Explain the process of finding confidence interval.
Implementation	Using Python solve the following questions given below :

1. *If Z follows standard normal distribution, then find*

$(i)\ P(Z < 1.5)$ $(ii)\ P(Z > 0.5)$ $(iii)\ P(|Z| < 1.5)$ $(iv)\ P(|Z| > 0.5)$

$(v)\ P(-2.2 < Z < 1)$

Code:

```
prob_1i = stats.norm.cdf(1.5)
print(f"(i) P(Z < 1.5) = {prob_1i:.4f}")

prob_1ii = 1 - stats.norm.cdf(0.5)
print(f"(ii) P(Z > 0.5) = {prob_1ii:.4f}")

prob_1iii = stats.norm.cdf(1.5) - stats.norm.cdf(-1.5)
print(f"(iii) P(|Z| < 1.5) = {prob_1iii:.4f}")

prob_1iv = 1 - (stats.norm.cdf(0.5) - stats.norm.cdf(-0.5))
print(f"(iv) P(|Z| > 0.5) = {prob_1iv:.4f}")

prob_1v = stats.norm.cdf(1) - stats.norm.cdf(-2.2)
print(f"(v) P(-2.2 < Z < 1) = {prob_1v:.4f}")
print("\n")
```

✓ 0.0s

```
(i) P(Z < 1.5) = 0.9332
(ii) P(Z > 0.5) = 0.3085
(iii) P(|Z| < 1.5) = 0.8664
(iv) P(|Z| > 0.5) = 0.6171
(v) P(-2.2 < Z < 1) = 0.8274
```

2. *If Z follows standard normal distribution, then find value of Z_0 satisfying the given equation*

$(i)\ P(Z < Z_0) = 0.90$ $(ii)\ P(Z < Z_0) = 0.95$ $(iii)\ P(Z < Z_0) = 0.99$

$(iv)\ P(|Z| < Z_0) = 0.90$ $(v)\ P(|Z| < Z_0) = 0.95$ $(vi)\ P(|Z| < Z_0) = 0.99$

Code:

```
z0_2i = stats.norm.ppf(0.90)
print(f"(i) Z0 for P(Z < Z0) = 0.90 is {z0_2i:.4f}")

z0_2ii = stats.norm.ppf(0.95)
print(f"(ii) Z0 for P(Z < Z0) = 0.95 is {z0_2ii:.4f}")

z0_2iii = stats.norm.ppf(0.99)
print(f"(iii) Z0 for P(Z < Z0) = 0.99 is {z0_2iii:.4f}")

z0_2iv = stats.norm.ppf(0.90 + (1 - 0.90) / 2)
print(f"(iv) Z0 for P(|Z| < Z0) = 0.90 is {z0_2iv:.4f}")

z0_2v = stats.norm.ppf(0.95 + (1 - 0.95) / 2)
print(f"(v) Z0 for P(|Z| < Z0) = 0.95 is {z0_2v:.4f}")

z0_2vi = stats.norm.ppf(0.99 + (1 - 0.99) / 2)
print(f"(vi) Z0 for P(|Z| < Z0) = 0.99 is {z0_2vi:.4f}")
print("\n")
```

✓ 0.0s

```
(i) Z0 for P(Z < Z0) = 0.90 is 1.2816
(ii) Z0 for P(Z < Z0) = 0.95 is 1.6449
(iii) Z0 for P(Z < Z0) = 0.99 is 2.3263
(iv) Z0 for P(|Z| < Z0) = 0.90 is 1.6449
(v) Z0 for P(|Z| < Z0) = 0.95 is 1.9600
(vi) Z0 for P(|Z| < Z0) = 0.99 is 2.5758
```

3. $If\ t\ follows\ students\ t\ distribution, then\ find$

$(i)\ P(t < 1.5)\ with\ d.o.f. = 20$ $\quad (ii)\ P(t > 0.5)\ with\ d.o.f. = 15$

$(iii)\ P(|t| < 1.5)\ with\ d.o.f. = 25$ $\quad (iv)\ P(|t| > 0.5)\ with\ d.o.f. = 35$

$(v)\ P(-2.2 < t < 1)\ with\ d.o.f. = 42$

Code:

```
df_3i = 20
prob_3i = stats.t.cdf(1.5, df=df_3i)
print(f"(i) P(t < 1.5) with df={df_3i} = {prob_3i:.4f}")

df_3ii = 15
prob_3ii = 1 - stats.t.cdf(0.5, df=df_3ii)
print(f"(ii) P(t > 0.5) with df={df_3ii} = {prob_3ii:.4f}")
df_3iii = 25
prob_3iii = stats.t.cdf(1.5, df=df_3iii) - stats.t.cdf(-1.5, df=df_3iii)
print(f"(iii) P(|t| < 1.5) with df={df_3iii} = {prob_3iii:.4f}")

df_3iv = 35
prob_3iv = 1 - (stats.t.cdf(0.5, df=df_3iv) - stats.t.cdf(-0.5, df=df_3iv))
print(f"(iv) P(|t| > 0.5) with df={df_3iv} = {prob_3iv:.4f}")

df_3v = 42
prob_3v = stats.t.cdf(1, df=df_3v) - stats.t.cdf(-2.2, df=df_3v)
print(f"(v) P(-2.2 < t < 1) with df={df_3v} = {prob_3v:.4f}")
print("\n")
```

✓ 0.0s

```
(i) P(t < 1.5) with df=20 = 0.9254
(ii) P(t > 0.5) with df=15 = 0.3122
(iii) P(|t| < 1.5) with df=25 = 0.8539
(iv) P(|t| > 0.5) with df=35 = 0.6202
(v) P(-2.2 < t < 1) with df=42 = 0.8218
```

4. *If t follows students t distribution, then find value of* t_0 *satisfying the given equation*

$(i)\ P(t < t_0) = 0.90\ with\ d.o.f. = 20$ $(ii)\ P(t < t_0) = 0.95\ with\ d.o.f. = 15$

$(iii)\ P(t < t_0) = 0.99\ with\ d.o.f. = 25$ $(iv)\ P(|t| < t_0) = 0.90\ with\ d.o.f. = 30$

$(v)\ P(|t| < t_0) = 0.95\ with\ d.o.f. = 42$ $(vi)\ P(|t| < t_0) = 0.99\ with\ d.o.f. = 10$

Code:

```python
df_4i = 20
t0_4i = stats.t.ppf(0.90, df=df_4i)
print(f"(i) t0 for P(t < t0) = 0.90 with df={df_4i} is {t0_4i:.4f}")

df_4ii = 15
t0_4ii = stats.t.ppf(0.95, df=df_4ii)
print(f"(ii) t0 for P(t < t0) = 0.95 with df={df_4ii} is {t0_4ii:.4f}")

df_4iii = 25
t0_4iii = stats.t.ppf(0.99, df=df_4iii)
print(f"(iii) t0 for P(t < t0) = 0.99 with df={df_4iii} is {t0_4iii:.4f}")

df_4iv = 30
t0_4iv = stats.t.ppf(0.90 + (1 - 0.90) / 2, df=df_4iv)
print(f"(iv) t0 for P(|t| < t0) = 0.90 with df={df_4iv} is {t0_4iv:.4f}")

df_4v = 42
t0_4v = stats.t.ppf(0.95 + (1 - 0.95) / 2, df=df_4v)
print(f"(v) t0 for P(|t| < t0) = 0.95 with df={df_4v} is {t0_4v:.4f}")

df_4vi = 10
t0_4vi = stats.t.ppf(0.99 + (1 - 0.99) / 2, df=df_4vi)
print(f"(vi) t0 for P(|t| < t0) = 0.99 with df={df_4vi} is {t0_4vi:.4f}")
print("\n")
```

✓ 0.0s

```
(i) t0 for P(t < t0) = 0.90 with df=20 is 1.3253
(ii) t0 for P(t < t0) = 0.95 with df=15 is 1.7531
(iii) t0 for P(t < t0) = 0.99 with df=25 is 2.4851
(iv) t0 for P(|t| < t0) = 0.90 with df=30 is 1.6973
(v) t0 for P(|t| < t0) = 0.95 with df=42 is 2.0181
(vi) t0 for P(|t| < t0) = 0.99 with df=10 is 3.1693
```

5. *If F follows Snedecor's F distribution, then find*

$(i)\ P(F < 1.5)\ with\ df_1 = 5, df_2 = 14$ $(ii)\ P(F > 2.5)\ with\ df_1 = 15, df_2 = 14$

$(iii)\ P(0.5 < F < 4.1)\ with\ df_1 = 13, df_2 = 17$

Code:

```python
df1_5i, df2_5i = 5, 14
prob_5i = stats.f.cdf(1.5, dfn=df1_5i, dfd=df2_5i)
print(f"(i) P(F < 1.5) with df1={df1_5i}, df2={df2_5i} = {prob_5i:.4f}")

df1_5ii, df2_5ii = 15, 14
prob_5ii = 1 - stats.f.cdf(2.5, dfn=df1_5ii, dfd=df2_5ii)
print(f"(ii) P(F > 2.5) with df1={df1_5ii}, df2={df2_5ii} = {prob_5ii:.4f}")

df1_5iii, df2_5iii = 13, 17
prob_5iii = stats.f.cdf(4.1, dfn=df1_5iii, dfd=df2_5iii) - stats.f.cdf(0.5, dfn=df1_5iii, dfd=df2_5iii)
print(f"(iii) P(0.5 < F < 4.1) with df1={df1_5iii}, df2={df2_5iii} = {prob_5iii:.4f}")
print("\n")
```

```
0.0s
(i) P(F < 1.5) with df1=5, df2=14 = 0.7481
(ii) P(F > 2.5) with df1=15, df2=14 = 0.0473
(iii) P(0.5 < F < 4.1) with df1=13, df2=17 = 0.8911
```

6. *If F follows Snedecor's F distribution, then find value of* F_0 *satisfying the given equation*

$(i)\ P(F < F_0) = 0.90\ with\ df_1 = 5, df_2 = 14$

$(ii)\ P(F < F_0) = 0.95\ with\ df_1 = 15, df_2 = 13$

$(iii)\ P(F < F_0) = 0.99\ with\ df_1 = 25, df_2 = 28$

Code:

```python
df1_6i, df2_6i = 5, 14
f0_6i = stats.f.ppf(0.90, dfn=df1_6i, dfd=df2_6i)
print(f"(i) F0 for P(F < F0) = 0.90 with df1={df1_6i}, df2={df2_6i} is {f0_6i:.4f}")

df1_6ii, df2_6ii = 15, 13
f0_6ii = stats.f.ppf(0.95, dfn=df1_6ii, dfd=df2_6ii)
print(f"(ii) F0 for P(F < F0) = 0.95 with df1={df1_6ii}, df2={df2_6ii} is {f0_6ii:.4f}")

df1_6iii, df2_6iii = 25, 28
f0_6iii = stats.f.ppf(0.99, dfn=df1_6iii, dfd=df2_6iii)
print(f"(iii) F0 for P(F < F0) = 0.99 with df1={df1_6iii}, df2={df2_6iii} is {f0_6iii:.4f}")
print("\n")
```

```
0.0s
(i) F0 for P(F < F0) = 0.90 with df1=5, df2=14 is 2.3069
(ii) F0 for P(F < F0) = 0.95 with df1=15, df2=13 is 2.5331
(iii) F0 for P(F < F0) = 0.99 with df1=25, df2=28 is 2.5060
```

7. $If\ X\ follows\ \chi^2 - distribution, then\ find$

$(i)\ P(X < 1.5)\ with\ df = 10$ $\quad$ $(ii)\ P(X > 2.5)\ with\ df = 5$

$(iii)\ P(0.5 < X < 4.1)\ with\ df = 2$

Code:

```
df_7i = 10
prob_7i = stats.chi2.cdf(1.5, df=df_7i)
print(f"(i) P(X < 1.5) with df={df_7i} = {prob_7i:.4f}")

df_7ii = 5
prob_7ii = 1 - stats.chi2.cdf(2.5, df=df_7ii)
print(f"(ii) P(X > 2.5) with df={df_7ii} = {prob_7ii:.4f}")

df_7iii = 2
prob_7iii = stats.chi2.cdf(4.1, df=df_7iii) - stats.chi2.cdf(0.5, df=df_7iii)
print(f"(iii) P(0.5 < X < 4.1) with df={df_7iii} = {prob_7iii:.4f}")
print("\n")
```

✓ 0.0s

```
(i) P(X < 1.5) with df=10 = 0.0011
(ii) P(X > 2.5) with df=5 = 0.7765
(iii) P(0.5 < X < 4.1) with df=2 = 0.6501
```

8. *If X follows χ^2- distribution, then find value of X_0 satisfying the given equation*

$(i)\ P(X < X_0) = 0.90\ with\ df = 1$ $(ii)\ P(X < X_0) = 0.95\ with\ df = 3$

$(iii)\ P(X < X_0) = 0.99\ with\ df = 2$ $(iv)\ P(X < X_0) = 0.05\ with\ df = 1$

$(v)\ P(X < X_0) = 0.025\ with\ df = 3$ $(vi)\ P(X < X_0) = 0.005\ with\ df = 2$

Code:

```python
df_8i = 1
x0_8i = stats.chi2.ppf(0.90, df=df_8i)
print(f"(i) X0 for P(X < X0) = 0.90 with df={df_8i} is {x0_8i:.4f}")

df_8ii = 3
x0_8ii = stats.chi2.ppf(0.95, df=df_8ii)
print(f"(ii) X0 for P(X < X0) = 0.95 with df={df_8ii} is {x0_8ii:.4f}")

df_8iii = 2
x0_8iii = stats.chi2.ppf(0.99, df=df_8iii)
print(f"(iii) X0 for P(X < X0) = 0.99 with df={df_8iii} is {x0_8iii:.4f}")

df_8iv = 1
x0_8iv = stats.chi2.ppf(0.05, df=df_8iv)
print(f"(iv) X0 for P(X < X0) = 0.05 with df={df_8iv} is {x0_8iv:.4f}")

df_8v = 3
x0_8v = stats.chi2.ppf(0.025, df=df_8v)
print(f"(v) X0 for P(X < X0) = 0.025 with df={df_8v} is {x0_8v:.4f}")

df_8vi = 2
x0_8vi = stats.chi2.ppf(0.005, df=df_8vi)
print(f"(vi) X0 for P(X < X0) = 0.005 with df={df_8vi} is {x0_8vi:.4f}")
print("\n")
```

✓ 0.0s

```
(i) X0 for P(X < X0) = 0.90 with df=1 is 2.7055
(ii) X0 for P(X < X0) = 0.95 with df=3 is 7.8147
(iii) X0 for P(X < X0) = 0.99 with df=2 is 9.2103
(iv) X0 for P(X < X0) = 0.05 with df=1 is 0.0039
(v) X0 for P(X < X0) = 0.025 with df=3 is 0.2158
(vi) X0 for P(X < X0) = 0.005 with df=2 is 0.0100
```

9. *Construct a 95 % confidence interval for population mean in an experiment that found the sample mean temperature for a certain city in August was 101.82, with a population standard deviation of 1.2. There were 6 samples in this experiment.*

Code:

```python
sample_mean_9 = 101.82
pop_std_dev_9 = 1.2
n_9 = 6
confidence_level_9 = 0.95
alpha_9 = 1 - confidence_level_9

z_critical_9 = stats.norm.ppf(1 - alpha_9 / 2)

margin_of_error_9 = z_critical_9 * (pop_std_dev_9 / math.sqrt(n_9))

ci_lower_9 = sample_mean_9 - margin_of_error_9
ci_upper_9 = sample_mean_9 + margin_of_error_9

print(f"Sample Mean: {sample_mean_9}")
print(f"Population Std Dev: {pop_std_dev_9}")
print(f"Sample Size: {n_9}")
print(f"Confidence Level: {confidence_level_9*100}%")
print(f"Z-critical value: {z_critical_9:.4f}")
print(f"Margin of Error: {margin_of_error_9:.4f}")
print(f"95% Confidence Interval: ({ci_lower_9:.4f}, {ci_upper_9:.4f})")
print("\n")
```

✓ 0.0s

```
Sample Mean: 101.82
Population Std Dev: 1.2
Sample Size: 6
Confidence Level: 95.0%
Z-critical value: 1.9600
Margin of Error: 0.9602
95% Confidence Interval: (100.8598, 102.7802)
```

10. *Construct a 98% Confidence Interval for population mean based on the following data:*

45, 55, 67, 45, 68, 79, 98, 87, 84, 82.

Code:

```python

data_10 = [45, 55, 67, 45, 68, 79, 98, 87, 84, 82]
n_10 = len(data_10)
sample_mean_10 = np.mean(data_10)
sample_std_dev_10 = np.std(data_10, ddof=1)
confidence_level_10 = 0.98
alpha_10 = 1 - confidence_level_10
df_10 = n_10 - 1

t_critical_10 = stats.t.ppf(1 - alpha_10 / 2, df=df_10)

margin_of_error_10 = t_critical_10 * (sample_std_dev_10 / math.sqrt(n_10))

ci_lower_10 = sample_mean_10 - margin_of_error_10
ci_upper_10 = sample_mean_10 + margin_of_error_10

print(f"Data: {data_10}")
print(f"Sample Mean: {sample_mean_10:.4f}")
print(f"Sample Std Dev: {sample_std_dev_10:.4f}")
print(f"Sample Size: {n_10}")
print(f"Degrees of Freedom: {df_10}")
print(f"Confidence Level: {confidence_level_10*100}%")
print(f"t-critical value: {t_critical_10:.4f}")
print(f"Margin of Error: {margin_of_error_10:.4f}")
print(f"98% Confidence Interval: ({ci_lower_10:.4f}, {ci_upper_10:.4f})")
print("\n")
```

✓ 0.0s

```
Data: [45, 55, 67, 45, 68, 79, 98, 87, 84, 82]
Sample Mean: 71.0000
Sample Std Dev: 18.1720
Sample Size: 10
Degrees of Freedom: 9
Confidence Level: 98.0%
t-critical value: 2.8214
Margin of Error: 16.2134
98% Confidence Interval: (54.7866, 87.2134)
```

11. 510 *people applied to the Bachelor's in Elementary Education program at Florida State College. Of those applicants,* 57 *were men. Find the* 90% *CI of the true proportion of men who applied to the program.*

Code:

```python
n_11 = 510
x_11 = 57
confidence_level_11 = 0.90
alpha_11 = 1 - confidence_level_11

p_hat_11 = x_11 / n_11

z_critical_11 = stats.norm.ppf(1 - alpha_11 / 2)

margin_of_error_11 = z_critical_11 * math.sqrt((p_hat_11 * (1 - p_hat_11)) / n_11)

ci_lower_11 = p_hat_11 - margin_of_error_11
ci_upper_11 = p_hat_11 + margin_of_error_11

print(f"Total Applicants (n): {n_11}")
print(f"Number of Men (x): {x_11}")
print(f"Sample Proportion (p_hat): {p_hat_11:.4f}")
print(f"Confidence Level: {confidence_level_11*100}%")
print(f"Z-critical value: {z_critical_11:.4f}")
print(f"Margin of Error: {margin_of_error_11:.4f}")
print(f"90% Confidence Interval for Proportion: ({ci_lower_11:.4f}, {ci_upper_11:.4f})")
```

✓ 0.0s

```
Total Applicants (n): 510
Number of Men (x): 57
Sample Proportion (p_hat): 0.1118
Confidence Level: 90.0%
Z-critical value: 1.6449
Margin of Error: 0.0229
90% Confidence Interval for Proportion: (0.0888, 0.1347)
```

Conclusion	Thus, we studied how to implement Confidence Interval using Python.
Colab Link	https://colab.research.google.com/github/SmayanKulkarni/AI-and-ML-Course/blob/master/SDS/exp_5.ipynb

Signature of Faculty