Q1 The empirical rule, also referred to as the three-sigma rule or 68-95-99.7 rule, is a statistical rule which states that for a normal distribution(bell-curve), almost all observed data will fall within three standard deviations (denoted by σ) of the mean or average (denoted by μ).

Normal distribution, is a probability distribution that is symmetric about the mean, without any outliers.

Q2. A measure used to indicate the extent to which two random variables change at the same time is known as covariance. (Direction) A measure used to represent how strongly two random variables are related known as correlation.(Strength)

Covariance is a measure of correlation

Q10. Confusion Matrix or Error Matrix is a specific table layout that allows visualization of the performance of an algorithm.

The template for any binary confusion matrix uses the four kinds of results - (true positives, false negatives, false positives, and true negatives)

False Positive: (Type 1 Error) Interpretation: You predicted positive and it's false Real time Example- Breath test predicts presence of Alcohol when person is not drunk.

False Negative: (Type 2 Error) Interpretation: You predicted negative and it's false. Example- Person diagnosed No Cancer , when it is actually there.

Q3 Delivery Boy

```
import io
import pandas as pd
from google.colab import files


uploaded = files.upload()
```

Choose files   No file chosen   Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving Delivery Boy.xlsx to Delivery Boy.xlsx

```
df = pd.read_excel(io.BytesIO(uploaded.get('Delivery Boy.xlsx')))
df
```

	deliveryboy1	deliveryboy2
0	12	40
1	13	10
2	17	31
3	21	59
4	24	13
5	24	50
6	26	27
7	27	33
8	27	53
9	30	34
10	32	16
11	35	10
12	37	34
13	38	30
14	41	33
15	43	42
16	44	34

```
df.std()
```

```
deliveryboy1    12.890225
deliveryboy2    13.477076
dtype: float64
```

delivery boy1 has lower standard deviation so he will be selected.

Q4. Load Tips Dataset

```
import seaborn as sns
sns.set_theme()
tips = sns.load_dataset("tips")
```

```
tips
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2

Q5. Correlation

244 rows × 7 columns

```
tips['total_bill'].corr(tips['tip'])
```

0.6757341092113641

Double-click (or enter) to edit

```
import io
import pandas as pd
from google.colab import files


import numpy as np
import scipy.stats as stats
from scipy.stats import chi2
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

```
uploaded = files.upload()
```

Choose files No file chosen   Upload widget is only available when the cell has been executed
this cell to enable.
Saving Adidas (1).xlsx to Adidas (1).xlsx

```
df1 = pd.read_excel(io.BytesIO(uploaded.get('Adidas  (1).xlsx')))
df1
```

	discount	city	sales
0	0.1	Delhi	133
1	0.1	Delhi	107
2	0.1	Delhi	114
3	0.1	Delhi	150
4	0.1	Delhi	85
5	0.2	Delhi	130
6	0.2	Delhi	86
7	0.2	Delhi	130
8	0.2	Delhi	112
9	0.2	Delhi	125
10	0.1	Mumbai	84
11	0.1	Mumbai	114
12	0.1	Mumbai	138
13	0.1	Mumbai	155
14	0.1	Mumbai	78
15	0.2	Mumbai	138
16	0.2	Mumbai	119
17	0.2	Mumbai	146
18	0.2	Mumbai	158
19	0.2	Mumbai	137
20	0.1	Bangalore	94
21	0.1	Bangalore	123
22	0.1	Bangalore	108
23	0.1	Bangalore	148
24	0.1	Bangalore	93
25	0.2	Bangalore	120
26	0.2	Bangalore	127
27	0.2	Bangalore	150
28	0.2	Bangalore	106
29	0.2	Bangalore	127