Q1: What is the relationship between stress and self-esteem?

Statistical test: Pearson's Correlation.

Reason: A correlation test measures the strength and direction of the association between two continuous variables. In this case, we are examining if there is a significant relationship between stress and self-esteem scores, which can help determine if higher stress is associated with higher or lower self-esteem.

Results:

Test 1: Correlation between Total Perceived Stress and Total Self-Esteem

```
from scipy.stats import pearsonr
```

```
corr, p_value = pearsonr(df1['tslfest'], df1['tpstress'])
```

```
print("Correlation Test Between Stress and Self-Esteem")
print("Correlation Coefficient:----->>", corr)
print("P-Value:----->>", p_value)
```

```
Correlation Test Between Stress and Self-Esteem
Correlation Coefficient:----->> -0.5835192552576107
P-Value:----->> 8.990356869439582e-32
```

- If the p-value is less than 0.05, we conclude that there is a statistically significant correlation between stress and self-esteem. The correlation coefficient (positive or negative) will indicate whether this relationship is positive (both increase together) or negative (one decreases as the other increases).
- If the p-value is greater than 0.05, we do not have enough evidence to conclude a significant correlation between stress and self-esteem.
- Based on the test results, with a p-value close to zero and a correlation coefficient of -0.58, there is a statistically significant negative correlation between stress and self-esteem. It indicates that higher stress is associated with lower self-esteem.

Q2: How well do stress and optimism predict life satisfaction?

Statistical test: Multiple Linear Regression.

Reason: A multiple linear regression analysis is used to examine how one or more independent variables predict the value of a dependent variable. Here, we are exploring if stress and optimism scores can significantly predict life satisfaction (dependent variable).

Results:

Test 2: Regression Analysis to Predict Total Life Satisfaction from Total Perceived Stress and Total Optimism

```
from sklearn.linear_model import LinearRegression
linreg = LinearRegression()
```

```
x = df1[['tpstress', 'toptim']]
y = df1['tlifesat']
```

```
linreg.fit(x,y)
```

```
▾ LinearRegression
LinearRegression()
```

```
coefficients = linreg.coef_
intercept = linreg.intercept_
```

```
print("Regression Analysis to Predict Life Satisfaction from Stress and Optimism")
print("Intercept:----->>", intercept)
print("Coefficient for Stress:------>>", coefficients[0])
print("Coefficient for Optimism:------>>", coefficients[1])
```

```
Regression Analysis to Predict Life Satisfaction from Stress and Optimism
Intercept:----->> 22.182762848092697
Coefficient for Stress:------>> -0.41873441256134125
Coefficient for Optimism:------>> 0.521941474614787
```

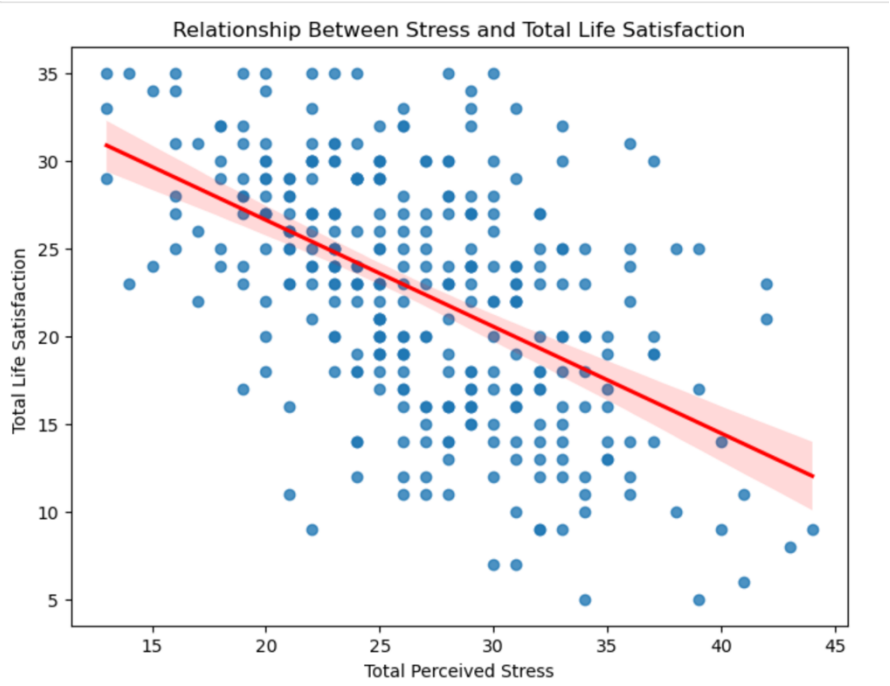
- The regression model indicates that for each unit increase in stress, life satisfaction decreases by approximately 0.42, while for each unit increase in optimism, life satisfaction increases by approximately 0.52.
- This suggests that higher optimism is positively associated with life satisfaction, whereas higher stress is negatively associated with it.
- The intercept of 22.18 represents the baseline level of satisfaction when both stress and optimism are zero.

Continued…

Scatter Plot with regression line for Total Perceived Stress vs Total Life Satisfaction:

The regression plot depicting the relationship between stress and life satisfaction shows a negative trend, indicating that as perceived stress levels increase, life satisfaction generally decreases.

The fitted regression line reinforces this negative correlation, suggesting that individuals with higher stress tend to report lower life satisfaction. This trend aligns with the expectation that stress negatively impacts overall well-being.

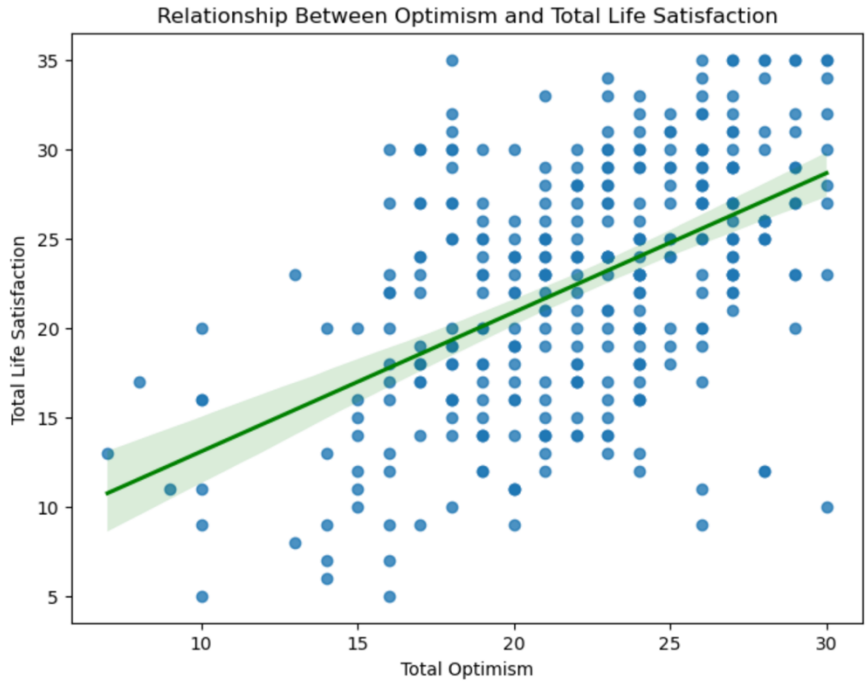


Scatter Plot with regression line for Total Optimism vs Total Life Satisfaction:

The regression plot for optimism versus life satisfaction shows a positive trend, implying that higher levels of optimism are associated with higher life satisfaction.

The fitted regression line highlights this positive correlation, suggesting that individuals who report greater optimism tend to experience greater life satisfaction.

This finding supports the idea that optimism can be a protective factor for well-being, contributing to higher satisfaction with life.



Q3: Is there a link between smoking status and age group?

Statistical test: Chi-Square Test.

Reason: A chi-square test is used to determine if there is a significant association between two categorical variables. In this case, we are checking if smoking status (smoker vs. non-smoker) is associated with age group or if smoking behaviour is independent of age group.

Results:

Test 3: Chi-Square Test for Smoking Status by Age Group

```
from scipy.stats import chi2_contingency
```

```
contingency_table = pd.crosstab(df1['smoke'], df1['agegp3'])
contingency_table
```

agegp3	1	2	3
smoke			
1.0	33	29	14
2.0	75	94	88

```
#1--> yes
#2--> no
```

```
chi2, p_value, dof, expected = chi2_contingency(contingency_table)
```

```
print("Chi-Square Test for Smoking Status by Age Group")
print("Chi-Square Statistic:----->>", chi2)
print("P-Value:---->>", p_value)
print("Degrees of Freedom:----->>", dof)
print("Expected Frequencies Table:\n------>>", expected)
```

```
Chi-Square Test for Smoking Status by Age Group
Chi-Square Statistic:----->> 8.498660752718344
P-Value:---->> 0.014273788775968396
Degrees of Freedom:----->> 2
Expected Frequencies Table:
------>> [[24.64864865 28.07207207 23.27927928]
 [83.35135135 94.92792793 78.72072072]]
```

- If the p-value is less than 0.05, we conclude that there is a statistically significant association between smoking status and age group. This means that smoking behaviour may vary significantly across age groups.
- If the p-value is greater than 0.05, we do not have enough evidence to conclude a significant association between smoking status and age group, indicating that smoking behaviour is likely independent of age group.
- Based on the test results, with a p-value of 0.014, we conclude that there is a statistically significant association between smoking status and age group. It suggests that smoking behaviour varies across different age groups.

Q4: Is there a significant difference in optimism scores between smokers and non-smokers?

Statistical test: T-test.

Reason: A t-test compares the means of two independent groups. Here, we want to see if smoking status (smokers vs. non-smokers) is associated with differences in optimism scores.

Results:

Test 4: T-Test for Total Optimism Scores by Smoking Status

```
from scipy.stats import ttest_ind
```

```
smokers_optimism = df1[df1['smoke'] == 1]['toptim']
non_smokers_optimism = df1[df1['smoke'] == 2]['toptim']
```

```
t_stat, p_value = ttest_ind(smokers_optimism, non_smokers_optimism)
```

```
print("T-Test for Optimism Scores by Smoking Status")
print("T-Statistic:----->>", t_stat)
print("P-Value:----->>", p_value)
```

```
T-Test for Optimism Scores by Smoking Status
T-Statistic:----->> -1.808191047217171
P-Value:----->> 0.07148425234695395
```

- If the p-value is less than 0.05, the two variables being compared are significant; if the value is more than 0.05, they are not significant.
- With a p-value of 0.071, we conclude that there is no statistically significant difference in optimism scores between smokers and non-smokers in this sample, suggesting that smoking status does not appear to be associated with optimism levels.

Q5: Does life satisfaction differ significantly across education levels?

Statistical test: ANOVA.

Reason: Analysis of Variance (ANOVA) compares the means of more than two independent groups. In this case, we use it to assess if life satisfaction varies significantly based on education levels.

Results:

Test 5: ANOVA for Total Life Satisfaction across Education Levels

```
from scipy.stats import f_oneway
```

```
group1 = df1[df1['educ'] == 1]['tlifesat']  #Education Level 1
group2 = df1[df1['educ'] == 2]['tlifesat']  #Education Level 2
group3 = df1[df1['educ'] == 3]['tlifesat']  #Education Level 3
group4 = df1[df1['educ'] == 4]['tlifesat']  #Education Level 4
group5 = df1[df1['educ'] == 5]['tlifesat']  #Education Level 5
```

```
f_stat, p_value = f_oneway(group1, group2, group3, group4, group5)
```

```
print("ANOVA for Life Satisfaction Across Education Levels")
print("F-Statistic:----->>", f_stat)
print("P-Value:------->>", p_value)
```

```
ANOVA for Life Satisfaction Across Education Levels
F-Statistic:----->> 3.6380768758575024
P-Value:------->> 0.006575475847309252
```

- If the p-value is less than 0.05, there is a statistically significant difference in life satisfaction across education levels, indicating that education might impact life satisfaction.
- If the p-value is greater than 0.05, we do not have enough evidence to suggest a significant difference, meaning life satisfaction is likely similar across education levels.
- Based on the test results, with a p-value of 0.006, there is a statistically significant relationship between education level and life satisfaction. It suggests that educational background impacts life satisfaction.