Excercises

A company tests two training programs (T1, T2) and considers gender (M/F) to see their impact on productivity.

Training	Gender	Scores
T1	M	80, 82, 85
T1	F	78, 76, 79
T2	M	88, 90, 87
T2	F	85, 84, 86

Task:

- Perform **two-way ANOVA** to determine:
  - If training type affects productivity
  - If gender affects productivity
  - If there's an interaction between training and gender

#CODE

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols

# Create the DataFrame
data = {
   'Training': ['T1']*6 + ['T2']*6,
   'Gender': ['M']*3 + ['F']*3 + ['M']*3 + ['F']*3,
   'Score': [80, 82, 85, 78, 76, 79, 88, 90, 87, 85, 84, 86]
}
df = pd.DataFrame(data)

# Two-Way ANOVA with interaction
model = ols('Score ~ C(Training) + C(Gender) + C(Training):C(Gender)', data=df).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
print(anova_table)
```

Exercise 1: Chi-Square Goodness of Fit

Scenario: A die is rolled 60 times. The outcomes are:

Face	Frequency
1	5
2	8

Face	Frequency
3	9
4	10
5	13
6	15

ask:

- Test whether the die is fair using the **Chi-Square Goodness of Fit** test.
- H_0: The die is fair (uniform distribution)
- $\alpha = 0.05$

```
#CODE
from scipy.stats import chisquare

# Observed frequencies from the die rolls
observed = [5, 8, 9, 10, 13, 15]

# Expected frequencies for a fair die (uniform distribution)
expected = [10] * 6  # Total rolls = 60, so each face expected 10 times

# Perform Chi-Square Goodness of Fit test
chi_stat, p_value = chisquare(f_obs=observed, f_exp=expected)

# Display results
print(f"Chi-Square Statistic = {chi_stat:.4f}")
print(f"P-Value = {p_value:.4f}")

# Conclusion
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis — the die is not fair.")
else:
    print("Fail to reject the null hypothesis — the die appears to be fair.")
```

Excercise-2

A candy company claims its packs have equal proportions of 5 colors.

Color	Observed
Red	22
Green	18
Blue	25
Yellow	20
Orange	15

- Use Chi-Square Goodness of Fit to test fairness.
- H_0: All colors occur with equal frequency.

#To test whether all 5 colors occur with **equal frequency**, we will use the **Chi-Square Goodness of Fit Test**.

Color	Observed (O)
Red	22
Green	18
Blue	25
Yellow	20
Orange	15

- Total = 22 + 18 + 25 + 20 + 15 = **100**
- Expected frequency for each color (if all are equally likely):
  E = 100 / 5 = 20

```
from scipy.stats import chisquare

# Observed frequencies
observed = [22, 18, 25, 20, 15]

# Expected frequencies (equal distribution)
expected = [20] * 5

# Chi-square test
chi2_stat, p_value = chisquare(f_obs=observed, f_exp=expected)

print(f"Chi-Square Statistic = {chi2_stat:.4f}")
print(f"P-Value = {p_value:.4f}")
```

Exercise 3:
You believe the average height of a population is **165 cm**. A sample of 10 people has the following heights:

[162, 167, 170, 160, 165, 168, 163, 164, 166, 161]

- Use one-sample t-test.
- H_0: $\mu = 165$ cm

```
from scipy.stats import ttest_1samp
```

```
# Sample data
sample = [162, 167, 170, 160, 165, 168, 163, 164, 166, 161]

# Perform one-sample t-test
t_stat, p_value = ttest_1samp(sample, popmean=165)

print(f"T-Statistic = {t_stat:.4f}")
print(f"P-Value = {p_value:.4f}")
```