Here are the answers to your assignment questions:

**Q1. Assumptions Required to Use ANOVA**

The assumptions for ANOVA include:

1. **Independence**: The observations within each group should be independent of each other.
   * *Violation example*: If measurements are taken from the same subjects repeatedly, it violates this assumption.
2. **Normality**: The data in each group should be approximately normally distributed.
   * *Violation example*: If the data has extreme skewness or outliers, it may not follow a normal distribution.
3. **Homogeneity of variance (Homoscedasticity)**: The variances within each group should be equal.
   * *Violation example*: If one group has much higher variance than others, it can lead to misleading results.

**Q2. Three Types of ANOVA**

1. **One-Way ANOVA**: Used to compare means of three or more independent groups based on one factor.
   * *Example*: Comparing the mean test scores of students taught using different teaching methods.
2. **Two-Way ANOVA**: Used to assess the impact of two factors simultaneously on the dependent variable and their interaction effect.
   * *Example*: Examining the effect of both gender and age on customer spending.
3. **Repeated Measures ANOVA**: Used when the same subjects are measured multiple times (i.e., within-subject design).
   * *Example*: Measuring the weight of the same group of participants before and after a treatment.

**Q3. Partitioning of Variance in ANOVA**

Partitioning variance in ANOVA refers to the process of breaking down the total variance into different components, such as:

* **Total Variance (SST)**: Overall variability in the data.
* **Explained Variance (SSE)**: Variability explained by the independent variable.
* **Residual Variance (SSR)**: Variability unexplained by the independent variable (random error).

Understanding partitioning is crucial because it helps determine whether the variability due to the independent variable is significant enough to conclude that there are differences between groups.

**Q4. Calculating SST, SSE, and SSR in a One-Way ANOVA Using Python**

import numpy as np

import scipy.stats as stats

# Sample data

group1 = [23, 25, 27, 28]

group2 = [30, 32, 33, 35]

group3 = [40, 42, 45, 47]

# Concatenate all groups

data = np.concatenate([group1, group2, group3])

# Calculate group means and overall mean

group\_means = [np.mean(group1), np.mean(group2), np.mean(group3)]

overall\_mean = np.mean(data)

# Calculate SST

SST = np.sum((data - overall\_mean) \*\* 2)

# Calculate SSE (explained variance)

SSE = np.sum([len(group1) \* (np.mean(group1) - overall\_mean) \*\* 2,

len(group2) \* (np.mean(group2) - overall\_mean) \*\* 2,

len(group3) \* (np.mean(group3) - overall\_mean) \*\* 2])

# Calculate SSR (residual variance)

SSR = SST - SSE

print(f"SST: {SST}, SSE: {SSE}, SSR: {SSR}")

**Q5. Calculating Main Effects and Interaction Effects in Two-Way ANOVA Using Python**

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

# Example data

data = pd.DataFrame({

'Software': ['A', 'A', 'B', 'B', 'C', 'C', 'A', 'A', 'B', 'B', 'C', 'C'],

'Experience': ['Novice', 'Experienced', 'Novice', 'Experienced', 'Novice', 'Experienced', 'Novice', 'Experienced', 'Novice', 'Experienced', 'Novice', 'Experienced'],

'Time': [45, 40, 50, 45, 60, 55, 47, 43, 52, 48, 62, 58]

})

# Fit model

model = ols('Time ~ C(Software) + C(Experience) + C(Software):C(Experience)', data=data).fit()

anova\_table = sm.stats.anova\_lm(model, typ=2)

print(anova\_table)

**Q6. Interpreting One-Way ANOVA Results**

With an F-statistic of 5.23 and a p-value of 0.02, you can conclude that there are significant differences between at least two of the groups. Since the p-value is less than the typical significance level (0.05), you would reject the null hypothesis, which states that all group means are equal.

**Q7. Handling Missing Data in Repeated Measures ANOVA**

Missing data in repeated measures ANOVA can be handled using:

1. **Listwise deletion**: Remove participants with missing data.
   * *Consequences*: Loss of information, potential bias if data is not missing at random.
2. **Multiple imputation**: Estimate missing values based on observed data.
   * *Consequences*: Less bias but increases complexity.

**Q8. Common Post-Hoc Tests After ANOVA**

1. **Tukey's HSD**: Used when comparing all pairs of group means.
2. **Bonferroni correction**: Used to control Type I error when making multiple comparisons.
3. **Dunnett's test**: Compares each treatment group to a control group.

*Example*: After a one-way ANOVA, if the F-test shows significance, a post-hoc test like Tukey's HSD can be used to determine which specific group pairs are different.

**Q9. One-Way ANOVA for Weight Loss**

Here’s an example of conducting a one-way ANOVA to compare mean weight loss between three diets:

import scipy.stats as stats

# Example data

diet\_A = [4.2, 4.5, 4.8, 5.1, 4.9]

diet\_B = [3.5, 3.8, 4.0, 3.9, 4.1]

diet\_C = [5.3, 5.6, 5.7, 5.9, 6.0]

# One-way ANOVA

f\_statistic, p\_value = stats.f\_oneway(diet\_A, diet\_B, diet\_C)

print(f"F-statistic: {f\_statistic}, p-value: {p\_value}")

**Q10. Two-Way ANOVA for Software Programs and Experience Levels**

# Data preparation and analysis would be similar to Q5's example.

**Q11. Two-Sample t-test for Test Scores Between Two Teaching Methods**

# Example of a two-sample t-test

control\_group = [75, 78, 72, 70, 68]

experimental\_group = [82, 85, 80, 79, 88]

t\_stat, p\_value = stats.ttest\_ind(control\_group, experimental\_group)

print(f"T-statistic: {t\_stat}, p-value: {p\_value}")

**Q12. Repeated Measures ANOVA for Sales Data**

# Example using `statsmodels` or `pingouin` for repeated measures ANOVA.

Let me know if you need any further details!