Q1. Explain the assumptions required to use ANOVA and provide examples of violations that could impact the validity of the results.

Assumptions for using ANOVA:

Independence: Observations are independent of each other. Homogeneity of variances (homoscedasticity): Variances within each group are roughly equal. Normality: The residuals (differences between observed and expected values) are approximately normally distributed. Violations:

Violations of independence: For example, if data points within a group are not independent, such as repeated measurements on the same subject. Violations of homoscedasticity: When the variances between groups significantly differ. Violations of normality: If the residuals are not normally distributed, especially in small sample sizes.

Q2. What are the three types of ANOVA, and in what situations would each be used?

Three types of ANOVA:

One-way ANOVA: Used when you have one independent variable with more than two levels or groups. Two-way ANOVA: Used when you have two independent variables (factors) and you want to test the main effects of each factor and their interaction. Three-way ANOVA: Used when you have three independent variables and want to analyze their main effects and interactions.

Q3. What is the partitioning of variance in ANOVA, and why is it important to understand this concept?

Partitioning of variance in ANOVA is the division of the total variability in the data into different sources, such as between groups and within groups variability. Understanding this concept is important because it helps identify the relative contributions of factors and errors to the total variation, allowing you to determine whether the differences between groups are statistically significant.

Q4. How would you calculate the total sum of squares (SST), explained sum of squares (SSE), and residual sum of squares (SSR) in a one-way ANOVA using Python?

In [ ]:

```
import scipy.stats as stats

# Assume 'data' is your data in the form of a list of groups.
f_statistic, p_value = stats.f_oneway(*data)

SST = sum((x - np.mean(data))**2 for group in data for x in group)
SSE = sum((x - np.mean(group))**2 for group in data for x in group)
SSR = SST - SSE
```

Q5. How would you calculate the total sum of squares (SST), explained sum of squares (SSE), and residual sum of squares (SSR) in a one-way ANOVA using Python?

To calculate main effects and interaction effects in a two-way ANOVA using Python, you would use a statistical library like statsmodels. The main effects and interaction can be obtained through the ANOVA table generated by the library. The code would depend on the specific data structure and library you're using, but it generally involves fitting a two-way ANOVA model and examining the table.

Q6. Suppose you conducted a one-way ANOVA and obtained an F-statistic of 5.23 and a p-value of 0.02. What can you conclude about the differences between the groups, and how would you interpret these results?

An F-statistic of 5.23 and a p-value of 0.02 in a one-way ANOVA indicates that there are significant differences between the groups. You reject the null hypothesis, implying that at least one group is different from the others. You can't determine which specific group(s) differ from each other with ANOVA alone, but post-hoc tests can be conducted for that purpose.

Q7. In a repeated measures ANOVA, how would you handle missing data, and what are the potential consequences of using different methods to handle missing data?

Handling missing data in repeated measures ANOVA can involve various methods such as imputation, exclusion, or modeling. Using different methods may lead to different results. Imputation can introduce bias, exclusion can reduce power, and modeling may require specific assumptions. The choice depends on the specific context, and it's important to consider the impact on results and validity.

Q8. What are some common post-hoc tests used after ANOVA, and when would you use each one? Provide an example of a situation where a post-hoc test might be necessary.

Common post-hoc tests used after ANOVA include Tukey's HSD, Bonferroni, Scheffé, and Dunnett's tests. They are used to determine which specific groups differ from each other. The choice depends on the research question and the significance level.

Q9. A researcher wants to compare the mean weight loss of three diets: A, B, and C. They collect data from 50 participants who were randomly assigned to one of the diets. Conduct a one-way ANOVA using Python to determine if there are any significant differences between the mean weight loss of the three diets. Report the F-statistic and p-value, and interpret the results.

In [ ]:
```
import scipy.stats as stats

# Create data for each diet group
diet_A = [list of weights]
diet_B = [list of weights]
diet_C = [list of weights]

# Perform one-way ANOVA
f_statistic, p_value = stats.f_oneway(diet_A, diet_B, diet_C)

# Interpret the results
if p_value < 0.05:
```

```
    print("There is a significant difference in mean weight loss between the die
else:
    print("There is no significant difference in mean weight loss between the di
```

Q10. A company wants to know if there are any significant differences in the average time it takes to complete a task using three different software programs: Program A, Program B, and Program C. They randomly assign 30 employees to one of the programs and record the time it takes each employee to complete the task. Conduct a two-way ANOVA using Python to determine if there are any main effects or interaction effects between the software programs and employee experience level (novice vs. experienced). Report the F-statistics and p-values, and interpret the results.

In [ ]:
```
import statsmodels.api as sm
from statsmodels.formula.api import ols

# Create a DataFrame with your data

model = ols('time ~ software * experience', data=df).fit()
anova_table = sm.stats.anova_lm(model, typ=2)

# Interpret the results
# Examine the F-statistics and p-values for main effects and interactions to det
```

Q11. An educational researcher is interested in whether a new teaching method improves student test scores. They randomly assign 100 students to either the control group (traditional teaching method) or the experimental group (new teaching method) and administer a test at the end of the semester. Conduct a two-sample t-test using Python to determine if there are any significant differences in test scores between the two groups. If the results are significant, follow up with a post-hoc test to determine which group(s) differ significantly from each other.

In [ ]:
```
import scipy.stats as stats

group_control = [test scores of control group]
group_experiment = [test scores of experimental group]

t_statistic, p_value = stats.ttest_ind(group_control, group_experiment)

if p_value < 0.05:
    print("There is a significant difference in test scores between the control
else:
    print("There is no significant difference in test scores between the groups.
```

Q12. A researcher wants to know if there are any significant differences in the average daily sales of three retail stores: Store A, Store B, and Store C. They randomly select 30 days and record the sales for each store on those days. Conduct a repeated measures ANOVA using Python to determine if there are any

significant differences in sales between the three stores. If the results are significant, follow up with a post- hoc test to determine which store(s) differ significantly from each other.

In [ ]:

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

# Create a DataFrame with your data

model = ols('sales ~ store', data=df).fit()
anova_table = sm.stats.anova_lm(model, typ=2)

# Interpret the results
# Examine the F-statistics and p-values to determine significance. If results ar
```