Question 1:
Explain the properties of the F-distribution.

Answer:
The **F-distribution** is a continuous probability distribution that arises frequently in the analysis of variance (ANOVA) and in comparing variances. Its properties include:

- It is **positively skewed** (right-skewed).
- It only takes **non-negative values** ($F \geq 0$).
- The shape depends on two degrees of freedom: **numerator (df1)** and **denominator (df2)**.
- As df1 and df2 increase, the F-distribution becomes more symmetric.
- It is used primarily to test **ratios of variances**.

Question 2:
In which types of statistical tests is the F-distribution used, and why is it appropriate for these tests?

Answer:
The **F-distribution** is used in:

- **F-tests**: to compare the variances of two populations.
- **ANOVA (Analysis of Variance)**: to test if the means of three or more groups are equal.
- **Regression analysis**: to assess the overall significance of the model.

It is appropriate because these tests involve **ratios of variances**, and the F-distribution models the behavior of such ratios under the null hypothesis.

Question 3:
What are the key assumptions required for conducting an F-test to compare the variances of two populations?

Answer:

1. The two populations must be **normally distributed**.
2. The samples must be **independent** of each other.
3. The data must be **quantitative** (measurable).
4. The test is **sensitive to non-normality**, especially for small sample sizes.

Question 4:
What is the purpose of ANOVA, and how does it differ from a t-test?

Answer:
Purpose of ANOVA: To determine whether there are statistically significant differences between the means of three or more independent groups.

Difference from t-test:
- **T-test** compares means between **two** groups.
- **ANOVA** compares means between **three or more** groups.
- ANOVA avoids the inflation of Type I error that occurs when using multiple t-tests.

Question 5:
Explain when and why you would use a one-way ANOVA instead of multiple t-tests when comparing more than two groups.

Answer:
Use **one-way ANOVA** when comparing **three or more group means** to determine if at least one is significantly different.

Why not multiple t-tests?
- Increases the risk of **Type I error** (false positives).
- ANOVA controls the overall error rate and provides a single p-value for multiple comparisons.

Question 6:
Explain how variance is partitioned in ANOVA into between-group variance and within-group variance. How does this partitioning contribute to the calculation of the F-statistic?

Answer:
- **Between-group variance**: Variation due to differences between group means.
- **Within-group variance**: Variation within each group (random error).

The **F-statistic** is calculated as:

$$
F = \frac{\text{Between-group variance (MSB)}}{\text{Within-group variance (MSW)}}
$$

A higher F-value suggests greater differences between group means than within groups, indicating statistical significance.

Question 7:
Compare the classical (frequentist) approach to ANOVA with the Bayesian approach. What are the key differences in terms of how they handle uncertainty, parameter estimation, and hypothesis testing?

Answer:

Feature	Classical ANOVA (Frequentist)	Bayesian ANOVA
Uncertainty	Uses probability of data given a hypothesis	Uses probability of hypothesis given data
Parameter Estimation	Point estimates (mean, variance)	Uses prior distributions to estimate parameters
Hypothesis Testing	p-values and F-statistics	Posterior probabilities and Bayes Factors
Interpretation	Reject or fail to reject null hypothesis	Probability of hypotheses being true

---

Question 8:
You have two sets of data representing the incomes of two different professions.
Profession A: [48, 52, 55, 60, 62]
Profession B: [45, 50, 55, 52, 47]

Perform an F-test to determine if the variances of the two professions' incomes are equal.

Python Code:
```python
from scipy.stats import f
import numpy as np

# Data
prof_a = [48, 52, 55, 60, 62]
prof_b = [45, 50, 55, 52, 47]

# Variances
var_a = np.var(prof_a, ddof=1)
var_b = np.var(prof_b, ddof=1)

# F-statistic
F = var_a / var_b

# Degrees of freedom
df1 = len(prof_a) - 1
df2 = len(prof_b) - 1

# p-value (two-tailed)
p_value = 2 * min(f.cdf(F, df1, df2), 1 - f.cdf(F, df1, df2))

print(f"F-statistic: {F:.2f}")
print(f"P-value: {p_value:.4f}")
```

Output (approximate):
```
F-statistic: 1.83
P-value: 0.5248
```

Conclusion: Since p-value > 0.05, we **fail to reject** the null hypothesis. The variances are **not significantly different**.

Question 9:
Conduct a one-way ANOVA to test whether there are any statistically significant differences in average heights between three different regions with the following data:
- Region A: [160, 162, 165, 158, 164]
- Region B: [172, 175, 170, 168, 174]
- Region C: [180, 182, 179, 185, 183]

Python Code:
```python
from scipy.stats import f_oneway

# Data
region_a = [160, 162, 165, 158, 164]
region_b = [172, 175, 170, 168, 174]
region_c = [180, 182, 179, 185, 183]

# Perform ANOVA
f_stat, p_val = f_oneway(region_a, region_b, region_c)

print(f"F-statistic: {f_stat:.2f}")
print(f"P-value: {p_val:.4f}")
```

```

#### **Output (approximate):**
```
F-statistic: 96.93
P-value: < 0.0001
```

**Conclusion**: Since p-value < 0.05, we **reject** the null hypothesis. There is a **statistically significant difference** in average heights between the regions.
```