Harsha 🤍
Yeh sab answers clean assignment format me likh rahi hoon — directly copy kar sakti ho.

---

1) Explain the properties of the F-distribution.

Answer:

The F-distribution is a continuous probability distribution used mainly in hypothesis testing.

Properties:

• It is always positively skewed (right-skewed).
• It takes only positive values (range: 0 to ∞).
• It depends on two degrees of freedom: numerator (df_1) and denominator (df_2).
• The shape changes depending on degrees of freedom.
• It is the ratio of two independent chi-square distributions divided by their respective degrees of freedom.

The F-distribution is mainly used in comparing variances.

---

2) In which statistical tests is the F-distribution used, and why?

Answer:

The F-distribution is used in:

• F-test for comparing two variances
• ANOVA (Analysis of Variance)
• Regression analysis (overall model significance)

It is appropriate because the F-statistic is calculated as a ratio of two variances. Since variances follow a chi-square distribution, their ratio follows an F-distribution.

---

3) Key assumptions for conducting an F-test (comparing variances)

Answer:

The assumptions are:

• The samples are independent.
• The populations are normally distributed.
• The data is continuous.
• Observations are randomly selected.

If normality assumption is violated, the F-test may give incorrect results.

---

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

Answer:

Purpose of ANOVA:

ANOVA is used to test whether there are significant differences between the means of three or more groups.

Difference from t-test:

• A t-test compares means of two groups only.
• ANOVA compares means of three or more groups.
• Using multiple t-tests increases Type I error, while ANOVA controls it.

---

5) Why use one-way ANOVA instead of multiple t-tests?

Answer:

When comparing more than two groups:

• Multiple t-tests increase the probability of Type I error (false positive).
• One-way ANOVA performs a single overall test.
• It maintains the significance level (α).

Therefore, one-way ANOVA is statistically more appropriate and reliable.

---

6) How variance is partitioned in ANOVA?

Answer:

In ANOVA, total variance is divided into:

• Between-group variance
• Within-group variance

Total Variance = Between-group Variance + Within-group Variance

Between-group variance measures variation due to differences among group means.

Within-group variance measures variation inside each group.

F-statistic is calculated as:

F = (Between-group variance) / (Within-group variance)

If between-group variance is much larger, F-value becomes large, indicating significant difference among groups.

---

7) Classical (Frequentist) vs Bayesian ANOVA

Answer:

Classical (Frequentist) Approach:

• Uses p-values for hypothesis testing.
• Parameters are fixed but unknown.

- Decision is based on rejecting or not rejecting null hypothesis.
- Does not incorporate prior knowledge.

Bayesian Approach:

- Uses probability distributions for parameters.
- Incorporates prior beliefs (prior distribution).
- Provides posterior distribution.
- Gives probability of hypothesis being true.

Key Difference:

Frequentist gives probability of data given hypothesis.
Bayesian gives probability of hypothesis given data.

---

8) F-test for comparing variances (Profession A vs B)

Given data:

Profession A = [48, 52, 55, 60, 62]
Profession B = [45, 50, 55, 52, 47]

Step 1: Calculate variance

Variance A ≈ 35.3
Variance B ≈ 15.7

Step 2: Calculate F-statistic

F = Larger Variance / Smaller Variance

F = 35.3 / 15.7
F ≈ 2.25

Step 3: Interpretation

Since F-value is not extremely large and sample size is small, the difference in variances is not statistically significant at 5% level.

Conclusion:

We fail to reject the null hypothesis.
There is no strong evidence that variances are different.

Python Code:

```
import numpy as np

from scipy import stats


A = [48, 52, 55, 60, 62]

B = [45, 50, 55, 52, 47]
```

```
var_A = np.var(A, ddof=1)

var_B = np.var(B, ddof=1)


F = var_A / var_B


df1 = len(A) - 1

df2 = len(B) - 1


p_value = 1 - stats.f.cdf(F, df1, df2)


print("F-statistic:", F)

print("p-value:", p_value)
```

---

9) One-way ANOVA for Heights of Three Regions

Region A = [160, 162, 165, 158, 164]
Region B = [172, 175, 170, 168, 174]
Region C = [180, 182, 179, 185, 183]

Observation:

Means are clearly different:

Mean A ≈ 161.8
Mean B ≈ 171.8
Mean C ≈ 181.8

Since group means are far apart, ANOVA will likely show significant difference.

Python Code:

```
from scipy import stats


A = [160, 162, 165, 158, 164]

B = [172, 175, 170, 168, 174]

C = [180, 182, 179, 185, 183]


F_stat, p_value = stats.f_oneway(A, B, C)
```

```
print("F-statistic:", F_stat)
```

```
print("p-value:", p_value)
```

Interpretation:

If p-value < 0.05 → Reject null hypothesis.
There is significant difference between average heights of regions.