Adding IQR, Degrees of Freedom (DOF), and Skewness to the Study Guide

---

1. Interquartile Range (IQR)

Definition

The Interquartile Range (IQR) measures the spread of the middle 50% of the data. It is the difference between the **third quartile (Q3)** and the **first quartile (Q1)**.

Formula

$$\text{IQR} = Q3 - Q1$$

Steps to Calculate IQR

1. Arrange the data in ascending order.
2. Divide the data into four equal parts (quartiles).
   - $Q1$: The median of the lower half of the data (excluding the overall median).
   - $Q3$: The median of the upper half of the data.
3. Subtract $Q1$ from $Q3$.

Step-by-Step Example

Dataset: $[3, 7, 8, 12, 14, 15, 18]$

1. **Step 1**: Arrange the data in ascending order (already done).

2. **Step 2**: Find the median (Q2):
   - Median = 12.
3. **Step 3**: Find $Q1$ (lower half: $[3, 7, 8]$):
   - $Q1$ = Median of $[3, 7, 8] = 7$.
4. **Step 4**: Find $Q3$ (upper half: $[14, 15, 18]$):
   - $Q3$ = Median of $[14, 15, 18] = 15$.
5. **Step 5**: Calculate IQR:

$$\text{IQR} = Q3 - Q1 = 15 - 7 = 8.$$

Answer: IQR = 8.

Use of IQR

1. **Outlier Detection**:
   - Outliers are data points that fall outside:

$$[Q1 - 1.5 \cdot \text{IQR}, Q3 + 1.5 \cdot \text{IQR}].$$

2. **Example**:
   - Lower bound = $Q1 - 1.5 \cdot \text{IQR} = 7 - 1.5 \cdot 8 = -5$.
   - Upper bound = $Q3 + 1.5 \cdot \text{IQR} = 15 + 1.5 \cdot 8 = 27$.
   - Any data point outside $[-5, 27]$ is an outlier.

---

2. Degrees of Freedom (DOF)

Definition

Degrees of Freedom (DOF) refer to the number of independent values in a dataset that are free to vary when calculating a statistic.

Formula

1. For variance or standard deviation:
$$\text{DOF} = n - 1,$$
where n is the total number of data points.
2. For other contexts, DOF depends on the specific formula or test used (e.g., in regression analysis or chi-square tests).

Step-by-Step Example (Variance Calculation)

Dataset: $[4, 8, 6, 10]$

1. **Step 1**: Calculate the mean:
$$\mu = \frac{4 + 8 + 6 + 10}{4} = 7.$$
2. **Step 2**: Find squared differences:
   - $(4 - 7)^2 = 9,$
   - $(8 - 7)^2 = 1,$
   - $(6 - 7)^2 = 1,$
   - $(10 - 7)^2 = 9.$
3. **Step 3**: Apply $\text{DOF} = n - 1$:

- $n = 4$,
- $\text{DOF} = 4 - 1 = 3$.

4. **Step 4**: Calculate variance:

$$\sigma^2 = \frac{\text{Sum of squared differences}}{\text{DOF}} = \frac{9 + 1 + 1 + 9}{3} = \frac{20}{3} \approx 6.67.$$

Answer:

- Variance = 6.67,
- DOF = 3.

Why Use DOF?

1. Adjusts for bias in sample variance and standard deviation.
2. Ensures accurate estimates for population parameters.

---

3. Skewness

Definition

Skewness measures the asymmetry of a dataset. It indicates whether the data distribution is tilted to the left (negative skew) or right (positive skew).

Formula

1. **Sample Skewness**:

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \mu}{\sigma}\right)^3,$$

where:

- n: Number of data points,
- μ: Mean,
- σ: Standard deviation.

2. A simpler estimate:

$$\text{Skewness} = \frac{\frac{1}{n}\sum (x_i - \mu)^3}{\sigma^3}.$$

Interpreting Skewness

1. **Symmetric Distribution**:
   - Skewness = 0 (e.g., normal distribution).
2. **Positive Skew**:
   - Skewness > 0.
   - Tail extends to the right.
   - Example: Income distribution (most people earn less, few earn much more).
3. **Negative Skew**:
   - Skewness < 0.
   - Tail extends to the left.

- Example: Exam scores (most people score high, few score very low).

Step-by-Step Example

Dataset: $[2, 3, 3, 4, 10]$

1. **Step 1**: Calculate the mean:
$$\mu = \frac{2+3+3+4+10}{5} = \frac{22}{5} = 4.4.$$
2. **Step 2**: Calculate the standard deviation:
   - $(2-4.4)^2 = 5.76,$
   - $(3-4.4)^2 = 1.96,$
   - $(3-4.4)^2 = 1.96,$
   - $(4-4.4)^2 = 0.16,$
   - $(10-4.4)^2 = 31.36.$
   - Variance:
$$\sigma^2 = \frac{5.76+1.96+1.96+0.16+31.36}{5} = 8.24.$$
   - $\sigma = \sqrt{8.24} \approx 2.87.$
3. **Step 3**: Calculate skewness:
   - Cube deviations:
     - $(2-4.4)^3 = -13.82,$
     - $(3-4.4)^3 = -3.43,$
     - $(3-4.4)^3 = -3.43,$

- $(4 - 4.4)^3 = -0.06$,
- $(10 - 4.4)^3 = 117.65$.

- Sum of cubed deviations:

$$-13.82 - 3.43 - 3.43 - 0.06 + 117.65 = 96.91.$$

- Skewness:

$$\text{Skewness} = \frac{96.91}{(5 \cdot 2.87^3)} \approx 0.47.$$

Answer:

- Skewness = 0.47 (slightly positively skewed).

Key Takeaways

- **IQR**: Measures the spread of the middle 50% of the data.
- **DOF**: Adjusts for bias in statistical calculations.
- **Skewness**: Indicates the asymmetry of the data distribution.