Calculation Practices

Okay, let's solidify your understanding by working through sample calculations and answers for each concept. This will help you see how the formulas are applied in practice.

Descriptive Statistics

1. **Measures of Central Tendency:**
    - **Example Data:** Test scores: 60, 65, 70, 72, 78, 80, 85, 90, 92, 95
    - **Mean:** (60+65+70+72+78+80+85+90+92+95) / 10 = 787 / 10 = 78.7
    - **Median:** Sorted data: 60, 65, 70, 72, 78, 80, 85, 90, 92, 95
        - (78 + 80) / 2 = 79
    - **Mode:** Shown the most frequent data points appear in the set of data
2. **Measures of Dispersion:**
    - **Same Example Data:** Test scores: 60, 65, 70, 72, 78, 80, 85, 90, 92, 95. Mean = 78.7
    - **Range:** 95 - 60 = 35
    - **Variance (sample):**

    1. Find squared differences: (60-78.7)^2 = 349.69, (65-78.7)^2 = 187.69, (70-78.7)^2 = 75.69, (72-78.7)^2 = 44.89, (78-78.7)^2 = 0.49, (80-78.7)^2= 1.69, (85-78.7)^2 = 39.69, (90-78.7)^2 = 127.69, (92-78.7)^2 = 176.89, (95-78.7)^2 = 265.69
    2. Sum of squared differences: 349.69 + 187.69 + 75.69 + 44.89 + 0.49 + 1.69 + 39.69 + 127.69 + 176.89 + 265.69 = 1269.9
    3. Divide by n-1: 1269.9 / (10 - 1) = 141.1.
        - Sample variance = 141.1

    - **Standard Deviation (sample):**
        - Square root of variance: √141.1 = 11.88

3. **Percentiles:**
    - **Same Example Data:** Test scores: 60, 65, 70, 72, 78, 80, 85, 90, 92, 95
    - **25th Percentile (Q1):** Sorted data: 60, 65, 70, 72, 78, 80, 85, 90, 92, 95
        - Median of the first half: (65 + 70) / 2 = 67.5
    - **75th Percentile (Q3):**
    - Median of the upper half: (90 + 92) / 2 = 91
4. **Five Number Summary:**
    - **Same Example Data:** Test scores: 60, 65, 70, 72, 78, 80, 85, 90, 92, 95
    - Minimum: 60
    - Q1: 67.5
    - Median: 79
    - Q3: 91
    - Maximum: 95
5. **Trimmed Mean**
    - **Example Data:** Exam scores: 70, 75, 78, 80, 82, 85, 88, 90, 92, 95, 96
    - **10% Trimmed Mean:**
        1. Sort: 70, 75, 78, 80, 82, 85, 88, 90, 92, 95, 96
        2. 10% of 11 = 1.1 , round down to one.
        3. Remove 1 smallest and 1 largest: 75, 78, 80, 82, 85, 88, 90, 92, 95
        4. Average: 775 / 9 = 86.1
6. **Geometric Mean**
    - **Example Data:** Growth rate: Year 1: 5%, Year 2: 10%, Year 3: 8%
    - Convert percentages to growth factors:

      Year 1: 1.05

      Year 2: 1.10

      Year 3: 1.08

- Multiply all values:
    1. 05 * 1.10 * 1.08 = 1.2474
- Take the nth root:

  3√1.2474 = 1.0768
- Convert back to percentages:

  (1.0768 - 1) * 100% = 7.68 %

7. **Stem-and-Leaf Plot**
    - **Example data**: 12, 15, 18, 21, 23, 25, 27, 30, 31

      ```
      Stem | Leaf
      1 | 2 5 8
      2 | 1 3 5 7
      3 | 0 1
      ```

      content_copy download Use code with caution.
8. **Box Plot**
    - Using the five number summary: Minimum: 60, Q1: 67.5, Median: 79, Q3: 91, Maximum: 95
    - Draw a box from Q1 to Q3. A line in the box to indicate median. Lines to maximum and minimum.

Inferential Statistics

1. **T-Test (One Sample)**
    - **Scenario:** A company claims their light bulbs last 1000 hours. A sample of 10 bulbs lasts on average 990 hours with a standard deviation of 25 hours. Test if the bulbs last less than 1000 hours with alpha = 0.05.
    - **Null Hypothesis (H_o):** $\mu = 1000$
    - **Alternative Hypothesis (H_1):** $\mu < 1000$
    - **T-Statistic:** t = (990 - 1000) / (25 / √10) = -1.26
    - **Degrees of Freedom:** df = 10 - 1 = 9
    - **Critical Value:** (one-tailed) ~ -1.83

- **Decision:** Since -1.26 is not less than -1.83, we fail to reject the null hypothesis.

2. **T-Test (Paired Samples)**
    - **Scenario:** A weight loss program measures 5 participants' weight before and after.
    - **Data:** Before: 100, 110, 95, 105, 115. After: 95, 105, 90, 100, 110
    - **Difference (Before - After):** 5, 5, 5, 5, 5. Mean diff = 5.
    - **Standard Deviation of Differences:** 0.
    - **T-Statistic:** t = 5 / (0 / √5) (undefined, need a real variance). Let's assume that std. of diff is 2. T = 5/(2/√5) = 5.59
    - **Degrees of Freedom:** df = 5 - 1 = 4
    - **P-Value:** (one tailed). very small p-value
    - **Decision:** Reject the null hypothesis
3. **T-Test (Independent Samples)**
    - **Scenario:** Two groups of patients are given different medications for the same condition. We want to test if the average improvement of these two groups are different.
    - **Data:**
        - Group A: 7, 8, 9, 6, 7. Mean: 7.4. Standard deviation = 1.14
        - Group B: 5, 6, 5, 7, 8. Mean: 6.2. Standard deviation = 1.30
    - **T-Statistic:** t = (7.4 - 6.2) / √(1.14²/5 + 1.30²/5) = 1.54
    - **Degree of Freedom:** df = 5 + 5 - 2 = 8
    - **P-Value:** Two-tailed. ~0.16
    - **Decision:** Fail to reject null hypothesis.
4. **ANOVA**

- **Scenario:** Testing the effect of different study methods (online, in-person, hybrid) on students' test scores

- **Data:**
- Online: 75, 80, 85
- In person: 88, 90, 92
- Hybrid: 70, 72, 74
- **Between-groups:** variance of group means compared to the overall mean
- **Within-groups:** average variability within each groups
- **F-Statistic:** F = MS(Between)/MS(Within)
  - F ~ 14.2
- **P-Value:** (alpha = 0.05). p value is small. $p<0.05$
- **Degrees of Freedom:** df(between) = 2, df(within) = 6
- **Decision:** Reject the null hypothesis

1. **Chi-Square Test**

- **Scenario:** Examining relationship between two categorical variable; education level and employment status.
- **Data:**

Education	Employed	Unemployed
High school	50	20
College	70	10

- **Expected frequencies (example):**
  - total employed: 50+70 = 120
  - total unemployed = 20 + 10 = 30
  - total highschool = 70
  - total college = 80
  - overall total = 150
  - Expected high school employed = (70 * 120) / 150 = 56

    - ... repeat

      | Education | Employed | Unemployed |

      |---|---|---|

      | High school | 56 | 14|

      | College | 64 | 16 |

    - (50-56)^2/56 + (20-14)^2/14 + (70-64)^2/64 + (10-16)^2/16= ~10.5
- **Chi-Square Statistic:** ~10.5
- **Degrees of Freedom:** df = (2-1)(2-1) = 1
- **P-Value:** small p-value <0.05
- **Decision:** Reject null hypothesis, there is an association.

Probability

1. **Permutations:**
    - **Problem:** How many ways can you arrange 3 letters from ABC?
    - **Calculation:** 3! = 3 × 2 × 1 = 6 (ABC, ACB, BAC, BCA, CAB, CBA)
2. **Combinations:**
    - **Problem:** How many ways can you choose 2 letters from ABC?
    - **Calculation:** 3! / (2! * 1!) = 3 (AB, AC, BC)
3. **Hypergeometric:**
    - **Problem:** A bag contains 5 red and 3 blue balls. What is the probability of drawing exactly 2 red balls in 3 draws without replacement?
    - **Calculation:** ((5C2) * (3C1))/(8C3) = (10*3)/56 = 15/28
4. **Binomial Distribution:**
    - **Problem:** What is the probability of getting 2 heads in 5 flips of a fair coin?
    - **Calculation:** (5C2)*(0.5^2)*(0.5^3) = 10*(0.25)*(0.125) = 0.3125
5. **Z-score and Normal Distribution:**

- **Problem:** A test has a mean of 70 and std.dev of 10. What is the Z score for a student who got a score of 85?
- **Calculation:** Z = (85-70)/10 = 1.5

6. **Independent Events:**
    - **Problem:** What is the probability of rolling a 6 on a die and getting tails on a coin flip?
    - **Calculation:** P(6)*P(Tails) = 1/6 * 1/2 = 1/12

Correlation

1. **Pearson Correlation:**
    - **Data:**

      | X | Y |

      |---|---|

      | 1 | 2 |

      | 2 | 4 |

      | 3 | 6 |

      | 4 | 8 |

      | 5 | 10 |
    - **Calculation:** First find the standard deviation of X and Y then the covariance of X and Y.

      r = cov(X,Y)/(std(X)*std(Y))

      r = 1
    - **Interpretation:** A perfect positive linear relationship between X and Y.

Important Notes:

- These are simplified examples. Real-world data sets can be more complex.
- Always check the assumptions of each test before you apply them.

I hope this detailed walkthrough helps! Let me know if you'd like even more examples or have any other questions.