Data Analysis Probability Explanation

You said:

Uploaded image
I have a problem with the solution, related to Data Analysis, as shown in the picture.
Please explain the context of the problem.
Please extract all specific terms, math expressions, and notations.
Then explain each specific term with an analogy and a quantitative example.

ChatGPT said:

Context of the Problem

The problem is related to finding the probability that Sara Elison's commute on a particular day, when there is a home game at Fenway stadium, exceeds 22 minutes. The commute times are modeled using normal distributions, which vary depending on whether there is a game day or a non-game day.

Specific Terms, Math Expressions, and Notations

1. X_G:
    - Represents the random variable of Sara's commute time on a game day.
    - Distribution: $X_G \sim \mathcal{N}(20, 9)$
    - 20: The mean or average commute time on a game day.
    - 9: The variance of the commute time on a game day.
    - $\mathcal{N}(20, 9)$: A normal distribution with a mean of 20 minutes and a variance of 9 minutes squared.
2. $P(X_G > 22)$:
    - Represents the probability that the commute time exceeds 22 minutes on a game day.
3. **Standardization**:
    - The process of converting a random variable to a standard normal distribution, which has a mean of 0 and a variance of 1.
    - Formula: $\frac{X_G - \mu}{\sigma}$, where μ is the mean, and σ is the standard deviation.
4. Z:
    - Represents the standard normal variable, where $Z \sim \mathcal{N}(0, 1)$.
5. $P(Z > \frac{2}{3})$:
    - Represents the probability that the standardized commute time is greater than $\frac{2}{3}$.
6. **Symmetry of the Normal Distribution**:
    - A property that allows calculating the probability of a value being greater than a certain threshold by subtracting the cumulative probability of being less than or equal to that threshold from 1.
7. **Cumulative Probability**:

- The probability that a random variable takes a value less than or equal to a particular value.

8. **Standard Normal Distribution Table**:
   - A table that provides the cumulative probability corresponding to different values of the standard normal variable Z.

Explanation of Specific Terms with Analogies and Quantitative Examples

1. **Random Variable (X_G)**:
   - **Analogy**: Imagine rolling a die, but instead of getting a number between 1 and 6, you get a commute time in minutes. The commute time is random but follows a specific pattern, where some times are more likely than others.
   - **Quantitative Example**: If X_G follows a normal distribution with a mean of 20 minutes and a variance of 9 minutes squared, on most days, the commute will be around 20 minutes, but it can vary from, say, 17 to 23 minutes.
2. **Probability ($P(X_G > 22)$)**:
   - **Analogy**: Imagine you're pulling a ticket from a bag. The bag has tickets with different commute times. The probability $P(X_G > 22)$ is like asking, "What are the chances I pull a ticket that says the commute will take more than 22 minutes?"
   - **Quantitative Example**: If out of 100 days, 25 days had a commute time greater than 22 minutes, then $P(X_G > 22)$ would be 0.25 or 25%.
3. **Standardization**:
   - **Analogy**: Think of standardization like converting all temperatures from Fahrenheit to Celsius, so you have a consistent scale to compare different temperatures.
   - **Quantitative Example**: If X_G is 22 minutes, you standardize it by subtracting the mean (20) and dividing by the standard deviation (3). So, $\frac{22-20}{3} = \frac{2}{3}$.
4. **Standard Normal Variable (Z)**:
   - **Analogy**: Z is like having a universal measuring stick. No matter the original unit, you convert everything to this stick's units.
   - **Quantitative Example**: If after standardizing, you get $Z = \frac{2}{3}$, it tells you that the original commute time was $\frac{2}{3}$ of a standard deviation above the mean.
5. **Symmetry of the Normal Distribution**:
   - **Analogy**: Think of the normal distribution like a perfectly balanced seesaw. Whatever happens on one side of the pivot is mirrored on the other side.
   - **Quantitative Example**: If there's a 25% chance of exceeding a certain time on the right side, there's a 75% chance of being below that time on the left side.
6. **Cumulative Probability**:
   - **Analogy**: Imagine filling a glass with water. The cumulative probability is like asking, "How full is the glass at this point?"
   - **Quantitative Example**: If the standard normal table tells you that $P(Z \leq \frac{2}{3}) = 0.7486$, it means that 74.86% of the time, the commute will be less than or equal to the corresponding time.

By understanding these terms and using the standard normal distribution, you can calculate the probability that Sara's commute on a game day exceeds 22 minutes.

You said:

What core concepts are involved to solve the problem?
Please show detailed steps for obtaining the final solution.

ChatGPT said:

To solve this problem, several core statistical concepts and steps are involved:

Core Concepts Involved:

1. **Normal Distribution**:
   - Understanding the properties of the normal distribution, which is symmetric and bell-shaped, with mean μ and variance σ^2.
2. **Standardization (Z-score)**:
   - Converting a normal random variable to a standard normal variable to use standard normal tables.
3. **Cumulative Distribution Function (CDF)**:
   - Using the cumulative distribution function of the standard normal distribution to find probabilities.
4. **Symmetry of the Normal Distribution**:
   - Utilizing the property that the total area under the curve of a normal distribution is 1, and areas can be split symmetrically.

Detailed Steps to Obtain the Final Solution:

1. **Identify the Distribution**:
   - The commute time on a game day is normally distributed as $X_G \sim \mathcal{N}(20, 9)$.
   - Mean (μ) = 20 minutes.
   - Variance (σ^2) = 9 minutes^2.
   - Standard Deviation (σ) = $\sqrt{9} = 3$ minutes.
2. **Formulate the Probability Statement**:
   - You need to find the probability that the commute time exceeds 22 minutes: $P(X_G > 22)$.
3. **Standardize the Random Variable**:
   - Convert the normal random variable X_G to a standard normal variable Z using the formula:

$$Z = \frac{X_G - \mu}{\sigma}$$

   - Substitute the values:

$$Z = \frac{22 - 20}{3} = \frac{2}{3} \approx 0.6667$$

   - Now, you need to find $P(Z > \frac{2}{3})$.
4. **Use the Standard Normal Distribution**:
   - The standard normal distribution is denoted as $Z \sim \mathcal{N}(0, 1)$, where 0 is the mean, and 1 is the variance.
   - Look up the value $\frac{2}{3} \approx 0.67$ in the standard normal distribution table to find $P(Z \leq 0.67)$.

5. **Find the Cumulative Probability**:
   - From the standard normal table, $P(Z \leq 0.67) = 0.7486$.
   - Therefore, $P(Z > 0.67) = 1 - P(Z \leq 0.67)$:

$$P(Z > 0.67) = 1 - 0.7486 = 0.2514$$

6. **Conclusion**:
   - The probability that Sara's commute time on a game day exceeds 22 minutes is approximately 0.2514, or 25.14%.
   - Since the problem asks for the answer rounded to two decimal places:

$$\text{Final Answer} = 0.25$$

Final Solution:

The final probability that the commute on a particular game day exceeds 22 minutes is **0.25** or **25%**.

You said:

Uploaded image

Still related, I have the second problem (with solution) as shown in this picture.
What core concepts are involved to solve the problem?
Please show detailed steps for obtaining the final solution.

ChatGPT said:

Core Concepts Involved:

1. **Independence of Events**:
   - When the outcome of one event does not affect the outcome of another, the events are independent. In this problem, the commute times on different days are independent events.
2. **Multiplication Rule for Independent Events**:
   - The probability that all independent events occur simultaneously is the product of their individual probabilities.
3. **Normal Distribution**:
   - As in the first problem, the commute time on a game day follows a normal distribution, and the solution involves understanding this distribution.
4. **Probability of Exceeding a Threshold**:
   - This involves calculating the probability that a random variable exceeds a certain value, using the cumulative distribution function (CDF) and its complement.

Detailed Steps to Obtain the Final Solution:

1. **Understanding the Problem**:

- The problem asks for the probability that the commute time exceeds 22 minutes on all three days of a 3-game homestand.

2. **Identify the Probability of Exceeding 22 Minutes on a Single Day**:
   - From the first problem, you already know that the probability that the commute time on a game day exceeds 22 minutes is $P(X_G > 22) = 0.2514$.
3. **Recognize the Independence of Events**:
   - Since the commute times on different days are independent, the probability of the commute exceeding 22 minutes on all three days can be found using the multiplication rule for independent events.
4. **Apply the Multiplication Rule**:
   - Calculate the probability that the commute time exceeds 22 minutes on all three days:

$$P(\text{Commute} > 22 \text{ minutes on all 3 days}) = P(X_1 > 22 \text{ AND } X_2 > 22 \text{ AND } X_3 > 22)$$

$$P(X_1 > 22) \times P(X_2 > 22) \times P(X_3 > 22) = (P(X_G > 22))^3$$

   - Substitute the known value:

$$= (0.2514)^3$$

$$= 0.01588897274$$

5. **Round the Result**:
   - The final probability is approximately 0.0159.
   - Since the problem asks for rounding to two decimal places, round this result:

$$\approx 0.02$$

Final Solution:

The probability that Sara's commute time exceeds 22 minutes on all three days of a 3-game homestand is approximately **0.02** or **2%**.

You said:

Uploaded image
Still related, I have the next problem (with solution) as shown in this picture.
What core concepts are required to solve the problem?
Please show detailed steps for obtaining the final solution.

ChatGPT said:

Core Concepts Required:

1. **Difference of Two Independent Normal Distributions**:
    - When you subtract one independent normal random variable from another, the result is also normally distributed, with the mean and variance derived from the original distributions.
2. **Standardization (Z-score)**:
    - Converting the difference between two random variables to a standard normal distribution to calculate probabilities.
3. **Cumulative Distribution Function (CDF)**:
    - Using the cumulative distribution function of the standard normal distribution to find probabilities.
4. **Complement Rule**:
    - The probability of an event occurring is 1 minus the probability of the event not occurring.

Detailed Steps to Obtain the Final Solution:

1. **Understand the Problem**:
    - You need to find the probability that the commute time on a game day (X_G) exceeds the commute time on a non-game day (X_N).
2. **Define the Distributions**:
    - Commute time on a game day: $X_G \sim \mathcal{N}(20, 9)$
    - Commute time on a non-game day: $X_N \sim \mathcal{N}(12, 4)$
3. **Formulate the Probability Statement**:
    - You want to calculate $P(X_G > X_N)$.
    - This is equivalent to $P(X_G - X_N > 0)$.
4. **Determine the Distribution of the Difference**:
    - The difference $X_G - X_N$ is normally distributed with:
        - Mean: $\mu_{X_G - X_N} = \mu_{X_G} - \mu_{X_N} = 20 - 12 = 8$
        - Variance: $\sigma^2_{X_G - X_N} = \sigma^2_{X_G} + \sigma^2_{X_N} = 9 + 4 = 13$
    - Therefore, $X_G - X_N \sim \mathcal{N}(8, 13)$.
5. **Standardize the Random Variable**:
    - Convert the difference to a standard normal variable Z:

$$Z = \frac{X_G - X_N - \mu_{X_G - X_N}}{\sigma_{X_G - X_N}}$$

    - Substitute the values:

$$Z = \frac{0 - 8}{\sqrt{13}} = \frac{-8}{\sqrt{13}} \approx -2.218$$

6. **Find the Cumulative Probability**:
    - Use the standard normal distribution table to find $P(Z \leq -2.218)$.
    - However, the problem is asking for $P(X_G > X_N)$, which is $P(Z > -2.218)$.
    - By the symmetry and properties of the normal distribution:

$$P(Z > -2.218) = 1 - P(Z \leq -2.218)$$

    - From the standard normal table, $P(Z \leq 2.218) = 0.9868$, hence:

$$P(Z > -2.218) = 1 - 0.0132 = 0.9868$$

7. **Conclusion**:
   - The probability that the commute on a particular game day exceeds the commute on a non-game day is approximately **0.99**.

Final Solution:

The final probability that Sara's commute on a game day exceeds her commute on a non-game day is **0.99** or **99%**.