

Fill In The Blank Problems

(PDF/CDF)

In a company's data center, maintaining an optimal humidity level is crucial for the performance and stability of the servers. While high humidity can lead to condensation and corrosion, low humidity can cause static electricity build-up, which can damage sensitive electronic components. The deviation of humidity X from the ideal operating humidity level (measured in percentage points) is analyzed to ensure the servers operate within safe limits. The deviation of humidity typically ranges from **-3 to 3** percentage points because the data center's environmental control system is designed to maintain humidity within this narrow range. Deviations beyond this range are extremely rare due to the robust controls in place and are assumed negligible.

The probability density function (**PDF**) of the humidity deviation is given by:

$$f_X(x) = \begin{cases} -\frac{3}{104}(x^2 - 9) & -3 \leq x \leq -1 \\ \frac{3}{13} & -1 \leq x \leq 1 \\ -\frac{3}{104}(x^2 - 9) & 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

- a) The corresponding cumulative distribution function (**CDF**) of the humidity deviation is partially given by:

$$F_X(x) = \begin{cases} 0 & x < -3 \\ -\frac{1}{104}[x^3 - 27x - 54] & -3 \leq x < -1 \\ \frac{3}{13}x + \frac{1}{2} & -1 \leq x < 1 \\ \text{[Unknown]} & 1 < x < 3 \\ 1 & x \geq 3 \end{cases}$$

Determine the missing part of the **CDF** labeled **[Unknown]**. Type the correct answer as either A, B, C, or D.

[A] $-\frac{1}{104}[x^3 - 27x + 26]$

[B] $-\frac{1}{104}[x^3 - 27x - 50]$

$$[C] - \frac{1}{104} [x^3 - 27x - 54]$$

$$[D] - \frac{1}{104} [x^3 - 27x - 26]$$

- b) Determine the probability that the humidity deviation at a randomly sampled time during a particular day falls between -2 and 0. **(Use 4 decimal places.)**

$$P(-2 < X < 0) = P(X < 0) - P(X < -2)$$

$$= \left(\frac{3}{13} \times 0 + \frac{1}{2} \right) - \left(-\frac{1}{104} [(-2)^3 - 27 \times (-2) - 54] \right)$$

$$= \frac{1}{2} - \frac{1}{13} = \frac{11}{26} = 0.4230769$$

- c) Determine the value x representing the 71st percentile.

The **71st percentile** will fall between the interval $[-1, +1]$ because below -1 is at most $\frac{7}{26}$ of the area under the curve, and after $+1$ we would have accumulated at least $\frac{7}{26} + \frac{12}{26} = \frac{19}{26} = 0.7307692$ of the area.

Setting the CDF for the region $[-1, +1]$ equal to 0.71 allows us to find the **71st percentile**.

$$\frac{3}{13}x + \frac{1}{2} = 0.71$$

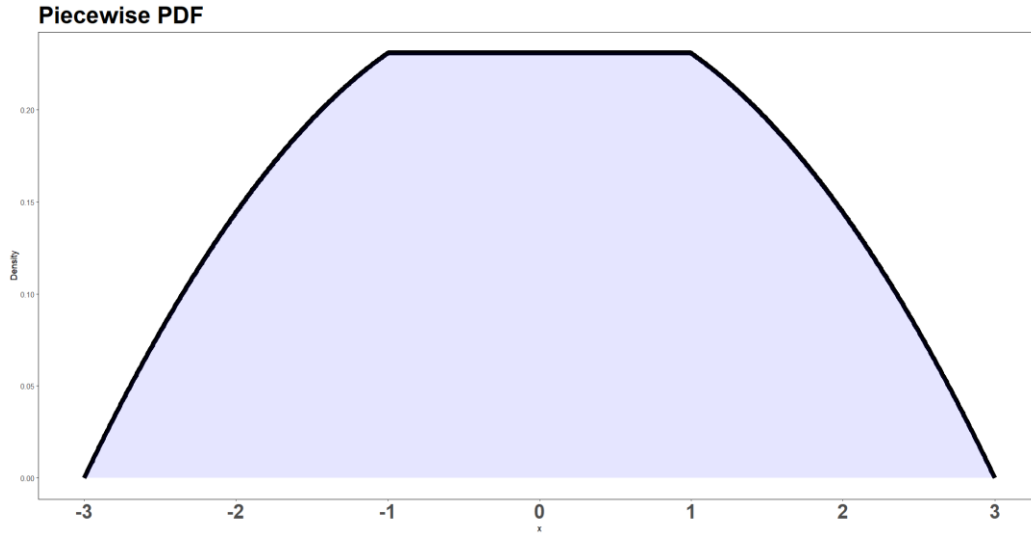
$$x = \left(\frac{13}{3} \right) (0.71 - 0.5)$$

$$x = 0.91$$

- d) Determine the expected value of the deviation of humidity for the company's data center.

The distribution is symmetrical about 0 as seen in the graph below and therefore both the median and expected value are 0.

$$E[X] = 0$$



- e) The efficiency loss Y of systems operating in the data center due to humidity deviation is modeled by the quadratic function:

$$Y = 0.1X^2 + 2$$

where:

- Y represents the efficiency loss (measured in percentage points).
- X is the humidity deviation from the ideal relative humidity (measured in percentage points).
- 0.1 is a constant that represents the sensitivity of efficiency loss to the square of the humidity deviation.
- 2 is a constant that represents the baseline efficiency loss when there is no deviation in humidity.

Studies have determined that the humidity deviation X has variance $\text{Var}(X) = \frac{121}{65}$. Using this knowledge determine the **expected** efficiency loss of systems operating in the data center, i.e., determine $E[Y]$.

$$\begin{aligned} E[Y] &= E[0.1X^2 + 2] = 0.1 \times E[X^2] + 2 = 0.1 \times (\text{Var}(X) + (E[X])^2) + 2 \\ &= 0.1 \times \frac{121}{65} + 2 = \frac{1421}{650} = 2.186154 \end{aligned}$$

(Conditional Probability Poisson Question) "Phony Solutions Inc." is a well-known call center that has been reported for fraudulent activities. The performance of the call center is monitored daily to track the number of failed scam attempts. The call center operates under two conditions: normal operations and under attack. If the call center is operating normally, there is a lower rate of failed scam attempts. However, if the call center is under attack by cybersecurity measures or law enforcement, the rate of failed scam attempts increases significantly.

- Event **N** is used to indicate that the call center is operating normally, and the probability of this event is known to be **$P(N) = 0.8$** .
- Event **F** is used to indicate the number of failed scam attempts in a day.
- Under normal operations the **number of failed scam attempts in a day follows a Poisson distribution** with a mean of $\lambda = 9$.

$$F|N \sim \text{Poisson}(\lambda = 9)$$
- When the call center is under attack (Event **N'**), the **number of failed scam attempts in a day follows a Poisson distribution** with a mean of $\lambda = 25$.

$$F|N' \sim \text{Poisson}(\lambda = 25)$$

- a) Determine the probability of exactly 18 failed scam attempts on a day in which the call center is operating normally. In other words find **$P(F = 18 | N)$** .

$$P(F = 18|N) = P(F = 18|\lambda = 9) = \frac{9^{18}e^{-9}}{18!} = 0.002893169 = 0.0029$$

- b) Determine the probability of exactly 18 failed scam attempts on a day in which the call center is under attack. In other words find **$P(F = 18 | N')$** .

$$P(F = 18|N') = P(F = 18|\lambda = 25) = \frac{25^{18}e^{-25}}{18!} = 0.03156582 = 0.0316$$

- c) Determine the probability of exactly 18 failed scam attempts on a day regardless of whether the call center is under attack or operating normally. In other words find **$P(F = 18)$** .

$$\begin{aligned} P(F = 18) &= P(\{F = 18\} \cap N) + P(\{F = 18\} \cap N') \\ &= P(F = 18|N)P(N) + P(F = 18|N')P(N') \\ &= 0.002893169 \times 0.8 + 0.03156582 \times 0.2 = 0.008627699 = 0.0086 \end{aligned}$$

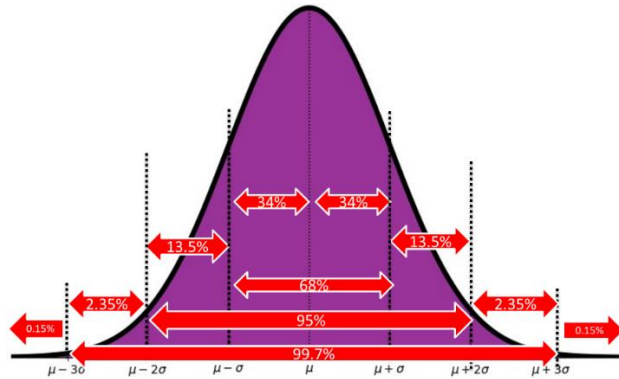
- d) On a particular day, the management of Phony Solutions Inc. observes that there were **18 failed scam attempts**. The management wants to determine the probability that the call center was operating normally on that day, given this observation.

$$\begin{aligned} P(N|F = 18) &= \frac{P(\{F = 18\} \cap N)}{P(F = 18)} = \frac{P(F = 18|N)P(N)}{P(F = 18)} \\ &= \frac{0.002893169 \times 0.8}{0.008627699} = \frac{0.002314535}{0.008627699} = 0.2682679 = 0.2683 \end{aligned}$$

(Normal Probability Question)

A tech company is evaluating the precision scores of its AI model used in an image recognition competition. Precision is defined as the ratio of true positive results to the total number of positive results (both true positives and false positives). The company conducted multiple runs of the model under different initializations for training. The precision scores from these runs are observed to approximately follow a **normal distribution** with a **mean** of **0.90** and a **standard deviation** of **0.02**. In what follows, assume that the precision scores **exactly** follow a **normal distribution** with a **mean** of **0.90** and a **standard deviation** of **0.02**.

- a) Using the **empirical rule (68-95-99.7 rule)**, determine the probability that a random initialization of this model results in a precision score **greater than 0.96 or less than 0.84**.



Let $X \sim N(\mu = 0.9, \sigma = 0.02)$

$$\begin{aligned} P(\{X < 0.84\} \cup \{X > 0.96\}) &= P(X < 0.84) + P(X > 0.96) \\ &= P(X < 0.9 - 3 \times 0.02) + P(X > 0.9 + 3 \times 0.02) \\ &\approx 0.0015 + 0.0015 = 0.003 \end{aligned}$$

- b) Determine the probability that a random initialization achieves a precision score **between 0.93 and 0.95**?

$$\begin{aligned} P(0.93 < X < 0.95) &= P\left(\frac{0.93 - 0.9}{0.02} < Z < \frac{0.95 - 0.9}{0.02}\right) \\ &= P(1.5 < Z < 2.5) = P(Z < 2.5) - P(Z < 1.5) \\ &= \Phi(2.5) - \Phi(1.5) = 0.9938 - 0.9332 = 0.0606 \end{aligned}$$

- c) Determine the precision score that represents the **99.66th percentile**.

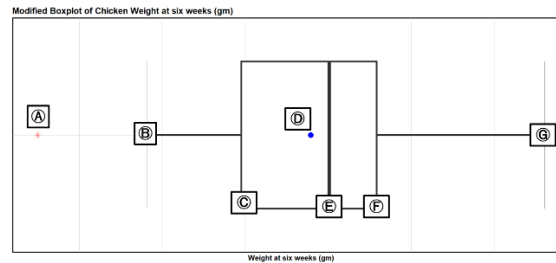
The **99.66th percentile** is $z = 2.71$ for the **standard normal distribution**. Convert this to the precision scores distribution:

$$x = \mu + \sigma \cdot z = 0.9 + 0.02 \times 2.71 = 0.9542$$

(Modified Boxplot) The following dataset records the weights of 24 chickens at six weeks of age, recorded to study the effects of various feeding regimes. Each chicken was fed one of two diets, Meatmeal, or Linseed. Below are the recorded weights (in grams) of the chickens:

75, 141, 148, 153, 169, 181, 203, 206, 213, 229, 242, 244,
257, 257, 258, 260, 263, 271, 303, 309, 315, 325, 344, 380

(Note: the data is stored in a dataframe `chick_weights` with variable `weight`)



The R code has provided the following statistical measures:

```
> quantile(chick_weights$weight)
 0%   25%   50%   75%  100%
75.0 197.5 250.5 279.0 380.0
> mean(chick_weights$weight)
[1] 239.4167
```

Refer to the modified boxplot provided above.

Based on the output from the R code below and the provided data, determine the numerical values indicated by the labels Ⓐ through Ⓔ.

Provide your answers in order (left to right) from Ⓐ through Ⓔ.

A, C, D, E, F, G are each worth 1.56 points and B is worth 2.64 points.

- Ⓐ 75.0
- Ⓑ 141
- Ⓒ 197.5
- Ⓓ 239.4167
- Ⓔ 250.5
- Ⓕ 279.0
- Ⓖ 380.0