Midterm 1 SUMMER 2024 - Requires Respondus LockDown Browser + We...

Attempt 1 of 1

Written Jul 3, 2024 11:43 AM - Jul 3, 2024 2:03 PM

Attempt Score **90.84 / 105 - B**

Overall Grade (Highest Attempt) **90.84 / 105 - B**

Question 1 **0 / 0.00000001 points**

Pre-approved links used in this exam:

z table

I attest that I have read and will follow all the instructions above honestly while taking this exam, and that the work I submit will be my own, produced without assistance from others (including other students in this class or AI technologies). Furthermore, I understand that if I share any information about the exam with another student before they have taken it, or if I use AI assistance of any kind, I will face actions including receiving a failing grade in the course and likely being reported to the Office of the Dean of Students for Academic Dishonesty

As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do.
Accountable together - We are Purdue.

- (x) True
- ( ) False

True/False Questions

Each **True/False question is worth 2 point**s. For each question, select **'True'** if you believe the statement is correct and **'False'** if you believe the statement is incorrect. Ensure that you carefully read each statement before making your selection.

Question 2 **2 / 2 points**

Consider a random variable **X** that follows a **Binomial distribution** with parameters **n = 5** and **p = 0.2**, and a random variable **Y** that follows a **Poisson distribution** with parameter **λ = 4**.

Consider a new random variable

$$Z = 8(Y+1) - X$$

if it is known that **X** and **Y** are independent, then it follows that

$$\mathbf{Var}(Z) = 320.8$$

- ( ) True

False

Question 3 **2 / 2 points**

Let X be a continuous random variable, then it follows that

$$\mathbf{P}(X < 5) \text{ is less than } \mathbf{P}(X \leq 5).$$

True

False

Question 4 **2 / 2 points**

Let **A** and **B** be events in a same sample space Ω. It is known that **P(A) > 0** and **P(B) > 0**. If **A** and **B** are mutually exclusive, then it must be true that the events **A** and **B** are complements of each other.

True

False

Question 5 **2 / 2 points**

Let **A** and **B** be events in a same sample space Ω, with **P(A) >0** and **P(B) >0**. It is known that

$$\mathbf{P(A \cap B) = P(A \mid B)P(B)},$$

therefore the events **A** and **B** must be independent.

True

False

Question 6 **2 / 2 points**

A marine biologist is studying the sizes of a specific species of fish in a lake. The lengths of the fish are **uniformly distributed** between **2 meters** and **15 meters**.

1. **Measurement A**: The biologist was interested in determining the probability that a fish selected at random was **less than 3.625 meters** given that it was **less than 5.25 meters**.
2. **Measurement B**: The biologist also wanted to determine the probability that a fish selected at random was **greater than 13.375 meters** given that it was **greater than 11.75 meters**.

Is it true that the probability calculated in **Measurement A** is **less** than the probability calculated in **Measurement B**?

True

False

Question 7 **2 / 2 points**

A wildlife conservation project is designed to monitor the population of various animal species in a national park. One of these variables measures the number of sightings of a particular species per day.

This variable is a numerical discrete variable.

- (•) True
- ( ) False

Multiple Choice Questions

Each **multiple-choice question is worth 3 points**. Please select the best choice for each problem.

Question 8 — 0 / 3 points

σ is an example of a

- ( ) a) None of the above
- ( ) b) population mean
- (•) c) measure of central tendency
- ( ) d) sample standard deviation
- ( ) e) sample statistic for central tendency
- ( ) f) location parameter

Question 9 — 3 / 3 points

Consider the following **dataset**:

Row Index	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5
1	NA	5	8	2	1
2	1	NA	NA	NA	9
3	7	5	NA	NA	4
4	2	1	4	7	5
5	NA	NA	NA	1	5
6	4	NA	7	9	3
7	NA	7	NA	2	8
8	6	3	8	4	1
9	5	6	2	3	7
10	8	2	9	3	6

What is the result of running the following command in **R** `sum(complete.cases(dataset))`?

a) 38

b) 6

c) 1

d) 4

e) 12

Question 10 **3 / 3 points**

Let $X_1, X_2, \ldots, X_{10}$ be $n = 10$ independent normal random variables with identical mean and standard deviation as given below.

$$X_i \sim N(\mu = 100, \sigma = 10), \text{for} | i = 1, \ldots, 10$$

Consider a new random variable constructed as the average of these 10 normal random variables as given below.

$$\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$$

Using the fact that the random variables $X_1, X_2, \ldots, X_{10}$ are independent and the rules of expectations and variances, we find that

$$\mathbf{E}[\bar{X}] = \frac{1}{10} \sum_{i=1}^{10} \mathbf{E}[X_i]$$

and

$$\mathbf{Var}(\bar{X}) = \frac{1}{10^2} \sum_{i=1}^{10} \mathbf{Var}(X_i)$$

It can be shown that the distribution of $\overline{X}$ will be normal with mean $E[\overline{X}]$and variance $\text{Var}(\overline{X})$. Using this fact and the formulas above for $E[\overline{X}]$ and $\text{Var}(\overline{X})$, determine the probability that $\overline{X}$ is greater than **95**.

In other words, find

$$\mathbf{P}(\bar{X} > 95)$$

a) 0.0606

b) 0.5636

c) 0.4364

d) 0.9394

e) 0

FIB Questions

DIRECTIONS: Fill in each blank with the numeric answer or letter only—do not use symbols like **"="** or **"$"**.

Follow these guidelines:

1. **Leading Zero**: Use a leading zero for any value between 0 and 1 (e.g., use 0.2463, not .2463).
2. **Intermediate Rounding:** Preserve at least eight digits in intermediate calculations.
3. **Final Rounding**: Round your answers to four decimal places. Do not add extra zeros (e.g., use 0.45, not 0.4500).
4. **No Extra Characters**: Do not type any extra characters. (This includes white space characters.)

Ensure you adhere to these instructions to avoid losing points.

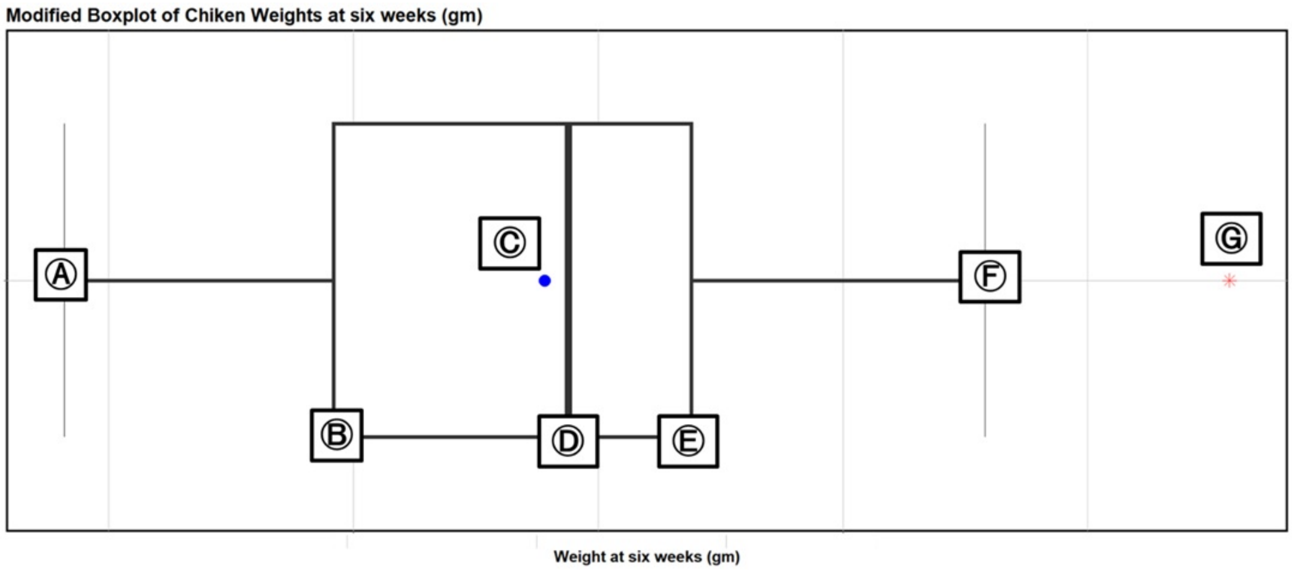
If you are unable to determine the final numerical solution but have gotten a mathematical expression for the solution you may type it in the answer box, but you will not get full points.

Question 11 — 10.44 / 12 points

The following dataset records the weights of **26 chickens** at six weeks of age, recorded to study the effects of various feeding regimes. Each chicken was fed one of two diets, including those containing Casein, Meatmeal, or Linseed. Below are the recorded weights (in grams) of the chickens:

$$141, 148, 158, 169, 171, 181, 193, 199, 203, 213, 229, 230, 243,$$
$$244, 248, 250, 257, 260, 267, 271, 271, 309, 316, 327, 329, 379$$

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



The R code has provided the following statistical measures:

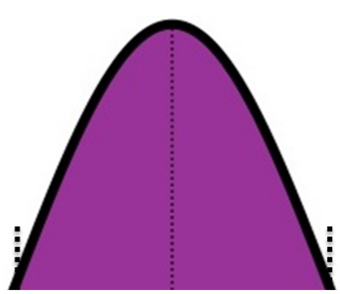
```
> quantile(chick_weights$weight)
   0%   25%   50%   75%  100%
 141.0 196.0 244.0 269.0 379.0
> mean(chick_weights$weight)
[1] 239.037
```

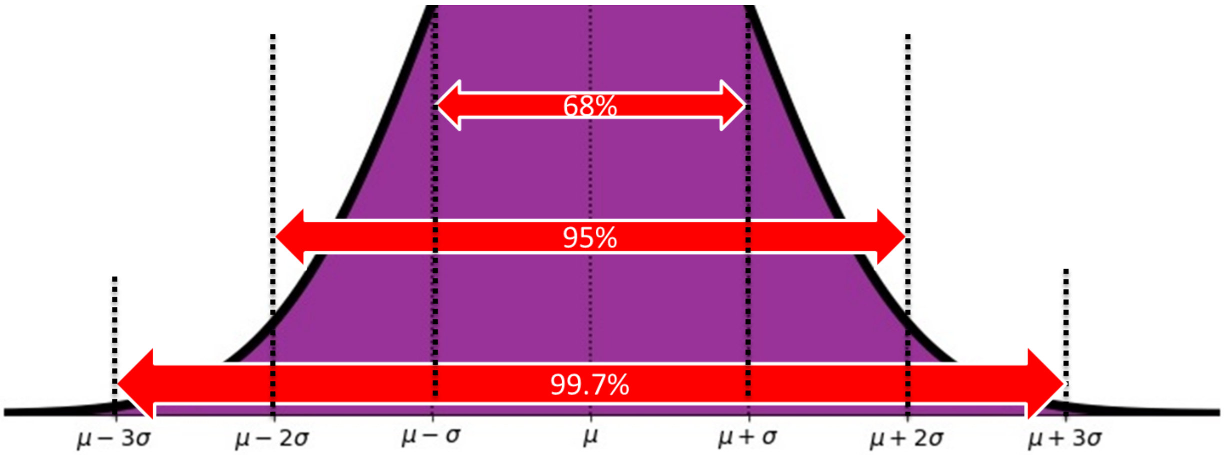
Refer to the modified boxplot provided above.
Based on the output from the R code below and the provided data, determine the numerical values indicated by the labels Ⓐ **through** Ⓖ.

Provide your answers in order (left to right) from Ⓐ **through** Ⓖ.
___86.5___ ✗ **(141)** ___196___ ✓(13 %) ___239.037___ ✓(13 %) ___244___ ✓(13 %) ___269___ ✓(13 %) ___329___ ✓(22 %) ___379___ ✓(13 %)

Question 12 — 16 / 16 points

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.**





The visualization above is to assist you with answering part **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 0.94 or less than 0.84**.

Note: You must use the empirical rule for this part, or you will be marked wrong. Answers should be in probabilities (decimal) not percentages.

___0.0265___ (25 %)

Determine the probability that a random initialization achieves a precision score **between 0.85** and **0.87**?

Note: You must use the **normal table for this part.**

___0.0606___ (37.5 %)

Determine the precision score that represents the **99.49th percentile**.

Note: You must use the **normal table for this part.**

___0.9512___ (0.9514)

Question 13 **20.4 / 30 points**

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 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:

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

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

$$F_X(x) = \begin{cases} 0 & x < -3 \\ -\frac{1}{104}\left[x^3 - 27x - 54\right] & -3 \le x < -1 \\ \frac{3}{13}x + \frac{1}{2} & -1 \le x < 1 \\ [\textbf{Unknown}] & 1 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

Determine the missing part of the **CDF** labeled **[Unknown]**.

Type the correct answer as either **A, B, C, or D**.

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

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

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

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

___B___ ✓(36 %)

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.)
___0.4231___ ✓(16 %)

Determine the value ***x*** representing the **71st percentile**.
___0.91___ ✓(16 %)

Determine the expected value of the deviation of humidity for the company's data center.
___0.153___ ✖ **(0, 0.0000, E[X]=0, E[X] = 0)**

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 **Var(*X*) = 121/65**.

Using this knowledge determine the **expected efficiency loss** of systems operating in the data center, i.e., determine **E[Y]**.

(Use 4 decimal places.)

________ ✖ (2.186154, 2.1862, 2.1861, 2.18615)

Question 14 — 26 / 26 points

"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 **λ=12**.

$$\mathbf{F|N} \sim \mathbf{Poisson}(\lambda = 12)$$

- 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 **λ=30**.

$$\mathbf{F|N'} \sim \mathbf{Poisson}(\lambda = 30)$$

Determine the probability of **exactly 24 failed scam attempts** on a day in which the **call center is under attack**.

$$\mathbf{P(F} = 24 \mid \mathbf{N'}) = ?$$

Use f**our decimal places for your final answer**, but **retain all decimal places in your calculations** as they will be **needed** for the **subsequent parts.**

___0.0426___ ✓(15 %)

Determine the probability of **exactly 24 failed scam attempts on a day** in which the call center is **operating normally**.

$$\mathbf{P(F} = 24 \mid \mathbf{N}) = ?$$

Use f**our decimal places for your final answer**, but **retain all decimal places in your calculations** as they will be **needed** for the **subsequent parts.**

___0.0008___ ✓(15 %)

Determine the probability of **exactly 24 failed scam attempts on a day** regardless of whether the call center is under attack or operating normally.

$$\mathbf{P}(\mathbf{F} = 24) = ?$$

Use f**our decimal places for your final answer**, but **retain all decimal places in your calculations** as they will be **needed** for the **subsequent parts**.

___0.0091___ ✓(30 %)

On a particular day, the management of Phony Solutions Inc. observes that there were **24 failed scam attempts**. The management wants to **determine the probability that the call center was under attack on that day**, **given this observation of 24 failed scam attempts**.

$$\mathbf{P}(\mathbf{N}' \mid \mathbf{F} = 24) = ?$$

(Use 4 decimal places.)

___0.9312___ ✓(40 %)

Done