MATH40005 Spring 2022 Blackboard Quiz 1

Question 1

Suppose you obtain the sample of measurements $\{4, 9, 2, 7, 8\}$. The sample mean of this sample is

(a) 6

(b) 5

(c) 5.2

(d) 10

(e) 7

Question 2

Suppose you obtain the sample of measurements $\{4, 9, 2, 7, 8\}$. The sample median of this sample is

(a) 6

(b) 5

(c) 5.2

(d) 10

(e) 7

Question 3

Consider the real-valued random variable X. Then the minimum value of the quantity $\mathrm{E}(|X-a|)$ always occurs when a has the value equal to

(a) the mean of X

(b) the mode of X

(c) the median of X

(d) the variance of X

(e) 0

Question 4

Suppose that the random variable X follows the distribution F_X (i.e. it has cumulative distribution function F_X), which is unknown. However, it is known that the mean of X is $\mathrm{E}[X] = \mu$ and that the variance of X is $\mathrm{Var}[X] = \sigma^2$. Then the smallest value of p for which the inequality

$$P(|X-\mu| \geq 5\sigma) \leq p$$

is true is:

(a) $\frac{1}{5}$

(b) $\frac{1}{10}$

(c) $\frac{1}{25}$

(d) 0.25

(e) 0.01

Question 5

Suppose that the random variable X has mean μ and variance σ^2. Then the quantity $E[X^2]$ is known as

(a) The second (raw) moment of X

(b) The second central moment of X

(c) The variance of X

(d) The mean-squared of X

(e) The first moment of X

Question 6

Suppose that the random variable X has mean μ and variance σ^2. Then the quantity $E[X^2]$ is equal to

(a) $\mu^2 + \sigma^2$

(b) $\mu + \sigma^2$

(c) $\mu^2 + \sigma$

(d) μ^2

(e) σ^2

Question 7

Suppose that the collection of n random variables $X_1, X_2, \ldots, X_n$ are independently and identically distributed according to distribution F_X and have mean denoted by θ, i.e. each X_i has cumulative distribution function F_X and $\mathrm{E}[X_i] = \theta$ for $i = 1, 2, \ldots, n$. Suppose that θ is unknown. You wish to estimate the mean θ, after observing realisations of the random variables, and read a blog post online that suggests using the estimator

$$\widehat{\Theta}(X_1, \ldots, X_n) = \frac{1}{n+1} \sum_{i=1}^{n} X_i.$$

This estimator looks a bit different to what was suggested in your notes, so you decide to compute its bias. The **bias** of $\widehat{\Theta}(X_1, \ldots, X_n)$, when considering it as an estimator of the mean θ, is:

(a) $\dfrac{1}{n+1}\theta$

(b) $\dfrac{1}{n}\theta$

(c) 0

(d) $\dfrac{-1}{n}\theta$

(e) $\dfrac{-1}{n+1}\theta$

Question 8

Suppose that the collection of n random variables $X_1, X_2, \ldots, X_n$ are independently and identically distributed according to distribution F_X which has unknown mean θ and known variance σ^2, where $\sigma > 0$. In other words, for $i = 1, 2, \ldots, n$, each X_i has cumulative distribution function F_X and $\mathrm{E}[X_i] = \theta$ and $\mathrm{Var}[X_i] = \sigma^2$. You decide to use the estimator

$$\widehat{\Theta}(X_1, \ldots, X_n) = \frac{1}{n+1} \sum_{i=1}^{n} X_i$$

to estimate the unknown mean θ. The **standard error** of this estimator is

(a) σ^2

(b) σ

(c) $\dfrac{\sigma^2 n}{(n+1)^2}$

(d) $\dfrac{\sigma\sqrt{n}}{n+1}$

(e) $\dfrac{\theta^2}{n+1}$

Question 9

Suppose someone decides to determine the most common colour of cars in South Kensington by recording the colour of cars driving down Exhibition Road between 12.00 and 13.00 on a particular day. The 1234 measurements are recorded as

$$\{R, B, G, S, R, R, G, W, B, \ldots, S, G, G, S, B, R\}$$

where R is red, B is blue, G is green, S is silver, W is white, etc.

Given this data set, which visualisation would most clearly indicate the **mode** of the data?

(a) Scatterplot

(b) Lineplot

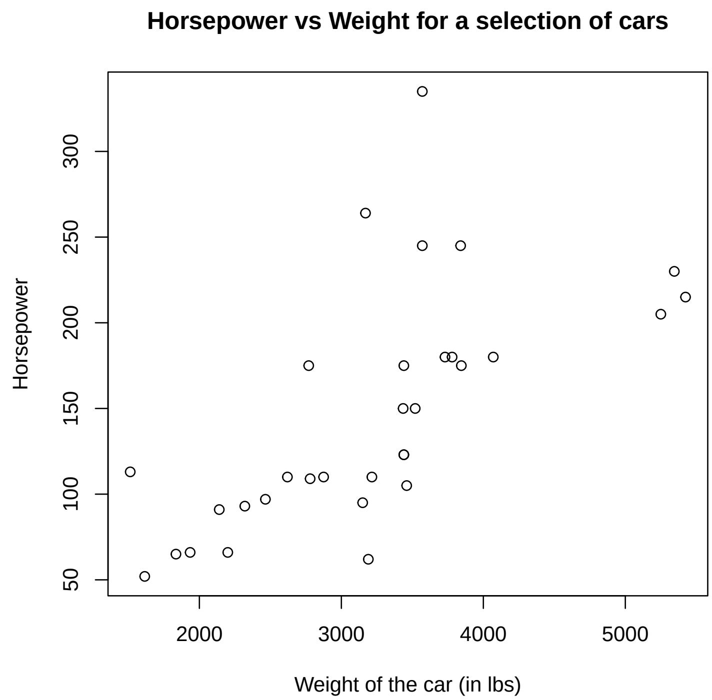
(c) Bar chart

(d) Cumulative distribution function

(e) None of the above

Question 10

Suppose the weight and horsepower of 32 cars have been measured, and the statistician decides to visualise the data using the scatterplot below.



Which of the following relationships between the horsepower and weight of these cars does the scatterplot seem to suggest is true? **Select all that apply**.

(a) Heavier cars tend to have lower horsepower.

(b) Lighter cars tend to have lower horsepower.

(c) Most cars have a horsepower higher than 200.

(d) Cars with a weight between 3000 lbs and 4000 lbs can have either low or high horsepower.

(e) The heavier a car is, the higher its horsepower must be.

Solution

1. **(a)**; $(4+9+2+7+8)/5 = 6$

2. **(e)**; middle value of ordered $(2, 4, 7, 8, 9)$ is 7.

3. **(c)**; Theorem 1.6.13 in the notes

4. **(c)**; Chebyshev's inequality, Corollary 1.3.5 in the notes

5. **(a)**; since by definition $\mathrm{E}[X^k]$ is the kth moment of X.

6. **(a)**; Using the identity $\mathrm{E}[X^2] = \mathrm{Var}[X] + (\mathrm{E}[X])^2 = \sigma^2 + \mu^2$.

7. **(e)**; $\mathrm{E}[\widehat{\Theta}] - \theta = \frac{n}{n+1}\theta - \theta = -\frac{1}{n+1}\theta$

8. **(d)**;

$$\begin{aligned}
\mathrm{Var}(\widehat{\Theta}) &= \mathrm{Var}\left(\frac{1}{n+1}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{(n+1)^2}\sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) \\
&= \frac{1}{(n+1)^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{n\sigma^2}{(n+1)^2} \\
\Rightarrow \mathrm{SE}_{\widehat{\Theta}} &= \sqrt{\mathrm{Var}(\widehat{\Theta})} = \frac{\sigma\sqrt{n}}{n+1}
\end{aligned}$$

9. **(c)** The tallest bar in a bar chart is the catgeory with the largest number of elements in the set, which is by definition the mode.

10. **(b) and (d)** no evidence for (a), (b) appears to be true, (c) is false because most cars have horsepower less than 200, (d) appears to be true, (e) is not true because the heaviest cars do not have the highest horsepower.