University of Texas at Austin, Department of Mathematics
M358K - Applied Statistics

The Mock In-term One

**Problem 1.1.** (5 points) What is the **R** output of the following command:

```
>pbinom(2,4,0.5)
```

(a) 0.375
(b) 0.5
(c) 0.6875
(d) 0.75
(e) None of the above.

**Solution: (c)**
This is exactly the probability that a random variable $X \sim Binomial(size = 4, p = 0.5)$ takes the value less than or equal to 2. We have

$$\begin{aligned}\mathbb{P}[X \leq 2] &= \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \mathbb{P}[X = 2]\\ &= \binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 + \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2\\ &= (1 + 4 + 6)\left(\frac{1}{2}\right)^4 = \frac{11}{16} = 0.6875.\end{aligned}$$

**Problem 1.2.** (5 points) Consider the following charts, focusing on the bar graph on the right-hand side, and choose which of the offered statements is **correct**.


(a) The shortest number of work days needed to purchase a visa is for Europe.
(b) The more affluent the region, the longer one needs to work to purchase the visa.

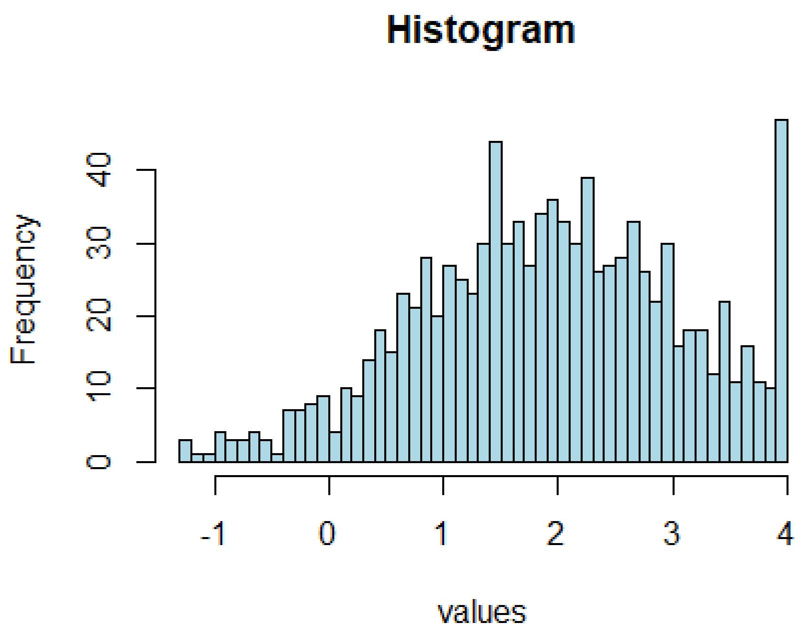
(c) One has to work more than three times as long in Sub-Saharan Africa than in the Middle East to purchase a visa.

(d) One has to work longer in Oceania than in Asia to purchase a visa.

(e) None of the above.

**Solution: (c)**

**Problem 1.3.** (5 points) Consider the following histogram:




The histogram is ...

(a) ...unimodal.

(b) ...left-skewed.

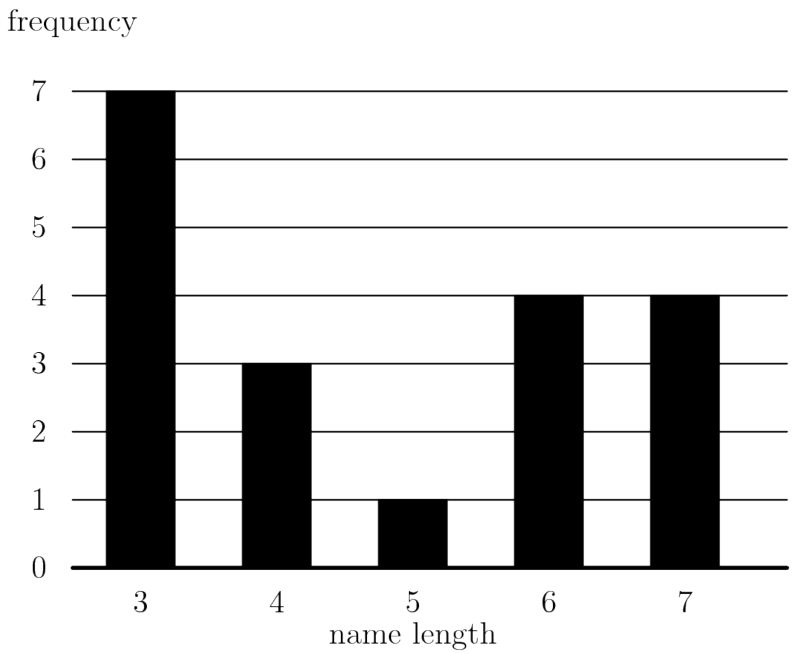
(c) ...right-skewed.

(d) ...symmetric.

(e) None of the above.

**Solution: (b)**
The ”hump” is on the right; see how the values ”max-out” at 4. So, it’s left-skewed.

**Problem 1.4.** (5 points) *Source: AMC8, 2016.* The following bar graph represents the length (in letters) of the names of 19 people. What is the median length of these names?




(a) 3

(b) 4

(c) 5

(d) 6

(e) None of the above.

**Solution: (b)**
According to the bar graph, there is a total of $7+3+1+4+4 = 19$ names. The median name length is, thus, in the $10^{th}$ spot. This means it's of length 4.

**Problem 1.5.** (5 points) Which one of the following statements is **false**?

(a) The observed sample variance is more sensitive than the sample average to a few observations with extreme values.

(b) The observed sample variance is equal to zero if and only if all of the observations are identical.

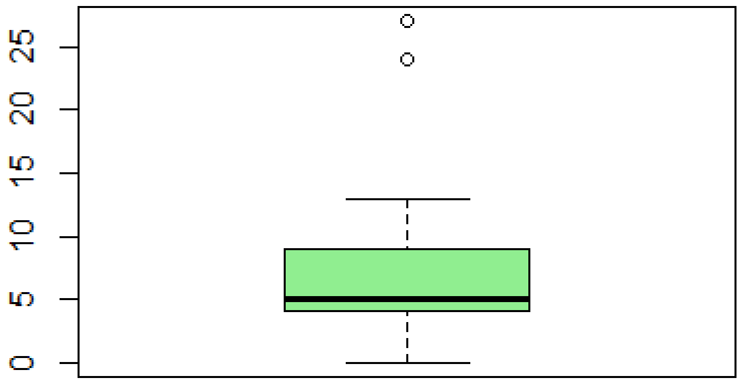
(c) When all observations are multiplied by a constant $\kappa \neq 0$, the IQR increases by $|\kappa|$.

(d) When all the observations are increased by 5, the IQR increases by 5.

(e) None of the above.

**Solution: d**

**Problem 1.6.** (5 points) Consider the following box plot:




Which summary statistics does it correspond to?

(a) Min=0.000, Q1=0.000, median=5.000, mean=7.103, Q3=9.000, Max=27.000

(b) Min=0.000, Q1=0.000, median=5.000, mean=7.103, Q3=12.500, Max=27.000

(c) Min=0.000, Q1=4.000, median=7.103, mean=5.000, Q3=9.000, Max=27.000

(d) Min=0.000, Q1=4.000, median=5.000, mean=7.103, Q3=9.000, Max=27.000

(e) None of the above.

**Solution: (d)**

**Problem 1.7.** (5 points) Ahead of the school year, the KMS Functional-Fitness coach plans to track the progress of her students. She will record the number of jumping jacks her students can make before collapsing on a weekly basis. What kind of a procedure is this?

(a) A prospective observational study.

(b) A retrospective observational study.

(c) An experiment.

(d) A survey.

(e) None of the above.

**Solution: (a)**

**Problem 1.8.** (5 points) **Post-It Thievery!**

Chris P. Bacon, the office manager at a large temp agency wants to figure out what proportion of his workforce has been pilfering Post-Its. Realizing the issues with conducting a survey which outright asks: *"Have you ever committed unauthorized removal of Post-Its from the premises?"*, he decides to use the randomized-response method.

He prompts a computer to display the question

*"Have you ever taken a Post-It home?"*

with probability 0.75. The rest of the time, a virtual fair coin is flipped on the screen and the subject is asked

*"Is the outcome heads?"*

In both cases, the subject is prompted to click the button with **Yes** or **No**. The interviewer did not know the actual question asked, just the ultimate response. So, there was no real reason for the subject to lie, and we assume that the subjects responded truthfully.

It turned out that 80% of the subjects answered "yes". Give an estimate of the proportion of *Post-It thieves* in this population.

(a) 0.75

(b) 0.8

(c) 0.85

(d) 0.9

(e) None of the above.

**Solution: (d)**

Now, we are given that $\mathbb{P}[Yes] = 0.80$. Our goal is to figure out $p = \mathbb{P}[Yes \mid Q]$ with the conditioning event $Q$ given by

$$Q = \{\text{the subject was asked the Post-It question}\}.$$

We are given that $\mathbb{P}[Q] = 0.75$.

By the *Law of Total Probability*,

$$\begin{aligned}\mathbb{P}[Yes] = \mathbb{P}[Yes \cap Q] + \mathbb{P}[Yes \cap Q^c] &= \mathbb{P}[Yes \mid Q]\mathbb{P}[Q] + \mathbb{P}[Yes \mid Q^c]\mathbb{P}[Q^c] \\ &= p(0.75) + 0.5(0.25) = 0.75p + 0.125.\end{aligned}$$

So,

$$0.75p = 0.8 - 0.125 = 0.675 \quad \Rightarrow \quad p = 0.9.$$

**Problem 1.9.** (5 points) There are three variants of a genetic marker for *goosepox*: **immune, middling,** and **susceptible**. In the population, 10% are **immune**, 70% are **middling**, and 20% are **susceptible**. Within each category, here are the chances of contracting *goosepox*:

- for **immune** it is 0%,
- for **middling** it is 50%, and
- for **susceptible** it is 90%.

Say that you learn that a randomly chosen individual contracted *goosepox*. What is the probability that this individual was **susceptible**?

(a) 0.18

(b) 0.34

(c) 0.53

(d) 0.9

(e) None of the above

**Solution: (b)**
By the Bayes' Theorem,

$$\mathbb{P}[Susc \,|\, Goose] = \frac{\mathbb{P}[Goose \,|\, Susc]\mathbb{P}[Susc]}{\mathbb{P}[Goose]}.$$

By the Law of Total Probability, we have

$$\begin{aligned}\mathbb{P}[Goose] &= \mathbb{P}[Goose \,|\, Imm]\mathbb{P}[Imm] + \mathbb{P}[Goose \,|\, Mid]\mathbb{P}[Mid] + \mathbb{P}[Goose \,|\, Susc]\mathbb{P}[Susc] \\ &= 0(0.10) + 0.5(0.7) + 0.9(0.2) = 0.53.\end{aligned}$$

So,

$$\mathbb{P}[Susc \,|\, Goose] = \frac{0.9(0.2)}{0.53} = 0.3396226.$$

**Problem 1.10.** (5 points) A piggy bank contains coins of three different types: $T_1, T_2$ and $T_3$.

There are twice as many type $T_1$ coins as type $T_2$ coins, and twice as many type $T_2$ coins as type $T_3$ coins. The coins are indistinguishable to touch.

Coins of type $T_1$ are fair, coins of type $T_2$ come up heads (H) when tossed with probability 3/10, and coins of type $T_3$ come up heads (H) when tossed with probability 1/10.

A coin is drawn from the piggy bank at random and tossed. What is the probability that the result of the coin toss was heads?

(a) 1/5

(b) 17/60

(c) 27/70

(d) 1/2

(e) None of the above.

**Solution: (c)**

Let the probability that the coin is of type $T_i$ be denoted by $p_i$ for $i = 1, 2, 3$. From the problem statement, we have that

$$p_1 = 2p_2 = 4p_3.$$

Since $p_1 + p_2 + p_3 = 1$, we have that $p_3 = 1/7$, $p_2 = 2/7$ and $p_1 = 4/7$.

By the rule of average conditional probabilities, we get

$$\begin{aligned}\mathbb{P}[H] &= \mathbb{P}[T_1]\mathbb{P}[H \mid T_1] + \mathbb{P}[T_2]\mathbb{P}[H \mid T_2] + \mathbb{P}[T_3]\mathbb{P}[H \mid T_3] \\ &= \frac{4}{7} \cdot \frac{1}{2} + \frac{2}{7} \cdot \frac{3}{10} + \frac{1}{7} \cdot \frac{1}{10} = \frac{27}{70}.\end{aligned}$$

**Problem 1.11.** (5 points) In Country X, the latest census has revealed the following:

- 40% of the population exercise regularly,
- 30% own a dog,
- 20% like cauliflower,
- 60% of all dog owners exercise regularly,
- 10% own a dog and like cauliflower,
- 4% exercise regularly, own a dog and like cauliflower.

A person is selected at random. Compute the probability that they are a dog owner who does not exercise regularly.

(a) 0.12

(b) 0.14

(c) 0.16

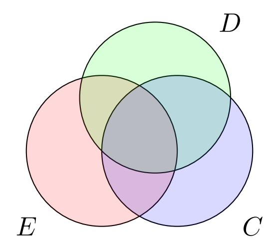
(d) 0.18

(e) None of the above.

**Solution: (a)**

Let $E$ denote the event that the person chosen exercises regularly, $D$ that he/she owns a dog and $C$ that he/she likes cauliflower. The problem states that

$$\mathbb{P}[E|D] = 0.6, \qquad \mathbb{P}[C \cap D] = 0.1, \qquad \mathbb{P}[C \cap D \cap E] = 0.04,$$
$$\mathbb{P}[E] = 0.4, \qquad \mathbb{P}[D] = 0.3, \qquad \mathbb{P}[C] = 0.2.$$



It follows that $\mathbb{P}[E \cap D] = \mathbb{P}[E|D] \times \mathbb{P}[D] = 0.6 \times 0.3 = 0.18$, so $\mathbb{P}[D \cap E^c] = \mathbb{P}[D] - \mathbb{P}[E \cap D] = 0.12$.

**Problem 1.12.** (5 points) Two coins are tossed. If $E_1$ is the event "heads on first coin", $E_2$ the event "head on the second coin", and $E_3$ the event "the coins match; both are heads or tails". Which of the following statement is **not** true?

**(a):** $E_1$ and $E_2$ are independent.
**(b):** $E_2$ and $E_3$ are independent.
**(c):** $E_1$ and $E_3$ are independent.
**(d):** $E_3$ and $E_1$ are independent.
**(e):** $E_1$, $E_2$ and $E_3$ are independent.

**Solution: (e)**