UNIVERSITY OF TEXAS AT AUSTIN, DEPARTMENT OF MATHEMATICS
M358K - Applied Statistics

THE PREREQUISITE IN-TERM EXAM

1.1. TRUE/FALSE QUESTIONS.

Problem 1.1. (2 pts)
Let E and F be any two events. If

$$\mathbb{P}[E] = \mathbb{P}[F] = \frac{2}{3},$$

then E and F cannot be mutually exclusive. *True or false?*

Solution: TRUE
We will argue by contradiction. Assume that E and F have given probabilities and that they are mutually exclusive. Then,

$$1 \geq \mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F] = \frac{4}{3}.$$

A contradiction!

Problem 1.2. (2 pts) If events A and B are mutually exclusive, they are necessarily independent. *True or false?*

Solution: FALSE
Let A and B be two events with strictly positive probabilities such that $A \cap B = \emptyset$. Then,

$$\mathbb{P}[A \cap B] = \mathbb{P}[\emptyset] = 0$$

while

$$\mathbb{P}[A]\mathbb{P}[B] > 0.$$

Problem 1.3. (2 points) If $Var[X] = 0$, then $\mathbb{P}[X = \mathbb{E}[X]] = 0$. *True or false?*

Solution: FALSE

Problem 1.4. (2 points)
Let X denote the outcome of a roll of a fair, regular dodecahedron (a polyhedron with 12 faces) with numbers $1, 2, \cdots, 12$ written on its sides. Then $\mathbb{E}[X] = 13/2$. *True or false?*

Solution: TRUE
Since each outcome is equally likely, by the definition of the expected value

$$\mathbb{E}[X] = \frac{1}{12} \cdot 1 + \frac{1}{12} \cdot 2 + \cdots + \frac{1}{12} \cdot 12 = \frac{1}{12}(1 + 2 + \cdots + 12) = \frac{1}{12} \cdot \frac{12 \cdot 13}{2} = \frac{13}{2}.$$

1.2. MULTIPLE CHOICE QUESTIONS.

Problem 1.5. (5 pts) A class contains 20 men and 10 women. You know that half the men and half the women have brown eyes. What is the probability that a person chosen at random from this class is a woman or has brown eyes?

(a) 1/3
(b) 2/3
(c) 7/18
(d) 7/9
(e) None of the above

Solution: (b)
Let

$$W := \{\text{the chosen person is a woman}\},$$
$$B := \{\text{the chosen person has brown eyes}\}.$$

Then, we need to calculate the probability

$$\mathbb{P}[W \cup B] = \mathbb{P}[W] + \mathbb{P}[B] - \mathbb{P}[W \cap B] = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2+3-1}{6} = \frac{4}{6} = \frac{2}{3}\,.$$

Problem 1.6. (5 pts) A pair of dice is thrown. Find the probability that the sum of the outcomes is 10 or greater if a 5 appears on the first die.

(a) 1/6
(b) 1/4
(c) 1/3
(d) 1/2
(e) None of the above

Solution: (c)
Let A_i denote the event that i was the outcome on the first die for $i = 1, 2, \dots 6$. Let E denote the event that the sum of the outcomes on both of the dies was greater than or equal to 10. Formally,

$$\begin{aligned}
E &= \{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j \geq 10\} \\
&= \{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j = 10\} \cup \{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j = 11\} \\
&\quad \cup \{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j = 12\}.
\end{aligned} \tag{1.1}$$

We want to find the probability $\mathbb{P}[E|A_5]$. Directly from the definition of conditional probability, we get

$$\mathbb{P}[E|A_5] = \frac{\mathbb{P}[E \cap A_5]}{\mathbb{P}[A_5]}.$$

From the representation in (1.1), we get that

$$\begin{aligned}
\mathbb{P}[E \cap A_5] &= \mathbb{P}[\{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j = 10\} \cap \{(i,j) : i = 5 \text{ and } 1 \leq j \leq 6\}] \\
&\quad + \mathbb{P}[\{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j = 11\} \cap \{(i,j) : i = 5 \text{ and } 1 \leq j \leq 6\}] \\
&\quad + \mathbb{P}[\{(i,j) : 1 \leq i,j \leq 6 \text{ and } i+j = 12\} \cap \{(i,j) : i = 5 \text{ and } 1 \leq j \leq 6\}] \\
&= \mathbb{P}[\{(i,j) : i = 5 \text{ and } i+j = 10\}] \\
&\quad + \mathbb{P}[\{(i,j) : i = 5 \text{ and } i+j = 11\}] \\
&\quad + \mathbb{P}[\{(i,j) : i = 5 \text{ and } i+j = 12\}] \\
&= \mathbb{P}[\{(5,5)\}] + \mathbb{P}[\{(5,6)\}] + \mathbb{P}[\emptyset] \\
&= \frac{1}{36} + \frac{1}{36} + 0 = \frac{1}{18}\,.
\end{aligned}$$

On the other hand, $\mathbb{P}[A_5] = \frac{1}{6}$, and so $\mathbb{P}[E|A_5] = \frac{1}{3}$.

Note: The above solution is very formal. You could have solved this problem correctly by straightforward counting of "good" outcomes quite fast and in many ways.

Problem 1.7. (5 points) What is the **R** output of the following command:

```
>dbinom(2,3,0.5)
```

(a) 0.375
(b) 0.5
(c) 0.725
(d) 0.75
(e) None of the above.

Solution: (a)
This is exactly the probability that a random variable $X \sim Binomial(size = 3, p = 0.5)$ takes the value 2. We have

$$\mathbb{P}[X = 2] = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) = \frac{3}{8} = 0.375.$$

Problem 1.8. (5 points) A test is used to determine whether people exhibiting green spots have the *duckpox* or not. It is believed that at any given time 4% of people exhibiting green spots actually have the *duckpox*. The test is 99% accurate if a person actually has the *duckpox*. The test is 96% accurate if a person does **not** have the *duckpox*. What is the probability that a randomly selected person who tests positive for the *duckpox* actually has the *duckpox*?

(a) About 4%.
(b) About 8%
(c) About 51%
(d) About 72%.
(e) None of the above.

Solution: (c)
The overall probability of obtaining a positive result on the *duckpox* test is

$$0.04(0.99) + 0.96(0.04) = 0.078.$$

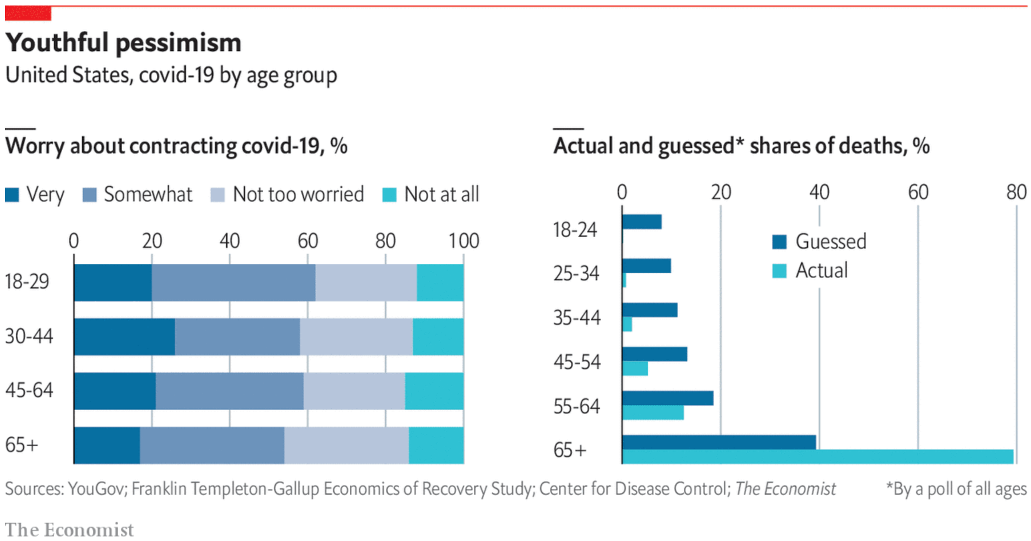
The probability of both having the *duckpox* and testing positive is

$$0.04(0.99) = 0.0396$$

So, the answer is

$$\frac{0.0396}{0.078} = 0.5076923.$$

Problem 1.9. (5 points) Consider the following charts:



Which of the following statements is **not** correct?

(a) More than three in five 18- to 29-year-old Americans are "very" or "somewhat" worried about contracting COVID-19.

(b) Young people do not fall ill with the virus as often as older people.

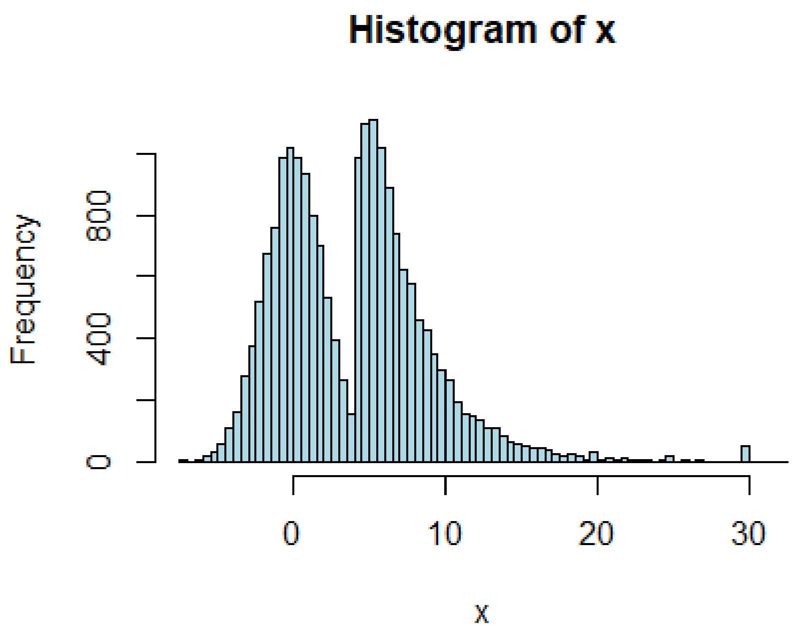
(c) The findings **suggest** that many people underestimate the age of a typical COVID-19 victim.

(d) Young people are **underestimating** the number of COVID-19 victims in their age group.

(e) People aged 65 and older account for more than a half of the COVID-19 victims.

Solution: (d)

Problem 1.10. (5 points) Consider the following histogram:



The histogram is ...

(a) ...unimodal.

(b) ...bimodal, symmetric.

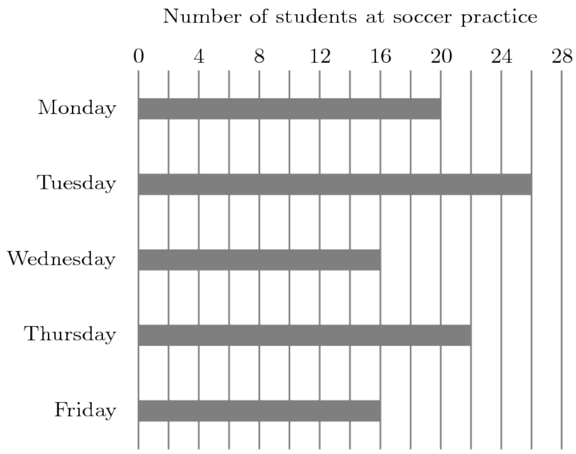
(c) ...bimodal, asymmetric.

(d) ...trimodal.

(e) None of the above.

Solution: (c)

Problem 1.11. (5 points) *Source: AMC8, 2019.* The diagram shows the number of students at soccer practice each weekday during last week. After computing the mean and median values, Coach discovers that there were actually 21 participants on Wednesday. Which of the following statements describes the change in the mean and median after the correction is made?



(a) The mean increases by 1 and the median does not change.

(b) The mean increases by 1 and the median increases by 1.

(c) The mean increases by 1 and the median increases by 5.

(d) The mean increases by 5 and the median increases by 1.

(e) The mean increases by 5 and the median increases by 5.

Solution: (b)
The sum of the number of students increased by 5 with the correction (from 16 to 21 on Wednesday. So the average increased by $5/5 = 1$. As for the median, the old data set was $16, 16, 20, 22, 26$ and the median was 20. Now, the data set is, in increasing order: $16, 20, 21, 22, 26$. So, the new median is 21; an increase of 1.

Problem 1.12. (5 points) Your sample consists of 50 sixth-graders from Kealing Middle School (KMS) and 60 sixth-graders from Murchison Middle school (MMS). The measures of center and spread of the students' heights are:

KMS : the mean of 60 inches with the standard deviation of 1 inch, and
MMS : the mean of 60 inches with the standard deviation of 2 inches.

What are the measures of center and spread for the pooled sample of 100 sixth-graders?

(a) The mean is 60; the standard deviation is 1.62.

(b) The mean is 60; the standard deviation is 1.5.

(c) The mean is 59.5; the standard deviation is 1.5.

(d) Not enough information is given.

(e) None of the above.

Solution: (a)
Let the KMS mean and standard deviation be denoted by $\bar{x}_K$ and SD_K and let the MMS mean and standard deviation be denoted by $\bar{x}_M$ and SD_M. Also, let the pooled mean and standard deviation be denoted by $\bar{x}$ and SD.

Then, the overall mean will be

$$\bar{x} = \frac{50\bar{x}_K + 60\bar{x}_M}{110} = 60.$$

As for the standard deviation, we have (with measurements from Kealing denoted by $\{x_i^K\}$ and measurements from Murchison denoted by $\{x_i^M\}$)

$$SD_K^2 = \frac{1}{50-1}\sum_{i=1}^{50}(x_i^K - \bar{x})^2;$$

$$SD_M^2 = \frac{1}{60-1}\sum_{i=1}^{60}(x_i^M - \bar{x})^2;$$

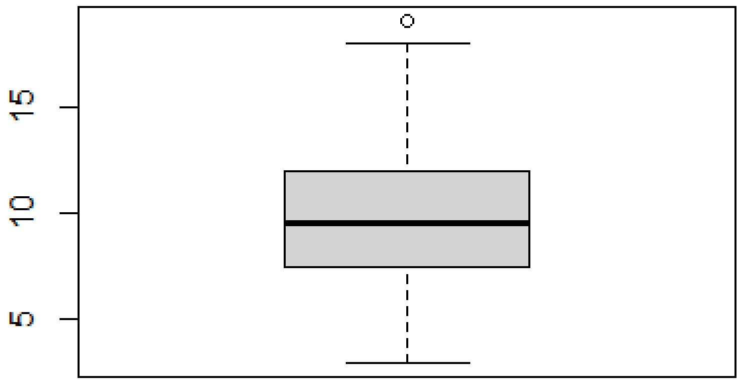
$$SD^2 = \frac{1}{50+60-1}\left(\sum_{i=1}^{50}(x_i^K - \bar{x})^2 + \sum_{i=1}^{60}(x_i^M - \bar{x})^2\right)$$

The key is to notice that $\bar{x}_K = \bar{x}_M = \bar{x} = 60$. Therefore, we have

$$\begin{aligned} SD^2 &= \frac{1}{50+60-1}\left(\sum_{i=1}^{50}(x_i^K - \bar{x})^2 + \sum_{i=1}^{60}(x_i^M - \bar{x})^2\right) \\ &= \frac{1}{109}\left(49(1)^2 + 59(2)^2\right) = \frac{285}{109}\,. \end{aligned}$$

So, the standard deviation is $\sqrt{\frac{285}{109}}$ which is about 1.616997.

Problem 1.13. (5 points) Consider the following box plot:



Which summary statistics does it correspond to?

(a)

Min.	Q1	Median	Mean	Q3	Max.
3.00	7.75	9.50	9.95	12.00	19.00

(b)

Min.	Q1	Median	Mean	Q3	Max.
3.00	7.75	9.50	9.95	12.00	17.00

(c)

Min.	Q1	Median	Mean	Q3	Max.
0.00	7.75	8.00	9.95	12.00	17.00

(d)

Min.	Q1	Median	Mean	Q3	Max.
3.00	7.75	9.50	9.95	10.00	17.00

(e) None of the above.

Solution: (a)

Problem 1.14. (5 points) Which of the following claims in **not** correct?

(a) Conclusions based on data obtained from **volunteer samples** can me misleading.

(b) It is important to be clear about the **target population** of a study.

(c) **Random choice** is the generally accepted method of drawing samples from a population.

(d) The **stratification method** only looks at one particular subset of the population of interest.

(e) Different random samples are **likely** to yield different values of the point estimates of the same parameter.

Solution: (d)

1.3. FREE-RESPONSE PROBLEMS.

Problem 1.15. Don't mess with Texas!

The *Anti-Littering League* wishes to gauge the success of the ingenious *Don't mess with Texas!* campaign.

Realizing the obvious problems with conducting a survey which outright asks the questions: "Are you or have you ever been a litterer?", they resort to the randomized-response method.

They prompt a computer to display the question

"Have you ever littered?"

with probability 0.6. 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.

i. (5 points)

It turned out that 50% of the subjects answered "yes". Give an estimate of the proportion of *litterers* in this population.

Solution: Now, we are given that $\mathbb{P}[D] = 0.50$ with the event D defined as

$$D = \{\text{the subject answered } \textbf{Yes}\}.$$

Our goal is to figure out $\mathbb{P}[D \mid C]$ with the conditioning event C given by

$$C = \{\text{the subject was asked the littering question}\}.$$

By the *Law of Total Probability*,

$$\mathbb{P}[D] = \mathbb{P}[D \cap C] + \mathbb{P}[D \cap C^c] = \mathbb{P}[D \mid C]\mathbb{P}[C] + \mathbb{P}[D \mid C^c]\mathbb{P}[C^c].$$

So,

$$0.6\,\mathbb{P}[D \mid C] = 0.50 - 0.5 \times 0.4 = 0.3 \quad \Rightarrow \quad \mathbb{P}[D \mid C] = 0.5.$$

ii. (5 points) What percentage of "yes" answers would you have obtained in an ideal world in which nobody ever litters?

Solution: Now, we are given that $\mathbb{P}[D \mid C] = 0$. So, using the same technique as above, i.e., the *Law of Total Probability*, we get

$$\begin{aligned}\mathbb{P}[D] &= \mathbb{P}[D \cap C] + \mathbb{P}[D \cap C^c]\\ &= \mathbb{P}[D \mid C]\mathbb{P}[C] + \mathbb{P}[D \mid C^c]\mathbb{P}[C^c]\\ &= 0 \times 0.6 + 0.5 \times 0.4 = 0.2.\end{aligned}$$