# Homework 5

PSTAT 5A: Spring 2023, with Ethan P. Marzban

**Instructions**

- Please submit your work to Gradescope by no later than **11:59pm on Wednesday, May 10**. As a reminder, late homework will not be accepted.
- Recall that you will be asked to upload a **single** PDF containing your work for *both* the programming and non-programming questions to Gradescope.
  - You can merge PDF files using either Adobe Acrobat, or using adobe's online PDF merger at this link.

## Problem 1: Browsing Habits

Suppose that the time a randomly-selected person spends on their phone in a day is found to be a random variable that follows the normal distribution with mean 195 minutes and standard deviation 50 minutes.

a. Define the random variable of interest, and call it $X$.

> **Solution:** Let $X$ denote the amount of time (in minutes) a randomly-selected person spends on their phone in a given day. From the problem statement, we have $X \sim \mathcal{N}(195,\ 50)$.

b. What is the probability that a randomly-selected person spends less than one hour per day on their phone?

> **Solution:** We seek $\mathbb{P}(X \leq 60)$ [note that 1 hour = 60 minutes!], which we compute using standardization:
>
> $$z = \frac{60 - 195}{50} = -2.7$$
>
> From the standard normal table, we see that if $Z \sim \mathcal{N}(0,\ 1)$ we have $\mathbb{P}(Z \leq -2.7) = 0.0035 = 0.35\%$

c. What proportion of the population spends greater than 173 minutes per day on their phone?

> **Solution:** We seek $\mathbb{P}(X \geq 173)$. By the complement rule, this is equivalent to $1 - \mathbb{P}(X < 173)$, which is a quantity we can compute by first standardizing and then consulting a normal table. The $z-$score associated with $x = 173$ is
>
> $$z = \frac{173 - 190}{50} = -0.34$$
>
> From a $z-$ table we see that, if $Z \sim \mathcal{N}(0,\ 1)$, $\mathbb{P}(Z \leq -0.34) = 0.3369$, meaning the

desired probability is $1 - 0.3669 = \boxed{0.6331 = 63.31\%}$.

d. What proportion of the population spends between 170 minutes and 200 minutes on their phone in a given day?

**Solution:** We seek $\mathbb{P}(170 \leq X \leq 200)$. We first break this up as

$$\mathbb{P}(Z \leq 200) - \mathbb{P}(X \leq 170)$$

Next, we compute the $z-$scores associated with $x = 200$ and $x = 170$, respectively:

$$z_1 = \frac{200 - 190}{50} = 0.2$$
$$z_2 = \frac{170 - 190}{50} = -0.4$$

From a standard normal table we have $\mathbb{P}(Z \leq 0.2) = 0.5793$ and $\mathbb{P}(Z \leq -0.4) = 0.3446$; hence, the desired probability is

$$0.5793 - 0.3446 = \boxed{0.2347 = 23.47\%}$$

## Problem 2: Trees!

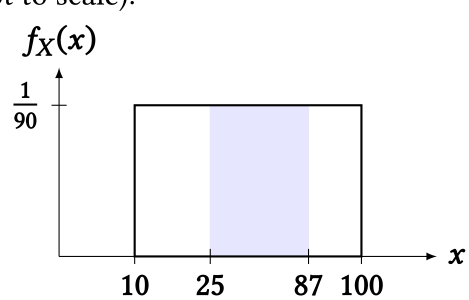
The heights of trees in a particular forest are found to vary uniformly between 10ft and 100ft. A park ranger is interested in the heights of randomly-selected trees.

a. Define the random variable of interest, and call it $X$.

**Solution:** Let $X$ denote the height of a randomly-selected tree from this forest. From the problem statment, we then have $X \sim \text{Unif}(10,\ 100)$.

b. What proportion of trees have heights between 25 and 87 feet?

**Solution:** We seek $\mathbb{P}(25 \leq X \leq 87)$. Let's sketch a picture of the area corresponding to this region (picture not to scale):

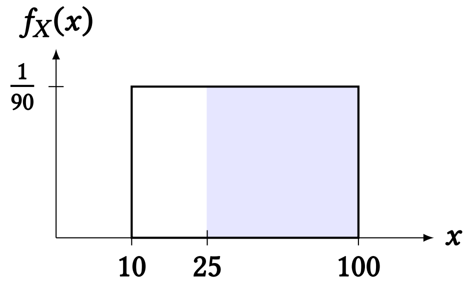


This is a rectangle with base $87 - 25 = 62$ and height $1/90$, meaning its area (and, consequently, the desired probability) is

$$(62) \cdot \frac{1}{90} = \frac{62}{90} = 68.\overline{8}\%$$

c. What proportion of trees have heights between 25 and 150 feet?

**Solution:** Now we seek $\mathbb{P}(25 \leq X \leq 150)$. Let's sketch a picture of the area corresponding to this region (picture not to scale): this time, we have to be a bit more careful. Specifically, notice that 150 is greater than 100 (which is our value of the parameter $b$); hence, our picture looks like

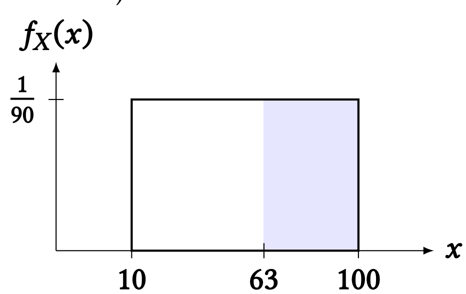



This is a rectangle with base $100 - 25 = 75$ and height $1/90$, meaning its area (and, consequently, the desired probability) is

$$(75) \cdot \frac{1}{90} = \frac{75}{90} = \frac{5}{6} = 83.\overline{3}\%$$

d. What proportion of trees have heights larger than 63 feet?

**Solution:** Now we seek $\mathbb{P}(X \geq 63)$. Let's sketch a picture of the area corresponding to this region (picture not to scale):




This is a rectangle with base $100 - 63 = 37$ and height $1/90$, meaning its area (and, consequently, the desired probability) is

$$(37) \cdot \frac{1}{90} = \frac{40}{90} = \frac{37}{90} = 41.\overline{1}\%$$

e. Suppose the ranger collects a sample of 50 different tree heights (assume the sample was taken *with* replacement). What is the probability that exactly 22 of these trees had heights larger than 63 feet? **Hint:** You may need to define another random variable, and, consequently, use *another* distribution!

**Solution:** Let $Y$ denote the number of trees in a sample of 50 (taken with replacement) that have heights greater than 60 feet. We suspect $Y$ follows a **binomial** distribution: to verify this, we check the binomial conditions:

1. Independent trials? Yes, because sampling was done with replacement.
2. Fixed number of trials? Yes, $n = 50$ trials.
3. Well-defined notion of success? Yes; 'success' = 'finding a tree with height greater than 60 feet'
4. Fixed probability of success? Yes; probability of success is the probability we computed in part (d) above; $p = 37/90$.

Therefore,

$$Y \sim \text{Bin}\left(50, \ \frac{37}{90}\right)$$

and so

$$\mathbb{P}(Y = 22) = \binom{50}{22}\left(\frac{37}{90}\right)^{22}\left(1 - \frac{37}{90}\right)^{50-22} \approx 10.38\%$$

## Problem 3: Parameter or Statistic?

In each of the parts below, determine whether the provided quantity is a population parameter or a sample statistic. Use this to further determine whether the quantity is a deterministic (i.e. non-random) constant, or a random variable.

a. The median score of 80 students, sampled from a class of 100.

**Solution:** Sample statistic; therefore, a random variable.

b. The maximum amount of time (in minutes) any human can hold their breath under water.

**Solution:** Population parameter; therefore, a deterministic constant.

c. The true IQR of incomes in Brazil.

**Solution:** Population parameter; therefore, a deterministic constant.

d. The standard deviation of the times it took 40 randomly-selected runners to complete a marathon.

**Solution:** Sample statistic; therefore, a random variable.

## Problem 4: Reducing Blood Pressure

A new drug is advertised to significantly reduce systolic blood pressure. To test these claims, a clinician takes a representative sample of 120 volunteers to whom she administers the drug. She records the difference in (systolic) blood pressure pre- and post- administration of the drug for each of the 120 volunteers, and finds that the volunteers had an average difference of $-8$ mm Hg (millimeters of mercury).

a. Identify the population of interest.

**Solution:** The population is: all people.

b. Identify the sample.

**Solution:** The sample is the 120 volunteers.

c. Is the mean difference of $-8$ mm HG a population parameter or an observed instance of a sample statistic?

**Solution:** The mean difference is an observed instance of a sample statistic, as it is tied to the specific sample of 120 volunteers.

## Problem 5: College Degrees

Suppose that 33% of a particular country's population has a college degree. A representative sample of 243 people is taken, and the proportion of these people who have a college degree is recorded.

a. Define the parameter, and use the notation discussed in Lecture 10.

**Solution:** Let $p$ = the proportion of the country's population that has a college degree.

b. Define the random variable of interest, and use the notation discussed in Lecture 10.

**Solution:** Let $\widehat{P}$ denote the proportion of people in a representative sample of 243 that have a college degree.

c. Check whether the success-failure conditions are satisfied.

**Solution:** In this case, we know the value of $p$: $p = 0.33$. Additionally, $n = 243$, so we check:

1. $np = (243)(0.33) = 80.19 \geq 10$
2. $n(1-p) = 162.81 \geq 10$

We see that both conditions are satisfied.

d. What is the probability that over 30% of the sample have college degrees?

**Solution:** We seek $\mathbb{P}(\widehat{P} \geq 0.3)$. By the Central Limit Theorem for Proportions (which we are able to invoke because the success-failure conditions are satisfied),

$$\widehat{P} \sim \mathcal{N}\left(0.33,\ \sqrt{\frac{(0.33)(1-0.33)}{243}}\right) \sim \mathcal{N}(0.33,\ 0.0302)$$

Therefore, we compute

$$\mathbb{P}(\widehat{P} \geq 0.3) = 1 - \mathbb{P}(\widehat{P} < 0.3) = 1 - \mathbb{P}\left(\frac{\widehat{P} - 0.33}{0.0302} < \frac{0.3 - 0.33}{0.0302}\right) = 1 - \mathbb{P}(Z \leq -0.99)$$

where $Z \sim \mathcal{N}(0,\ 1)$. From a normal table, we therefore see that the desired probability is

$$1 - 0.1611 = \boxed{0.8389 = 83.89\%}$$

e. What is the probability that the proportion of people in the sample who have college degrees lies within 5% of the true proportion of 33%.

**Solution:** We now seek $\mathbb{P}(0.28 \leq \widehat{P} \leq 0.38)$, which we compute as

$$\begin{aligned}
\mathbb{P}(0.28 \leq \widehat{P} \leq 0.38) &= \mathbb{P}(\widehat{P} \leq 0.38) - \mathbb{P}(\widehat{P} \leq 0.28) \\
&= \mathbb{P}\left(\frac{\widehat{P} - 0.3}{0.0302} \leq \frac{0.38 - 0.3}{0.0302}\right) - \mathbb{P}\left(\frac{\widehat{P} - 0.3}{0.0302} \leq \frac{0.28 - 0.3}{0.0302}\right) \\
&= \mathbb{P}\left(Z \leq \frac{0.05}{0.0302}\right) - \mathbb{P}\left(Z \leq -\frac{0.05}{0.0302}\right) \\
&= \mathbb{P}\left(Z \leq 1.66\right) - \mathbb{P}\left(Z \leq -1.66\right) \\
&= 0.9515 - 0.0485 = \boxed{0.903 = 90.3\%}
\end{aligned}$$

***There is no programming part on this homework.***