Name:______________________________ **Date:**_______________

PSTAT 5A: Midterm 2 Review Problems

Spring 2023, with *Ethan P. Marzban*

MORE PROBLEMS WILL BE POSTED TO THE COURSE WEBSITE SHORTLY.

1. The Intelligence Quotient (IQ) has historically been used as a measure of a person's reasoning ability. Studies have shown that IQs are normally distributed with mean 100 and standard deviation 15.

   (a) If a person has been selected at random, what is the probability that their IQ is lower than 90?

   Solution: Let X denote the IQ of a randomly-selected person; then $X \sim \mathcal{N}(100, 15)$ and

   $$\mathbb{P}(X \leq 90) = \mathbb{P}\left(\frac{X - 100}{15} \leq \frac{90 - 100}{15}\right) = \mathbb{P}\left(\frac{X - 100}{15} \leq -0.67\right) = 0.2514$$

   (b) According to the Wechsler Intelligence Scale, an IQ of between 120 and 129 is classified as "Superior". What proportion of the population have "Superior" IQs?

   Solution:

   $$\begin{aligned} \mathbb{P}(120 \leq X \leq 129) &= \mathbb{P}(X \leq 129) - \mathbb{P}(X \leq 120) \\ &= \mathbb{P}\left(\frac{X - 100}{15} \leq \frac{129 - 100}{15}\right) - \mathbb{P}\left(\frac{X - 100}{15} \leq \frac{120 - 100}{15}\right) \\ &= \mathbb{P}\left(\frac{X - 100}{15} \leq 1.93\right) - \mathbb{P}\left(\frac{X - 100}{15} \leq 1.33\right) \\ &= 0.9732 - 0.9082 = 0.065 = 6.5\% \end{aligned}$$

   (c) Mensa is an organization open only to those with extremely high IQs. The minimum IQ one must possess in order to be eligible for Mensa membership is 130- what proportion of the population qualify for Mensa memberships?

   Solution:

   $$\begin{aligned} \mathbb{P}(X \geq 130) &= 1 - \mathbb{P}(X < 130) = 1 - \mathbb{P}\left(\frac{X - 100}{15} < \frac{130 - 100}{15}\right) \\ &= 1 - \mathbb{P}\left(\frac{X - 100}{15} < 2.00\right) = 1 - 0.9772 = 0.0228 = 2.28\% \end{aligned}$$

Name:______________________________ **Date:**________________

(d) Caoimhe knows that 68% of the population have an IQ lower than hers. What IQ does Caoimhe have?

> **Solution:** Definitionally, Caoimhe's score is at the 68^{th} percentile of scores. Using Python, we can compute this to be 107.0155.

2. Around 1 out of every 3000 calico cats is born male. Suppose a sample of one thousand calico cats is taken (with replacement), and the number of male cats is recorded.

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

> **Solution:** Let X denote the number of male calico cats in a sample of 1000 calico cats.

(b) Show that X follows the Binomial Distribution, and identify its parameters.

> **Solution:**
>
> **1) Independent Trials?** Yes, because sampling was done with replacement.
>
> **2) Fixed Number of Trials?** Yes; $n = 1{,}000$ trials.
>
> **3) Well-defined notion of "success"?** Yes; "success" = "the calico cat in question is male"
>
> **4) Fixed probability of success?** Yes; $p = 1/3000$.
>
> Since all conditions are met, we conclude
>
> $$X \sim \text{Bin}\left(1000, \frac{1}{3000}\right)$$

(c) What is the probability that the sample contains at least 2 male cats? (You will need to use Python to compute this.)

> **Solution:** We can use either
>
> ```
> 1 - sum(sps.binom.pmf(np.arange(0, 2), 1000, 1/3000))
> ```
>
> or
>
> ```
> sum(sps.binom.pmf(np.arange(2, 1001), 1000, 1/3000))
> ```
>
> (there will be very minor differences due to computational over/underflow) to see that the probability is around 4.46%.

(d) What is the expected number of male cats in the sample?

Solution:

$$\mathbb{E}[X] = np = (1000) \cdot \left(\frac{1}{3000} \right) = \frac{1}{3}$$

(e) What is the standard deviation of the number of male cats in the sample?

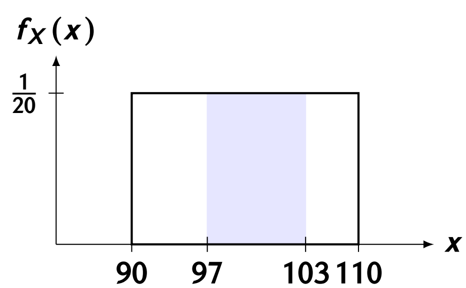
Solution:

$$\mathrm{SD}(X) = \sqrt{np(1-p)} = \sqrt{(1000) \cdot \left(\frac{1}{3000} \right) \cdot \left(\frac{2999}{3000} \right)} = \approx 0.58$$

3. Steel rods manufactured at a particular plant are meant to be 100 meters in length. Due to minor errors in the production cycle, rod lengths are actually uniformly distributed between 90 and 110 meters.

(a) What is the probability that a randomly selected rod is between 97 and 103 meters in length?

Solution: Let Y denote the length of a randomly-selected rod, so that $Y \sim \mathrm{Unif}(90,\ 110)$. We seek $\mathbb{P}(97 \leq Y \leq 103)$, which corresponds to the following area (remember to **always** sketch a picture, even if it's not perfectly to scale!)

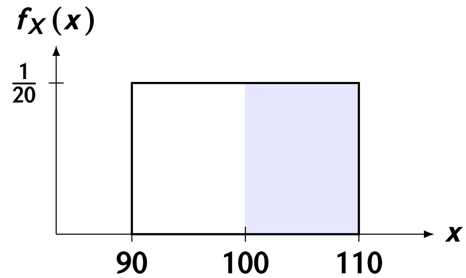


This is a rectangle with base $(103 - 97) = 6$ and height $1/20$, meaning its area (and, consequently, the desired probability) is

$$(103 - 97) \times \frac{1}{20} = \frac{6}{20} = \frac{3}{10} = 30\%$$

(b) What is the probability that a randomly selected rod is longer than 100 meters?

Solution: Now, we seek $\mathbb{P}(Y \geq 100)$, which corresponds to the following area:



$$(110 - 100) \times \frac{1}{20} = \frac{10}{20} = \frac{1}{2} = 50\%$$

(c) If a sample of 120 rods is to be taken (with replacement) and the number of rods over 100 meters in length is recorded, what is the probability that this sample contains exactly 60 rods that are longer than 100 meters?

Solution: Let X denote the number of rods, in a sample of 120 (taken with replacement), that are longer than 100 meters. We have a suspicion that X is Binomially distributed; to verify this, we check the Binomial Criteria:

1) **Independent trials?** Yes, since sampling is done with replacement.
2) **Fixed number of trials?** Yes; $n = 120$ trials.
3) **Well-defined notion of "success"?** Yes; "success" = "selected rod is longer than 60 meters"
4) **Fixed probability of success?** Yes; $p = 0.5$, as found in the previous part.

Hence, we see that $X \sim \text{Bin}(120,\ 0.5)$ and so

$$\mathbb{P}(X = 60) = \binom{120}{60}(0.5)^{60}(0.5)^{60} = 0.0727 = 7.27\%$$

Name:_______________________________ **Date:**____________________

4. A shopowner believes that 22% of his customers prefer to pay with cash. To test this belief, he takes a representative sample of 150 customers and finds that 20 of these sampled customers prefer to pay with cash. **Use a two-sided alternative, and an $\alpha = 0.05$ level of significance**.

   (a) Write down the null and alternative hypotheses for this test. Be sure to use mathematical notation, and define any variables you use.

   > **Solution:** Let p denote the true proportion of customers that prefer to pay with cash; then $H_0 : \ p = 0.22$. Because we are using a two-sided alternative hypothesis, we have $H_A : p \neq 0.22$; i.e.
   >
   > $$\left[\begin{array}{ll} H_0 : & p = 0.22 \\ H_A : & p \neq 0.22 \end{array} \right.$$

   (b) Compute the value of the test statistic.

   > **Solution:**
   >
   > $$\mathrm{TS} = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\left(\frac{20}{150}\right) - 0.22}{\sqrt{\frac{0.22\cdot(1-0.22)}{150}}} = \boxed{-2.56}$$

   (c) What is the distribution of the test statistic, assuming the null hypothesis is correct? Be sure to check any relevant conditions!

   > **Solution:** We would like to invoke the Central Limit Theorem for Proportions; in order to do so, we must first check:
   >
   > **1)** $np_0 = (150)(0.22) = 33 \geq 10$ ✓
   >
   > **2)** $n(1 - p_0) = (150) \cdot (1 - 0.22) = 117 \geq 10$ ✓
   >
   > Since both conditions are met, we use the CLT for proportions to conclude
   >
   > $$\mathrm{TS} \stackrel{H_0}{\sim} \mathcal{N}(0, \ 1)$$

   (d) Find the critical value of the test.

   > **Solution:** Since we are using an $\alpha = 0.05$ level of significance, we use a critical value of $\boxed{1.96}$.

   (e) Now, conduct the hypothesis test and state your conclusions in the context of the problem.

Solution: In this case, $|\text{TS}| = |-2.56| > 1.96$ meaning we reject the test; that is,

At an $\alpha = 0.05$ level of significance, there was sufficient evidence to reject the null that 22% of people prefer to pay with cash, in favor of the two-sided alternative that the true proportion was *not* 22%.

5. Mary would like to determine the true average speed with which cars drive on a particular stretch of Highway 101. As such, she takes a representative sample of 40 cars; the average speed of these cars is 65mph and the standard deviation of the speeds of these 40 cars is 10mph.

   (a) Define the random variable of interest, $\overline{X}$

   Solution: Let $\overline{X}$ denote the average speed of a sample of 40 cars, taken from a strecth of Highway 101.

   (b) What distribution would we use to construct a 95% confidence interval for the true average speed of cars along this stretch of Highway 101? Be sure to include any/all relevant parameter(s), and check any/all conditions!

   Solution:

   - Is the population normally distributed? No.
   - Is our sample size large enough? Yes; $n = 40 \geq 30$.
   - Do we have access to σ (the true population standard deviation) or s, the sample standard deviation? We have access to s, not σ.

   As such, we will use the $t-$distribution; specifically, with $n - 1 = 40 - 1 = 39$ degrees of freedom: t_{39}

   (c) Construct a 95% confidence interval for the true average speed of cars along this stretch of Highway 101, and interpret your interval.

   Solution: Our CI will be of the form

   $$\overline{x} \pm t_{39,\ \alpha} \cdot \frac{s}{\sqrt{40}}$$

   From the $t-$table, we take $t_{39,\ \alpha}$ to be 2.02 meaning our confidence interval is

   $$65 \pm (2.02) \cdot \frac{10}{\sqrt{40}} = 65 \pm (3.1939) = [61.8061\ ,\ 68.1939]$$

Name:____________________________________ **Date:**__________________

6. Consider a random variable X with the following p.m.f. (probability mass function):

k	-0.1	0.1	0.2	1
$\mathbb{P}(X = k)$	0.1	0.1	a	0.1

(a) Find the value of a.

Solution: We know that the probabilities in a p.m.f. must sum to 1; as such,

$$0.1 + 0.1 + a + 0.1 = 1 \implies a = 0.7$$

(b) Compute $\mathbb{P}(\{X = 0\} \cup \{X = 1\})$.

Solution: By the Addition Rule,

$$\mathbb{P}(\{X = 0\} \cup \{X = 1\}) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) - \mathbb{P}(\{X = 0\} \cap \{X = 1\})$$

It is impossible for X to simultaneously be 0 and 1, meaning $\mathbb{P}(\{X = 0\} \cap \{X = 1\}) = 0$; additionally, since 0 is not in the state space of X we see that $\mathbb{P}(X = 0) = 0$, meaning

$$\mathbb{P}(\{X = 0\} \cup \{X = 1\}) = \mathbb{P}(X = 1) = 0.1$$

(c) Compute $\mathbb{P}(X \leq 0.5)$.

Solution: Using direct computations, we would have

$$\mathbb{P}(X \leq 0.5) = \mathbb{P}(X = -0.1) + \mathbb{P}(X = 0.1) + \mathbb{P}(X = 0.2) = 0.1 + 0.1 + 0.7 = 0.9$$

We could have equivalently used the Complement Rule, to conclude

$$\mathbb{P}(X \leq 0.5) = 1 - \mathbb{P}(X > 0.5) = 1 - [\mathbb{P}(X = 1)] = 1 - (0.1) = 0.9$$

(d) Find $\mathbb{E}[X]$.

Solution:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{\text{all } k} k \cdot \mathbb{P}(X = k) \\
&= (-0.1) \cdot \mathbb{P}(X = 0.1) + (0.1) \cdot \mathbb{P}(X = 0.1) + (0.2) \cdot \mathbb{P}(X = 0.2) + (1) \cdot \mathbb{P}(X = 1) \\
&= (-0.1) \cdot (0.1) + (0.1) \cdot (0.1) + (0.2) \cdot (0.7) + (1) \cdot (0.1) = 0.24
\end{aligned}$$

Name:____ **Date:**____________________

(e) Find $\mathrm{SD}(X)$.

Solution: We need to first find the variance of X. Recall that we can find this using either of the two variance formulas we have encountered: using the second formula, we have

$$\sum_{\text{all } k} k^2 \cdot \mathbb{P}(X = k) = (-0.1)^2 \cdot \mathbb{P}(X = 0.1) + (0.1)^2 \cdot \mathbb{P}(X = 0.1) + (0.2)^2 \cdot \mathbb{P}(X = 0.2) + (1)^2 \cdot \mathbb{P}(X = 1)$$

$$= (-0.1)^2 \cdot (0.1) + (0.1)^2 \cdot (0.1) + (0.2)^2 \cdot (0.7) + (1)^2 \cdot (0.1) = 0.13$$

$$\mathrm{Var}(X) = \sum_{\text{all } k} k^2 \cdot \mathbb{P}(X = k) - (\mathbb{E}[X])^2 = (0.13) - (0.24)^2 = 0.0724$$

Using the first formula, we would have computed

$$\mathrm{Var}(X) = \sum_{\text{all } k} (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k)$$

$$= (-0.1 - 0.24)^2 \cdot \mathbb{P}(X = 0.1) + (0.1 - 0.24)^2 \cdot \mathbb{P}(X = 0.1) + (0.2 - 0.24)^2 \cdot \mathbb{P}(X = 0.2) + (1 - 0.24)^2 \cdot \mathbb{P}(X = 1)$$

$$= (-0.1 - 0.24)^2 \cdot (0.1) + (0.1 - 0.24)^2 \cdot (0.1) + (0.2 - 0.24)^2 \cdot (0.7) + (1 - 0.24)^2 \cdot (0.1)$$

$$= 0.0724$$

Either way, we have $\mathrm{Var}(X) = 0.0724$ meaning

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{0.0724} \approx 0.2691$$

