STA101 Problem Set 4 - KEY

Summer I, 2021, Duke University

Exercises from the OpenIntro book - 100pts total

Problems include: Chapter 4 exercises 4.4, 4.20, and 4.24 and 3 additional problems

4.4 (22pts total)

(a) **(4pts)** Let X denote the finishing times of *Men, Ages 30-34* and Y denote the finishing times of *Women, Ages 25-29*. Then

$$X \sim N(\mu = 4313, \sigma = 583)$$
$$Y \sim N(\mu = 5261, \sigma = 807)$$

(b) **(2pts for z-scores, 2pts for explanation)**The Z scores can be calculated as follows:

$$Z_{Leo} = \frac{x - \mu}{\sigma} = \frac{4948 - 4313}{583} = 1.09$$
$$Z_{Mary} = \frac{y - \mu}{\sigma} = \frac{5513 - 5261}{807} = 1.09$$

Leo finished 1.09 standard deviations above the mean of his group's finishing time and Mary finished 0.31 standard deviations above the mean of her group's finishing time.

(c) **(2pts)** Mary ranked better since she she has a lower Z score indicating that her finishing time is relatively shorter.

(d) **(4pts)**Leo:

$$\begin{aligned}\mathbb{P}(Z > 1.09) &= 1 - \mathbb{P}(Z \leq 1.09)\\ &= 1 - 0.8621\\ &= 0.1379\end{aligned}$$

(e) **(4pts)** Mary:

$$\begin{aligned}\mathbb{P}(Z > 0.31) &= 1 - \mathbb{P}(Z \leq 0.31)\\ &= 1 - 0.6217\\ &= 0.3783\end{aligned}$$

(f) **(2pts for unchanged z-scores, 2pts for noting that we could not do parts c-e)** Answer to part (b) would not change as Z scores can be calculated for distributions that are not normal. However, we could not answer parts (c)-(e) since we cannot use the Z table to calculate probabilities and percentiles without a normal model.

4.20 (16pts total)

(a) **(2pts for distribution, 2pts for calculations)** Since we are asked for the expected number of successes in a given number of trials, we use the binomial distribution with $n = 120$ and $p = 0.9$:

$$\mu = np = (120)(0.9) = 108$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{120(0.9)(1-0.9)} = 3.29$$

(b) **(4pts for "not unusual")** Since 105 is less than 2 standard deviations away from the mean we wouldn't consider this to be an unusual observation or be surprised.

(c) **(4pts for checking conditions, 4pts for probability. Continuity correction not needed for full marks.)**

$$\mathbb{P}(X \leq 105) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \cdots \mathbb{P}(X = 105).$$

Since this is a bit tedious to solve using binomial distribution, we can instead use the normal approximation to binomial to estimate this probability. But first we must verify that np and $n(1-p)$ are at least 10.

$$np = (120)(0.9) = 108 > 10 \quad \text{and} \quad n(1-p) = (120)(0.10) = 12 > 10$$

Since the conditions are met, we can use the normal model $N(\mu = 108, \sigma = 3.29)$.

$$\begin{aligned} \mathbb{P}(X < 105) &= \mathbb{P}\left(Z < \frac{105 - 108}{3.29}\right) \\ &= \mathbb{P}(Z < -0.91) \\ &= 0.1814 \end{aligned}$$

In part (b) we had determined that it would not necessarily be considered unusual to observe 105 American adults who have had chickenpox among a random sample of 120. Here we calculated a somewhat high probability for this event, so the results from parts (b) and (c) agree.

If we were to apply a 0.5 correction, the calculations would change very slightly, still yielding a high probability.

$$\begin{aligned} \mathbb{P}(X < 105 + 0.5) &= \mathbb{P}(X < 105.5) \\ &= \mathbb{P}\left(Z < \frac{105.5 - 108}{3.29}\right) \\ &= \mathbb{P}(Z < -0.76) \\ &= 0.2236 \end{aligned}$$

4.24 (20pts total)

(a) **(4pts)** Using the binomial distribution with $n = 3$ and $p = 0.25$,

$$\mathbb{P}(X = 2) = \binom{3}{2}(0.25)^2(0.75)^1 = 3(0.25)^2(0.75) = 0.1406$$

(b) **(4pts)** Using the binomial distribution with $n = 3$ and $p = 0.25$,

$$\mathbb{P}(X = 0) = \binom{3}{0}(0.25)^0(0.75)^3 = 0.4219.$$

(c) **(4pts)** Using the binomial distribution with $n = 3$ and $p = 0.25$,

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - 0.4219 = 0.5781$$

(d) **(4pts)** Let X be the trial at which the first success (disease) occurs. Then, using a geometric distribution with $p = 0.25$,

$$\mathbb{P}(X = 3) = (0.75)^2(0.25) = 0.1406$$

Problem 4 (12pts total)

(a) **(2pts for distribution, 2pts for probability)** This is a geometric distribution with the probability of having blonde hair given by $p = 0.7$, and the probability not having blonde hair $1 - p = 0.3$. Thus,

$$\mathbb{P}(1^{\text{st}} \text{ blonde child is } 3^{\text{rd}} \text{ child}) = (1-p)^2 p = (0.3)^2(0.7) = 0.063.$$

(b) **(2pts for distribution, 3pts for mean, 3pts for standard deviation)** This is a geometric distribution with the probability of having brown hair given by $p = 0.2$, and the probability of not having brown hair $1 - p = 0.8$. We can then use the formulas for the mean and standard deviation for a geometric random variable:

$$\mu = \frac{1}{p} = \frac{1}{0.2} = 5$$

$$\sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{0.8}{0.2^2}} \approx 4.47$$

The parents would expect to have 5 children before they have a child with brown hair with a standard deviation of 4.47.

Problem 5 (16pts total)

(a) **(4pts for "Poisson")** We could use a Poisson distribution since we are interested in the count of the (independent) buses that appear at this bus stop over a short period of time.

(b) **(2pts for mean, 2pts for standard deviation)** Assume we model the number of buses that appear at the bus stop as a Poisson distribution for the remaining parts of the problem. The rate parameter is given by $\lambda = 25$, so the mean is $\mu = \lambda = 25$, and the standard deviation is given by $\sigma = \sqrt{\lambda} = 5$.

(c) **(4pts for "unusually high")** If 35 buses appeared at this bus stop on a particular day, this event would be 2 standard deviations away from the mean, so it would be unusually high.

(d) **(4pts)**

$$\mathbb{P}(\text{20 buses arrive within an hour}) = \frac{e^{-25}25^{20}}{20!} \approx 0.0519$$

Problem 6 (14pts total)

(7pts for stating the conditional probability, 4pts for calculation, 3pts for final answer. Z-scores not needed for full credit.) Let X denote the students' GRE scores. Then $X \sim N(\mu = 150, \sigma = 8.5)$. An awardee satisfies $X > 160$, and we are interested in the probability that $X > 165$ given we choose an awardee.

$$\begin{aligned}
\mathbb{P}(X > 165 \mid X > 160) &= \frac{\mathbb{P}(X > 165 \text{ and } X > 160)}{\mathbb{P}(X > 160)} \\
&= \frac{\mathbb{P}(X > 165)}{\mathbb{P}(X > 160)} \\
&= \frac{1 - \mathbb{P}(X \leq 165)}{1 - \mathbb{P}(X \leq 160)}.
\end{aligned}$$

We can compute $\mathbb{P}(X \leq 165)$ using the `R` command `pnorm(165, mean = 150, sd = 8.5)`. Similarly, we can compute $\mathbb{P}(X \leq 160)$ using the `R` command `pnorm(160, mean = 150, sd = 8.5)`. We get about 0.961 and 0.88, respectively. Substituting this back into the above equation gives us

$$\mathbb{P}(X > 165 \mid X > 160) = \frac{1 - 0.961}{1 - 0.88} = 0.325.$$