Name: Ivan Leung **Student ID:** 67485846 **Section:** ___

Instructions

Please read the following instructions carefully:

1. Please show all notation for probability statements.
2. Box your final answers.
3. Please verify that your scans are legible.
4. Please assign pages the the questions when submitting to gradescope.
5. This assignment is due via gradescope on the due date.

1. A baby's weight at birth is strongly associated with mortality risk during the first year and, to a lesser degree, with developmental problems in childhood and the risk of various diseases in adulthood. The weight of a newborn baby is assumed to be normally distributed with a mean of 3.39kg, and a standard deviation of 0.55kg.

$X \sim Normal(\bar{x} = 3.39, \theta = 0.55)$

(a) Could we calculate the probability that a randomly selected a newborn's weight is less than 3.58kg? If so, find this probability. If not, explain why this would not be possible.

$P(X < 3.58) = pnorm(3.58, 3.39, 0.55)$

$= \boxed{0.6351}$

(b) A data analyst obtains the weights of 30 randomly selected newborns. What is the probability that the mean weight of randomly selected newborns in this sample is less than 3.2kg?

$\bar{X} \sim Normal(\mu_{\bar{x}} = 3.39, \theta_{\bar{x}} = \frac{\theta}{\sqrt{n}} = \frac{0.55}{\sqrt{30}} = 0.1004)$

$P(\bar{X} < 3.2) = pnorm(3.2, 3.39, 0.1004)$

$= \boxed{0.0292}$

(c) Another data analyst obtains the weight of 100 randomly selected newborns. Will there be a lower or higher probability that the mean weight of the 100 newborns in this sample is less than 3.2kg, compared to the sample in part (b)? Explain.

As the sample size increases, the standard deviation will get smaller, which means the probability of the sample mean is closer to the population mean. Therefore, there will be lower probability that the sample mean weight is less than 3.2 kg.

(d) Is the probability that the mean weight of newborns in the sample in (b) is between 3.2kg and 3.58kg greater than the probability that the mean weight of newborns in the sample in (c) is between 3.2kg and 3.58kg? Explain. $P(3.2 < X < 3.58)$

$P(3.2 < \bar{X}_{30} < 3.58) = pnorm(3.58, 3.39, \frac{0.55}{\sqrt{30}}) - pnorm(3.2, 3.39, \frac{0.55}{\sqrt{30}}) = 0.9415$

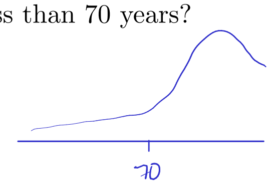
$P(3.2 < \bar{X}_{100} < 3.58) = pnorm(3.58, 3.39, \frac{0.55}{\sqrt{100}}) - pnorm(3.2, 3.39, \frac{0.55}{\sqrt{100}}) = 0.9994$

No, with smaller sample size in (b), (b) has larger standard deviation and (b) has its bell shaped more spreaded out which means has lower probability than (c)

2. Unlike the newborn's weight from question 1, it is assumed that the distribution of the age of death from natural causes (heart disease, cancer, etc.) is left-skewed. Most such deaths happen at older ages, with fewer cases happening at younger ages. Let's suppose that the mean age of death from natural causes is 70 years with a standard deviation of 6 years.

   (a) Since the distribution of the age of death from natural causes is left-skewed, would the majority of people's age of death from natural causes that is greater than or less than 70 years?

   It is greater than 70 years.



   (b) Suppose that we randomly select a people who died of natural causes. Could we calculate the probability that this individual's age is greater than 75 years? If so, find this probability. If not, explain why this would not be possible.

   No, since approximately normal distribution requires a sample size of 30 or greater, we cannot find the probability with only one sample.

   (c) Suppose 120 people who died of natural causes are randomly selected and the average of their death age is calculated. What is the standard deviation of the mean death age in the sample of 120 people?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{120}} = \boxed{0.5477}$$

(d) Why is the standard deviation of the average death age in the sample of 120 people who died of natural causes computed in part c (~~d~~) much lower than the population standard deviation of 6 years?

Since the sample size is much smaller than the population size, it makes sense that the sample standard deviation is much smaller than the population standard deviation.

(e) Suppose 120 people who died of natural causes are randomly selected and the mean death age in the sample is computed. Based on the Central Limit Theorem, what is the approximate probability that the average death age in the sample of 120 is greater than 71 years?

$$x \sim AN\left(\mu_{\bar{x}} = 70,\ \sigma_{\bar{x}} = \frac{6}{\sqrt{120}}\right)$$

$$P(x > 71) = 1 - P(x < 71)$$

$$= 1 - \text{pnorm}\left(71, 70, \frac{6}{\sqrt{120}}\right)$$

$$= \boxed{0.0339}$$

3. A fashion retailer conducted a nationwide study to understand the favorite clothing choices of American consumers. The results revealed that among the general population, 53% prefer wearing casual attire, 32% prefer dressing formally, and 15% favor sportswear.

   Now, the retailer intends to investigate a random sample of 70 American consumers' fashion preferences.

   (a) What is the sampling distribution for the sample proportion of individuals who prefer wearing **casual** attire?

   $$\hat{p} \sim AN\left(\mu_{\hat{p}} = p = 0.53,\ \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.53(0.47)}{70}} = 0.0597\right)$$

   (b) What is the sampling distribution for the sample proportion of individuals who prefer wearing **sportswear**?

   $$\hat{p} \sim AN\left(\mu_{\hat{p}} = p = 0.15,\ \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.15(0.85)}{70}} = 0.0427\right)$$

   (c) What is the probability that the sample proportion of individuals who prefer wearing **sportswear** is lower than 20%?

   $$P(\text{sportswer} < 0.2) = \text{pnorm}(0.2, 0.15, 0.0427)$$
   $$= \boxed{0.8793}$$

   (d) What is the probability that the sample proportion of individuals who prefer wearing **casual** attire is higher than 60%?

   $$P(\text{casual} > 0.6) = 1 - P(\text{casual} < 0.6)$$
   $$= 1 - \text{pnorm}\left(0.6, 0.53, \sqrt{\frac{0.53(0.47)}{70}}\right)$$
   $$= \boxed{0.1203}$$

(e) What is the probability that sample proportion of individuals who prefer wearing **formal** attire is between 0.3 and 0.4? $\hat{p} \sim AN(\mu_{\hat{p}} = p = 0.32, \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.32(0.68)}{70}})$

$$P(0.3 < \text{formal} < 0.4) = P(\text{formal} < 0.4) - P(\text{formal} < 0.3)$$
$$= \text{pnorm}(0.4, 0.32, \sqrt{\frac{0.32(0.68)}{70}}) - \text{pnorm}(0.3, 0.32, \sqrt{\frac{0.32(0.68)}{70}})$$
$$= \boxed{0.5644}$$

(f) Comparing the sample proportion of those who prefer wearing casual attire and those who prefer wearing sportswear, which one has a lower standard deviation? Why do you think this is the case?

casual	sportswear
$\sigma_{\hat{p}}$	$\sigma_{\hat{p}}$
0.0597	0.0427

Sportswear has a lower standard deviation because Sportswear has a smaller sample proportion compare to casual attire

4. According to the Centers for Disease Control and Prevention (CDC), 24% of Americans met the Physical Activity Guidelines for aerobic and muscle-strengthening activities in 2020. Suppose we take a random sample of 110 Americans, and we assume that the 'true' proportion of Americans meeting the Physical Activity Guidelines is the same as the CDC estimation.

   (a) Based on our assumption, what is the exact distribution of the **number** of Americans in the sample who meet the physical activity guidelines?

   $n = 110$

   $p = 0.24$

   $X \sim \text{Binomial}(110, 0.24)$

   (b) Based on our assumption, what is the exact probability that the number of surveyed Americans who meet the physical activity guidelines is less or equal to 20?

   $P(X \leq 20) = \text{pbinom}(20, 110, 0.24)$

   $= \boxed{0.09109977}$

   (c) Based on our assumption, what is the exact probability that the number of surveyed Americans who meet the physical activity guidelines is strictly less than 20 (does not include 20)?

   $P(X < 20) = P(X \leq 19)$

   $= \text{pbinom}(19, 110, 0.24)$

   $= \boxed{0.05796336}$

(d) Under our assumption, use the Normal approximation to binomial, what is the sampling distribution of the **proportion** of surveyed Americans who meet the physical activity guidelines?

$\hat{p} \sim AN(\mu_{\hat{p}} = p = 0.24, \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.24)(0.76)}{110}} = 0.0407)$

(e) Using the sampling distribution you obtained in part (d), what is the approximate probability that the number of Americans who meet the physical activity guidelines is at most 20?

$P(Y \leq \frac{20}{110}) = \text{pnorm}(\frac{20}{110}, 0.24, \sqrt{\frac{(0.24)(0.76)}{110}})$

$= \boxed{0.0765}$

(f) Using the sampling distribution you obtained in part (d), what is the approximate probability that the number of Americans who meet the physical activity guidelines is strictly less than 20?

$P(Y < \frac{20}{110}) = \boxed{0.0765}$

(g) Compare the exact probabilities from part (b) and (c) to the approximate probabilities from (e) and (f), do you think the Normal approximating is a good approximation?

If we add the probabilities from (b) and (c) then divided by 2, we get 0.0745 which is very close to the probabilities in (e) and (f). Therefore, the normal approximating is a good approximation.

5. For the multiple choice questions, circle the best answer. You can assume that all random samples mentioned are drawn independently from the same population.

   (a) What is true about the standard deviation of the sample mean $\bar{X}$?
      A. It increases along with the sample size.
      B. As the sample size n increases, standard deviation decreases.
      C. It will always be greater than the population standard deviation.
      D. The sample size n increasing or decreasing has no effect on the standard deviation.

   (b) Heights for men are normally distributed with a mean of 70 inches and standard deviation of 3 inches. If a random sample of 27 men is taken, what is the mean of the sampling distribution of $\bar{X}$?
      A. It depends on the value of the sample mean.
      B. $3/\sqrt{27} = 0.5774$ inches
      C. 3 inches
      D. 70 inches

   (c) Is the following statement True or False: The sampling distribution becomes approximately normal, for all statistics, if there is a large enough sample size.
      A. False
      B. True

   (d) If we took a small sample from a normally distributed population, the distribution of the sample mean would be
      A. not normally distributed.
      B. approximately a standard normal distribution.
      C. a skewed distribution.
      D. approximately a normal distribution.

   (e) Using the approximation based on the normal distribution, the standard deviation of the sampling distribution for a sample proportion, depends on
      A. true population proportion only.
      B. the sample size and standard deviation.
      C. the sample size and the true population proportion.
      D. the true population proportion and population standard deviation.

   (f) The Rule for Sample Means would **not** be applicable for which scenario?
      A. We draw a random sample of 20 from a normally distributed population.
      B. We draw a random sample of 20 from a skewed population.
      C. We draw a random sample of 60 from a normally distributed population.
      D. We draw a random sample of 60 from a skewed population.