HW 08 Written Work

Graded

Student

Sangwon Ji

Total Points

36 / 36 pts

Question 1

Q1.1

6 / 6 pts

✔ **+ 3 pts** Correct answer

✔ **+ 3 pts** Valid explanation: Since we cannot make assumptions about the distribution of ages, we must use Chebyshev to bound the variability. The range given is +/- 4 standard deviations, which gives us 93.75%.

+ 0 pts Incorrect/blank

Question 2

Q1.2

6 / 6 pts

✔ **+ 3 pts** Correct answer: 6.25

✔ **+ 3 pts** Valid explanation: If *at least* 93.75% fall within the range +/- 4 SDs, then *at most* 6.25% will fall outside that range

+ 0 pts Incorrect/blank

Question 3

Q1.3

6 / 6 pts

✔ **+ 3 pts** Correct answer: 6.25

✔ **+ 3 pts** Valid explanation: In the most extreme case, all of the people not inside the range [25, 65] would be older than 65 years old. Since there is at most 6.25% of people outside that range, the maximum percentage of people older than 65 must be 6.25%

+ 0 pts Incorrect/blank

Question 4

Q2.2

6 / 6 pts

✔ **+ 3 pts** Correct answer: 2

✔ **+ 3 pts** Valid explanation: In terms of standard deviations, the width of a 90% confidence interval is smaller than a 95% confidence interval. However, if we want both intervals to be the same width (0.06), we would have to take a larger sample size for our 95% CI. Thus, with a 90% CI, we could take a smaller sample size than what we needed for our 95% CI.

+ 0 pts Incorrect/blank

Question 5

Q2.4 **6** / 6 pts

✔ **+ 6 pts** If we take +/- 1.65 SDs from the mean, we will get the “middle 90%” of the data. We need to look at the “middle 90%” of data in order to construct a 90% CI

+ 0 pts Incorrect/blank

Question 6

Q3.2 **Resolved** **6** / 6 pts

✔ **+ 3 pts** Correctly identifies that the proportion of yes votes can be interpreted as a sample mean. We are thus generating a collection of resample means (from samples taken with replacement), so the CLT applies.

✔ **+ 3 pts** Makes the connection that the CLT implies that the resampled means will have an approximately normal distribution, so that’s why the histogram above is bell-shaped.

+ 0 pts Incorrect/blank

Regrade Request **Submitted on: Aug 08**

Hi,
I mentioned that proportion of yes votes is close to the normal, which is similar to the sample mean I guess? I think it also says that CLT is applied in this case.. Thank you!

resolved

Reviewed on: Aug 08

Question 1.1. At least _____% of the people are between 25 and 65 years old. **(6 Points)**

At least 93.75% of the people are between 25 and 65 years old. Using the Chebychev's Bounds, since we don't know the distribution, not an exact answer, but we can get a bound. In a standard deviation of 5 years, 65-45, 20, dividing that by 5, we have average += 4 Sds. In a formula 1 - (1/z **2), 4 for z, we get 0.9375, 93.75 percent.

Question 1.2. At most _____% of the people have ages that are not in the range 25 years to 65 years. **(6 Points)**

At most 6.25% of the people have ages that are not in the range 25 years to 65 years. This is because we know that the bound of pecrcentage of people in range of 25 years to 65 years old is 93.75. So, from the tota, 100 percent, we subtaract the number, and we will be able to get the population outside of the range. That is 100 - 93.75 = 6.25.

Question 1.3. At most _____% of the people are more than 65 years old. **(6 Points)**

Hint: If you're stuck, try thinking about what the distribution may look like in this case.

At most 6.25% of the people are more than 65 years old. we know that the range that is not in the 25 to 65 range are 6.25%. In this case, talking about there is no value located under 25, there is a chance that it is at most 6.25%, which is all the population outside of 25 to 65 range.

Question 2.2. Suppose the data science class decides to construct a 90% confidence interval instead of a 95% confidence interval, but they still require that the width of the interval is no more than 6% from left end to right end. Will they need the same sample size as in 2.1? Pick the right answer and explain further without calculation. **(6 Points)**

1. Yes, they must use the same sample size.
2. No, a smaller sample size will work.
3. No, they will need a bigger sample.

2: No, a samller sample size will work. This is because in 95% confidence interval, we calculate the width with 4SDs, and using 90 pecent confidence interval would have a smaller sample size since we are using less SDs in a calculation. In this essence, the sample size will go down.

Question 2.4. This shows that the percentage in a normal distribution that is at most 1.65 SDs above average is about **95%**. Explain why 1.65 is the right number of SDs to use when constructing a **90%** confidence interval. **(6 Points)**

The percentage of 1.65 SD, you can get by subtracting, 100-95 which is 5 percent. The percent below 1.65SD will also be 5 percent. The percentage in the middle is 90 percent, between above 1.65 SD and below 1.65 SD. That will sum up to a 100%, which indicates that 1.65 is the correct when using for constructing 90% confidence interval.

Question 3.2. Why does the Central Limit Theorem (CLT) apply in this situation, and how does it explain the distribution we see above? **(6 points)**

According to the Central Limit Theorem (CLT), as the sample size simulation was done 10000 times, a larage amount of number, the distribution of random variable x is approaching a normal distribution. So, we can first not think about the distribution of the population where this sample is drawn. The distribution shown above, the proportion of people who voted yes is close to normal and shows a bell shaped graph. This case the CLT is applied which explains how the distribution is close to normal.