**Question 1**

**New medication and side effects**

A systematic review of drug safety shows that 5% of adults who take a new medication experience negative side effects.

For the surveillance of drug safety, your hospital is interested in the chance of seeing more than 10 side effects among 100 randomly selected patients who need to take this medication now.

1. Find the probability that more than 10 patients in the random sample of 100 will experience negative side effects. You can use the binomial formula or normal approximation.
2. Find the expected number of side effects in this random sample.
3. The cost for the side effects is high from the patient's QoL (Quality of Life) and monetary perspective, but the hospital will have 1000 patients who need this medication next year and need to be certain about the potential loss. The hospital cannot use this medication if over 5% of people develop the side effects. What is the probability that less than 5% of patients will have side effects next year?

For all problems, write down the probability you are trying to calculate, to simplify the problems.

E.g., P(X > 5) = 1 – P(X<=5)

**Question 2 – Confidence Intervals**

Your client is interested in estimating the self-reported health score (ranging from 0-100; see the pet and pandemic study in Part 1) among low-income adults in Montreal.

From a large-scale population-representative survey conducted recently, the mean health score of adults in Montreal was 70, and the standard deviation was 20.

You randomly sampled 50 participants from the low-income population in Montreal, and found the sample mean among the low-income groups as 60.

a) State your population of interest and write down -values (lower and upper side) corresponding to 95% of the normal curve.

b) Calculate the margin of error at a 95% confidence level. Use the population standard deviation of 20 (in the next class, you will see how to calculate this from your sample)

c) Make an interpretation for the 95% confidence interval for the health score.

d) What are the approximate bounds on the values of health scores that would include 68% of the sampling distribution among the low-income population in Montreal?

e) Calculate the margin of error at a 90% confidence level.

f) Calculate the margin of error when the sample size is n = 200.

g) Do you think that the low-income population in Montreal has the same average health score as all adults in Montreal? Explain at 95% confidence level.

h) State two assumptions to construct CI (hint, one has something to do with sampling and the other is related to the Central Limit Theorem)

1. Finally, you want to set the confidence level to 99%. To keep the same width of the margin of error as question b), do you need a larger or smaller sample size? Explain in one line.

FYI, the z-scores for the commonly used confidence level are 90%, 95%, and 99%. This implies that the green areas take 10%, 5%, and 1% of the extreme side of the curve for CI bands that are expected to hit the true value 90%, 95%, and 99% of the time in the long run.

A diagram of a normal curve

Description automatically generated

**Questions about hypothesis testing**

Refer to the problem setting above i.e., health score. Our research question is: whether the health score among the low-income population in Montreal is the same as the mean score from all adults in Montreal or not. We will use a hypothesis test. Again, the mean score among all adults is 70, and the population standard deviation is 20. The sample mean obtained from low-income participants from random sampling is with *n* = 50.

1) State the null Hypothesis and the alternative hypothesis.

*H0*:

*Ha*:

2) Do you think that the two-sided test is more appropriate than the one-sided test? If so, why?

3) Compute the test statistic (Z-score corresponding to your sample mean, standardized by the sampling distribution assuming that the null hypothesis is true). Show your work to derive the Z-score.

4) Calculate the p-value from the resulting Z-score. Note that the value depends on whether your hypothesis is based on a 2-sided test or not.

5) Provide an interpretation for the computed p-value.

6) Draw the location of your Z-score on the plot of Z-distribution and shade the p-value (the area as extreme or more extreme than the obtained Z-value from the center).

7) On a separate image, point to the location of the Z-score (both sides of the curve) that indicates the center 68%, 95%, and 99% of the Z-distribution.

8) Based on the calculated p-value, do you think you will reject the null hypothesis or retain it at the threshold probability of rejection for 0.32, 0.05, and 0.01?

9) Provide an interpretation for the p-value and state your conclusion based on the 0.05 threshold for the rejection of the null hypothesis.

10) Does this conclusion agree with that of 95% CI?