Class Exercise Solutions

1. **Fasting Blood Glucose**
   Clinical data indicates that the fasting blood glucose levels in a specific population of healthy adults are approximately normally distributed with a mean $\mu = 90$ mg/dL and a standard deviation $\sigma = 16$ mg/dL.

   Compute the probability that the average fasting blood glucose level will be greater than 92 mg/dL for:

   a) A randomly selected individual ($n = 1$).
   b) A random sample of $n = 64$ individuals.
   c) A random sample of $n = 512$ individuals.
   d) Explain conceptually why the probability of observing a sample mean strictly greater than 92 mg/dL decreases so drastically as the sample size increases.

   Solution:
   Let X be the blood glucose level. We are given $X \sim N(90, 16^2)$.

   a) For $n = 1$, we want $P(X > 92)$.

   $$Z = \frac{92 - 90}{16} = 0.125$$

   $$P(X > 92) = P(Z > 0.125) = 1 - 0.5497 = 0.4503$$

   b) For $n = 64$, the sample mean follows $\bar{X} \sim N\left(90, \left(\frac{16}{\sqrt{64}}\right)^2\right) \to N(90, 2^2)$.

   $$Z = \frac{92 - 90}{2} = 1.00$$

   $$P(\bar{X} > 92) = P(Z > 1.00) = 1 - 0.8413 = 0.1587$$

   c) For $n = 512$, the sample mean follows $\bar{X} \sim N\left(90, \left(\frac{16}{\sqrt{512}}\right)^2\right) \to N(90, 0.7071^2)$.

   $$Z = \frac{92 - 90}{0.7071} \approx 2.83$$

   $$P(\bar{X} > 92) = P(Z > 2.83) = 1 - 0.9977 = 0.0023$$

   d) As the sample size n increases, the standard error of the mean ($\sigma/\sqrt{n}$) decreases. This means the sampling distribution of $\bar{X}$ becomes much tighter and highly concentrated around the true population mean ($\mu = 90$). Consequently, observing a sample mean even slightly far from the true mean (such as 92) becomes increasingly rare.

2. **Genetic Marker Prevalence**
A specific genetic marker associated with an increased risk of a rare autoimmune condition is known to be present in 15% of a given demographic. A clinical research team selects a random sample of $n = 400$ patients from this demographic. Let X be the number of patients in the sample who carry the genetic marker.

   a) Use the normal approximation to the binomial distribution **with** the appropriate continuity correction to estimate the probability that between 50 and 70 patients (inclusive) carry the marker. That is, estimate $P(50 \le X \le 70)$.

   b) Now, let $\hat{p} = X/n$ be the sample proportion of patients carrying the marker. Using the Central Limit Theorem, compute the probability that the sample proportion lies between 0.125 and 0.175. Do **not** use a continuity correction for this part. Compare your result to part (a).

Solution:
We have $X \sim \text{Binom}(400, 0.15)$. The mean is $\mu_X = np = 400(0.15) = 60$. The standard deviation is $\sigma_X = \sqrt{np(1-p)} = \sqrt{400(0.15)(0.85)} = \sqrt{51} \approx 7.141$.

a) We want $P(50 \le X \le 70)$. Applying the continuity correction, we extend the interval by 0.5 on each side: $P(49.5 \le X \le 70.5)$.

$$Z_1 = \frac{49.5 - 60}{7.141} \approx -1.47$$

$$Z_2 = \frac{70.5 - 60}{7.141} \approx 1.47$$

$$P(-1.47 \le Z \le 1.47) = 0.9292 - 0.0708 = 0.8584$$

b) We want $P(0.125 \le \hat{p} \le 0.175)$. The mean of the sample proportion is $\mu_{\hat{p}} = p = 0.15$. The standard error is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.15(0.85)}{400}} \approx 0.01785$.

$$Z_1 = \frac{0.125 - 0.15}{0.01785} \approx -1.40$$

$$Z_2 = \frac{0.175 - 0.15}{0.01785} \approx 1.40$$

$$P(-1.40 \le Z \le 1.40) = 0.9192 - 0.0808 = 0.8384$$

Comparison: The results differ slightly (0.8584 vs 0.8384). The calculation in part (a) is generally considered a better approximation for the true binomial probability because the continuity correction actively compensates for using a continuous distribution (the normal curve) to estimate a discrete variable (counts of patients).