STAT 521: Assignment 3

Make sure to show your computation and/or attach appropriate output.

Problem 1

Do Example 8 of Lecture 3 (page 12).

Problem 2 (Computer exercise)

Read my note on the central limit theorem ("Central Limit Theorem.pdf") and answer following questions using SAS, R, or Python (Show all the relevant output. You do not need to include the SAS/R code):

a) Using the same population, draw 1000 samples of size $n = 5$ and produce a histogram of sample means. How does it look like? Report the mean and standard deviation of the sampling distribution.

b) Repeat part (a) for 1000 samples of size $n = 10$ and $n = 60$. For each case, produce a histogram of sample means and report the mean and standard deviation of the sampling distribution.

c) Compare the three histograms you produced in part (a) and (b) above. As the sample size n changes, are there any changes in the shape of the sampling distribution? Which one is closer to the normal distribution? How do the mean and standard deviation of the sample means change? Describe your findings.

Problem 3

The population mean blood level of lead of children who live in a city is 12 with a standard deviation of 3. If you randomly select 36 children from this population, what is the probability that the sample mean of blood lead level will be:

a) Below 10.8

b) Above 12.8

c) Between 11.2 and 12.8

Problem 4

Refer to Problem 3 above. For a sample of $n = 36$, what is the probability that the sample mean of blood lead level will differ from the true population mean by no more than 0.5?

Problem 5

For a population of 17-year-old boys and girls, the means and standard deviations, respectively, of their subscapular skinfold thickness values are as follows: boys, 9.7 and 6.0; girls, 15.6 and 9.5. Simple random samples of 40 boys and 35 girls are selected from the populations. What is the probability that the difference between sample means $\bar{x}_{girls} - \bar{x}_{boys}$ will be greater than 10?

Problem 6

It is known that 35 percent of the members of a certain population suffer from one or more chronic diseases. What is the probability that in a sample of 200 subjects drawn at random from this population 80 or more will have at least one chronic disease?

Problem 7 (Biostat students only)

Let X be a continuous random variable with the probability density function $f(x)$. The expected value of X is:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Now suppose $Y = g(X)$. The expected value of Y is then defined as:

$$E(Y) = E\left(g(x)\right) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Also note that $Var(X) = E[(x-\mu)^2] = E(X^2) - \mu^2$ where $E(X) = \mu$ (See Assignment #2).

Example: Let X be a continuous random variable with the PDF: $f(x) = \begin{cases} \frac{3}{8}x^2, & 0 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$.

The expected value of X is:

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^2 x\left(\frac{3}{8}x^2\right)\, dx = \int_0^2 \frac{3}{8}x^3\, dx = \frac{3}{8(4)}x^4\Big]_0^2 = \frac{3}{2}$$

Similarly, you can find $E(X^2)$:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^2 x^2\left(\frac{3}{8}x^2\right)\, dx = \int_0^2 \frac{3}{8}x^4\, dx = \frac{3}{8(5)}x^5\Big]_0^2 = \frac{12}{5}$$

The variance of X is then:

$$Var(X) = E(X^2) - \mu^2 = \frac{12}{5} - \left(\frac{3}{2}\right)^2 = \frac{3}{20}$$

Question: Find the expected value and variance for the following continuous random variables:

a) A random variable X with $f(x) = \begin{cases} 6x(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$

b) A random variable Y with $f(y) = \begin{cases} \frac{1}{3}e^{-y/3}, & y > 0 \\ 0, & y \le 0 \end{cases}$

c) A random variable W with $f(w) = \begin{cases} w, & 0 < w < 1 \\ 2-w, & 1 \le w < 2 \\ 0, & \text{elsewhere} \end{cases}$

Problem 8 (Biostat students only)

Let X be a continuous random variable with the probability density function $f(x)$. Define the cumulative density function of X as $P(X \leq x) = F(x)$. By definition,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

where t is a dummy variable of integration. Thus it follows that: $f(x) = \frac{d}{dx}F(x) = F'(x)$

Also note that $F(x)$ is a non-decreasing function of x and satisfies: $F(-\infty) = 0$ and $F(\infty) = 1$.

Example 1: Let X be a continuous random variable with the PDF: $f(x) = \begin{cases} \frac{3}{8}x^2, & 0 \leq x \leq 2 \\ 0, & \text{elsewhere} \end{cases}$. We wish to find $F(x)$.

For $x < 0$, $F(x) = 0$ because $f(x) = 0$ in this range. For $0 \leq x \leq 2$, $F(x) = \int_0^x \frac{3}{8}t^2\,dt = \frac{1}{8}x^3$. For $x > 2$, $F(x) = 1$. Thus,

$$F(X) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}x^3, & 0 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$

Example 2: Let X be a continuous random variable with the CDF: $F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/3}, & x \geq 0 \end{cases}$. We wish to find $f(x)$.

For $x < 0$, $f(x) = 0$. For $x \geq 0$, $f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(1 - e^{-x/3}\right) = \frac{1}{3}e^{-x/3}$. Thus,

$$f(x) = \begin{cases} \frac{1}{3}e^{-x/3}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Question: The length of time required by students to complete a one-hour exam is a random variable with the PDF:

$$f(x) = \begin{cases} \frac{2}{3}x^2 + x, & 0 \leq x \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

a) Find $F(x)$.

b) Use $F(x)$ above to find $F(0)$, $F(0.5)$, and $F(1)$.