STATISTICS FOR MANAGEMENT (IDS 570)

HOMEWORK 7
DUE DATE: THURSDAY, OCTOBER 30 AT 11:59 PM

Problem 1. (The Normal Distribution). Let $X \sim N(-2, 2)$.

(a) [3 pts] Find the mean, median and standard deviation of X.
(b) [3 pts] Let $Y = a \cdot X + b$. For what values of a and b will Y be a standard normal random variable?

Problem 2. (Linear Combinations of Random Variables). Let X_1 and X_2 be independent and identically distributed (i.i.d.) random variables. Consider the random variable $X = 2X_1 - X_2$. For each of the following 3 scenarios, find $E[X]$ and $P(X \geq 4)$.

(a) [3 pts] X_1 and X_2 are uniform (continuous) [0,2] random variables ($X_1, X_2 \sim U[0, 2]$).
(b) [3 pts] X_1 and X_2 are discrete random variables taking values $\{0, 2\}$ with equal probability.
(c) [3 pts] X_1 and X_2 are normally distributed with mean 1 and variance 1.

Problem 3. Assume that the number of telephone calls arriving at a switchboard is described by a Poisson Process. We observe that, on average, one call comes in every 10 minutes.

(a) [4 pts] What is the probability that no calls arrive in the first 20 minutes of the day?
(b) [4 pts] The switchboard has been live for 24 minutes and 4 calls have come in. What is the probability that the 5^{th} call arrives after 44 minutes of the switchboard being live?
(c) [4 pts] What is the probability that exactly one call arrives in the first ten minutes of the day and no calls arrive in the following 5 minutes?
(d) [4 pts] What is the probability that between 10:00 am and 10:20 am the switchboard receives exactly 1 call and that between 10:10 am and 10:30 am the switchboard receives exactly 3 calls?

Problem 4. [5 pts] Jones figures that the total number of thousands of miles that an auto can be driven before it would need to be junked is an exponential random variable with parameter $\lambda = \dfrac{1}{20}$. Smith has a used car that he claims has been driven only 10,000 miles. If Jones purchases the car, what is the probability that she would get at least 20,000 additional miles out of it? Repeat under the assumption that the lifetime mileage of the car is not exponentially distributed but rather is (in thousands of miles) uniformly distributed over (0, 40).

Problem 5. (Combining Continuous and Discrete Distribution). The amount of milk that a cow produces each day is normally distributed with mean 10 gallons and standard deviation 1.4 gallons. Farmer Joe considers a cow to be good if it produces over 9 gallons of milk in a day.

(a) [4 pts] What is the probability that a cow is good?
(b) [5 pts] Suppose that Joe has 5 cows and the milk production for each cow is independent of the milk production for all other cows. What is the probability that Joe has 2 good cows or less?

Problem 6. Given a population with a mean $\mu = 100$ and a variance of $\sigma^2 = 900$, the central limit applies when the sample size $n \geq 25$. A random sample of size $n = 30$ is obtained

(a) [3 pts] What are the mean and variance of the sampling distribution for the sample means?
(b) [5 pts] What is the probability that $96 \leq \bar{x} \leq 110$?

Problem 7. (Confidence Interval)

(a) [6 pts] A sample of size $n = 100$ produced the sample mean of $\bar{x} = 16$. Assuming the population standard deviation $\sigma = 3$, compute a 95% confidence interval for the population mean μ.

(b) [5 pts] Assuming the population standard deviation $\sigma = 3$, how large should a sample be to estimate the population mean μ with a margin of error not exceeding 0.5?

(c) [5 pts] We observed 28 successes in 70 independent Bernoulli trials. Compute a 90% confidence interval for the population proportion p.

Problem 8. Use R to answer the following question.

(a) [5 pts] Pull 50 random samples of size 4 from a population which is $N(10, 2)$ in R. Calculate the mean and standard deviation.

(b) [3 pts]Pull 50 random samples of size 25 from a population which is $N(10, 2)$. Calculate the mean and standard deviation.

(c) [3 pts] Pull 50 random samples of size 100 from a population which is $N(10, 2)$. Calculate the mean and standard deviation

(d) [5 pts] Create a table showing the actual figures and the figures predicted by the central limit theorem.

Create a histogram of all three distributions using bins of 7, 9, 11, 13, 15. Dont worry if you cannot see the distribution from (b) and (c) very well Comment on your results. Are your mean and standard deviation figures close to what was predicted by the Central Limit Theorem? Does the graph show that the standard deviation from (b) and (c) is smaller (which the distributions appear to have less spread)?

Hint: To generate 50 random samples you can use a "for" loop: For example: "for(i in 1:50) {...}". In addition, you can store the values of each sample in a column of a matrix. For example, for part (a) you can define a matrix "rand.samp = matrix(0, nrow = 4, ncol = 50)" and fill each column "rand.sample[, i]" in the for loop with the values from samples.

Problem 9. (Sampling Distribution in R)

(a) [5 pts] Suppose we have a random sample of size 100 from a Binomial distribution with the population proportion of 0.3. What are the expected mean $E[\hat{p}]$ and variance $Var[\hat{p}]$ of the sample proportion $\hat{p}$? What is the sampling distribution of the sample proportion $\hat{p}$?

(b) [3 pts] Generate 1000 samples of size 100 from a Binomial distribution with population proportion of 0.3, calculate the sample proportion for each sample, and save them in the vector "sample.p".

(c) [3 pts] Calculate the mean and variance of the 1000 sample proportions, and compare to the results in (a).

(d) [4 pts] Plot the density of the sample proportions, and compare it to the curve of Normal distribution with mean E.p and variance V.p, where E.p is the expected mean $E[\hat{p}]$ and V.p is the variance $Var[\hat{p}]$ from (a), using the following commands:
plot(density(sample.p), main="Density Plot of Sample Proportions", xlab="Sample Proportions", xlim=c(0.1, 0.5), col="red")
curve(dnorm(x, mean=E.p, sd=sqrt(V.p)), from=0.1, to=0.5, add=TRUE, col="green")