STATISTICS FOR MANAGEMENT (IDS 570)

HOMEWORK 7 SOLUTION

Problem 1

(a) Mean: -2
Median: Since normal distribution are symmetric about their mean, the median is also -2
Standard Deviation: $Var(X) = 4$ and so the standard deviation is 2.

(b) Recall that since X is a normal random variable, Y is also a normal random variable, with $E[Y] = E[a \cdot X + b] = a \cdot E[X] + b = a \cdot (-2) + b = -2a + b$. Furthermore, $Var[Y] = Var[a \cdot X + b] = a^2 \cdot Var[X]$. Therefore, $Var[Y] = a^2 \cdot 4 = 4a^2$.

We desire $Y \sim N(0, 1)$ and so we want $E[Y] = 0$ and $Var[Y] = 1$. Since $Var[Y] = 4a^2$ we can set $4a^2 = 1$. Solving for a yields $a = \pm\frac{1}{2}$.

For $a = \frac{1}{2}$, using $E[Y] = -2 \cdot a + b = -2 \cdot \frac{1}{2} + b = -1 + b$ and since we want $E[Y] = 0$, we can solve for b to get $b = 1$.

For $a = -\frac{1}{2}$, using $E[Y] = -2 \cdot a + b = -2 \cdot (-\frac{1}{2}) + b = 1 + b$ and since we want $E[Y] = 0$, we can solve for b to get $b = -1$.

Problem 2

(a) $E[X] = 2 \cdot E[X_1] - E[X_2] = 2 \cdot 1 - 1 = 1$. The only outcomes for which $X = 2 \cdot X_1 - X_2 = 4$ is when $X_1 = 2, X_2 = 0$. Therefore, $\mathrm{P}(X \geq 4) = P(X_1 = 2, X_2 = 0) = 0$ since the probability that a continuous random variable equals a single value is zero.

(b) In this case $E[X_1] = E[X_2] = 0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = 1$. Therefore $E[X] = 2 \cdot E[X_1] - E[X_2] = 2 \cdot 1 - 1 = 1$, as in part (a). In contrast to part (a), however, $P(X \geq 4) = P(X_1 = 2, X_2 = 0) = P(X_1 = 2) \cdot P(X_2 = 0) = \frac{1}{4}$.

(c) X will be a normally distributed random variables with mean $E[X] = 2 \cdot E[X_1] - E[X_2] = 1$ and the variance $Var[X] = 4 \cdot Var[X_1] + Var[X_2] = 5$. And so, $P(X \geq 4) = 1 - P(X \leq 4) = 0.0899$.

Problem 3

(a) Let T be the time it takes for the first call of the day to arrive. T follows an exponential distribution, with $E[T] = 10$ minutes, so that $T \sim exp(\frac{1}{10})$. We desire $P(T \geq 20) =$ 1−EXPONDIST(20, 0.1, 1)= 0.1353.

Alternatively, let N be the number of calls received in the first 20 minutes of the day. Since the calls arrive according to a Poisson Process, with average one per 10 minutes, the number of calls that arrive in the first 20 minutes follows a Poisson distribution with average two per 20 minutes. We desire $P(N = 0) = \dfrac{e^{-2} \cdot 2^0}{0!} = 0.1353$.

(b) Let T_5 be the time between the 4^{th} and 5^{th} calls and suppose that the fourth call occurred at time t_4, with $t_4 \leq 24$ because the 4^{th} call came before the 24^{th} minute. $T_5 \sim exp(1/10)$. The question asks for $P(T_5 \geq (44-24)+(24-t_4) \mid T_5 \geq (24-t_4))$. By the memoryless-ness property of the exponential distribution this is exactly $P(T_5 \geq (44-24)) = P(T_5 \geq 20)$ which, by part (a), is 0.1353.

(c) Let N_1 be the number of calls that arrive in the first ten minutes of the day and let N_2 be the number of calls that arrive in the following 5 minutes. Therefore we desire $P(N_1 = 1, N_2 = 0)$. Note that these two random variables are independent because of the memorylessness property of the exponential distribution and the fact that these time intervals do not intersect. Therefore, following the reasoning in part (a), N_1 follows a Poisson distribution with mean 1 and N_2 follows a Poisson distribution with mean 0.5, and so $P(N_1 = 1, N_2 = 0) = P(N_1 = 1) \cdot P(N_2 = 0)$ =POISSON(1, 1, 0) $\cdot$ POISSON(0, 0.5, 0) = 0.2231.

(d) Since the intervals in this question intersect, we have to break up the time between 10:00 am and 10:30 am into 3 intervals, 10:00 - 10:10, 10:10 - 10:20, 10:20 - 10:30. Let N_1(respectively N_2, N_3) be the number of calls received between 10:00 - 10:10 (respectively 10:10 - 10:20, 10:20 - 10:30). Each follows a Poisson distribution with mean 1. We desire $P(N_1+N_2 = 1, N_2+N_3 = 3)$. There are two ways that this can occur: $N_1 = 1, N_2 = 0, N_3 = 3$ or $N_1 = 0, N_2 = 1, N_3 = 2$. Therefore, $P(N_1 + N_2 = 1, N_2 + N_3 = 3) = P(N_1 = 1, N_2 = 0, N_3 = 3) + P(N_1 = 0, N_2 = 1, N_3 = 2) =$ POISSON(1, 1, 0) $\cdot$ POISSON(0, 1, 0) $\cdot$ POISSON(3, 1, 0) + POISSON(0, 1, 0) $\cdot$ POISSON(1, 1, 0) $\cdot$ POISSON(2, 1, 0) =0.0083 + 0.0249 = 0.0332.

Problem 4

Let X be the number of miles the car can be driven before being junked.

(a) First, let $X \sim exp(\frac{1}{20})$. We know that the car has been driven 10,000 miles and we want to find the probability that it will last an additional 20,000 miles. We use the memoryless property of exponential distribution, where $t = 10000$ and $s = 20000$. Then $P(X \geq 10000 + 20000 \mid X > 10000) = P(X \geq 20000)$.

$$\begin{aligned} P(X \geq 30000 \mid X > 10000) &= P(X \geq 20000) \\ &= \int_{20}^{\infty} e^{-\frac{1}{20}x} dx \\ &= e^{-1} \end{aligned}$$

(b) Let $X \sim U(0, 40)$. We want to find $P(X > 30 \mid X > 10)$. By Bayes Rule,

$$\begin{aligned} P(X > 30 | X > 10) &= \frac{P(X > 30)}{P(X > 10)} \\ &= \frac{\int_{30}^{40} \frac{1}{40} dx}{\int_{10}^{40} \frac{1}{40} dx} \\ &= \frac{40 - 30}{40 - 10} \\ &= \frac{1}{3} \end{aligned}$$

Problem 5

(a) Let X be the amount of milk produced by a given cow. We desire $P(X \geq 9) = 1-$NORMDIST(9, 10, 1.4) = 0.7625.

(b) Let Y be the number of good cows that Joe has. Then Y is a binomial random variables with $n = 5$, $p = 0.7625$. Therefore, $P(Y \leq 2)$ =BINOMDIST(2, 5, 0.7625) = 0.0908.

Problem 6

(a) Based on the central limit theorem, the sampling distribution is approximately a Normal distribution with mean $\mu = 100$ and the standard deviation $\sigma/\sqrt{n} = 30/\sqrt{30} = 5.45$.

(b)

$$P(96 \le \bar{x} \le 110) \quad = \quad P(\frac{96-100}{5.45} \le Z \le \frac{110-100}{5.45})$$

Problem 7

(a) A 95% confidence interval for μ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

where $z^* = z_{0.025} = 1.96$ from the table of Normal distribution.
Then, the 95% confidence interval for μ is

$$16 \pm (1.96)\frac{3}{\sqrt{100}} = 16 \pm 0.588 = [15.412, 16.588]$$

(b) $n \ge (z^*\sigma/m)^2$, where the critical value $z^* = z_{0.025} = 1.96$, and the required margin $m = 0.5$. Then we need a sample size of

$$n \ge \left(\frac{(1.96)(3)}{0.5}\right)^2 = 138.3$$

i.e., we need a sample of size at least 139.

(c) A 90% confidence interval for p is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where the sample proportion $\hat{p} = 28/70 = 0.4$ and the sample size $n = 70$. Then, the 90% confidence interval for p is

$$0.4 \pm (1.645)\sqrt{\frac{(0.4)(0.6)}{70}} = 0.4 \pm 0.096 = [0.304, 0.496].$$

Problem 8

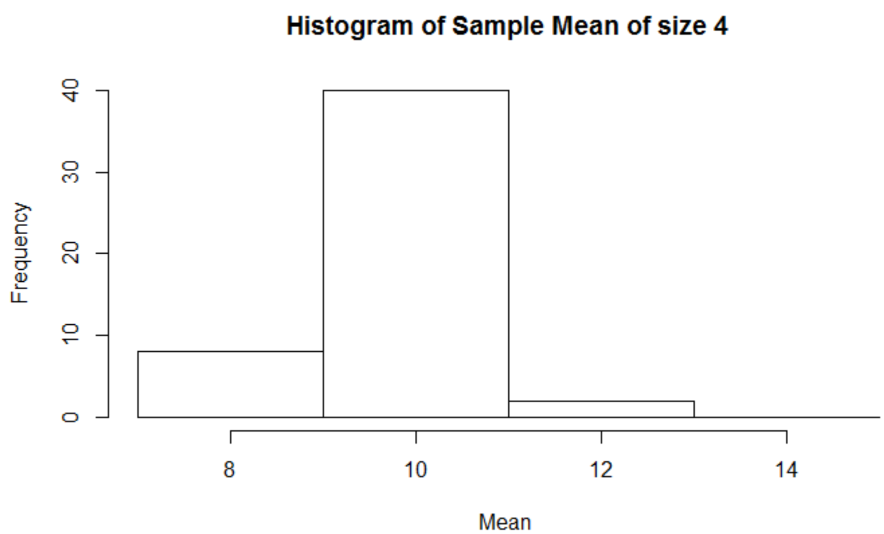
(a) To generate random sample of size 4 from $N(10, 2)$, you can use the code:
sample ← rnorm(n = 4, mean =10, sd = 2).
To generate 50 random sample of this type we can use:
sample.data ← matrix(0, nrow = 4, ncol = 50),
for(1 in 1:50){sample.data[, i] ← rnorm(n = 4, mean =10, sd = 2)}.
Each column of the matrix sample.data contains a sample of size 4 from the population with distribution $N(10, 2)$.
To obtain the mean and standard deviation of the sample data we can use the following code:
for(i in 1:50) {mean[i] ← mean(sample.data[, i]), stdv[i] ← sd(sample.data[, i])}

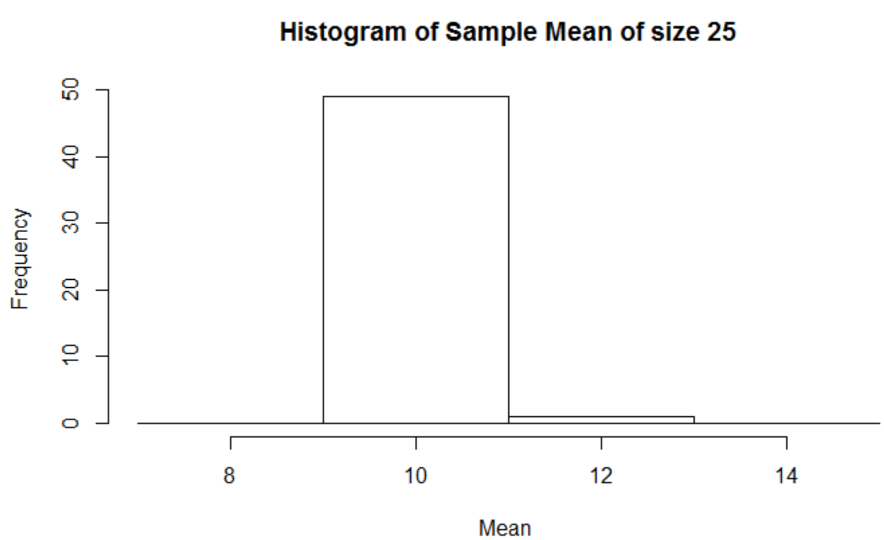
(b) Similar to (a) except nrow is 25.

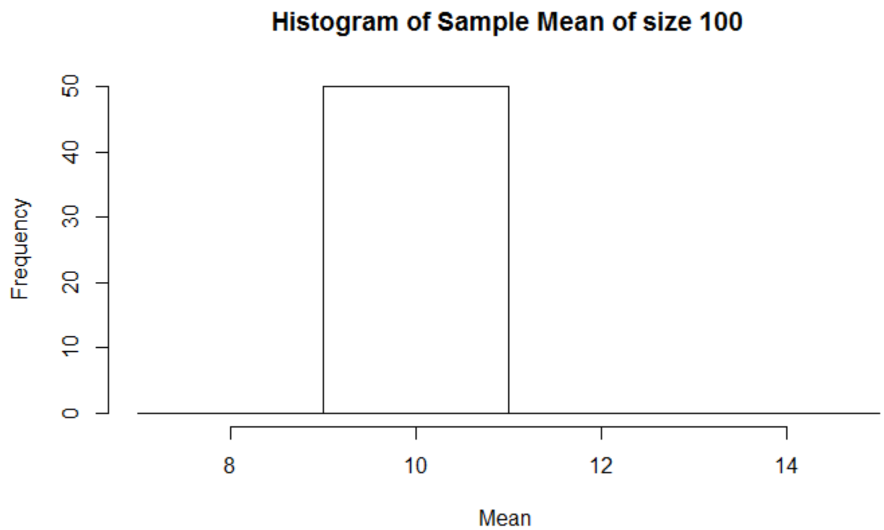
(c) Similar to (a) except nrow is 100.

(d) We can see that the mean of the distribution of sample means appears to be the same as the mean of the "population" from which we selected our samples This can be easily checked by taking the average of means. In other words if we take the average of all the values in the mean and stdv vectors we see that the average of samples' means is close to 10 and the average of samples standard deviation is close to 0.4. In addition, when we select samples from a normal distribution, then the distribution of sample means is also "normal" in shape.

By increasing the size of our samples, the histogram of the sample means becomes "less spread out" or "narrower," as is clearly seen when contrasting the spread of the histograms. This behavior (increasing the sample size decreases the spread) seems quite reasonable. We would expect a more accurate estimate of the mean of the population if we take the mean of larger sample size. All these observations follow the central limit theorem.







Problem 9

(a) The expected mean is $E(\hat{p}) = p = 0.3$, and the variance is $Var(\hat{p}) = pq/n = 0.3 \times (1 - 0.3)/100 = 0.0021$. The sampling distribution of $\hat{p}$ is approximately normal with mean 0.3 and variance 0.0021.

(b) We can generate the 1000 sample proportions in R: sample.p ← rbinom(1000, 100, 0.3) / 100

(c) Calculate the mean and variance of the 1000 sample proportions in R:
mean(sample.p)
var(sample.p)
You may get different answers given different simulations, but the mean and variance of the 1000 sample proportions should be very close to the expected mean and variance in (a), respectively.

(d) From the plot, we can see that the shape of the density of simulated sample proportions matches very well with the shape of the normal distribution with mean p and variance pq/n. The sampling distribution of the sample proportion is approximately normal.