**Group # 9**

**Group Leader and ID: HARISH RAGURU (02121168)**

**Member Names and IDs: BANOTH JEEVAN KUMAR (02105145) and HANUMA VENKATA VIJAY KAMAL KONDURI (02128290)**

**HW1 answers**

**1)**

The table provides us with the joint probabilities of being male or female (X) and having a risk level of low, medium, or high (Y).

X=0 represents male, X=1 represents female.

Y=0 represents low risk, Y=1 represents medium risk, Y=2 represents high risk.

The table entries are the probabilities for each combination of X and Y.

**a)**

p(X = female, Y = high risk) = p(X=1, Y=2)

We look at the cell where X=1 and Y=2 and read the probability directly from the table, which is 1/10 or **0.1**.

**b)**

This is the marginal probability of X being female (X=1), which is the sum of the probabilities across all Y values given X=1. From the table, we add:

p(X=1, Y=0) + p(X=1, Y=1) + p(X=1, Y=2) = 2/5 + 4/25 + 1/10

When we reduce this fraction, we get **0.66**.

**c)**

p(Y = high risk | X = female) = p(Y=2 | X=1)

This is the conditional probability of Y being high risk given that X is female. We use the definition of conditional probability:

p(Y=2 | X=1) = p(X=1, Y=2) / p(X=1)= 0.1 / 0.66 = **0.1515**

**2)**

**I)**

The Central Limit Theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed.

* np ≥ 5 and n(1-p) ≥ 5: In your case, np = 30 \* 0.2 = 6 and n(1-p) = 30 \* 0.8 = 24. Both values are greater than 5, so this suggests the CLT might be applicable.
* n ≥ 30: This is another commonly used rule, and your sample size n = 30 satisfies it.

Based on these considerations, the CLT is likely to hold true for our scenario of drawing samples of size n = 30 from a binomial distribution with p = **0.2**.

**II)**

Mean and Standard Deviation of Sample Means (assuming CLT holds):

If CLT holds then:

* Mean of sample means (μ\_̅) = population mean (μ) = p = 0.2
* Standard deviation of sample means (σ\_̅) = σ / √n, where σ is the population standard deviation. For a binomial distribution, σ = √(np(1-p)) = √ (30 \* 0.2 \* 0.8) ≈ 2.77. Therefore, σ\_̅ ≈ 2.77 / √30 ≈ 0.073

*μ* =*n* × *p*

where,

*n = 30 (sample size)*

*p = 0.2 (Probability of success)*

*So,*

*μ = 30 \* 0.2 = 6*

**Standard Deviation of the Sample Means (σ\_x̄):**

The standard deviation of sample means is given by:

*σx*ˉ = *σ / sqrt(N)*

For a binomial distribution, the standard deviation (σ) of the population is given by:

*σ* = sqrt(*n*×*p*×(1−*p*))

So,

= sqrt (30 \* 0.2 \* (1 - 0.2))

= sqrt (4.8)

= 2.19

Then, the standard deviation of the sample means (σ\_x̄) is:

*σx*ˉ = *σ* / sqrt(n)

*σx*ˉ ≈0.4

Therefore, if the CLT holds true, the mean of these sample means is **6**, and the standard deviation of these sample means is approximately **0.4**.

**3)**

For a Poisson distribution, the mean (often denoted as λ) is equal to the variance. Therefore, when we are given that the mean of the sample means is 6, we can deduce that the population mean (λ) is also 6 because the mean of the sample means is an unbiased estimator of the population mean.

The Central Limit Theorem (CLT) tells us that for sufficiently large sample sizes, the distribution of sample means will be approximately normally distributed with a mean (μ\_x̄) equal to the population mean (λ) and a standard deviation (σ\_x̄) equal to the population standard deviation (σ) divided by the square root of the sample size (n).

In the case of the Poisson distribution, the population standard deviation (σ) is the square root of the population mean (λ). So, for a Poisson distribution with a mean of 6, the standard deviation is also the square root of 6.

Now, since we are drawing samples of size n=40, the standard deviation of the sample means (also called the standard error) will be the population standard deviation divided by the square root of n.

Let's calculate the population standard deviation based on the CLT:

The population standard deviation (σ) for a Poisson distribution with a mean (λ) of 6 is approximately 2.45, since the standard deviation is the square root of the mean for a Poisson distribution.

Using the Central Limit Theorem (CLT), the standard deviation of the sample means (also known as the standard error) for samples of size n=40 is approximately 0.3873. This is calculated by dividing the population standard deviation by the square root of the sample size (σ/√n).

To summarize, with the mean of sample means being 6 (which is also the population mean for the Poisson distribution), the population standard deviation is approximately **2.45**, and the standard deviation of the sample means is approximately **0.3873**. ​

**4)**

**a)**

set.seed(530)

> mean\_vector <- c(2, 4, 6)

> cov\_matrix <- matrix(c(4, 3, 2,

+ 3, 9, 5,

+ 2, 5, 36), nrow = 3, byrow = TRUE)

> library(MASS)

> data <- mvrnorm(n = 10, mu = mean\_vector, Sigma = cov\_matrix)

> print(data)

[,1] [,2] [,3]

[1,] 3.0120331 9.488371 11.325643

[2,] 1.6990853 1.793820 13.731936

[3,] 2.3514593 6.417456 11.908199

[4,] 4.5419218 7.112661 14.075809

[5,] 0.3526126 -2.722047 4.312137

[6,] 5.4535370 6.597597 2.509881

[7,] 3.4812633 3.585767 3.892062

[8,] -0.7022051 2.133685 -7.147663

[9,] 1.7487346 4.047091 8.173315

[10,] 0.9503228 3.957287 8.685950

**b)**

> mean\_x1 <- mean(data[,1])

> mean\_x2 <- mean(data[,2])

> mean\_x3 <- mean(data[,3])

> median\_x1 <- median(data[,1])

> median\_x2 <- median(data[,2])

> median\_x3 <- median(data[,3])

> skewness\_x1 <- skewness(data[,1])

> skewness\_x2 <- skewness(data[,2])

> skewness\_x3 <- skewness(data[,3])

> kurtosis\_x1 <- kurtosis(data[,1])

> kurtosis\_x2 <- kurtosis(data[,2])

> kurtosis\_x3 <- kurtosis(data[,3])

> print(c(mean\_x1, mean\_x2, mean\_x3))

[1] 2.288876 4.241169 7.146727

> print(c(median\_x1, median\_x2, median\_x3))

[1] 2.050097 4.002189 8.429632

> print(c(skewness\_x1, skewness\_x2, skewness\_x3))

[1] 0.1246822 -0.4244045 -0.8307757

> print(c(kurtosis\_x1, kurtosis\_x2, kurtosis\_x3))

[1] -1.2317698 -0.6100751 -0.3325631

**c)**

> cov\_matrix <- cov(data)

> cor\_matrix <- cor(data)

> print("Covariance matrix:")

[1] "Covariance matrix:"

> print(cov\_matrix)

[,1] [,2] [,3]

[1,] 3.574107 4.310075 4.680237

[2,] 4.310075 11.725862 8.500448

[3,] 4.680237 8.500448 42.154763

> print("Correlation matrix:")

[1] "Correlation matrix:"

> print(cor\_matrix)

[,1] [,2] [,3]

[1,] 1.0000000 0.6657768 0.3812947

[2,] 0.6657768 1.0000000 0.3823368

[3,] 0.3812947 0.3823368 1.0000000

**Summary of our group meet time and duration:**

* Meeting format: In person
* Group meet time and duration: 12pm-4pm, Feb 6th and 2pm-7pm, Feb 7th.
* Average time in communication and discussion regarding assigned group work: we formed a what’s app group and constantly communicated about HW and discussed around 5hrs on avg about HW.
* Participants: HARISH RAGURU (02121168), BANOTH JEEVAN KUMAR (02105145) and HANUMA VENKATA VIJAY KAMAL KONDURI (02128290)

**Contribution report:**

* Each member contributes equally.
* Participants: HARISH RAGURU(02121168), BANOTH JEEVAN KUMAR(02105145) and HANUMA VENKATA VIJAY KAMAL KONDURI(02128290)