1. (10 points) The blood type among Americans is approximately distributed as: 37% type A, 13% type B, 44% type O, and 6% type AB. Suppose the blood types are distributed the same in both male and female populations. And assume that the blood types are independent of marriage.

(A) An individual with type B blood can safely receive transfusions only from persons with type B or type O blood. What is the probability of a husband having type B or type O blood? If a woman has type A blood, what is the probability that her husband is NOT an acceptable donor for her?

(B) What is the probability that, in a randomly chosen couple, the husband has type A blood and the wife has type AB blood?

Solution: Given, $P(A) = 0.37, P(B) = 0.13, P(O) = 0.44, P(AB) = 0.06$

A. Probability of husband having type B or type O blood is:
$P(B) + P(O) = 0.13 + 0.44 = 0.57$ (Since both are mutually exclusive)

Probability of husband Not being and acceptable donor for a woman with type A blood = Probability of husband having type B or type AB, which is:
$P(B) + P(AB) = 0.13 + 0.06 = 0.19$ (Since both are mutually exclusive)

B. Probability of a randomly chosen couple, husband has type A and wife has type AB is:
$P\ (husband\ having\ type\ A) * P(wife\ having\ type\ AB) = P(A) * P(AB) = 0.37 * 0.06 = 0.0222$ (Since both are independent events)

2. (10 points) To overcome the difficulty in getting truthful answers on sensitive issues, "randomized response" is proposed. As an example, we ask the students whether they have gone to a high COVID-risk event. However, they are asked to toss a coin first in private (which has a 50% chance landing on heads and 50% chance landing on tails). The coin toss result is known only to the students themselves. If a "head" is tossed, they are to answer "Yes" regardless of whether they went or not. If a "tail" is tossed, they are to answer truthfully. This way, the surveyor has no way of telling whether a particular individual went to the event or not. Suppose that in fact 20% of the students have gone to this event.

A. What is the probability of a student answering "No" (that is, he/she did not go AND tossed a "tail")?

B. What is the probability that a student did not go and answered "yes" (because a "head" toss)?

C. What is the conditional probability of a student did not go given that he/she answered "yes"?

Solution: Given , $P(tails) = P(heads) = 0.5$;

$P(Student\ attending\ Event) = 0.2$; $P(Student\ not\ attending\ Event) = 0.8$

A. $P(Student\ answering\ No) = P(Student\ not\ attending\ event) * P(tails)$
$= 0.8 * 0.5 = 0.40$

$$P(Student\ answering\ Yes) = 1 - P(Student\ answering\ No) = 1 - 0.40 = 0.60$$

B. $P(Student\ who\ didn't\ attend\ event\ and\ said\ yes) = P(Student\ not\ attending\ event) * P(Student\ answering\ Yes) = 0.8 * 0.6 = 0.48$

C. $P(Student\ didn't\ attend\ event\ |\ Student\ answering\ yes)\ =\ P(Student\ Answering\ Yes)/$ $P(Student\ not\ attending\ event)\ =\ 0.1/\ 0.8\ =\ 0.125$

3. (10 points each) Exercises 11, 12, 17 in section 7.6.

Sec 7.6 Ex. 11. According to the Youth Risk Behavior Surveillance System, 20.7% of all American high school students watch television for three or more hours on a typical school day [169].

(a) If you select repeated samples of size 20 from the population of high school students, what would be the mean number of individuals per sample who watch television for three or more hours per day? What would be the standard deviation?

(b) Suppose that you select a sample of 20 individuals and find that 18 of them watch at least three hours of television per day. Assuming that the Surveillance System is correct, what is the probability that you would have obtained results as extreme as or even more extreme than those you observed?

(c) Suppose that you select a sample of 20 individuals and find that 8 of them watch at least three hours of television per day. Assuming that the Surveillance System is correct, what is the probability that you would have obtained results as extreme as or even more extreme than those you observed?

Solution: X ~ # students out of 20 who watch television for at least 3 hours a day.

X is Binomial distribution with n = 20 and p = probability of success = 0.207

(a) The mean number of students per sample who watch television for at least 3 hours is

$$E(X)\ =\ np\ =\ (20)\ (0.207)\ =\ 4.14\ \approx\ 4$$

Probability mass function of X:

$P(X\ =\ x) = \binom{n}{x}\ p^x\ q^{n-x} = \binom{20}{x}\ 0.207^x\ (1-0.207)^{20-x}$ where x = 0,1, 2,….,20

The standard deviation is $\sigma\ =\ \sqrt{np\ (1\ -\ p)}\ =\ \sqrt{20\ \times\ 0.207\ \times\ 0.793}\ =\ 1.81$

(b) $P(X\ =\ 18) = \binom{20}{18}\ 0.207^{18}\ (1-0.207)^2 = 5.82\ \times\ 10^{-11}$

(c) $P(X\ =\ 8) = \binom{20}{8}\ 0.207^8\ (1-0.207)^{12} = 0.026$

In part(b) that we found 18 such students is very very less likely to happen and in part(c) getting 8 such students is also less likely but more probable than part (b).

Sec 7.6 Ex. 12. The number of cases of tetanus reported in the United States during a single month has a Poisson distribution with parameter = 4.5 [164].

(a) What is the probability that exactly one case of tetanus will be reported during a given month?

(b) What is the probability that at most two cases of tetanus will be reported?

(c) What is the probability that four or more cases will be reported?

(d) What is the mean number of cases of tetanus reported in a one-month period? What is the standard deviation?

Solution:

(a) Given, X ~ Poisson (mean = 4.5)

$$P(X = x) = \frac{4.5^x \, e^{-4.5}}{x!}$$

So, $P(X = 1) = \frac{4.5^1 \, e^{-4.5}}{1!} = 0.04999048$

(b) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$=> \frac{4.5^0 \, e^{-4.5}}{0!} + \frac{4.5^1 \, e^{-4.5}}{1!} + \frac{4.5^2 \, e^{-4.5}}{2!} = 0.1735781$$

(c) $P(X \geq 4) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3)$

$$=> 1 - \frac{4.5^0 \, e^{-4.5}}{0!} - \frac{4.5^1 \, e^{-4.5}}{1!} - \frac{4.5^2 \, e^{-4.5}}{2!} - \frac{4.5^3 \, e^{-4.5}}{3!} = 0.657704$$

(d) $mean = 4.5$

$$standard\ deviation = \sqrt{4.5} = 2.12132$$

Sec 7.6 Ex. 17. The distribution of weights for the population of males in the United States is approximately normal with mean μ = 172.2 pounds and standard deviation σ = 29.8 = pounds [165].

(a) What is the z-score associated with a weight of 130 pounds?

(b) What is the probability that a randomly selected male weighs less than 130 pounds?

(c) What is the probability that he weighs more than 210 pounds?

(d) What is the probability that among five males selected at random from the population, exactly two will have a weight outside the range 130 to 210 pounds?

(e) What is the probability that among five males selected at random from the population, at least one will have a weight outside the range 130 to 210 pounds?

Solution: $\mu = 172.2; \ \sigma = 29.8$

(a) To find Z value for $X = 130$

$$Z = \frac{X - \mu}{\sigma} = \frac{130 - 172.2}{29.8} = -1.4161$$

$$| Z | = 1.4161$$

(b) To find $P(X < 130)$:

From Table Area Under Standard Normal Curve corresponding to Z = 1.4161 is area = 0.4222

$P(X < 130) = 0.5 - 0.4222 = 0.0778.$

(c) To find $P(X > 210)$:

$$Z = \frac{210 - 172.2}{29.8} = 1.2685$$

From Table, area = 0.3980

$P(X > 210) = 0.5 - 0.3980 = 0.1020$

(d) $P(exactly\ 2) = \binom{5}{2} 0.0778^3\ 0.9222^2 + \binom{5}{2} 0.1020^3\ 0.898^2$

$$= 8.5092 + 8.0747 = 16.5839$$

(e) Probability of 1 male between 130 to 210 pounds $= 0.0778 + 0.1020 = 0.1798$

Probability all the 5 within range $= (0.1798) \times 5 = 0.899$

Probability at least 1 outside range $= 1 - 0.899 = 0.101$

4. (5 points) Use R to do this problem.

(a) Compute P(980 < Z < 1032.6) for the random variable Z following a Chi-squared distribution with degrees of freedom 1000.

(b) Use a normal approximation to calculate this probability.

(c) Include in the submission those R commands calculating probabilities in (a) and (b).

Solution:

(a) $P(980 < Z < 1032.6) = P(Z < 1032.6) - P(Z > 980)$

(R Code)

= `pchisq( q = 1032.6, df = 1000, lower.tail = TRUE) - pchisq( q = 980, df = 1000, lower.tail = TRUE)`

$$=> 0.7691504 - 0.3316733 = 0.4374772$$

(b) Using normal approximation:

(R Code)

```
= pnorm(q = 1032.6, mean = 1000, sd = sqrt(2 * 1000), lower.tail = TRUE)
- pnorm(q = 980, mean = 1000, sd = sqrt(2 * 1000), lower.tail = TRUE)
```

$$=> 0.7669864 - 0.3273604 = 0.4396259$$

5. (10 points)

20 employees of a company drive to work. During the time-period of 11am to 11:10am, each employee has a 6% chance of driving in front their office building, and they do this independent of each other. Other cars driving by during the same time-period follows an independent Poisson distribution with mean 1. Let X and Y denote, respectively, the number of employee cars and other cars during this time-period. W=X+Y denote the total number of cars driving by.

(a) Find the mean and variance of W.

(b) Find the probability that exactly W=1 car drove by.

(c) Use the normal approximation to calculate probability in part (b) instead.

(d) Submit R codes calculating the probabilities in parts (b) and (c)

Solution: X is binomial distribution, Y is Poisson distribution.

(a) $W = X + Y$ (X & Y are independent events)

$$Mean(W) = Mean(X) + Mean(Y)$$
$$=> np + 1 = (20)(0.06) + 1 = 2.2$$

$$Var(W) = Var(X) + Var(Y)$$
$$=> np(1-p) + 1 = (20)(0.06)(0.94) + 1 = 2.128$$

(b) $P(W = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 0)$

(R Code)

```
Prob_W1 = [dbinom(x = 0, size = 20, prob = 0.06) * dpois(x = 1, lambda =
1)] + [dbinom(x = 1, size = 20, prob = 0.06) * dpois(x = 0, lambda = 1)]
```

$$=> 0.1067241 + 0.1362436 = 0.2429677$$

(c) Using Normal Approximation:

(R Code)

```
std_w = sqrt(2.128)
Prob_Normal_W1 = pnorm(q = 1.5, mean = 2.2, sd = std_w) - pnorm(q = 0.5,
mean = 2.2, sd = std_w)
```

$$=> \ 0.3156644 - 0.121935 = 0.1937294$$