Homework 2

Mari Sanders

Problem 1

a)

This can be solved using a binomial distribution because there is a fixed number of trials and there are only two possible outcomes, either the individual had at least 1 dental checkup or not.

To solve this problem, we then know that X ~ Binomial(56,0.73). The formula of a binomial distribution is $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

In this case, we have to do $P(X = 40) = \binom{56}{40} 0.73^{40} (1-(0.73))^{56-40} = \frac{56!}{40!(56-40)!} (0.73)^{40} (1-0.73)^{56-40} = 0.1133$

```
exactly_40 <- dbinom(40,size = 56, prob = 0.73)
```

b)

```
at_least_40 <- pbinom(q = 39, size = 56, prob = 0.73, lower.tail = FALSE)
```

$P(X \geq 40) = 0.6678734$

c)

It would not be good to approximate this with a Poisson distribution because n = 56, which is not greater than 100. Also the probability of success is 0.73, and it should be less than 0.01 for Poisson to be a good approximation.

However, a normal approximation would be valid for this situation because $np = (0.73)(56) = 40.88 \geq 10$ and $n(1-p) = 56(1-0.73) = 15.12 \geq 10$. Therefore, we can approximate this as a normal distribution.

We need to find the expected value and standard deviation of the normal distribution

$E(X) = \mu = np = (0.73)(56) = 40.88$

$Var(x) = \sigma = \sqrt{np(1-p)} = \sqrt{(0.73)(56)(1-0.73)} = 12.9185337$

```
mu <- 0.73 * 56
sd <- sqrt(.73 * 56 * .27)

exactly_40_normal <- dnorm(40, mean = mu, sd =  sd)
```

```r
at_least_40_normal <- pnorm(40, mean = mu, sd = sd, lower.tail = FALSE)
```

The difference between the normal distribution and the binomial distribution probability for $P(X = 40) =$ 0.0026128. The difference between the normal distribution and the binomial distribution for $P(X \geq 40) =$ 0.0634252. .

The normal approximation for exact probabilities are closer than those for $P(X \geq x)$ but still is a good approximation.

d)

$E(X) = \mu = np = (0.73)(56) = 40.88$

I expect about 41 individuals to have at least one dental checkup.

e)

$Var(x) = \sigma = \sqrt{np(1-p)} = \sqrt{(0.73)(56)(1-0.73)} = 12.9185337$

The standard deviation of the number of individuals who will have at least one dental checkup is about 14.

Problem 2

a)

Poisson Distribution: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

Then $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$

$P(X = 0) = \frac{6^0 e^{-6}}{0!} = .0025$

$P(X = 1) = \frac{6^1 e^{-6}}{1!} = 0.0145$

$P(X = 2) = \frac{6^2 e^{-6}}{2!} = 0.045$

$P(X < 3) = 0.0025 + 0.0145 + 0.045 = 0.062$

Want to find $P(X < 3)$

```r
fewer_than_3_tornadoes <- ppois(2, lambda = 6)
```

The probability of seeing less than 3 tornadoes is 0.0619688

b)

Want to find $P(X = 3)$

```r
exactly_3_tornadoes <- dpois(3, lambda = 6)
```

The probability of having exactly 3 tornadoes is 0.0892351

c)

Want to find $P(X \geq 3)$

```
more_than_3_tornadoes <- ppois(3, lambda = 6, lower.tail = FALSE)
```

The probability of having more than three tornadoes is 0.8487961

Problem 3

a)

$H_0 : \mu = 137$

$H_A : \mu > 137$

```
pop_mean <- 128
pop_sd <- 10.2
probability_greater_137 <-  1 - pnorm(137, mean = pop_mean, sd = pop_sd)
```

The probability that a randomly selected American male between 20 and 29 has a systolic blood pressure above 137.0 is 0.188793

b)

```
n <- 50
sample_se <- (10.2/sqrt(50))
prob_less_125 <- pnorm(125, mean = pop_mean, sd = sample_se)
```

The probability that the sample mean for blood pressure of 50 males between 20 and 29 years old will be less than 125 is 0.0187753.

c)

```
z_90 <- qnorm(0.90)

n <- 40

se <- pop_sd/sqrt(40)

percentile_90 <-
  pop_mean + z_90 * se
```

The 90th percentile of the sampling distribution of the sample mean X for a sample size of 40 is 130.0668372

Problem 4

a)

```
sample_mean <- 80
sample_sd <- 10
n <- 40

SE <- sample_sd / sqrt(n)

critical_val <- qt(1 - (0.05/2),df = 39)

margin_of_error <- critical_val*SE

lower_bound <- sample_mean - margin_of_error
lower_bound
```

```
## [1] 76.80184
```

```
upper_bound <- sample_mean + margin_of_error
upper_bound
```

```
## [1] 83.19816
```

b)

We are 95% confident that the mean pulse of young women suffering from Fibromyalgia falls between (76.8018448, 83.1981552)

c)

$H_0 : \mu = 70$ $H_A : \mu \neq 70$

```
mu <- 80
standard_error <- 10
se <- (standard_error/sqrt(40))
critical_val_2 <- qt(1 - (0.01/2), df = 39)

t_value <- (80 - 70) / se

p_value <- 2 * pt(t_value, 39, lower.tail = FALSE)
```

Reject the null hypothesis that $\mu = 70$ because the p-value is small. This means that the mean pulse of young women suffering from Fibromyalgia is not equal to 70.