Max Greene HW9

Question 1

Sample size $n = 400$, proportion estimate $\hat{p} = 0.52$. Use this to find confidence level of interval $(0.5, \infty)$ for population parameter p.

Normal approximation:

$$X \approx N(np, np(1-p)) = N(208, 99.84)P(X > 200) = P\Big(\frac{X-200}{\sqrt{99.84}} > 0\Big) = \boxed{0.50}$$

Question 2

Random sample of 50 chocolates. $\hat{\mu} = 20$mg and $\hat{\sigma} = 4$mg. Find 90, 95, 99 percent confidence intervals for unknown mean μ using normal approximation.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

is approximately $Norm(0, 1)$

```
p_hat <- 0.52
alpha <- c(.90,.95,.99)
z_half_alpha <- qnorm(1-(1-alpha)/2,0,1)
lower <- p_hat-z_half_alpha*sqrt(p_hat*(1-p_hat)/400)
upper <- p_hat+z_half_alpha*sqrt(p_hat*(1-p_hat)/400)
data.frame(alpha=alpha,lower=lower,upper=upper)
```

```
##   alpha     lower     upper
## 1  0.90 0.4789116 0.5610884
## 2  0.95 0.4710401 0.5689599
## 3  0.99 0.4556558 0.5843442
```

Question 3

Simulate sampling distribution of mean for $X_1, X_2, \ldots, X_{50}$, where each X_i is a Poisson r.v. with $\lambda = 2.3$.

```
mean(rpois(50,2.3))
```

```
## [1] 2.52
```

Question 4

Do the following problems from the book.

4.3.22

```
numChild <- 4000
costPerChild <- 1750
Pl80 <- pnorm(80,100,16)
Pg135 <- 1-pnorm(135,100,16)
numChild*costPerChild*(Pl80+Pg135)
```

```
## [1] 840019.6
```

4.3.38

Let $Y_1, Y_2, ..., Y_9$ be random sample of size 9 from normal distribution with $\mu = 2$ and $\sigma = 2$.
Let $X_1, X_2, ..., X_9$ be random sample of size 9 from normal distribution with $\mu = 1$ and $\sigma = 1$. $(Y_i^* = X_i)$
Find $P(\hat{Y} \geq \hat{X})$

First, define another random variable, $D = \hat{Y} - \hat{X}$. Mean and variance of $\hat{Y}$ and $\hat{X}$ are $\mu = 2, \sigma = \frac{2}{\sqrt{9}}$ and $\mu = 1, \sigma = \frac{1}{\sqrt{9}}$, respectively given by the CLT. By combining X, Y the mean and variance of D are $\mu = 1, \sigma = \frac{3}{3}$. Then, for $Y \geq X, D \geq 0$. Find $P(D \geq 0)$

```
pnorm(Inf,1,1)-pnorm(0,1,1)
```

```
## [1] 0.8413447
```

5.3.2

```
data <- c(.61,.7,.63,.76,.67,.72,.64,.82,.88,.82,.78,.84,.83,.82,.74,.85,.73,.85,.87)
trueMean <- 0.80
sampleSD <- sd(data)
sampleMean <- mean(data)

qnorm(c(0.025,0.975),sampleMean,sampleSD)
```

```
## [1] 0.5979400 0.9346916
```

Yes, it is believable. The true average of persons with no lung dysfunction fits well inside the upper and lower bounds of the 95%confidence interval.

5.3.6

(a)

```
pnorm(2.33,0,1)-pnorm(-1.64,0,1)
```

```
## [1] 0.9395943
```

(b)

```
pnorm(2.58,0,1)-pnorm(-Inf,0,1)
```

```
## [1] 0.99506
```

(c)

```
pnorm(0,0,1)-pnorm(-1.64,0,1)
```

```
## [1] 0.4494974
```

5.3.8

The particular interval

$$\left(\bar{y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{y} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

is unique because it shares a mean with the distribution. That is, it is centered exactly on the mean with bounds at the 95% confidence interval bounds.

5.3.22

$n = 350$, "yes"$= 126$ (a) Find 90% confidence interval. 90% confidence interval for the true probability of the binomial variable given by

$$[\frac{126}{350} - z_{\alpha/2}\sqrt{\frac{(126/350)(1-126/350)}{350}}, \frac{126}{350} + z_{\alpha/2}\sqrt{\frac{(126/350)(1-126/350)}{350}}]$$

```r
frac <- 126/350
z <- qnorm(c(0.05,0.95),0,1)
100*(frac+z*sqrt((frac)*(1-frac)/350))
```

```
## [1] 31.77979 40.22021
```

5.4.12

Suppose μ is known, not to be estimated by $\bar{y}$. Show that $\hat{\sigma}^2 = \frac{1}{n}\Sigma_{i=1}^n(\bar{Y_i} - \mu)^2$ is unbiased for μ.

$$\begin{aligned}
E(\hat{\sigma}^2) &= E\big[\frac{1}{n}\Sigma_{i=1}^n(Y_i - \mu)^2\big] \\
&= E\big[\frac{1}{n}\Sigma_{i=1}^n Y_i^2 - 2\mu Y_i + \mu^2\big] \\
&= \frac{1}{n}\big(\Sigma_{i=1}^n E(Y_i^2)\big) - E(\mu^2) \\
&= E(\bar{Y}^2) - \mu^2 \\
&= \sigma^2
\end{aligned}$$

Therefore, $\hat{\sigma}^2$ is unbiased for σ^2

5.4.18

$\hat{\theta}_1$ would be a better estimator, because the variance is lower. Given that $f_Y(y;\theta)$ is symmetric, meaning the max and min values are expected to be equally extreme, the variance of $\hat{\theta}_2$ is greater than the variance of $\hat{\theta}_1$ due to their coefficients.

Yes, this answer makes sense on intuitive grounds. With a greater coefficient, the variance will be greater because the random variables will be more spread out. Another way to think of this is the greater the coefficient to a r.v., the greater the differences (variation) between points becomes.