Joseph Blubaugh
STAT 630
HW09

1) 6.3.1 $\mu = 5$ is not outside of the 95% confidence interval so we do not reject the null hypothesis

```
data = c(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3)
ci = 2.228*(sqrt(.5) / sqrt(10))

# 95% confidence interval
mean(data) - ci; mean(data) + ci
```

```
## [1] 4.381804
```

```
## [1] 5.378196
```

```
# pvalue for estimated mean
z = (abs(mean(data) - 5)) / (.5 / sqrt(10))
1 - pnorm(-z) * 2
```

```
## [1] 0.5521155
```

2) 6.3.2 $\mu = 5$ is not outside of the 95% confidence interval so we still do not reject the null hypothesis

```
data = c(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3)
ci = 2.228*(sqrt(var(data)) / sqrt(10))

# 95% confidence interval
mean(data) - ci; mean(data) + ci
```

```
## [1] 4.38984
```

```
## [1] 5.37016
```

```
# pvalue for estimated mean
z = (abs(mean(data) - 5)) / (var(data) / sqrt(10))
1 - pnorm(-z) * 2
```

```
## [1] 0.5669809
```

3) 6.3.8 $\theta = .65$ is inside the confidence bounds of the Wald Interval, but outside of the bounds of the score confidence interval. For the Wald test we would not reject the null hypothesis. For the Score confidence interval we have enough information to reject the null hypothesis

```
# Wald test with 90% confidence interval
n = 250; prop.y = .62
.62 - 2.65 * sqrt(.62 * (1 - .62) / n); .62 + 2.65 * sqrt(.62 * (1 - .62) / n)
```

```
## [1] 0.5386489
```

```
## [1] 0.7013511
```

```r
# Score confidence
(prop.y + (2.65^2 / 2)/(2 * n) - (2.65/2) *
    sqrt(((prop.y * (1 - prop.y))/n) + ((2.65^2 / 2) / (4 * n^2)))) / (1 + 2.65^2 / n)
```

```
## [1] 0.5700327
```

```r
(prop.y + (2.65^2 / 2)/(2 * n) + (2.65/2) *
    sqrt(((prop.y * (1 - prop.y))/n) + ((2.65^2 / 2) / (4 * n^2)))) / (1 + 2.65^2 / n)
```

```
## [1] 0.6497486
```

4) 6.4.18   All 3 confidence interval estimates seem reasonably close, but the 2 bootstrap estimates appear to be better estimators than straight confidence interval

```r
data = c(3.27, -1.24, 3.97, 2.25, 3.47, -.09, 7.45, 6.20, 3.74, 4.12,
         1.42, 2.75, -1.48, 4.97, 8, 3.26, .15, -3.64, 4.88, 4.55)

# 95% confidence interval
ci = 1.96 * (sqrt(var(data)) / sqrt(20))
mean(data) - ci; mean(data) + ci
```

```
## [1] 1.586368
```

```
## [1] 4.213632
```

```r
# nonparametric bootstrap confidence interval
temp.par = as.numeric()
for (i in 1:1000) {
  temp.par[i] = mean(sample(data, replace = TRUE))
}

ci.boot = 1.96 * (sqrt(var(temp.par)))
mean(data) - ci.boot; mean(data) + ci.boot
```

```
## [1] 1.58914
```

```
## [1] 4.21086
```

```r
# parametric bootstrap interval
ci.boot = 1.96 * (sqrt(var(temp.par)))
mean(temp.par) - ci.boot; mean(temp.par) + ci.boot
```

```
## [1] 1.602197
```

```
## [1] 4.223916
```

5) 6.5.1

$$I(\theta) = Var_\theta(S(\theta|X)) = E_\theta \frac{d^2 l(\theta|X)}{d\theta^2}$$

$$= \frac{d^2(\frac{-n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2)}{d\theta^2}$$

$$= \frac{n}{2\sigma^4}$$

6) 6.5.3

$$l(\alpha|x) = nlog(\alpha) + (-\alpha - 1)\sum_{i=1}^{n}(1 + x_i)$$

$$\frac{dl(\alpha|x)}{d\alpha} = \frac{d(\frac{n}{\alpha})}{d\alpha}$$

$$= \frac{n}{\alpha^2}$$

7) 6.5.4

$$Poisson(\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

$$WaldInterval = \lambda \pm \sqrt{\lambda/n}$$

$\lambda = 11$ is just inside the confidence interval around $\hat{\lambda}$ so we do not reject the null hypothesis

```r
data = c(9, 10, 8, 12, 11, 12, 5, 13, 9, 9, 7, 5, 16, 13, 9, 5, 13, 8, 9, 10)

# Wald Test
mean(data) - 1.96 * sqrt(mean(data) / 20); mean(data) + 1.96 * sqrt(mean(data) / 20)
```

```
## [1] 8.28854
```

```
## [1] 11.01146
```

```r
# Score
mean(data) - abs((mean(data) - 11)) / sqrt(11 / 20); mean(data) + abs((mean(data) - 11)) / sqrt(11/20)
```

```
## [1] 7.82966
```

```
## [1] 11.47034
```

```r
# Simulation
temp = as.numeric()
for (i in 1:1000) {
  temp[i] = mean(sample(data, replace = TRUE))
}

ci.boot = 1.96 * (sqrt(var(temp)))
mean(data) - ci.boot; mean(data) + ci.boot
```

```
## [1] 8.369367
```

```
## [1] 10.93063
```

```
# The Wald Interval has better coverage
```

8) 6.5.5

$$l(\theta|X) = \frac{1}{aN}\sum_{x=i}^{N}X_i$$
$$= \frac{1}{2(27)}43950.75$$
$$\hat{\theta} = 813.9028$$

```
data = c(336.87, 2750.71, 2199.44, 292.99, 1835.55, 1385.36, 2690.52, 710.64, 2162.01,
         1865.47, 2225.68, 3524.23, 2618.51, 361.68, 979.54, 2159.18, 1908.94, 1397.96,
         914.41, 1548.48, 1801.84, 1016.16, 1666.71, 1196.42, 1225.68, 2422.53, 753.24)

theta.hat = 813.9028
theta.hat - 1.645 * (sqrt(var(data)) / sqrt(20)); theta.hat + 1.645 * (sqrt(var(data)) / sqrt(20))
```

```
## [1] 512.2262
```

```
## [1] 1115.579
```

9) 6.5.6

$$l(\theta|X) = \frac{1}{aN}\sum_{x=i}^{N}X_i$$
$$= \frac{1}{1(27)}43950.75$$
$$\hat{\theta} = 1627.806$$

```
data = c(336.87, 2750.71, 2199.44, 292.99, 1835.55, 1385.36, 2690.52, 710.64, 2162.01,
         1865.47, 2225.68, 3524.23, 2618.51, 361.68, 979.54, 2159.18, 1908.94, 1397.96,
         914.41, 1548.48, 1801.84, 1016.16, 1666.71, 1196.42, 1225.68, 2422.53, 753.24)

theta.hat = 1627.806
theta.hat - 1.645 * (sqrt(var(data)) / sqrt(20)); theta.hat + 1.645 * (sqrt(var(data)) / sqrt(20))
```

```
## [1] 1326.129
```

```
## [1] 1929.483
```

They have the same interval range because the variance in the data are the same

10) Additional A   This is an example of a negative binomial distribution where $r = 1$ which simplifies to...

$$P(X) = \binom{x + r - 1}{x} \theta^x (1 - \theta)^r$$
$$= \theta^x (1 - \theta)$$
$$L(\theta|X) = \sum_{i=1}^{n} x_i log(\theta) + log(1 - \theta)$$
$$dl(\theta|X) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{1}{1 - \theta} = 0$$
$$\frac{1 - \theta}{\theta} = \frac{1}{\sum_{i=1}^{n} x_i}$$

11) Additional B   The MLE for σ^2 is ...

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

This assumes that $\bar{x} \neq \sum_{i=1}^{n} x_i$ in which case the estimator would be equal to 0 which makes it inconsistent.

12) Additional C   To show variance is asymptotically normal, we need to show how the estimate approaches $\frac{1}{\sqrt{n}}$ as n increases

$$I(\theta) = \frac{n}{2\sigma^4}$$

```
results = as.numeric()
for (i in 2:50) {
  rv = rnorm(i, 4, 1)
  x = ((mean(rv) - 5)^2)/length(rv)
  results = rbind(results, x)
  }
plot(results, type = "l", col = "blue")
lines(1/2:50, col = "red")
```