Homework 4 Solutions

1. Chapter 8, exercise 3 Suppose X has a binomial distribution with parameters n and p. Then X is approximately normally distributed with mean np and variance np(1-p). For each of the following, answer either A or E, for "approximately" or "exact," respectively:

   (a) the distribution of X is normal - Approximately
   (b) E(X) is np - Exact
   (c) Var(X) is np(1-p) - Exact

2. a. Suppose X is a random variable that follows a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. What is $\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$, $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$, and $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$?

```
pnorm(1) - pnorm(-1)
```

```
## [1] 0.6826895
```

```
pnorm(2) - pnorm(-2)
```

```
## [1] 0.9544997
```

```
pnorm(3) - pnorm(-3)
```

```
## [1] 0.9973002
```

   b. Now suppose that X follows an exponential distribution with $\lambda = 1$. What is $\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$, $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$, and $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$?

```
pexp(2, 1) - pexp(0, 1)
```

```
## [1] 0.8646647
```

```
pexp(3, 1) - pexp(-1, 1)
```

```
## [1] 0.9502129
```

```
pexp(4, 1) - pexp(-2, 1)
```

```
## [1] 0.9816844
```

   c. Now suppose that X follows a uniform distribution on $(0, 1)$. What is $\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$, $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$, and $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$?

```
punif(1/2 + sqrt(1/12), 0, 1) - punif(1/2 - sqrt(1/12), 0, 1)
```

```
## [1] 0.5773503
```

```r
punif(1/2 + 2*sqrt(1/12), 0, 1) - punif(1/2 - 2*sqrt(1/12), 0, 1)
```

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

```r
punif(1/2 + 3*sqrt(1/12), 0, 1) - punif(1/2 - 3*sqrt(1/12), 0, 1)
```

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

3. Suppose that test scores are normally distributed with average score of 80 and standard deviation of 7.

   a. What is the 90th percentile score? Meaning, what is the lowest test score that is larger than 90% of all scores? What is the 95th percentile score?

```r
qnorm(0.9, 80, 7)
```

```
## [1] 88.97086
```

```r
qnorm(0.95, 80, 7)
```

```
## [1] 91.51398
```

   b. For a class of 20 students, what is the probability the average class score is less than 75? The central limit theorem says that the average of 20 values is normal with mean 80 and standard deviation $7/\sqrt(20)$. Therefore the probability that the class average is less than 75 is

```r
pnorm(75, 80, 7/sqrt(20))
```

```
## [1] 0.0007006508
```

4. MENSA is an organization whose members possess IQs in the top 2% of the population.

   a. If IQs are normally distributed, with mean 100 and a standard deviation of 16, what is the minimum IQ required for admission to MENSA?

```r
qnorm(0.98, 100, 16)
```

```
## [1] 132.86
```

   b. If three individuals are chosen at random from the general population what is the probability that all three satisfy the minimum requirement for MENSA?

```r
0.02^3
```

```
## [1] 8e-06
```

5. At a ketchup factory, the amount of ketchup that goes into bottles are supposed to be normally distributed with mean 36 oz. and standard deviation 0.1 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of the bottle goes below 35.8 oz. or above 36.2 oz., then the bottle will be declared defective.

   a. If the process is in control, meaning $\mu = 36$ oz. and $\sigma = 0.1$ oz., what is the probability that a bottle will be declared defective?

```r
1 - (pnorm(36.2, 36, 0.1) - pnorm(35.8, 36, 0.1))
```

```
## [1] 0.04550026
```

b. In the situation of a above, what is the probability that the number of bottles found defective in eight-hour day (i.e. 16 bottles inspectioned) will be zero? This is binomial with p equal to the probability that a single bottle will be defective, as calculated above.

```r
dbinom(0, size = 16, prob = 1 - (pnorm(36.2, 36, 0.1) - pnorm(35.8, 36, 0.1)))
```

```
## [1] 0.4746932
```

c. In the situation of a above, what the probability that the number of bottles found defective in an eight-hour day (16 inspections) will be exactly one?

```r
dbinom(1, size = 16, prob = 1 - (pnorm(36.2, 36, 0.1) - pnorm(35.8, 36, 0.1)))
```

```
## [1] 0.3620521
```

d. If the process shifts so that $\mu = 37$ oz and $\sigma = 0.4$ oz, what is the probability that a bottle will be declared defective?

```r
1 - (pnorm(36.2, 37, 0.4) - pnorm(35.8, 37, 0.4))
```

```
## [1] 0.9785998
```

6. State whether the following statements are true or false and explain why.

   a. The probability that the average of 20 observations will be within 0.5 standard deviation of the population mean exceeds the probability that the average of 40 observations will be within 0.5 standard deviations of the population mean. FALSE
   b. If $\bar{X}$ is the average of $n > 1$ observations, then $\Pr(\bar{X} > 4)$ is larger than $\Pr(X > 4)$ if $X \sim N(8, \sigma)$. TRUE
   c. If $\bar{X}$ is the average of $n > 1$ observations from a $N(\mu, \sigma)$ distribution and c is a positive number, then $\Pr(\mu - c \leq \bar{X} \leq \mu + c)$ decreases as n increases. TRUE

7. Suppose that the number of tickets purchased by a student standing in line at the ticket window for an event follows a distribution with mean $\mu = 2.4$ and standard deviation $\sigma = 2$. Suppose that few hours before the start of one of these matches there are 100 students standing in line to purchase tickets. If only 250 tickets remain, what is the approximate probability that all 100 students will be able to purchase the tickets they desire? Let X_i be the number of tickets that the ith person buys. The total number of tickets bought is $\sum_{i=1}^{100} X_i$. We want $\Pr(\sum_{i=1}^{100} X_i \leq 250)$, which equals $\Pr(\bar{X} \leq 250/100)$ with $\bar{X}$ approximately normally distributed with mean 2.4 and standard deviation $2/\sqrt{100} = 0.2$.

```r
pnorm(2.5, 2.4, 0.2)
```

```
## [1] 0.6914625
```

8. Assuming a population standard deviation of $\sigma = 5$, how large of a sample should you take to estimate the population mean μ with a margin of error not exceeding 0.5 with 95% confidence? The margin of error is $1.959964 \cdot \sigma/\sqrt{n} = 0.5$. Solving for n gives $n = 384.1459$ and therefore we must have at least 385 people (rounding up is necessary).

9. To assess the accuracy of a laboratory scale, a standard weight that is known to weigh 100 grams is repeatedly weighed 10 times. The resulting measurements (in grams) are: 94, 107, 102, 90, 104, 105, 101, 110, 96, and 109. Assume that the weighings by the scale when the true weight is 100 grams is normally distributed. Compute a 95% confidence interval for the population mean μ. Does the data indicate any bias in the scale?

```
obs = c(94, 107, 102, 90, 104, 105, 101, 110, 96, 109)
xbar = mean(obs)
xbar
```

```
## [1] 101.8
```

```
xsd = sd(obs)
xsd
```

```
## [1] 6.629899
```

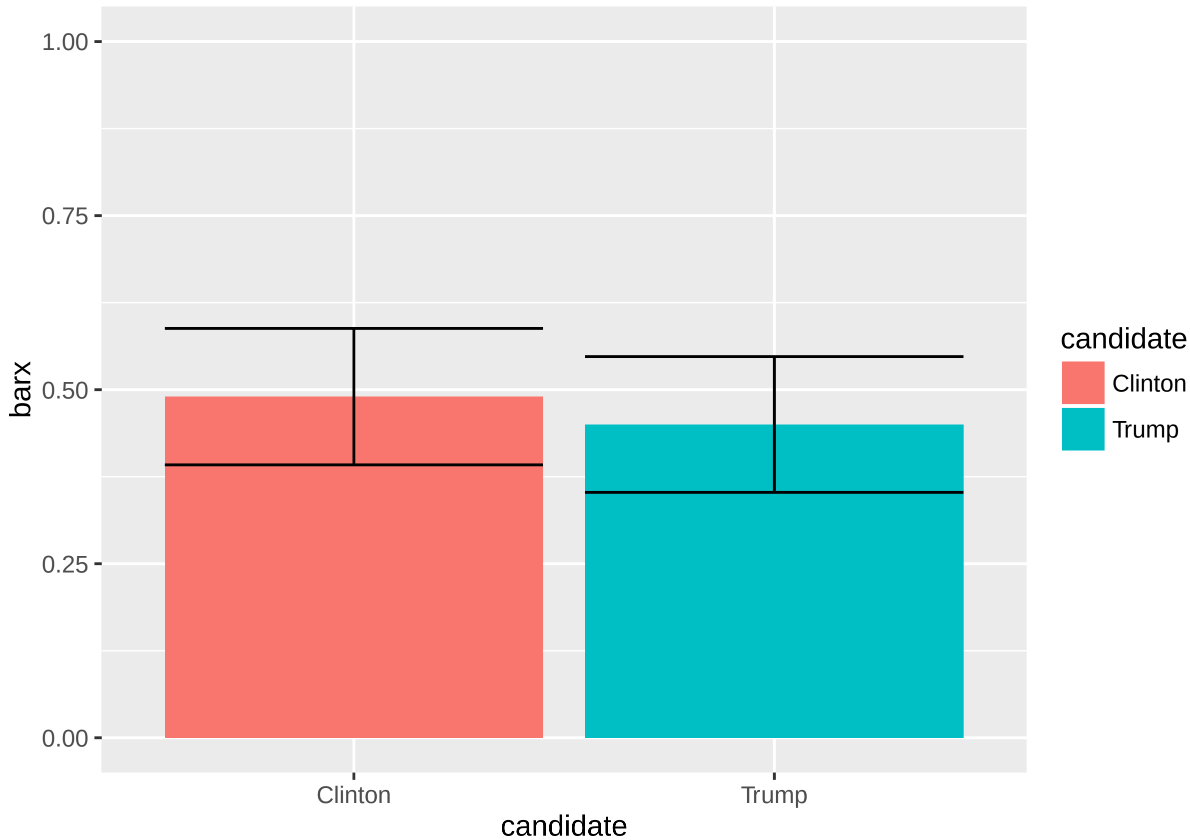
```
c(xbar - qnorm(0.975)*xsd/sqrt(10), xbar + qnorm(0.975)*xsd/sqrt(10))
```

```
## [1]  97.69082 105.90918
```

10. Suppose a poll for the presedential election finds that out of 100 people polled, 49 support Clinton, 45 Trump, and 6 other candidates.

    a. Calculate 95% confidence intervals for the two major candidates and plot them side by side. This can be done cleanly using the package ggplot2. Make a data frame with columns candidate, barx, upperlimit, and lowerlimit and then plot a bar plot with error bars with the following commands, inserting your own numbers where appropriate:

```
#install.packages("ggplot2")
library(ggplot2)
clinton_avg = 0.49
trump_avg = 0.45
clinton_sd = sqrt(0.49*(1 - 0.49))/sqrt(100)
trump_sd= sqrt(0.45*(1 - 0.45))/sqrt(100)
x = data.frame(candidate = c("Clinton", "Trump"), barx = c(clinton_avg, trump_avg),
               upperlimit = c(clinton_avg + qnorm(0.975)*clinton_sd,
                              trump_avg + qnorm(0.975)*trump_sd),
               lowerlimit = c(clinton_avg - qnorm(0.975)*clinton_sd,
                              trump_avg - qnorm(0.975)*trump_sd))
ggplot(x, aes(x = candidate, y = barx, fill = candidate)) +
  geom_bar(position = position_dodge(), stat="identity") +
  geom_errorbar(aes(ymin=lowerlimit, ymax=upperlimit),
                position=position_dodge(.9)) + ylim(0, 1)
```



b. The difference of two normal random variables with means μ_1 and μ_2 and standard deviations σ_1 and σ_2 is also normal with mean $\mu_1 - \mu_2$ and standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$. Use this to calculate a 95% confidence interval for the difference in proportions of the two major candidates. The difference between the means is $0.49 - 0.45 = 0.04$. The standard deviation is

```
sqrt(clinton_sd^2 + trump_sd^2)
```

```
## [1] 0.07052659
```

The 95% confidence interval is

```
c(0.04 + qnorm(0.975)*sqrt(clinton_sd^2 + trump_sd^2),
  0.04 - qnorm(0.975)*sqrt(clinton_sd^2 + trump_sd^2))
```

```
## [1]  0.17822958 -0.09822958
```

11. There is a theory that powerul men have a higher frequency of male children. We can test this by looking at the children of the wives presidents (ignoring children from mistresses, e.g. Warren Harding). From [http://www.infoplease.com/ipa/A0194051.html], the number of sons is 88 and the number of daughters is 65. Compute a 95% confidence interval for the observed fraction of sons, assuming a true population proportion of 50% males.

```
n_sons = 88
n_daught = 65
sons_prop = n_sons/(n_sons + n_daught)
s = sqrt(0.5*0.5/(n_sons + n_daught))
c(sons_prop - qnorm(0.975)*s, sons_prop + qnorm(0.975)*s)
```

```
## [1] 0.4959366 0.6543902
```