Textbook Exercises
6.12, 6.20, 6.38, 6.50, 6.64, 6.70, 6.84, 6.97, 6.120, 6.130, 6.145, 6.150

Computer Exercises
For each R problem, turn in answers to questions with the written portion of the homework. Send the R code for the problem to Katherine Goode. The answers to questions in the written part should be well written, clear, and organized. The R code should be commented and well formatted.

R problem 1 Ideally, a 95% confidence interval will be as tightly clustered around the true value as possible, and will have a 95% coverage probability. When the possible data values are discrete, (such as in the case of sample proportions which can only be a count over the sample size), the true coverage or capture probability is not exactly 0.95 for every \( p \). This problem examines the true coverage probability for three different methods of making confidence intervals.

To compute the coverage probability of a method, recognize that each possible value \( x \) from 0 to \( n \) for a given method results in a confidence interval with a lower bound \( a(x) \) and an upper bound \( b(x) \). The interval will capture \( p \) if \( a(x) < p < b(x) \). To compute the capture probability of a given \( p \), we need to add up all of the binomial probabilities for the \( x \) values that capture \( p \) in the interval. For a sample size \( n \) and true population proportion \( p \), this coverage probability is

\[
P(p \text{ in interval}) = \sum_{x:a(x)<p<b(x)} \binom{n}{x} p^x (1-p)^{n-x}
\]

To compute this in R, you need to find the lower and upper bounds of the confidence interval for each possible outcome \( x \), and add the probabilities of the outcomes that capture \( p \). Here is some code to get you started using the textbook method for an example where \( x = 10 \) and \( p = 0.3 \).

```r
x = 0:10
p.hat = x/10 # will be a vector
se = sqrt(p.hat*(1-p.hat)/10) # also a vector
z = qnorm(0.975)
a = p.hat - z*se # also a vector
b = p.hat + z*se # also a vector
x[ (a < 0.3) & (0.3 < b) ] # x that capture p
```

# [1] 2 3 4 5 6

For each of the following methods, find which outcomes \( x \) result in confidence intervals that capture \( p \) and compute the coverage probability from a sample of size \( n = 60 \) when \( p = 0.4 \).

1. Normal from maximum likelihood estimate, \( \hat{p} = X/n \), SE = \( \sqrt{\hat{p}(1-\hat{p})/n} \), with the interval

\[
\hat{p} \pm 1.96\text{SE}
\]

2. Normal from adjusted maximum likelihood estimate, \( \hat{p} = (X+2)/(n+4) \), SE = \( \sqrt{\hat{p}(1-\hat{p})/(n+4)} \), with the interval

\[
\hat{p} \pm 1.96\text{SE}
\]
3. Within $z^2/2$ of the maximum likelihood loglikelihood. For this method, the file logl.R has a function logl.ci.p() which returns the lower and upper bounds of a 95% confidence interval given $n$ and $x$. You can graph the loglikelihood using glogl.p() for $n$, $x$, and $z = 1.96$ to see if the returned values make sense.

**R Problem 2** Repeat the previous problem, but for $n = 60$ and $p = 0.1$.

**R Problem 3** This problem examines a $t$ distribution with 4 degrees of freedom.

1. Draw a graph of a $t$ distribution with 4 degrees of freedom and a standard normal curve from $-4$ to 4.
2. Find the area to the right of 2 under each curve.
3. Find the 0.975 quantile of each curve.

**R Problem 4** Repeat the previous problem, but for a $t$ distribution with 20 degrees of freedom.

**R Problem 5** Repeat the previous problem, but for a $t$ distribution with 100 degrees of freedom.

Here is some sample code to draw graphs of continuous distributions.

```r
x = seq(-4,4,0.001)
z = dnorm(x)
y.10 = dt(x, df=10)
d = data.frame(x,z,y.10)
require(ggplot2)
## Loading required package: ggplot2

ggplot(d) +
geom_line(aes(x=x,y=y.10),color="blue") + 
geom_line(aes(x=x,y=z),color="red") + 
ylab('density') +
ggtitle("t(10) distribution in blue, N(0,1) in red")
```

![Graph of t(10) distribution in blue, N(0,1) in red.](image)