1. Exercise 9.4 in the text. Do part (a) and assume that the given sample standard deviation of 6.9 is the true standard deviation of the entire population. Replace part (b) with: Is your confidence interval you constructed in (a) exact or approximate? What information did you use about the distribution of the sample mean of student heights to construct your confidence interval?

   Solution:

   Calculate the interval as

   $$174.5 \pm 2.33 * 6.9/\sqrt{50} = (172.23, 176.77).$$

   We do not know the distribution of the data points $X_1, ..., X_n$, so we use the central limit theorem (CLT) to conclude that $\bar{X}_n$ is approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. Our use of the CLT means that the interval is only approximately of 95% confidence, not exactly 95% confidence. That means

   $$P(z_{\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}) \approx 1 - \alpha.$$

   See the probability is approximately $1 - \alpha$, not exactly equal.

2. Exercise 9.8 in the text.

   Solution:

The confidence interval is

$$(\bar{x}_n - 1.96\frac{40}{\sqrt{n}}, \bar{x}_n + 1.96\frac{40}{\sqrt{n}}).$$

The interval is centered at $\bar{x}_n$, so if we want the sample mean to be within 15 seconds of the true mean, we need

$$1.96\frac{40}{\sqrt{n}} \leq 15.$$

Solve this to get $n \geq 28$; remember that n must be a whole number.

3. Exercise 9.10 in the text. Assume that the given sample standard deviation of 7.8 is the true standard deviation of the entire population.

   Solution:

$$79.3 \pm 1.96\frac{7.8}{\sqrt{12}} = (74.89, 83.71).$$

4. Exercise 9.12 in the text. Assume that the given sample standard deviation of 15 is the true standard deviation of the entire population. Additionally, use the confidence interval you construct to test the hypothesis $H_0 : \mu = 250$. What is the Type I error of your test?

   Solution:

$$230 \pm 2.575 * \frac{15}{\sqrt{10}} = (217.79, 242.21).$$

Since 250 is not in the interval, we are confident that the true population mean calories is less than 250. The Type I error is $\alpha = 1\%$.

5. Exercise 10.2 in the text.

   Solution:

   In both cases (a) and (b), she is testing H_0 : The training course is effective in increasing use of seat belts. Logically, the more serious error occurs when the course is truly effective, but our test says it is not effective. In this case, people will not take the course and safety belts use will be lessened. The less serious error is to say the course is effective when it is not. Some people will take the course, and it will waste their time, but ultimately no one will be less safe when driving because of the course.

6. Exercise 10.20 in the text. Assume that the weights are normally distributed and that the given sample standard deviation of 0.24 is the true population standard deviation.

   Solution:

   If H_0 is true, then
   $$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
   and $P(\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \geq -1.645) = 0.95$. But, $\frac{\bar{x}_n-\mu}{\sigma/\sqrt{n}} = \frac{5.23-5.5}{.24/8} = -9$. Since $-9 < -1.645$, it does not seem plausible that $\mu = 5.5$. Reject H_0.

7. Guided `R` exercise. This exercise will help you to understand the Central Limit Theorem (CLT) using computer simulations. Below I have written some `R` code to illustrate the CLT in action. The CLT says that

the sample mean behaves approximately like a normal random variable when the sample size is large, regardless of the distribution of the data. Let's test this out using distributions that are very different from the normal distribution. The first distribution we will use is the Gamma distribution. Run the following code to see a picture of this particular Gamma distribution:

```
x = seq(from = 0, to = 10, by =0.01)
y = dgamma(x,2,4)
plot(x,y,type = 'l')
```

Not very normal, is it? The following code simulates 1000 datasets of size n from this Gamma distribution, computes the sample means of each data set, and creates a histogram of these sample means. This histogram approximates the distribution of the sample means. for large enough n, we should see a histogram that looks like a normal distribution. Change the value of n in the code below until the historgram looks normal.

```
# Simulation of sample means of Gamma data


n=1
xbar = matrix(0,1000,1)
for(i in 1:1000){
sample = rgamma(n,2,4)
xbar[i,1] = mean(sample)
```

```
}
hist(xbar)
```

The uniform distribution is also very non-normal. It is flat and only nonzero on a finite interval. We will try the same experiment as above using the uniform distribution. Look at the plot of the uniform distribution and then change the value of n in the code below until the histogram looks normal.

```
x = seq(from = 0, to = 1, by =0.01)
y = dunif(x,0,1)
plot(x,y,type = 'l')
```

```
# Simulation of sample means of Uniform data
n=1
xbar = matrix(0,1000,1)
for(i in 1:1000){
sample = runif(n,0,1)
xbar[i,1] = mean(sample)
}
hist(xbar)
```

We have seen the normal approximation to the binomial before. It all comes from the CLT. Even though the binomial is discrete and the normal is continuous, the CLT still applies! Try the same experiment with the binomial

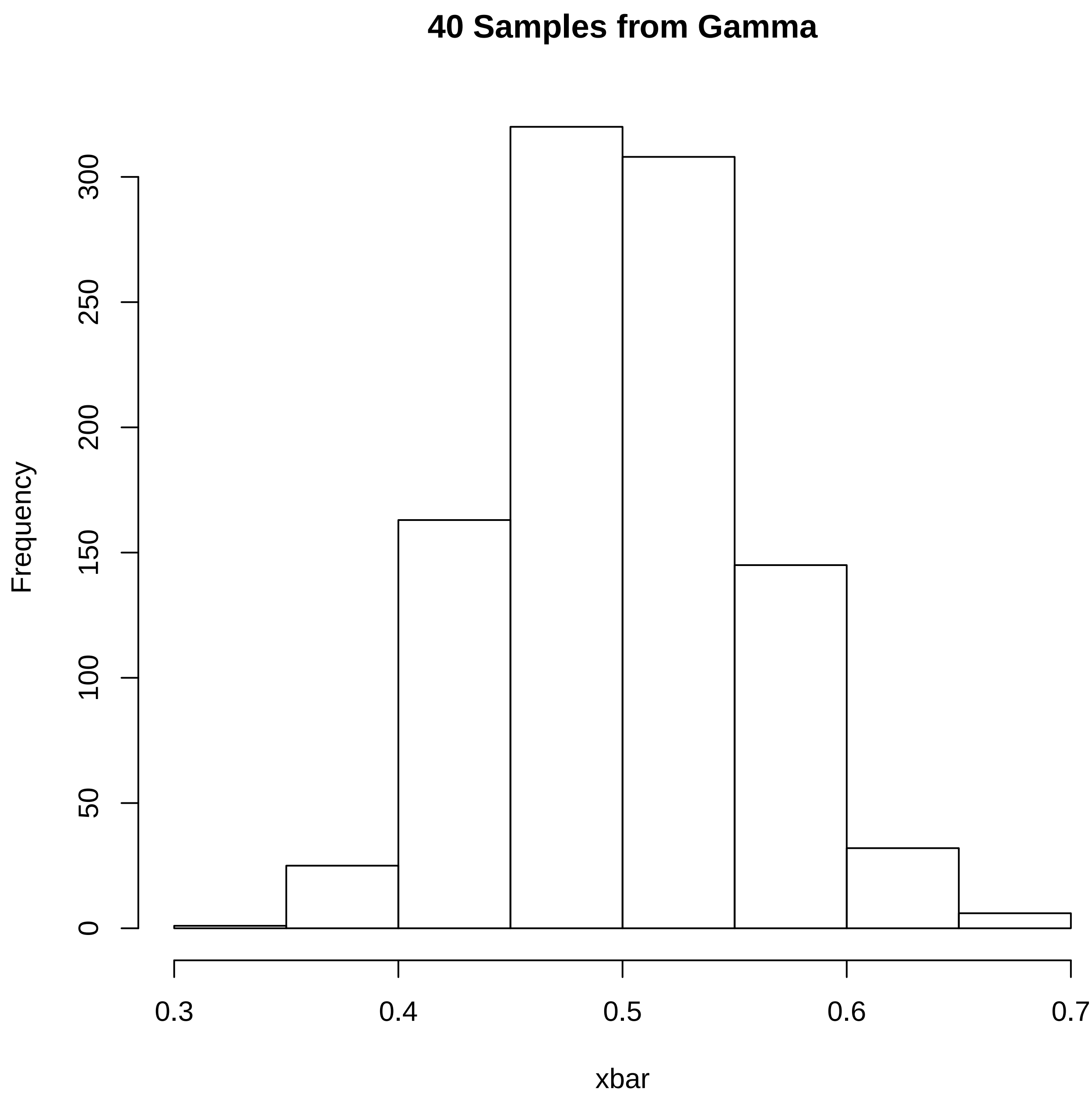
```
x = seq(from = 0, to = 20, by =1)
y = dbinom(x,20,0.1)
plot(x,y,type = 'p')
```

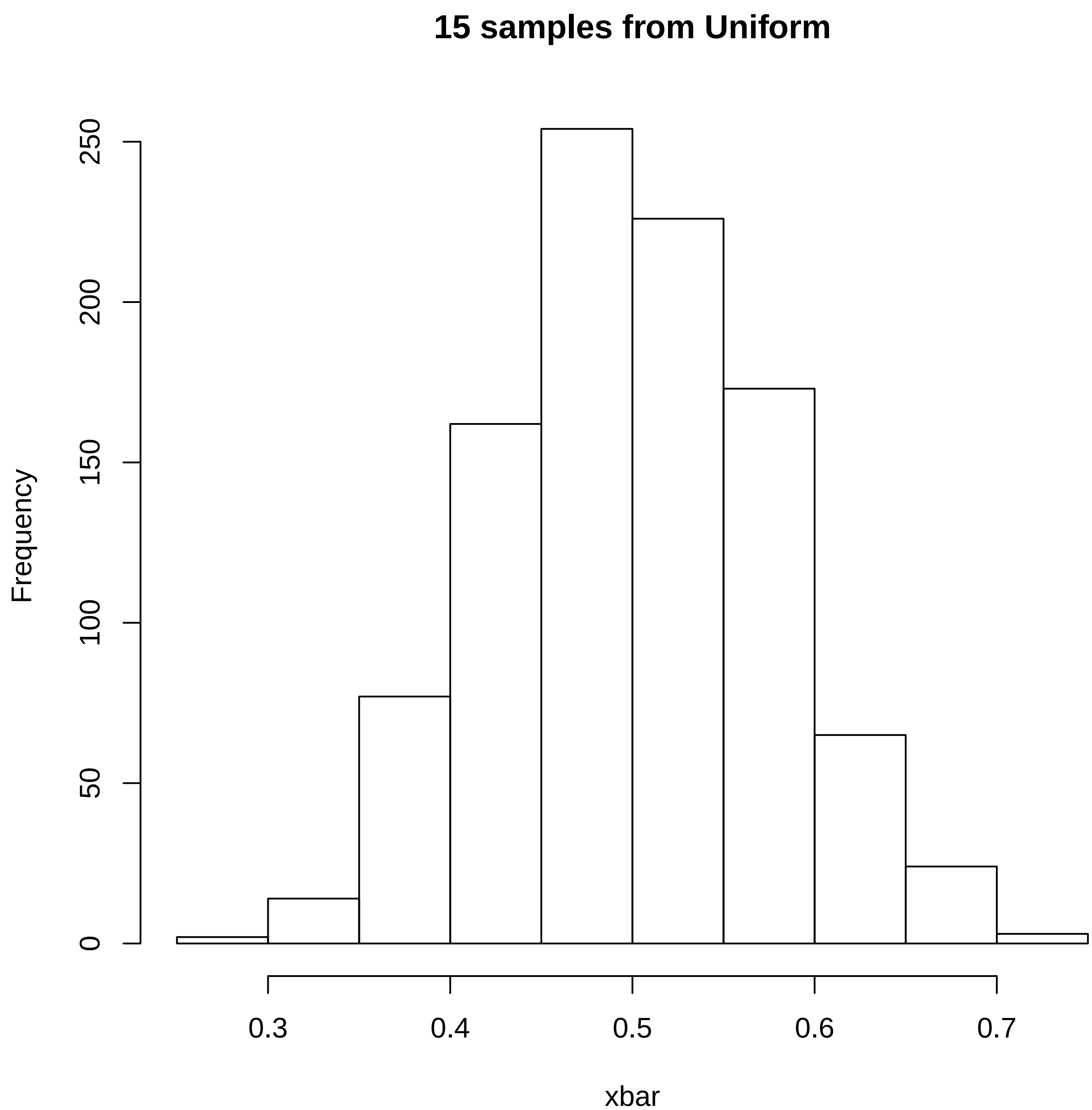
```
# Simulation of sample means of binomial data
n=1
xbar = matrix(0,1000,1)
for(i in 1:1000){
sample = rbinom(n,20,.1)
xbar[i,1] = mean(sample)
}
hist(xbar)
```

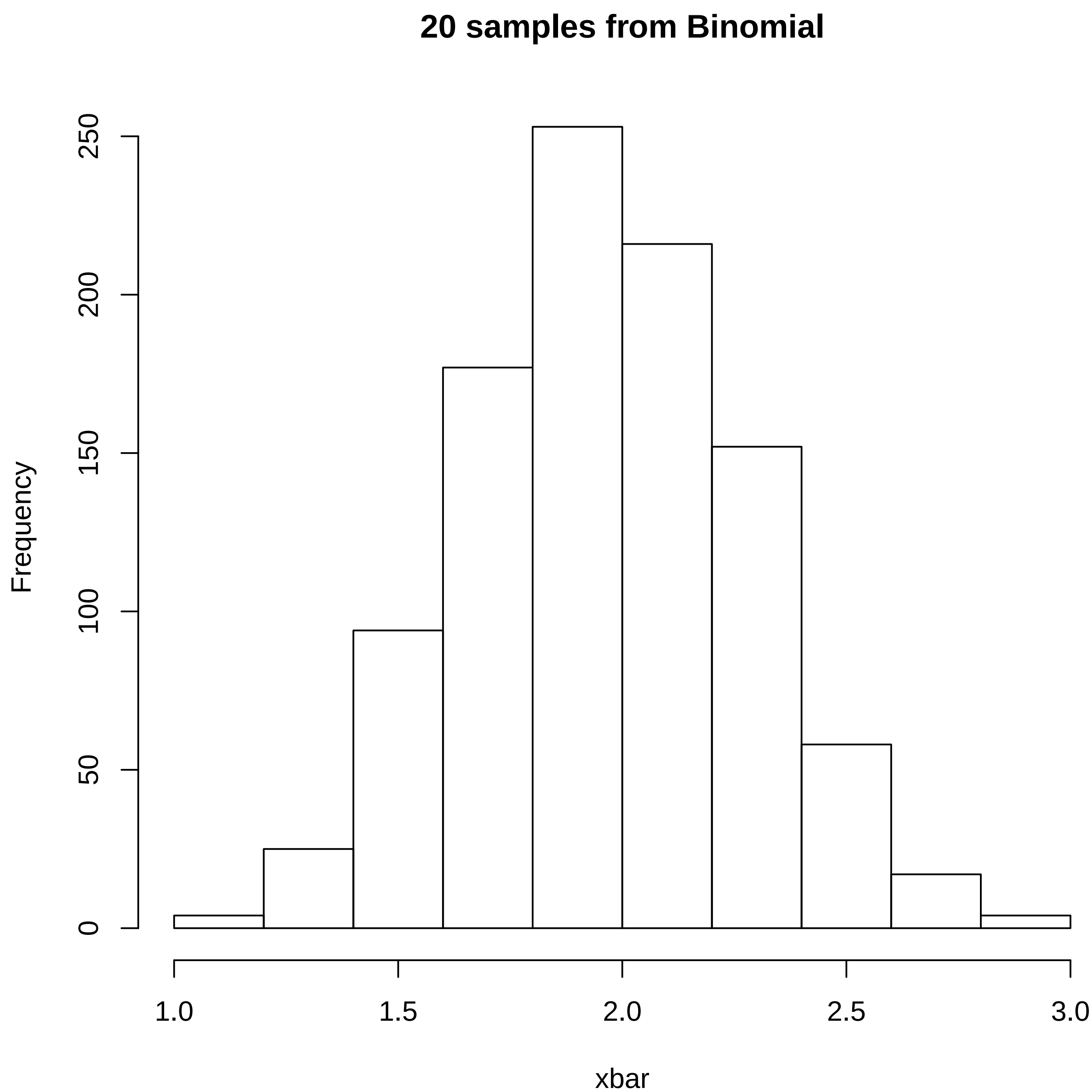
Tell me how large a value of n you needed for each distribution. Your book says CLT works fine for $n \geq 30$. Based on your experiments, do you agree?

Solution:

Based on my simulations, this specific Gamma distribution needed 35-40 samples before it looked "normal". I think the high degree of skewness is what caused a need for a larger sample size. The uniform distribution already looked "normal" for about 15-20 observations. And, the particular binomial I used seemed pretty "normal" even for just 20 observations. Here are some histograms I used from the simulations:


40 Samples from Gamma
Frequency
0
50
100
150
200
250
300
0.3
0.4
0.5
0.6
0.7
xbar


15 samples from Uniform
Frequency
0
50
100
150
200
250
0.3
0.4
0.5
0.6
0.7
xbar


20 samples from Binomial
Frequency
0
50
100
150
200
250
1.0
1.5
2.0
2.5
3.0
xbar