Test 2 will cover the material that was discussed in class and in the homework from chapters 4, 5, and 7. The basic topics are probability, sampling distributions and confidence intervals.

A brief, but not complete summary follows:

Probability We learned how random variables can be used to model a data set through a *random sample*. With this we make precise what we mean by a sample from a population, what a parameter is and what a statistic is.

Some language that we learned: The distribution of a random variable, and the mean and standard deviation of a population.

Two key distributions are the binomial distribution and the normal distribution. Each of these is specified completely by two parameters. These are n and p for the binomial and μ and σ for the normal.

The quantile-normal plot can help us decide if a sample comes from a normal population.

Sampling distributions Once a random sample is understood as n independent RVs with the population distribution then a statistic is just a numeric summary of these. The fact that the statistic depends on a random sample means that a statistic is too random. Hence it is described by a distribution and summarized with a mean and standard deviation.

We learned that the sampling disributions of all of the following are approximately normal (sometimes we need n to be large):

$$\widehat{p} = x/n, \quad \frac{\widehat{p} - p}{\sqrt{p(1-p)/n}}, \quad \frac{\widehat{p} - p}{\sqrt{\widehat{p}(1-\widehat{p})/n}};$$

$$\bar{x} = \frac{1}{n}\sum x_i, \quad \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

However, if the population is normally distributed, then

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\text{observed} - \text{expected}}{\mathsf{SE}}$$

has a t-distribution with $n-1$ degrees of freedom

Confidence intervals We learned about these two $(1-\alpha)100\%$ confidence intervals:

$$\widehat{p} - z^* \mathsf{SE}(\widehat{p}), \quad \bar{x} - t^* \mathsf{SE}(\widehat{p}).$$

The formulas are straightforward, but how we interpret them is subtle.

Some sample problems I might ask would be:

1. If X is binomial with $n = 5$ and $p = 1/4$ find all of the following:

$$E(X), \quad \mathsf{SD}(X), \quad P(X = 3), \quad P(X \leq 3)$$

2. Let Z be a standard normal Find the following:

$$P(Z < 1), \quad P(Z \leq 2.3), \quad P(Z \geq 1.23), \quad P(-1 \leq Z \leq 1/2)$$

3. Again, let Z be a standard normal. Find z for each

$$P(Z \leq z) = .32, \quad P(Z \geq z) - 0.10$$

4. Let Y be a normal random variable with mean 10 and standard deviation 20. Find

$$P(Y > 10), \quad P(Y > 20), \quad P(Y > 31), \quad P(15 < Y < 25)$$

5. Let X_1, X_2, ..., X_16 is random sample for a normal population with mean 10 and standard deviation 20. Find the following

$$P(\bar{x} > 10), \quad P(\bar{x} > 20), \quad P(\bar{x} > 31), \quad P(15 < \bar{x} < 25)$$

6. Suppose waist sizes are normally distributed with a mean of 92 cm and standard deviation of 11cm. Let Y denote a randomly chosen waist, find

   (a) $P(Y \geq 100)$.

   (b) $P(Y \geq y) = 0.80$

7. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a population with mean μ and standard deviation σ. Which of these statements actually makes sense?

   (a) The sample mean is the population mean.

   (b) The mean of the sample mean is the population mean.

   (c) the standard deviation of the sample mean is the population standard deviation.

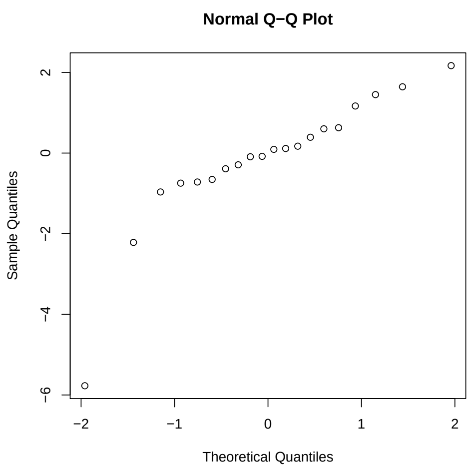
   (d) The distribution of the sample mean (for large n) is not the population distribution but the normal distribution.

8. A random sample of 50 people finds 46% agree with some proposition. Is .5 in the 95% CI based for π based on this sample?

   How large would n need to be so that if $\widehat{p} = .46$ you would be sure that $p = .5$ is not in the 95% CI for p given by $\widehat{p}$.

9. A random sample has the following summaries:

```
> length(x)

[1] 20

> mean(x)

[1] -0.1742689

> sd(x)

[1] 1.656451

> qqnorm(x)
```

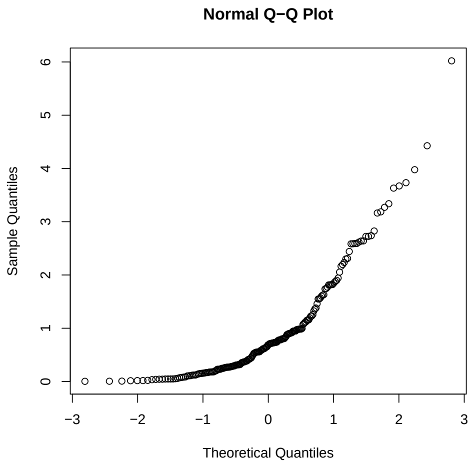


If this data set is appropriate for finding a 95% CI for μ based on $\bar{x}$ say why and then find the CI. Otherwise, say why not.

10. A random sample has the following summaries:

```
> length(x)

[1] 200

> mean(x)

[1] 0.9622028

> sd(x)

[1] 0.9940424
```

```
> qqnorm(x)
```



If this data set is appropriate for finding a 95% CI for μ based on $\bar{x}$ say why and then find the CI. Otherwise, say why not.