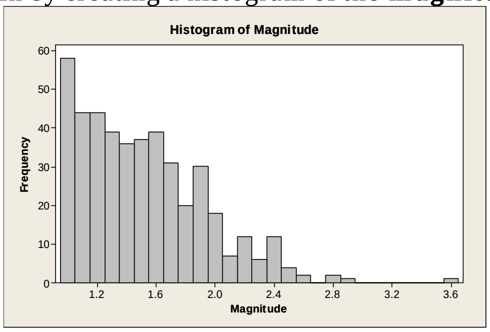
Chapter 23: Inference about Means

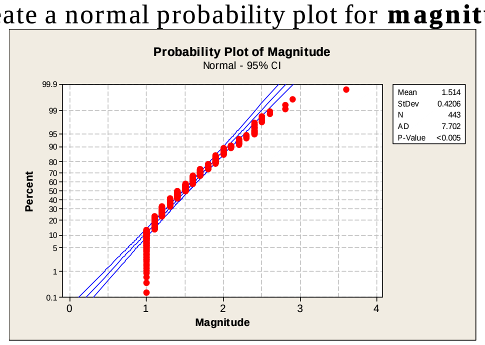
1. As the sample size increases, the t-distribution gets closer to ____NORMAL DISTRIBUTION_____.

2. Get the earthquake dataset from http://www.canyons.edu/faculty/morrowa/140/datasets/. This dataset consists of earthquakes in California from a seven day period in 2009.

   a) Begin by creating a histogram of the **magnitudes**.



   b) Create a normal probability plot for **magnitudes**.



   c) Does the data pass the nearly normal condition?

   The data is definitely not normal. The histogram shows us that the distribution is skew right, with definite outliers. The probability plot confirms these observations since we do not observe a straight line.

   Despite this, though, the sample size is quite large (443). Therefore, we are still okay to pass the nearly normal condition.

   d) Check the rest of the conditions for inference on the mean.

   Since earthquakes have aftershocks, examining earthquake data over a given time period likely means that the sample is not independent. Despite that, though, we hope that this sample is representative of all earthquakes in Southern California.

   Disclosing the lack of normality (in conjunction with the large sample size) and the origin of the data is enough to proceed with inference.

   e) Does this data set provide evidence that the earthquakes in California are below magnitude 2.0? Conduct an appropriate test. Write out all steps.

   $H_0: \mu = 2$
   $H_A: \mu < 2$

   Assumptions checked above.

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	T	P
Magnitude	443	1.5144	0.4206	0.0200	1.5474	-24.30	0.000

   Test Statistic: –24.30
   P-value: 0.000

   Because the p-value is so low, we reject the null hypothesis.
   Therefore, there is significant evidence to suggest that the mean earthquake magnitude is below 2.0.

f) If your conclusion is wrong, what type of error have you made?

Type I error – we concluded that earthquake magnitude is less than 2.0 on average, but in reality it might not be.

g) For members of California who are scared of earthquakes, which type of error is worse?

Type I Error – If someone is scared of earthquakes, they do not want you to be wrong about the magnitude being less than 2.

h) Construct (and interpret) an appropriate confidence interval. Use the interval to verify your conclusion in (e).

Since no significance was given (and the p-value was so low), we can use a confidence level of our choosing.

```
Variable     N    Mean   StDev  SE Mean        95% CI
Magnitude  443  1.5144  0.4206   0.0200  (1.4752, 1.5537)
```

We are 95% confident that the mean magnitude for earthquakes in California is between 1.5 and 1.6. Because both bounds are below 2.0, we can conclude that the mean magnitude is below 2.0, which coincides with our conclusion in (e).

3. In a sample of 20 individuals, the mean income was $60,000 with a standard deviation of $10,000.

a) Construct (and interpret) a 95% confidence interval for the mean income of Americans.

```
N   Mean  StDev  SE Mean      95% CI
20  60.00  10.00     2.24  (55.32, 64.68)
```

We are 95% confident that the mean income is between $55,320 and $64,680.

It is important to note that while Minitab is happy to give us answers, we know nothing about the data collection or about the shape of the distribution. We have no reason to believe any of the necessary conditions are met, and our **confidence interval is relatively meaningless**.

b) Write hypotheses to conduct an appropriate test to see if the mean income of Americans was above $43,000.

H_0: μ = $43,000

H_A: μ > $43,000

c) Using the confidence interval, test your hypotheses in (b).

The interval is above $55,000, and is therefore well above $43,000. Therefore, we would reject H_0.

d) The median income for Americans in 2004 was approximately $43,000. What does this suggest about the shape of the distribution of income? (Right skew, left skew, or symmetric?)

The mean is more than the median. This suggests that the distribution is right skew.

e) What is it about income that would lead to that kind of a distribution?

Two things:

1) There is a boundary at 0. Incomes can be as far positive as possible, but NEVER below 0.

2) While most Americans make little, there are a few (Bill Gates) pulling the mean higher. The median is resistant to outliers and will be close to 'all Americans'.

f) If you were a government official, trying to ease fears that income was decreasing, would you want to disclose the shape of the distribution? Explain.

The confidence interval provides a higher estimate of income compared to what the 'typical' American is earning. Therefore, if you wanted people to think Americans earn more than they do, you would report the confidence interval for the mean, but hide the fact that it is worthless.

4. Many people believe that students gain weight as freshmen. Suppose we plan to conduct a study to see if this is true.

a) Describe a study design that uses paired samples.

Select a random sample of freshmen. Weigh them when college starts in the fall. Weigh them again at the end of spring. Examine the differences in their weights.

b) Describe a study design that uses independent samples.

Weigh a random sample of freshmen when college starts in the fall. Choose a new sample to weight in the spring. Compare the average weights of the two samples.