Stat 218 - Day 12
Confidence Intervals: Checking technical conditions, assessing normality

- *t*-procedure for estimating a population mean μ: $\bar{y} \pm t^* \frac{s}{\sqrt{n}}$
  - This procedure is valid as long as:
    - The sample is random
    - Either the population is normal or the sample size is large ($n \geq 30$)
  - The interpretation is:
    - We are C% confident that the CI contains the actual value of the population mean μ.
    - This "confidence" means that if random samples were repeatedly drawn from the population and a CI constructed from each sample, then in the long run C% of those intervals would succeed in capturing the population mean μ.

Assessing Normality
How do we decide if sample data appear to come from a normally distributed population? Three ways:
1) See if a histogram of the sample data looks roughly bell-shaped.
2) See if the sample data roughly follow the empirical (68/95/99.7) rule.
3) See if a **normal probability plot** follows a general linear pattern.

A normal probability plot compares the ordered sample data values to what would be expected for the ordered values from a standard normal distribution. A perfect normal distribution would produce a normal probability plot that follows a straight line. (`Graph> Probability plot`)

Examples:
For each of the following variables, produce a histogram and a normal probability plot. Then comment on whether a normal distribution is a reasonable model for the distribution of that variable. For the last one, create separate probability plots for the two groups (choose the "multiple" option).

- Head breadths of Etruscan skulls (`etruscans.mtw`)

- Rower weights (`rowers04.mtw`)

- Body temperatures (`BodyTemps.mtw`)

What to do if the data are clearly not normally distributed?

Transformation: Take some function of the data.

For example, with a distribution that is skewed to the right, a log or square root transformation pulls in the upper tail and makes the distribution more symmetric.

Example: Cloud seeding

The data in `CloudSeeding.mtw` are from an experiment in a plane either injected silver iodide into a cloud (seeding) or not. The measurements are the rainfall amounts (in acre-feet) over the appropriate region in the next 24 hours.

(a) Examine graphical displays to determine whether a normal model would be appropriate for the rainfall amounts.

(b) Take the log (base 10) of the rainfall amounts:

```
MTB> let c5=logt(c3)
```

(c) Examine graphical displays to determine whether a normal model would be appropriate for these transformed rainfall amounts.

Example: *Price is Right*

We'll examine data on a sample of 208 prizes used on the game show *The Price is Right* from several years ago.

(a) Identify the observational units and the variable. Is the variable categorical or quantitative?

(b) Identify the population and the sample.

(c) Identify (in words and with a symbol) the parameter to be estimated.

(d) Use Minitab (`gameshow.mtw`) to determine a 95% CI for the population mean (`Stat> Basic statistics> 1-sample t…`).

(e) Are the technical conditions required for the validity of the procedure satisfied here? Explain.

(f) Do 95% of the sample prices fall in this interval? Explain why this makes sense.

(g) Would you expect 95% of the prices in the population to fall within this interval? Explain.

Example: UFO sighters
In a study of characteristics of people who claim to have had intense experiences with UFO's, one variable measured was their IQ. In the sample of 25 people, the mean IQ was 101.6 and the standard deviation was 8.9.

(a) Use these sample statistics to construct a 90% confidence interval for μ. Then verify this with Minitab (`Stat> Basic statistics> 1 sample t, summarized data`).

(b) Write a sentence describing what μ represents here.

(c) Write a sentence interpreting the CI.

(d) Write a sentence explaining what the phrase "90% confidence" means here.

(e) Does this CI include the value 100? Would you say that it's plausible that the mean IQ among people who claim to have had intense UFO experiences is 100? Explain.

(f) Can you check whether the technical conditions that underlie the validity of the *t*-procedure are satisfied here? Explain.

Example: House prices
Suppose that you want to estimate the mean price of a house in San Luis Obispo County. Some students collected data on a random sample of eight houses taken from realestate.com. The sample data, in thousands of dollars, are: 255, 349, 399, 469, 545, 649, 799, 1195.

(a) Construct a 95% CI for the population mean house price.

(b) Do you think that the *t*-procedure is valid in this situation? Explain.

Example: Body temperatures
Recall that a sample of 130 healthy adults had their body temperatures (in degrees Fahrenheit) recorded. The sample mean was found to be 98.25 degrees, and the sample standard deviation was calculated to be 0.733 degrees.

(a) Calculate the standard error of the sample mean.

(b) If researchers want to conduct a follow-up study and have the standard error equal to .04 degrees, how many subjects should they sample?

(c) If they want to cut that standard error in half, will sampling twice as many people accomplish that? Explain.

(d) If they want a 95% CI to have a margin-of-error of about .10 degrees, about how many subjects should they sample?