***Curriculum Topic IV- STATISTICAL INFERENCE***

***Chapters 10-15***

Statistical inference is the process of generalizing the results of sample to a larger population (or generalizing about cause and effect in the case of experiments). It requires that students combine all the topics they have already learned, collecting data, analyzing data, and probability.

**Conditions for inference**

The goal of inference is to make a generalization from a sample to a larger population. Due to sampling variability, the rules of probability must be used, so that any conclusions are stated with a reference to likelihood. In order for these probability rules to apply, however, certain characteristics of the sample and population must be present. These characteristics are different depending on the type of measurements being made and sometimes the size of the sample. These characteristics are called the conditions for inference, and students should check for compliance with these conditions in any inference setting (hypothesis tests and confidence intervals).

**Confidence Intervals**

A random sample drawn from a population can be used to make estimates about some of the characteristics of that population. For numerical variables, the mean value in a sample often provides a good point estimate of the mean of a population. To account for sampling variability, a more informative report from the sample is a confidence interval, which includes many plausible values for the population mean, assuming proper sampling has occurred. We prefer interval estimates to point estimates because they are more likely to be correct!

The language of confidence intervals is difficult for students. Many students confuse the “interpretation of the interval” with the “interpretation of the confidence level.” Let’s assume that a sample produces a mean of 20, satisfies the conditions needed for inference, and a calculation produces a 95% confidence interval that includes values from 16 to 24.

A proper interpretation of the interval would be: “I am 95% confident that the interval from 16 to 24 captures the true mean.”

A proper interpretation of the confidence level would be: “In the long run, 95% of intervals calculated in this manner will capture the true mean.”

A common incorrect interpretation of a confidence interval is “95% of the time the true mean will be between 16 and 24.” This implies that the true mean is a random variable instead of a fixed, population parameter.

To get full credit on the AP exam, the answer to a confidence interval question should include the following 3 parts:

1. State the type of interval that will be calculated and confirm that the conditions for this interval have been met.

2. Calculation of the interval. It is generally a good habit to reproduce the formula for the interval and indicate what values are being substituted into this formula instead of relying solely on the TI-83.

3. Interpretation of the interval in context.

**Hypothesis Tests**

A test of hypotheses, or significance test, is a method for using data from a sample to decide between two competing hypotheses about a population characteristic. All hypothesis tests have a similar form, and to get full credit on the AP exam, students should be sure to attend to these 4 parts:

1. State the problem in the form of two mutually exclusive hypotheses. The skeptical null hypothesis expresses the conservative view that the conjecture of some difference or change is not true. The null hypothesis will be assumed to be true, unless the result from the sample is so different from what is expected as to be very unlikely to occur due to random sampling variability. If this happens, the null hypothesis will be rejected in favor of the alternative hypothesis. The statements of the null and alternative hypotheses should be always in the context of the problem setting.

2. State which type of test will be done, and confirm that the conditions for this test have been met.

3. Carry out the mechanics of calculating the test statistic and p-value for the appropriate test. It is generally a good habit to reproduce the formula for the test statistic and indicate what values are being substituted into this formula instead of relying solely on the TI-83.

4. Based on the results of the calculations, two conclusions are possible. I will either fail to reject the null hypothesis (if the results might reasonably be due to random variation) or reject the null hypothesis (if random variation is unlikely to explain the results). The conclusion should be about the population, stated in the specific context of the problem. The student should clearly link the conclusion to the specific numbers in the calculation, usually the p-value (probability that the result is a random variation).

**Errors in significance testing and power**

Having made a conclusion in a significance test, there are two possibilities: you are either right or you are wrong. Because the decision is based on probability calculations and random variation, the likelihood of a correct decision can sometimes be determined.

If the null hypothesis is true, but you have rejected it (due to a somewhat unlikely sample), then you have made a Type I error. The probability of this error can be determined – it’s , the cutoff value for statistical significance.

If the null hypothesis is false, but you have failed to reject it (since your results were within what might be explained by random variation), then you have made a Type II error. Calculating the probability of this type of error is much more difficult, since this probability depends on the results of the sample as well as unknown values from the population. This probability is identified as .

The two probabilities,  and  are in opposition to one another. You have control over , but if you make it smaller, then  increases, and vice-versa. Part of the art of statistics is balancing these two competing probabilities. To help in this, calculations of  are sometimes made with assumed values for the population parameters. What makes this hard for students is that each of the assumed values that you might pick results in a different value for .

The power of a significance test is defined as the probability of correctly rejecting the null hypothesis when it is false. This is the exact opposite of making a Type II error, so the power is defined as 1 - . Like , the power depends on the unknown characteristics of the population. Power calculations are best seen as evaluating the sampling or experimental procedure, and not as evaluating the particular results obtained.

Students should understand the concepts of Type II error and power, but they do not have to be able to do any calculations with power.

**IV. Statistical Inference**

You must be able to decide which statistical inference procedure is appropriate in a given setting. Working lots of review problems will help you.

You need to know the difference between a population parameter, a sample statistic, and the sampling distribution of a statistic.

**On any hypothesis testing problem:**

1. State hypotheses in words and symbols.
2. Identify the correct inference procedure and verify conditions for using it.
3. Calculate the test statistic and the *P*-value (or rejection region).
4. Draw a conclusion in context that is directly linked to your *P*-value or rejection region.



**On any confidence interval problem:**

1. Identify the population of interest and the parameter you want to draw conclusions about.
2. Choose the appropriate inference procedure and verify conditions for its use.
3. Carry out the inference procedure.
4. Interpret your results in the context of the problem.

You need to know the specific conditions required for the validity of each statistical inference procedure -- confidence intervals and significance tests.

**Be familiar with the concepts of Type I error, Type II error, and Power of a test.**



**Type II error:** Accepting a null hypothesis when it is false.

**Power of a test:** Probability of correctly rejecting a null hypothesis
Power = 1 - P(Type II error).

You can increase the power of a test by increasing the sample size or increasing the significance level (the probability of a Type I error).