AP Statistics – Chapter 8 Notes: Estimating with Confidence

8.1 – Confidence Interval Basics

Point Estimate
A ***point estimator*** is a statistic that provides an estimate of a population parameter. The value of that statistic from a sample is called a ***point estimate***.

The Idea of a Confidence Interval
A ***C% confidence interval*** gives an interval of plausible values for a parameter. The interval is calculated from the data and has the form: **point estimate ± margin of error**

The difference between the point estimate and the true parameter value will be less than the margin of error in C% of all samples.

The ***confidence level C*** gives the overall success rate of the method for calculating the confidence interval. That is, in C% of all possible samples, the method would yield an interval that captures the true parameter value.

Interpreting Confidence Intervals
To interpret a C% confidence interval for an unknown parameter, say, "We are C% confident that the interval from _____ to _____ captures the actual value of the [population parameter in context]."

Interpreting Confidence Levels
To say that we are *95% confident* is shorthand for "If we take many samples of the same size from this population, about 95% of them will result in an interval that captures the actual parameter value."

8.2 – Estimating a Population Proportion

Conditions for Inference about a Population Proportion

- **SRS** - The data are a simple random sample (SRS) from the population of interest.
- **Normality** - Counts of successes and failures must be 10 or more.
- **Independence** - The population size is at least 10 times greater than the sample size

Standard Error of a Sample Proportion $\hat{p}$ is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

One-Proportion z-interval
The form of the confidence interval for a population proportion is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Sample size for a desired margin of error
To determine the sample size (n) for a given margin of error m in a 1-proportion z interval, use formula

$$n = p^*(1-p^*)\left(\frac{z^*}{m}\right)^2$$

where p* = 0.5, unless another value is given.

Remember, that we will always round up to ensure a smaller margin of error.

8.3 – Estimating a Population Mean

Conditions for Inference about a Population Mean

- **SRS** - Our data are a simple random sample (SRS) of size *n* from the population of interest. This condition is very important.
- **Normality** - Observations from the population have a normal distribution or the sample size is large ($n \geq 30$).
- **Independence** - Population size is at least 10 times greater than sample size

Standard Error

When the standard deviation of a statistic is estimated from the data, the result is called the standard error of the statistic. The standard error of the sample mean is

$$\frac{s}{\sqrt{n}}$$

The form of the confidence interval for a population mean with $n-1$ degrees of freedom is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

Paired Differences t-interval

To compare the responses to the two treatments in a paired data design, apply the one-sample *t* procedures to the observed differences.

For example, suppose that pre and post test scores for 10 individuals in a summer reading program are:

Subject	1	2	3	4	5	6	7	8	9	10
Pre-test	25	31	28	27	30	31	22	18	24	30
Post-test	28	30	34	35	32	31	26	16	28	36
Difference	3	-1	6	8	2	0	4	-2	4	6

We would use the data in the differences row and perform one-sample *t* analysis on it.

Robustness of the t Procedures

Except in the case of small samples, the assumption that the data are an SRS from the population of interest is more important than the assumption that the population distribution is normal.

- *Sample size less than 15*. Use *t* procedures if the data are close to normal. If the data are clearly nonnormal or if outliers are present, do not use *t*.
- *Sample size at least 15*. The *t* procedures can be used except in the presence of outliers or strong skewness.
- *Large samples*. The *t* procedures can be used even for clearly skewed distributions when the sample is large, roughly $n \geq 30$