AP Statistics - some things to remember:

1. Statistical inference (hypothesis tests, estimation and confidence intervals) based on data from observational studies uses a SAMPLE selected RANDOMLY from a POPULATION to make inferences about the population.
2. Hypotheses in tests of means or proportions are statements about unknown population parameters μ or p, NOT sample statistics $\bar{x}$ or $\hat{p}$.
3. Cause and effect relationships can NOT be inferred from observational studies.
4. Cause and effect relationships CAN be inferred from well designed experiments.
5. An experiment usually involves volunteers who are not randomly selected; the volunteers are randomly ASSIGNED to two or more groups (a CONTROL group that receives no treatment or a the standard existing treatment, and one or more TREATMENT groups).
6. Hypothesis tests and confidence intervals involving MEANS are based on the assumption that the sample mean $\bar{x}$ is exactly or approximately NORMALLY distributed. This is true if the sample is selected from a normal population or if the sample size is large enough (say 30), in which case the CENTRAL LIMIT THEOREM tells us that $\bar{x}$ is approximately normal even if the population is not normal.
7. If we need to check that a sample has been selected from a normal population, a quick and easy way is to use the normal probability plot on the TI-84+; this is the final type of statplot in the statplot menu: store your sample values in a list, create this statplot and then graph it using ZOOM 9. If the points appear close to collinear, you can assume the sample came from a normal population. The alternative method for checking normality is to produce a dotplot on paper and check that it is unimodal, and roughly symmetric with no or few outliers.
8. Hypothesis tests and confidence intervals involving PROPORTIONS are based on the assumption that the sample proportion $\hat{p}$ is exactly or approximately NORMALLY distributed. This is true if $np \geq 10$ and $nq \geq 10$. As we don't know p and q, this condition becomes $np_0 \geq 10$ and $nq_0 \geq 10$ in the case of a hypothesis test in which the null hypothesis is H_0: $p = p_0$, and it becomes $n\hat{p} \geq 10$ and $n\hat{q} \geq 10$ (the number of successes and the number of failures in the sample must both be at least 10) otherwise.

8\. Don't confuse means and proportions. Means are average MEASUREMENTS (weights, heights etc); proportions are obtained by COUNTING "successes" and dividing by the number of trials.

9\. Chi-squared tests are based on actual COUNTS (frequencies), NOT percentages or proportions. The condition that EXPECTED frequencies must be at least 5 does NOT apply to OBSERVED frequencies.

10\. In all hypothesis tests for means or proportions, the null hypothesis includes an equals sign (=); the alternative hypothesis includes an inequality sign ($\neq$ if it is a two-tailed test, < or > if it is a one-tailed test).

11\. A symmetrical confidence interval can be used to perform a two-tailed test, not a one-tailed test.

12\. The initial conclusion of a hypothesis test should clearly state "we have sufficient evidence to reject the null hypothesis" if the p-value is less than the significance level (α) or "we do not have sufficient evidence to reject the null hypothesis" if the p-value is greater than the significance level. **You should also state the conclusion in the context of the problem.**

13\. You may use the words "significant" or "significantly" in the CONCLUSION of a hypothesis test but NOT in the hypotheses. The result is significant, at the stated significance level, if the null hypothesis is rejected. For example in a one-tailed test of a single mean, you might write "The mean test score is significantly GREATER than 80." if you rejected $H_0 : \mu = 80$ in favor of $H_1 : \mu > 80$. However, in a two-tailed test, you would write "The mean test score is significantly DIFFERENT from 80." if you rejected $H_0 : \mu = 80$ in favor of $H_1 : \mu \neq 80$.

14. When interpreting a 95% confidence interval for μ, write" I am 95% confident that the true value of the population mean lies in this interval."; you could also write "If I repeated this procedure a large number of times, the true value of the population mean would lie in about 95% of the intervals." Avoid using the word "probability" when interpreting a confidence interval.

15. Types of error: **Type 1 Error**: This can only occur if the null hypothesis is true. A Type 1 error occurs when a true null hypothesis is mistakenly rejected. The probability of a Type 1 Error = α = the significance level,.

    Type 2 Error: This can only occur if the alternative hypothesis is true. A Type 2 error occurs when a false null hypothesis is mistakenly not rejected. The probability of a Type 2 Error = β; this varies depending on the true value of the population parameter that is the subject of the test.

16. The **POWER** of a hypothesis test = $1-\beta$ = the probability of NOT making a Type 2 Error. You can think of the power of a test as measuring the ability of the test to detect a true alternative hypothesis. **You can increase the power of a test by increasing the sample size. You can also increase the power of a test by increasing the significance level**. (α and β are inversely related; if one increases, the other decreases, provided all other things remain the same. So it follows that if α increases then β decreases and $1-\beta$ increases.)