Statistical Inference

Confidence Interval	Significance Test

A ____ is a formal procedure for comparing ______________________ ______________ with a claim (or ____________________) whose truth we want to assess. The claim is a statement about a _____________________ such as population proportion or population __________.

The results of a significance test are given in terms of a ______________________ that measure how well the data and claim __________.

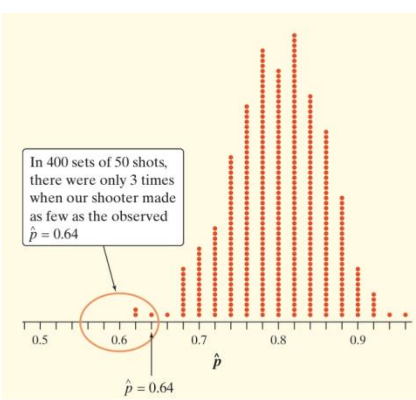
<u>Stating a Hypothesis</u>:

A high school basketball player claims to be an 80% free-throw shooter.

His claim: p = _______ (This is the claim we seek evidence against, we call it the NULL HYPOTHESIS H_o)

Our belief: p ________ (This is the claim we hope to be true instead, we call it the ALTERNATIVE HYPOTHESIS H_a)

The reasoning behind SIGNIFICANCE TESTS: What can we conclude about the player's claim based on sample data? Let's say that he makes 32/50 shots.



a) How would you simulate the results of 50 shots for a shooter who makes?

b) Determine how strong the evidence AGAINST the player's claim is by calculating the probability that he would make as few as 32 out of 50 free throws if he really makes 80% in the long run.

***An outcome that is unlikely to happen if the H_o were true is good evidence that H_o is not true

Definition: *Null Hypothesis and Alternate Hypothesis.*

The claim tested by a statistical test is called the _________________ (H_o). The test is designed to assess the strength of the evidence _____________ the _______ hypothesis. Often the _____ hypothesis is a statement of "____________________".
The claim about the population that we are trying to find evidence ______ is the _______________ hypothesis (H_a).

**A hypothesis is formed before the data is collected.

The H_a is ________________________ if it states that a parameter is larger than the H_o value or it states that the parameter is smaller than the null value. It is _________________ if it states that the parameter is different from the null hypothesis value. (could be larger or smaller)

CYU: Pg.532

- P-value:
  The probability, computed assuming H_o is true, that the statistic (   ) would take a value as extreme as or more extreme that the one actually observed (   ) is called the p-value of the test. The smaller the p-value, the stronger the evidence against H_o provided by the data.

α = Significance Level

What is a 'small p-value'? We can compare the P-value with a fixed value that we regard as decisive. This value is the significance level, ____. If we choose _____________, we are requiring that the data give evidence against H_o so strong that it would happen less than 5% of the time just by chance when H_o is true.

- If the p-value is smaller than alpha, we say that the data are statistically significant at level α. In that case, we reject the null hypothesis and conclude that there is convincing evidence in favor of H_a. (statistically significant = not likely to happen just by chance)

If p-value $< \alpha \rightarrow$ reject $H_o \rightarrow$ conclude H_a (in context)

If p-value $\geq \alpha \rightarrow$ fail to reject $H_o \rightarrow$ cannot conclude H_a (in context)

Writing Conclusions

Since the P-value (   ) is less than α =' ', we reject H_o. There is sufficient evidence to conclude that (H_a in context).

Or

Since the P-value (   ) is greater than α =' ', we fail to reject H_o. There is not sufficient evidence to conclude that (H_a in context).

**The most commonly used significance level is ________________.

Example: Tasty Chips

For his second semester project in AP Statistics, Zenon decided to investigate if students at his school prefer name-brand potato chips to generic potato chips. He randomly selected 50 students and had each student try both types of chips, in random order. Overall, 34 of the 50 students preferred the name-brand chips. Zenon performed a significance test using the hypotheses:

$$H_0 : p = 0.5 \qquad\qquad H_a : p > 0.5$$

where p = the true proportion of students at his school that prefer name-brand chips. The resulting *P*-value was 0.0055.

Problem: What conclusion would you make at each of the following significance levels?

(a) $\alpha = 0.01$

(b) $\alpha = 0.001$

Errors:

	H_o is true	H_o is false (H_a is true)
Reject H_o	TYPE 1 ERROR	Correct Conclusion
Fail to reject H_o	Correct Conclusion	TYPE 2 ERROR

See Example "Perfect Potatoes" (pg.538 and 539)
CYU: Pg.539

Definition: The _______________ of a test against a specific alternative is the probability that the test will reject H_o at a chosen significance level α when the specified alternative value of the parameter is true.

- the higher the power, the more sensitive the test is
- the power of a test against a specific alternative value of the parameter is between 0 and 1.
- A power close to zero means the test has almost no chance of detecting that H_o is false.
- A power close to one means the test is very likely to reject H_o when H_o is false.
- The significance level of a test is the probability of reaching the wrong conclusions when the null hypothesis is true.
- The power of a test to detect a specific alternative is the probability of reaching the right conclusion when that alternative is true.

Power and Type II Error:

The power of a test against any alternative is 1 minus the probability of a Type II Error for that alternative.

Power is affected by:

- Sample size:
- Significance Level:

**Increasing the sample size helps us be more certain that the difference between the observed proportion of successes and the hypothesized proportion of successes is not due to sampling variability.

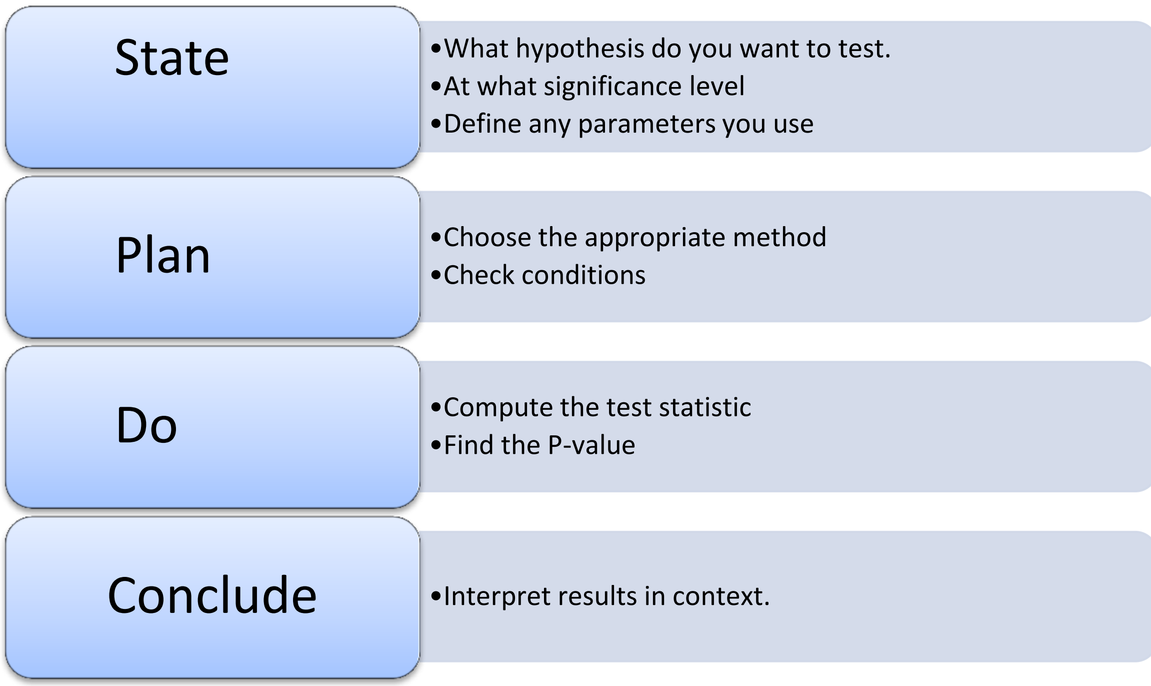
**It is easier to reject a hypothesized proportion of successes when there is a greater difference between the hypothesized proportion of successes and the actual proportion of successes. That is the power of a test will be higher when the true value of a parameter is farther from the hypothesized value of a parameter.

A significance test uses sample data to measure the strength of evidence against H_o.

Test Statistic: (measures how far a sample statistic diverges from what we would expect if the null hypothesis H_o were true, in standard units)

Formula:

Significance Tests: A Four-Step Process



Example: One Potato, Two Potato

The potato-chip producer has just received a truckload of potatoes from its main supplier. Recall that if the producer determines that more than 8% of the potatoes in the shipment have blemishes, the truck will be sent away to get another load from the supplier. A supervisor selects a random sample of 500 potatoes from the truck. An inspection reveals that 47 of the potatoes have blemishes. Carry out a significance test at the $\alpha = 0.10$ significance level. What should the producer conclude?

CYU: Pg.555

Example: Nonsmokers (A two-sided test)
According to the CDC, 50% of HS students have never smoked a cigarette. Taeyeon wonders whether this national result holds true in his large, urban high school. For his class project, Taeyeon surveys an SRS of 150 students from his school. He gets responses from all 150 students, and 90 say that they have never smoked a cigarette. What should Taeyeon conclude? Give appropriate evidence to support your answer.

CYU: Pg.558

Significance Test and Confidence Intervals

The result of a significance test helps us decide whether to reject or fail to reject H_o. But, what is the true proportion p? A confidence interval helps us answer this question.

Formulas:

Test statistic: Confidence Interval:

CYU: Pg.561

Confidence intervals and significance tests for a population proportion p are based on z-values from the standard Normal distribution. Inference about a population mean μ uses a t-distribution with n-1 degrees of freedom, except in the rare case when the population standard deviation is known.

Test Statistics (mean):

See Example: Better Batteries (pg.566-567)

CYU: Pg.570

Example: Healthy Streams (one sample t-test)
The level of dissolved oxygen (DO) in a stream or river is an important indicator of the water's ability to support aquatic life. A researcher measures the DO level at 14 randomly chosen locations along a stream. Here are the results in milligrams per liter (mg/l):

A DO level below 5 mg/l puts aquatic life at risk.

4.53	5.04	3.29	5.23	4.13	5.50
4.83	4.40	5.42	6.38	4.01	4.66
2.87	5.73	5.55			

a) Can we conclude that aquatic life in this stream is at risk? Carry out a test at the $\alpha = 0.05$ significance level to help you answer this question.
b) Given your conclusion in part (a), which kind of mistake –a Type I error or a Type II error –could you have made? Explain what his mistake would mean in context.

CYU: Pg.574

Example: Juicy Pineapple (A two-sided test)
At the Hawaii Pineapple Company, managers are interested in the sizes of the pineapples grown in the company's fields. Last year, the mean wight of the pineapples harvested from one large field was 31 ounces. A new irrigation system was installed in this field after the growing season. Managers wonder whether this change will affect the mean weight of future pineapples grown in the field. To find out, they select and weigh a random sample of 50 pineapples from this year's crop. The minitab output below summarizes the data (see pg. 574)

Descriptive Statistics: Weight (oz)

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Weight (oz)	50	31.935	0.339	2.394	26.491	29.990	31.739	34.115	35.547

a) Determine whether there are any outliers. Show your work.
b) Do these data suggest that the mean weight of pineapples produced in the field has changed this year? Give appropriate statistical evidence to support your answer.
c) Can we conclude that the new irrigation system caused a change in the mean weight of pineapples produced? Explain your answer.

CYU: Pg.577

Comparative Studies: Paired t-procedures

Paired Data: Study designs that involve making two observations on the same individual, or one observation on each of two similar individuals
**Used in comparative studies and very common.

Example: Is Caffeine Dependence Real?

Researchers designed an experiment to study the effects of caffeine withdrawal. They recruited 11 volunteers who were diagnosed as being caffeine dependent to serve as subjects. Each subject was barred from coffee and other substances with caffeine for the duration of the experiment. During one two-day period, subjects took capsules containing their normal caffeine intake. During another two-day period, they took placebo capsules. The order in which subjects took caffeine and the placebo was randomized. At the end of each two-day period, a test for depression was given to all 11 subjects. Researchers wanted to know whether being deprived of caffeine would lead to an increase in depression.

Results of a caffeine deprivation study

Subject	Depression (C)	Depression (P)	Difference (P – C)
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	- 1

Carry out a test to investigate the researchers' question.