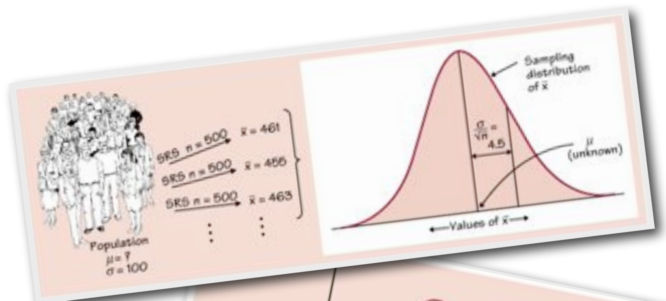
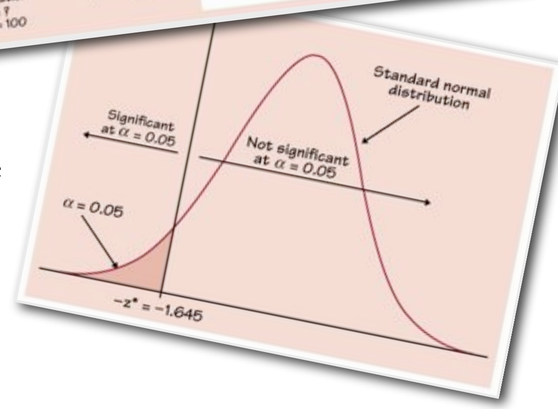
Chapter 10

Introduction to Inference





When we select a sample, we want to infer some conclusion about the population that the sample represents. In this chapter, we will be introduced to the two most common types of formal statistical inference: Confidence Intervals and Tests of Significance. Both types of inference are based on the sampling distributions of statistics. The purpose of this chapter is to describe the reasoning used in inference--we will study specific procedures in later chapters.

Introduction to Inference:

- 10.1: Estimating with Confidence
- 10.2: Tests of Significance
- 10.3: Making Sense of Statistical Significance
- 10.4: Inference as a Decision

AP Statistics Chapter 10: Introduction to Inference

"A statistical analysis, properly conducted, is a delicate dissection of uncertainties, a surgery of suppositions."
~M.J. Moroney

Tentative Lesson Guide					
Date		**Stats**	**Lesson**	**Assignment**	**Done**
Mon	1/29	10.1	Estimating with Confidence	Rd 535-541 **Do 1-4**	
Tues	1/30	10.1	Confidence Intervals	Rd 543-554 **Do 8-10, 20-21, 24-25**	
Wed	1/31	10.2	Tests of Significance	Rd 559-566 **Do 27-32**	
Thu	2/1	10.2	Hypotheses - Procedure	Rd 567-581 **Do 38-40,46-48**	
Fri	2/2	10.3	Using Significance Tests	**Practice Sheet**	
Mon	2/5	**Quiz**	**Quiz 10.1-10.2**	Rd 586-592 **Do 57-61**	
Tues	2/6	10.4	Inference as a Decision	Rd 593-598 **Do 66-68**	
Wed	2/7	10.4	TypeI,II Errors - Power	Rd 599-602 **Do 71-72**	
Thu	2/8	**Rev**	**Review Chapter 10**	Rd 606-608 **ReviewSheet**	
Fri	2/9	**Ex**	**Exam Chapter 10**	**Online Quiz 10 Due**	

Note:
The purpose of this guide is to help you organize your studies for this chapter.
The schedule and assignments may change slightly.

Keep your homework organized and refer to this when you turn in your assignments at the end of the chapter.

Class Website:
Be sure to log on to the class website for notes, worksheets, links to our text companion site, etc.

http://web.mac.com/statsmonkey

Don't forget to take your online quiz!. Be sure to enter my email address correctly!
http://bcs.whfreeman.com/yates2e

My email address is:

jmmolesky@isd194.k12.mn.us

Chapter 10 Objectives and Skills:

These are the expectations for this chapter. You should be able to answer these questions and perform these tasks accurately and thoroughly. Although this is not an exhaustive review sheet, it gives a good idea of the "big picture" skills that you should have after completing this chapter. The more thoroughly and accurately you can complete these tasks, the better your preparation.

Estimating with Confidence

- ☐ C% confidence Interval for a Parameter: An interval computed from sample data by a method that has probability C% of producing an interval containing the true value of the parameter.
- ☐ A confidence interval for an unknown population mean μ, calculated from a sample of size n with mean x(bar), has the form x(bar) plus/minus z*('sigma'/sqrt(n)) where z* is obtained from the normal distribution table and 'sigma' is the standard deviation of the population.
- ☐ The margin of error formula has the form z*('sigma'/sqrt(n)). The margin of error actually determines the length of the confidence interval.
- ☐ Determine a sample size necessary for a given margin of error.

Tests of Significance

☐ The reasoning of a test of significance is as follows: Suppose a null hypothesis is true. If we were to repeat our sampling numerous times, what would the chances be that we would observe a sample value as extreme as the one that was observed?

(1) A null hypothesis is basically a hypothesis of no change. Basically, if one does not reject a null hypothesis, then the test results are not statistically significant.

(2) Failing to reject a null hypothesis at a low level of significance is not strong evidence that it is true. It simply means that it is not unreasonable to assume that the hypothesis is true.

(3) Rejecting a null hypothesis is equivalent to saying that test statistic is statistically significant.

(4) The null and alternate hypotheses are both stated in terms of population parameters, not sample statistics. You are attempting to use sample statistics to come to reasonable conclusions about population parameters.

(5) A P-value is the probability that one would obtain a statistic as "extreme" as that which was calculated from the sample.

(6) The level of significance for a test is usually set beforehand. If a P-value is smaller than the level of significance, then the test statistic is statistically significant.

(7) A statistical test could be 2-tail or 1-tail. Which one is used depends on the purpose and nature of the test. If both positive and negative deviations from a parameter are important, then one should use a 2-tail test. If only positive (or negative) deviations are important, then one should use a 1-tail test.

Inference as a Decision

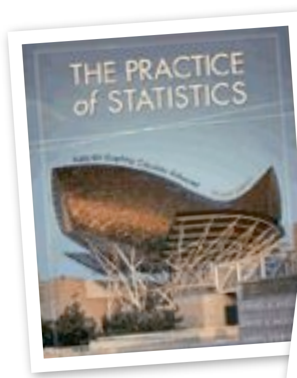
☐ Real-life situations frequently involve Type I and Type II errors. Consider the legal world and a null hypothesis "This accused man is innocent." A Type I error would be determining the man is guilty when he is innocent.. A Type II error would involve declaring the man innocent when he is guilty. In real world situations, one must often decide which error type is more important to minimize.

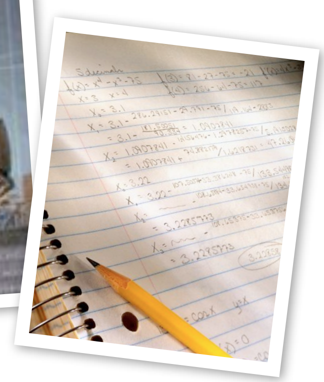
Things to remember:

(1) A Type I error can only occur when a null hypothesis is true.

(2) A Type II error can only occur when a null hypothesis is false.

(3) The Power of a test is 1 - probability (Type II error). This is the probability that you correctly reject a false null hypothesis.





10.1: Estimating with Confidence - m&m Intervals

"Statistical Inference":

☑

☑

Suppose we want to know what proportion of m&m's are RED. Theoretically, we could gather all the m&m's in the world and calculate the proportion. Or, considering that may take a while, we could take a sample of the population and base our estimate of the population proportion on the sample proportion. However, as we learned in the last chapter, due to sampling variability there's a good chance our sample proportion will not be the same as the population proportion. Since we don't know whether or not our sample proportion is equal to the population proportion, we can construct a **Confidence Interval** for our sample proportion that can be used in estimating the true population proportion.

1) Get two bags of m&m's from your teacher. This is your SRS. **Sample Size** n=___________

2) Calculate the proportion of RED m&m's in your SRS. **Sample Proportion** $\hat{p}$=___________

3) Sample proportions will vary from sample to sample--the sampling distribution of $\hat{p}$ will be approximately normal given certain conditions. However, we don't know what the standard deviation of that distribution is because we don't know π. However, we can *estimate* it from our sample using the

   Sample Standard Deviation $s=\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ =_______________

4) Since the sampling distribution of $\hat{p}$ is approximately normal, we know that 95% of sample proportions should be within 2 standard deviations of the true population proportion π. So, to estimate a confidence interval for π, we simply take the $\hat{p}$ from our SRS and construct an interval that stretches 2 standard deviations above and below $\hat{p}$. We are 95% confident the true π falls between and $\hat{p}$-2s and $\hat{p}$+2s.

95% Confidence Interval: We are 95% confident the true π falls between _____ and _____.

There is no way to know whether or not your interval contains the true population proportion, but we do know that 95% of all confidence intervals constructed in this manner will contain the true π That is, the method we used to construct the interval will succeed in capturing the true parameter 95% of the time. *Important Note*: The "95%" has nothing to do with probability...it is an indication of our *confidence* in the procedure used to estimate the parameter...hence the term "confidence interval."

The true proportion of RED m&m's is π=____________. Did your interval capture it?

5) Sketch this confidence interval on the board above the number line provided. Indicate your $\hat{p}$ with a dot and draw two whiskers to represent your interval. Do most of the class' intervals overlap? If so, what values are contained in the overlapping intervals? Did any intervals miss the true π?

10.1: Confidence Intervals

As we learned in Chapter 9, because of sampling variability, the statistic calculated from a sample is rarely equal to the true parameter of interest. Therefore, when we are trying to estimate a parameter, we must go beyond our statistic value to construct a reasonable range that captures the true parameter value.

Confidence Interval:

☑

☑

Sampling Distribution of p-hat Proportions	Sampling Distribution of x-bar Means
Conditions:	**Conditions:**

Confidence Interval for π Single Proportion	Confidence Interval for μ Single Mean (*sigma known*)
Conditions:	**Conditions:**
If an SRS of size n is chosen from a population with unknown proportion π, then a level C Confidence Interval for π is:	**If an SRS of size n is chosen from a population with unknown mean μ, then a level C Confidence Interval for μ is:**

Common Critical Values for Confidence Levels:

Confidence Level	Single Tail Area	z^*
90%		
95%		
99%		

10.1: Estimating with Confidence - Interval Behavior

When we estimate a parameter, we would like high confidence, but also a small margin of error.

Margin of Error =

What are some ways in which we could lower our margin of error?

☑

☑

☑

Inference Toolbox: *Confidence Intervals*

To Construct a Confidence Interval for a Parameter:

☑

☑

☑

☑

<u>Example</u>:

A manufacturer of high-resolution video terminals must control the tension of the viewing screen to avoid tears and wrinkles. The tension is measured in millivolts (mV). Careful study has shown that when the process is operating properly, the standard deviation of the tension reading is 43mV.

1. Given a sample of 20 screens, with $\bar{x} = 306.3\ mV$, from a single day's production, construct a 90% confidence interval for the mean tension μ of all the screens produced on this day. Follow your 4 steps.

2. Suppose the manufacturer wants 99% confidence rather than 90%. Using your data from problem 1, construct a 99% confidence interval. How does it compare to the 90% CI? Why?

3. Company management wants a report of the mean screen tension for the day's production accurate within ±5 mV with 95% confidence. How large a sample of video monitors must be measured to comply with this request?

10.2: Tests of Significance

We have learned that Confidence Intervals can be used to estimate a parameter. Often in statistics we want to use sample data to determine whether or not a claim or hypothesis about a parameter is plausible. A test of significance is a procedure in which we can use sample data to test such a claim. We will focus on tests about a population mean μ in this chapter. We'll study proportions and means in more detail in the next few chapters.

The Reasoning of Tests of Significance

The reasoning of tests of significance, like confidence intervals, is based on a consideration of sampling distributions. Consider the following situation.

According to the article "Credit Cards and College Students: Who Pays, Who Benefits?" the credit card industry asserts at most 50% of college students carry a credit card balance from month to month.

The **Parameter of Interest** in this problem is the proportion of college students who carry a balance month to month:

π=____

The industry claims 50% of students carry a balance. This claim is called the **Null Hypothesis**:

H_o: ___________

Suppose we think the actual proportion is higher. This would be an **Alternative Hypothesis:**

H_a: ___________

To test the claim, suppose we take a SRS of 310 college students and find that 217 carry a balance. Our Sample Proportion, p-hat, is equal to:_______________

What does our sample proportion suggest about π? Supposing π really did equal 50%, would it be possible to observe our p-hat in a sample of 310 students? How likely would it be?

What would the **Sampling Distribution** of p-hat look like for samples of size 310 if $\pi = 0.50$? Be sure to check your conditions before sketching the sampling distribution!

Calculate the probability of observing p-hat = 217/310 = ________ assuming $\pi = 0.50$.

This probability is called a **P-value**. P-value=__________ **Significance Level**: ____________

If the P-value is small, we have evidence to suggest the null hypothesis may be false. If the P-value is large, then we should not be surprised by our statistic value and we have no reason to question the null hypothesis.

Conclusion:

10.2: Inference Activity

Our goal in this activity will be to determine whether or not trial size bags of m&m's weigh as much as their advertised claim suggests. According to the label, the bags should contain _______oz or _______g.

To test the claim that μ=_____ g, we will need to collect data from an SRS of trial size m&m bags. With your classmates, carefully weigh 14 bags and record the data below. We'll assume weights of m&m trial size packages are normally distributed in the population with standard deviation = 0.25 g.

Enter the sample data below:

Calculate the sample mean $\bar{x}$ =__________

What is the population characteristic of interest? That is, what are we testing?

What are your null and alternative hypotheses?

We need to consider the sampling distribution of $\bar{x}$ …be sure to check conditions for normality. Sketch the sampling distribution of $\bar{x}$ and indicate your observed sample value.

Calculate the likelihood of observing $\bar{x}$ or something more extreme…this should be a **"two-tailed"** test. Why?

P-Value=______________ How does this compare to our significance level alpha=______.

Conclusion:

10.1-10.2: Introduction to Inference Practice

1. Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate the average number of unoccupied seats per flight. An SRS of 225 flights shows a sample average of 11.6 seats per flight. Suppose σ=4.1 and create a 90% confidence interval for the true average number of unoccupied seats per flight. Interpret this interval in the context of the problem.

2. Suppose you wish to conduct a study on the average length of a hospital stay. You would like to estimate the average length to within 0.5 days of the true mean at 90% confidence. Assuming σ=3.7, how many patients should you randomly select? What if you wanted 95% confidence?

3. Historically, science textbooks have taught students that the Earth is 93 million miles from the Sun. In *Experimentation and Measurement*, WJ Youden reported the following measurements of the average distance between the Earth and Sun, including Newcomb's original measurement from 1895, in millions of miles:

93.28	92.83	92.91	92.87	93.00	92.91	92.84	92.98
92.91	92.87	92.88	92.92	92.96	92.96	92.81	

Do these data provide significant evidence at the 1% level to suggest the true average distance is *not* 93 million miles? Assume σ=.112.

4. The average human body temperature is thought to be 98.6° with σ=.555. An SRS of 10 males produced an average temperature of 97.88°. What does this sample suggest about the commonly accepted body temperature? A different SRS of 10 females produced an average body temperature of 98.52°. What can you conclude about female body temperatures?

5. The effects of drugs and alcohol on the nervous system have been the subject of considerable research. Suppose a neurologist is testing the effect of a drug on response time by injecting 100 rats with a dose, subjecting each to a stimulus, and recording the response time. The average response time for the 100 injected rats was 1.05 seconds. Assuming the mean response time for a rat that has not been injected with the drug is 1.2 seconds with σ=.5, test the hypothesis that the drug has an effect on response time.

6. Spinifex pigeons, one of the few bird species that inhabit the desert of Western Austrailia, rely on seeds for food. The article "Field metabolism and water requirements of spinifex pigeons (Geophaps plumifera) in Western Austrailia" reported the following Minitab analysis of the weight of seeds (in grams) in the stomach contents of spinifex pigeons. Use the analysis to construct and interpret a 95% confidence interval for the average weight of seeds in this type of pigeon's stomach and compare your results to the hypothesis that the mean seed amount is 1g for all spinifex pigeons.

```
TEST OF MU = 1.000 VS MU N.E. 1.000
            N    MEAN       STDEV   SE MEAN  T      P VALUE
WEIGHT      16   1.373      1.034   0.258    1.44   0.17
```

10.3: Making Sense of Statistical Significance

Significance tests are widely used in reporting the results of research in many fields. Carrying out a test of significance is often very simple. It is your job to explain the test as completely as possible *in the context of the situation*. We will focus on tests about a population mean μ in this chapter. We'll study proportions and means in more detail in the next few chapters.

Inference Toolbox: ***Test of Significance***
To Test a Claim about an unknown Population Parameter: ☑ ☑ ☑ ☑ ☑ ☑

To determine the significance at a set level, alpha, one needs to calculate the P-value for the observed statistic. Careful consideration of the test statistic, z, can also be used to determine significance:

One-Tailed $\mu>k$	**One-Tailed** $\mu<k$	**Two-Tailed** $\mu \neq k$

When conducting a two-tailed test, you can also construct Confidence Interval. A level, alpha, two-sided significance test rejects a null hypothesis $\mu=k$ when k falls outside a level (1-alpha) confidence interval for μ.

Example:

10.4: Inference as a Decision

Tests of significance assess the strength of evidence against a null hypothesis. Based on the strength of evidence, we may reject or fail-to-reject a claim about a parameter. Since it is possible to observe extreme statistic values in a given sample, it is possible we may mistakenly reject a null hypothesis that was really true. Further, it is possible we may fail-to-reject a null hypothesis that is actually false.

	Ho is True	Ho is False
Reject Ho		
Fail-to-Reject Ho		

Type-I Error:

Type-II Error:

Suppose a new law was enacted in which cigarette companies were mandated to keep nicotine content at or below 1.5mg per cigarette. Suppose a manufacturer claimed their cigarettes contained 1.5 mg of nicotine on average. You have been hired to test whether or not this claim is true. If the cigarette manufacturer's claim is found to be false, they are subject to federal fines and a possible class-action lawsuit.

1) Define a Type I and Type II error in the context of the problem.

Type I:

Type II:

2) Suppose we are testing the claim at alpha=0.01. You are going to sample 36 cigarettes and know that σ=0.20 mg.

a) At what sample mean value would you reject the manufacturer's claim that the cigarettes contain 1.5mg of nicotine on average? Hint: Consider the sampling distribution assuming their claim is true…at what point would you reject the claim?

3) Let's pretend that (unbeknownst to us) the true mean content is 1.6mg. That is, in reality, the average is greater than the acceptable limit and the manufacturer's claim is false. Remember, we are assuming the claim is true. What is the probability that we'd correctly reject the manufacturer's claim?

What is the probability we'd fail-to-reject the false null? What are the implications of this Type II error?