Chapter 6: Inference

Read Chapter 6 first omit the power section**,** then read the notes and try the WEBCT assignment questions. If you need more practice, try the practice questions with answers available on the web.

Chapter 6 - Introduction

Statistical inference is a mathematical technique used by scientists to make conclusions about a population from the sample data they have collected. Two common forms of inference are introduced in this chapter, confidence intervals and tests of significance. We introduce a simple form of inference concerning the mean of a normal population with known variance.

Background

It is rare that we can know exactly the distribution of a variable X over a population π.

Based on a sample $X_1, X_2, \ldots, X_n$ from this distribution we want to make inferences about the true values of various characteristics of this distribution. For example: μ_x, σ_x, etc.

There are **3 Forms of Inference**

(1) estimation
(2) confidence intervals
(3) tests of significance

(1) Estimation

For example: **Consider quantitative variables**

Let X_1, X_2,...X_n be a sample from a distribution with unknown mean μ and unknown variance σ^2. We want to estimate:

$$\mu \text{ by } \bar{x}$$

and estimate

$$\boldsymbol{\sigma}^2 \quad \text{by} \quad \mathbf{s}^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

We have seen previously (chapter 3 and 5) that the sampling distribution of $\bar{X}$ is centered at the mean of the population:

$$\mu_{\bar{x}} = \mu$$

and the variability of the sampling distribution about μ is measured by

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

Hence, we estimate the accuracy of $\bar{x}$ as an estimate of μ by the **standard error of the estimate** below:

$$\frac{S}{\sqrt{n}}$$

Example: Here is a **simulation** to help you understand estimation. Estimation is illustrated here with a computer simulation with the population given below for a variable Y. Suppose Y ~ N (50, 15) and we wish to take 40 samples each of size 25. We want to estimate the mean.

SPSS EXAMPLE ONE

In my simulation we obtain:

$$\bar{X} = 419.866$$

$$\frac{S}{\sqrt{n}} = .4359$$

which is close to the population mean with small standard error. The researcher usually knows what accuracy would be required for the estimate. Greater accuracy can be achieved by increasing the sample size.

View the Video – Confidence Intervals

(2) <u>Confidence Intervals</u>

How do we interpret the standard error of an estimate?

Recall the Central Limit Theorem from Chapter 5. If X_1, X_2,...,X_n is a sample from a distribution_with mean μ and variance σ^2 then the mean "x" is distributed approximately:

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

We then can derive the confidence interval as follows:

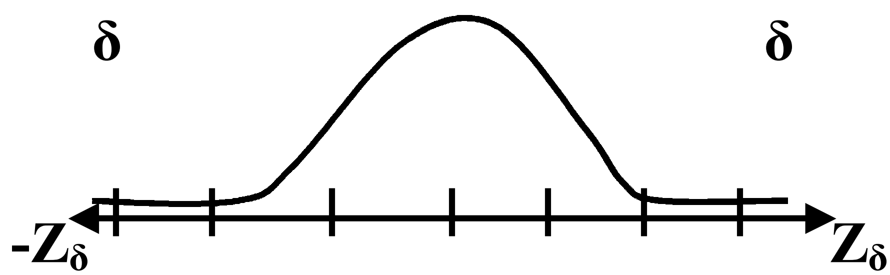
$$P(\bar{X} \leq X) \quad = \quad P(\bar{X} - \mu \leq X - \mu)$$

$$= P\left(\frac{\bar{x} - \mu}{\sigma\sqrt{n}} \leq \frac{x - \mu}{\sigma\sqrt{n}}\right)$$

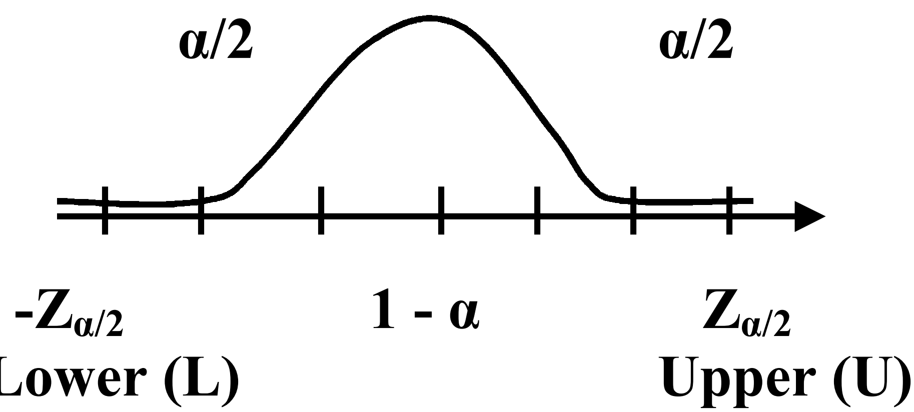
by the CLT we have

$$= P\left(Z \leq \frac{x - \mu}{\sigma\sqrt{n}}\right)$$

where $Z \sim N(0,1)$. We now choose α, which is the total amount of probability in the tails of the distribution and 1- α is the probability in the center of the distribution. A common value used in research is .05. Graphically we have N(0,1):



Note that α is between 0 and 1.
We have that δ + δ = α or 2δ = α. We can write δ=α/2. Graphically we have:



<u>We assume σ is known</u> and construct an interval such that the center probability is 1 - α.

This is called a (1-α) confidence interval (C.I.) for μ or P (L ≤ μ ≤ U) = 1 – α where L = Lower CI and U = Upper CI.

After observing X_1, X_2,...,X_n our inference takes form. We say μ is contained in [L,U] with confidence (1-α).

If (1-α) is close to 1, and the length U-L is small we have very precise inference about μ.

Now by CLT

$$= \mathbf{P} \left(Z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma\sqrt{n}} \leq Z_{\alpha/2} \right) = \mathbf{1-\alpha}$$

for known σ and fixed α is a (1-α). C.I. for μ of length $2\,\sigma/_{\sqrt{n}}\, Z\alpha/_2$ is:

$$[\bar{x} - \sigma/\sqrt{n}\, Z_{\alpha/2}\, ,\, \bar{x} + \sigma/\sqrt{n}\, Z_{\alpha/2}]$$

Note a common choice is 1 - α = .95

If we have a 95% C.I. then α/2 = .025. Using standard Z tables $Z_{.025} = 1.96$. The choice of Z depends on the choice of confidence interval (90%, 95%, 99%, etc.).

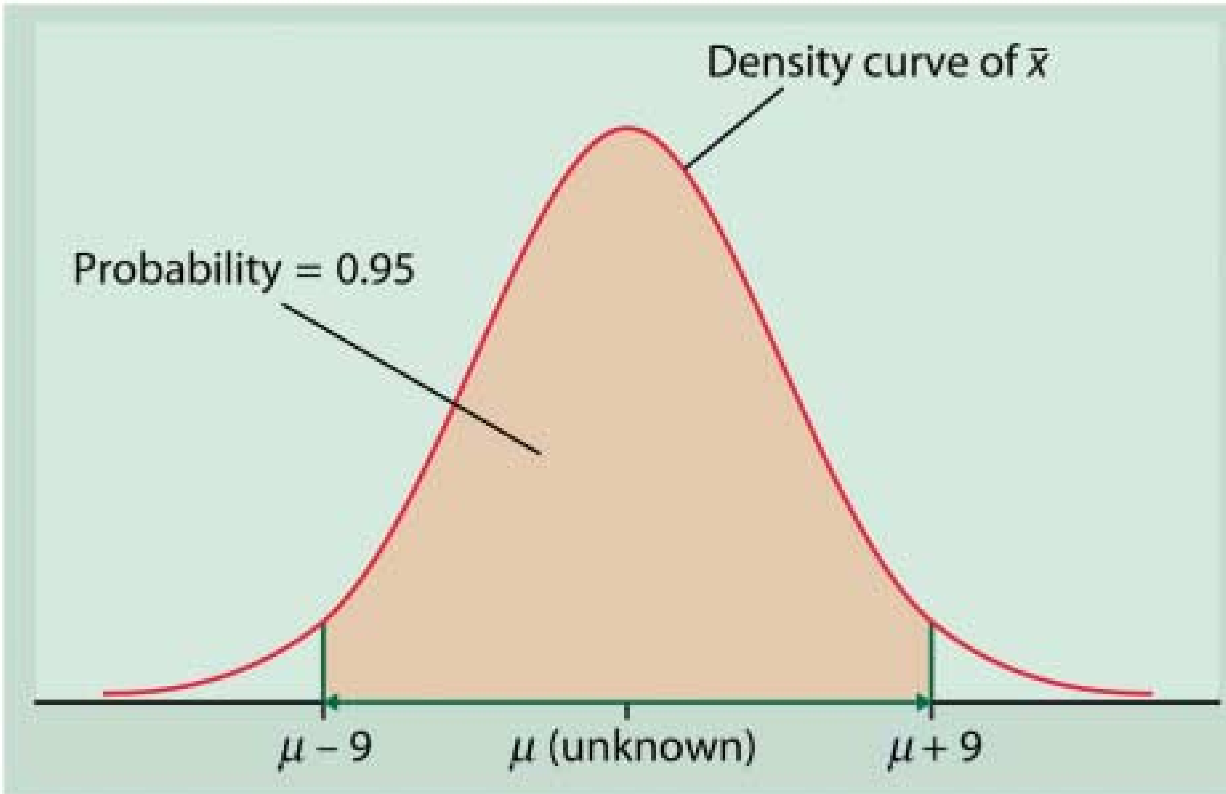


Figure 6.1 x(bar) lies within +/- 9 of μ in 95% of all samples, so μ also lies within +/- 9 of x(bar) in those samples.

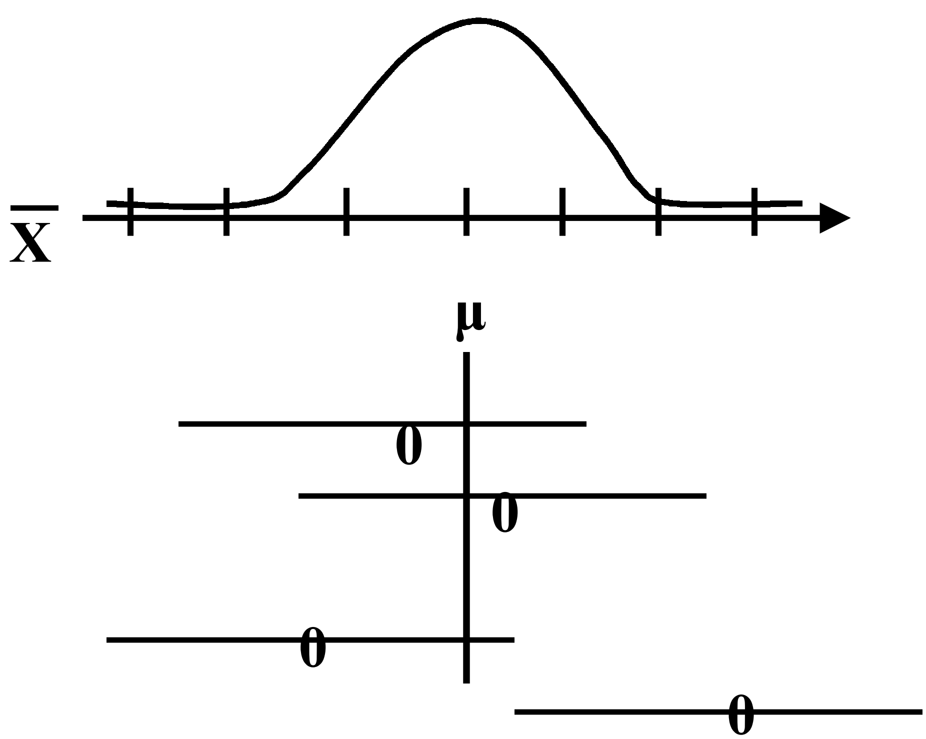
<u>Example</u>: Simulation

We will use SPSS to simulate how a confidence interval behaves and consequently better understand this concept. We generate 20 samples, each of size 25 from a population which is N (50,15). For each sample calculate a 90% confidence interval. How many of the intervals contain the true value 50?

In my simulation of the 20 intervals, 3 do not cover the mean of 50. We expect approximately 10% (or 2) not to cover mean since we have a 90% C.I. If we were to run the simulation over many samples in the long run, 10% would not contain the mean.

SPSS EXAMPLE TWO

Click above to view the C.I. simulation. If you wish you can repeat the simulation for a (1 - α) = 95% C.I. using the instructions at the end of the above example.



in the above figure 3 confidence intervals capture the population mean however one does not. In the long run if we calculate a 95% confidence interval for each sample, 95% of them will contain the mean of the population.

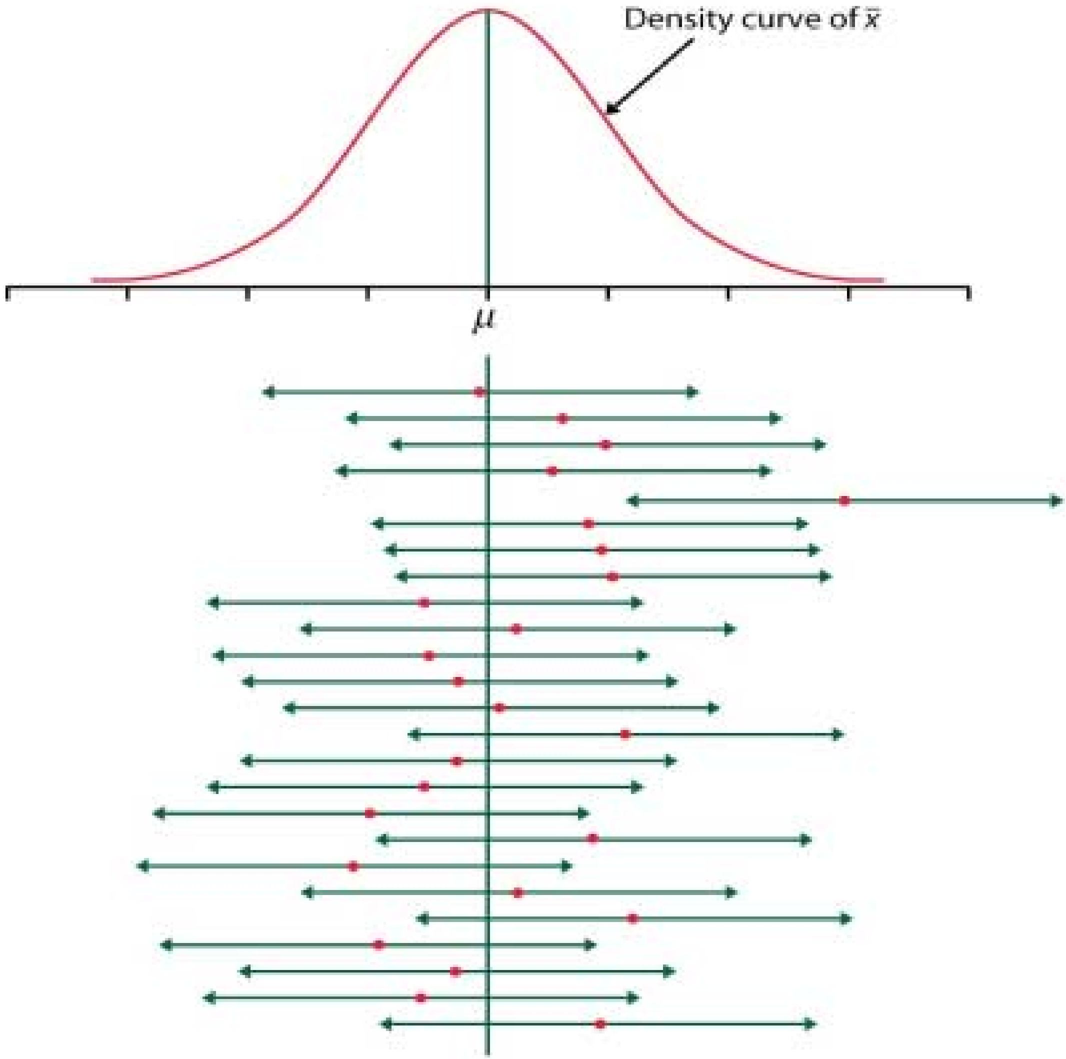


Figure 6.2 Twenty-five samples from the same population gave these 95% confidence intervals. In the long run, 95% of all samples give an interval that covers μ.

The above picture shows how a confidence interval behaves. The population distribution of a variable x has a mean (μ) which is unknown to the researcher.

The researcher, however, wants to know the mean of the population. The confidence interval gives the researcher a quantification of uncertainty with respect to the location of the mean.

For each of 4 samples(hand drawn figure above) the sample mean is presented with the corresponding C.I. The 0's denote the sample means and the solid line the confidence interval. Note the first three samples CI contain the population mean whereas the 4th does not.

Note most intervals will contain the population mean but some will not. In the long run, over many samples, a 90% C.I. will contain the population mean 90% of the time. Remember in real life you will usually calculate only one interval. In the above context we calculate many to give you a better understanding of the CI concept.

The confidence interval is not the probability that the true mean lies in the interval but the percent of the time the method gives correct answers. For example: 95% of all samples give an interval that covers μ for a 95% confidence interval for the mean. Examine the simulations and convince yourself this is true.

Example

A researcher makes 3 independent measurements on the concentration of a drug. They are:

.8403 .8363 .8447

You are given σ =.0068, find a 99% CI for μ, the mean concentration. Calculate $\bar{x}$ =.8404, z =2.576 for .005 in each tail (2 x .005 = 0.01). A 99% CI for μ is:

$$\bar{x} \pm Z\,\sigma/\sqrt{n}$$

$$.8404 \pm 2.576\,(.0068)\,/\,\sqrt{3}$$

$$.8404 \pm .0101$$

We see that (.83030, .85057) is a 99% confidence interval for the mean drug concentration. 1 - α = .99, therefore α = .01. The confidence interval is

symmetric, therefore α/2 = .005. The Z for .005 probability in the tail is 2.576. α is divided by two because the C.I. is symmetric.
How does the sample size and % confidence effect the length of the interval?

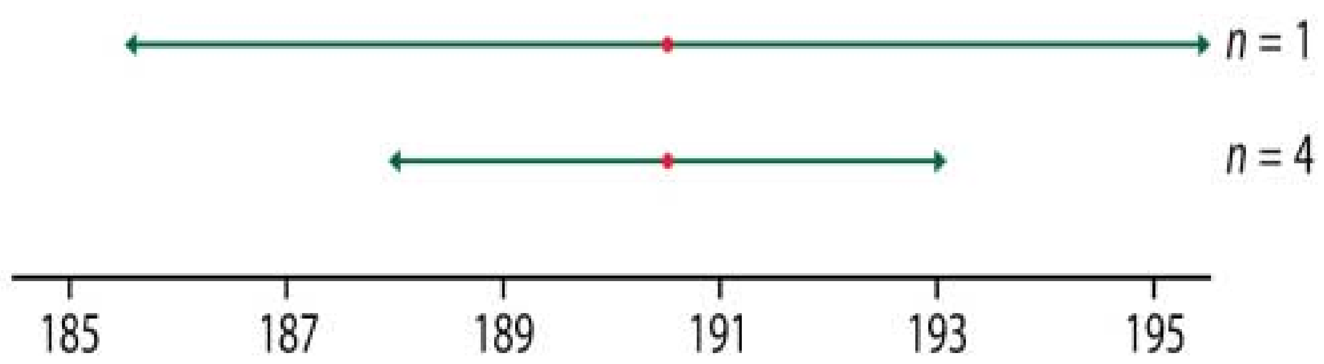


Figure 6.4 Confidence intervals for n = 4 and n = 1.

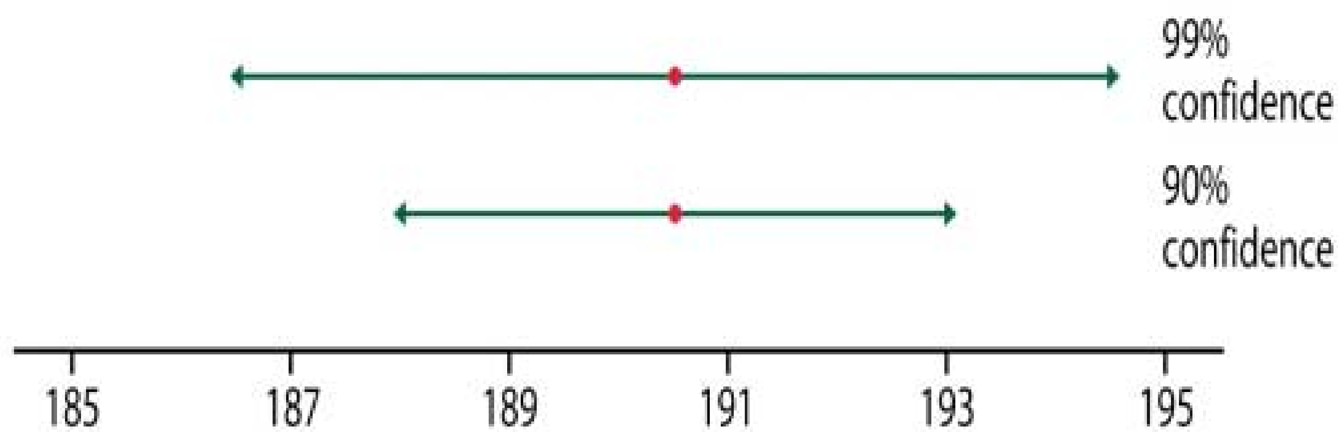


Figure 6.5 Confidence intervals for different confidence.

View the Video – Significance Tests

TEST OF SIGNIFICANCE

X quantitative variable

Suppose $X_1, X_2,...,X_n$ independent with unknown σ^2. Suppose some theory prescribes a value μ_o for the unknown.

Based on the observed data we want to test the

null hypothesis: $H_o: \mu=\mu_o$

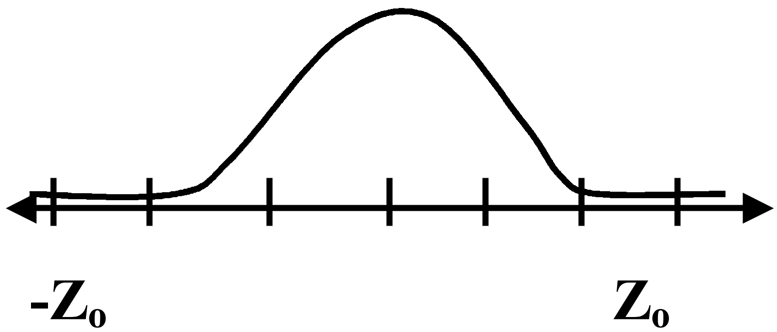
against the

alternative hypothesis: $H_a: \mu\neq\mu_o$

By 'test' we mean we want to decide whether or not the data $X_1, X_2,...,X_n$ provide evidence that H_o is not true. How should we go about testing H_o? If H_o is true then CLT says:

$$Z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}} \sim N(0,1)$$

If H_o is true the observed value Z_o of Z should be a reasonable value for the N(0,1) distribution.



If Z_o is far out in the tails of the N(0,1) distribution then we have a value which provides evidence against H_o.

We measure how far Z_o is out in the tails by computing the approximate **<u>p-value</u>** or **<u>observed level of significance (OLS).</u>** SPSS calls this **<u>SIG.</u>**

This is the probability that Z is bigger/as big as Z_o (use Z tables) or smaller/as small as $-Z_o$.

If the P-value is small (e.g., $p \leq .05$) then Z_o is providing evidence against H_o since this is a rare event given the mean value. Note, the mean value in this case comes from the null hypothesis; researchers assume the null hypothesis is true and they calculate the chance of obtaining the data they have collected.

The above test is called a **2 sided Z test.**

<u>Example</u>: A developmental psychologist wants to see if handling during infancy affects infant weight. n=16 infants and their caregivers are involved in a special handling course.

The infants are weighed at 2 yrs old $\bar{x}$ = 28.5 lbs. We know the weight distribution is N (26,4) for untreated infants. Does the new handling course have an impact on the weight of the infants?

Test this hypothesis:

H_o: $\mu = 26$ (mean is equal to 26)
H_a: $\mu \neq 26$ (mean can be greater than 26 or less than 26)

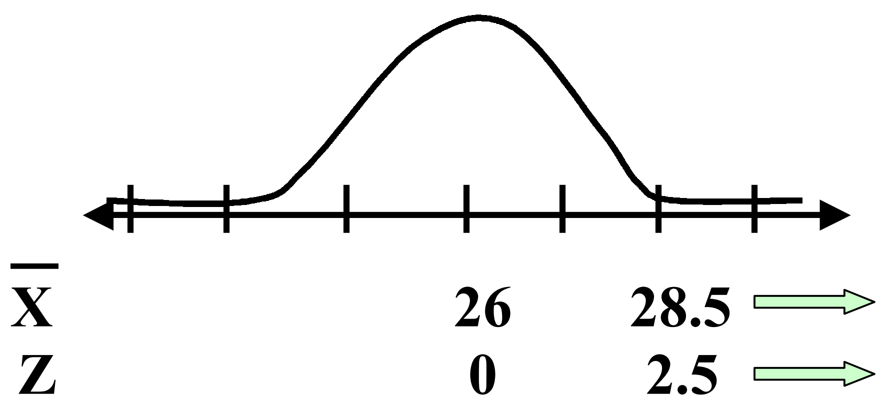
We assume the mean of the population is 26 (that H_0 is true). Does the data support this assumption?

$$\bar{X} = 28.5 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{16}} = 1$$
$$\mu = 26$$

We know the location and spread of the sampling distribution for the mean using the statistics above.

Since we know the location and spread of the sampling distribution under the assumption $\mu = 26$, we can find the probability of obtaining the sample mean under this assumption. Substitute into Z.

$$Z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}} = \frac{28.5 - 26}{1} = 2.5$$



The sample mean of 28.5 is 2.5 standard deviations above the mean. The probability of getting a value greater than or equal to 2.5 is:

P (Z > 2.5)= .0062

but note H_a is not equal, because the alternative states that the mean could be bigger or smaller than

26, we double value to include the possibility of less than (P (Z < -2.5)).

Therefore: **p – value = 2(.0062) = .0124.**

This is a small probability or rare event. In other words, if the population mean is 26 the chance of obtaining the observed mean (28.5) is small (p-value=.0124). We conclude that there is strong evidence against H_o. Handling during infancy affects the weight of the child.

One-Sided Test

If we are only concerned with alternatives to H_o: $\mu = \mu_o$ that are directional. For example:

$$H_a : \mu > \mu_o \text{ (greater than)}$$

Then the value of our test statistic:

$$Z_0 = \frac{\bar{x} - \mu}{\sigma\sqrt{n}}$$

will tend to lie far into the right tail when H_o is false. In this case the p-value is: P $(Z > Z_o)$ and this is called the **<u>one sided Z-test</u>.**

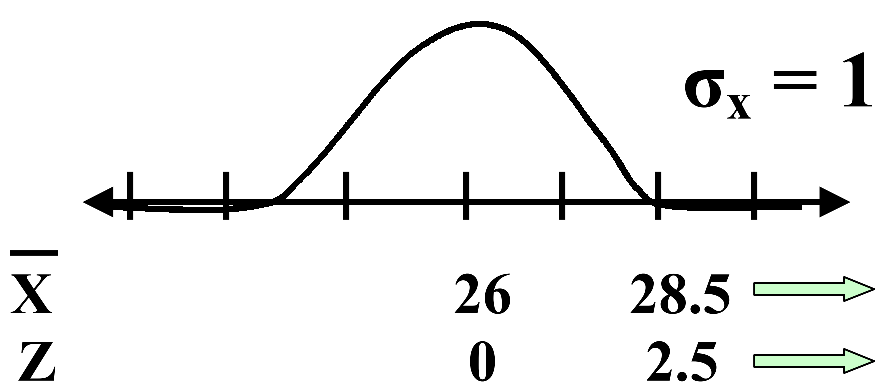
<u>Example</u>:
Consider our previous example with children and handling. The researcher is only concerned with gains in weight (μ>26).

In other words, the test is one-sided.

$$\mathbf{H_o: \mu = 26}$$
$$\mathbf{H_a: \mu > 26}$$

The calculation of the statistic is identical to our previous example.

$$\mathbf{Z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}} = \frac{28.5 - 26}{1} = 2.5}$$



The mean 28.5 is 2.5 standard deviations above the mean.

P (Z > 2.5)= .0062 (very small probability)

Since test is one-sided (mean is greater than 26) we do not double the probability to take into account the left side(mean is less than than 26). The p value is .0062 and provides very strong evidence against H_o. Handling does increase weight again.

<u>Example</u>: DRP (Degree of Reading Power scores). The data for 44 students who have taken the DRP test is given below.

Data	40 26 19
	47 19 46
	52 25 52
	47 35 45

There are n= 44 students. The standard deviation is given in the question as σ=11 as the population of all students in the district. The national average is 32 on this test. We want to test if this group of students comes from a population with a higher average.

$$H_o: \mu = 32$$
$$H_a: \mu > 32 \text{ (directional test)}$$

$\bar{X} = 35.091$ so we substitute into out test statistic

$$Z = \frac{35.091 - 32}{11/\sqrt{44}} = 1.86$$

We use the Z tables to find the probability to the right of 1.86 and obtain p-value .031 (small probability).

Conclusion:
There is strong evidence that the children have a mean score higher than the national average. A 95% confidence interval is given below:

$$35.091 \pm 1.96\ 11/\sqrt{44}$$

(32.84, 38.35)

The researcher now knows with 95% confidence the population mean of the scores is between 32.84 & 38.35.

Tests with a Fixed Significance Level

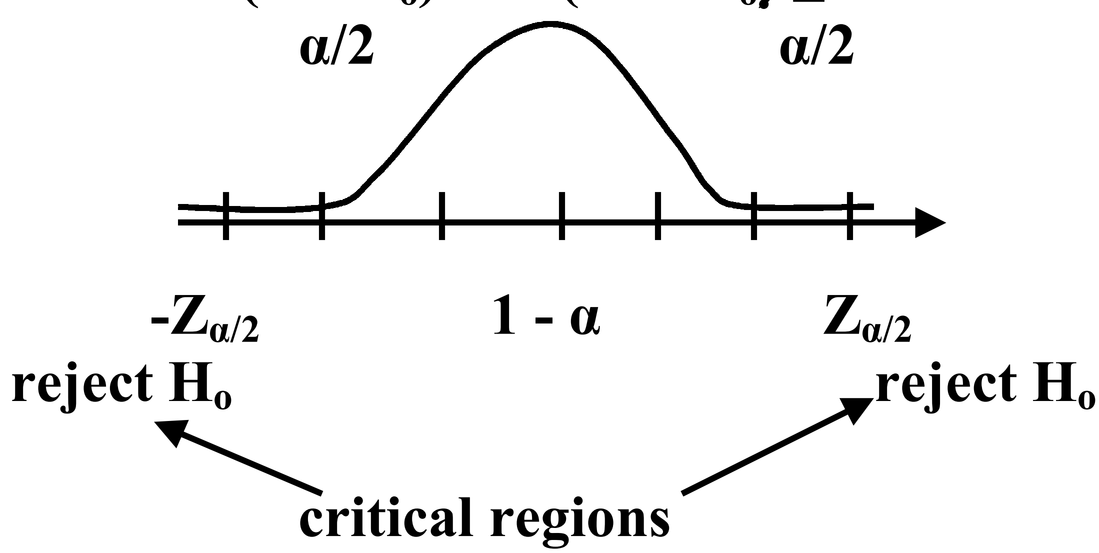
Suppose we test: $H_o: \mu = \mu_o$

vs

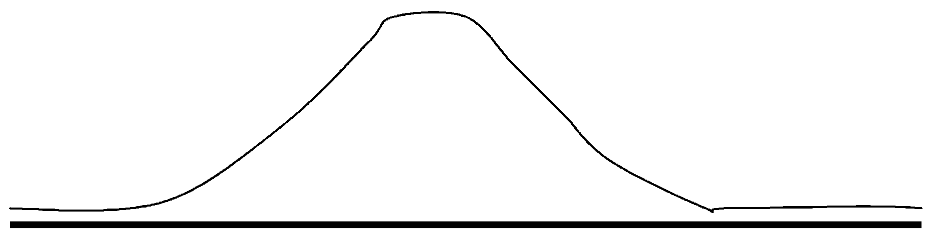
$H_a: \mu \neq \mu_o$

and decide to reject whenever

$$P(Z > Z_o) \text{ or } P(Z < -Z_o) \leq \alpha$$



The regions in the tails of the distribution are called the critical regions since, if the statistic falls in them, the decision is to reject the null hypothesis. Researchers usually use $\alpha = .05$ therefore $\alpha/2 = .025$. The corresponding Z from tables Z = 1.96. If the Z statistic is greater than 1.96 or less than -1.96 we reject the null hypothesis.



-1.96 **0** **1.96**

reject H_o **reject H_o**

Note: With statistical software readily available the fixed level test is not used often by researchers since the p-value is easily computed for an exact probability rather than the reject/accept at alpha=some value language used in a fixed test. The fixed level test was much more popular in the past however the mechanics of conducting the test are important for the calculation of power.

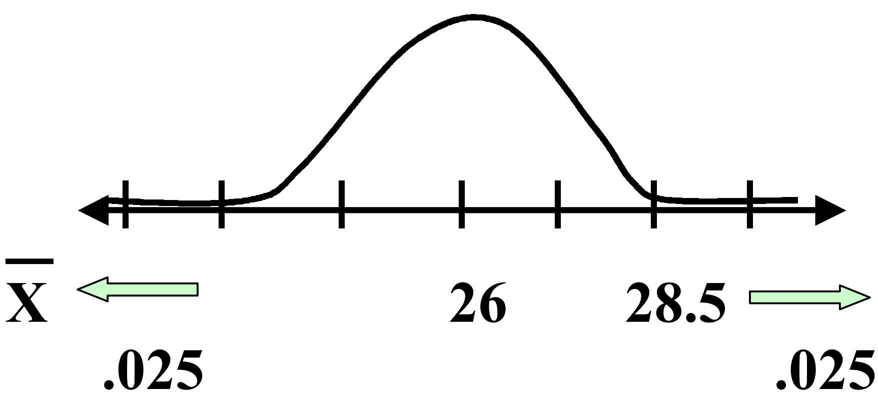
Example: Use the previous infant data (continued) and perform a fixed level test at α=.05.

$$H_o: \mu = 26$$
$$H_a: \mu \neq 26$$

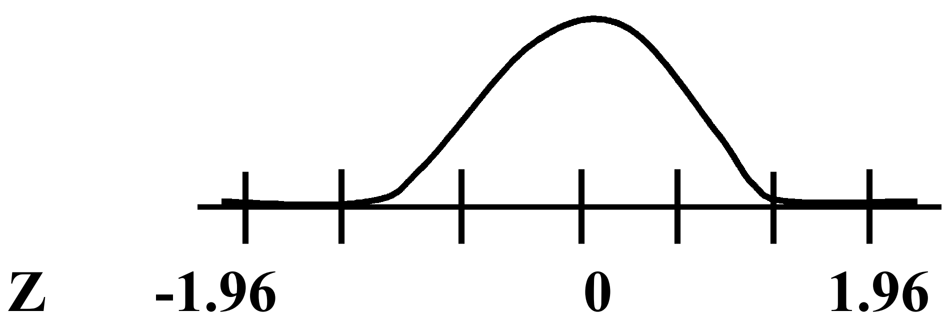
The calculation of the Z statistic is identical to our previous example.

$$Z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}} = \frac{28.5 - 26}{1}$$

$$= 2.5$$



Given α=.05 and the test is 2-sided, then α/2=.025. We look up Z for .025 probability in the tail and find Z=1.96 as the critical value.



Now is the Z statistic of the data larger than 1.96 or smaller than -1.96? Does it fall in the rejection region? Yes! (Z = 2.5). Note you do not have an exact probability(p-value) only that it is significant at the alpha specified.

We reject H_o and conclude that handling affects weight. An equivalent way to carry out a 2-sided test at level α for μ is to compute a $(1-\alpha)$ CI for μ and reject if it does not contain μ_o.

<u>For example</u>: A 95% CI for the infant data is given below:

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}$$

$$= 28.5 \pm \frac{4}{\sqrt{16}} 1.96$$

$$= 28.5 \pm 1.96$$

The 95% C.I. is (26.54,30.46). Does this interval contain $\mu=26$?

NO!!! THEREFORE, we reject H_o at $\alpha=.05$, and conclude handling affects weight.

Practical Significance vs Statistical Significance

When we test H_o: $\mu=\mu_o$ at level α and have evidence against H_o we say the results are **statistically significant.** Note, if we take n large enough we will reject H_o where:

$$\mathbf{H_o: \mu = \mu_o + \varepsilon}$$

no matter how small ε

ε may not be a difference of **practical significance**. We look at the (1-α) CI for practical significance. Both the p-value/OLS and fixed-level tests of hypothesis are reported by researchers. With the advent of readily available computers and software, the most popular method of reporting is the p-value or OLS.

Chapter 6 Summary

The confidence interval is used to estimate an unknown **population parameter.** The interval gives the researcher an idea of the accuracy of the estimate. The confidence level (95%, 99%, etc.)

gives the probability that the method will provide the correct answer.

A test of **significance** weighs the evidence in the data against a **null -hypothesis** (H_o) and indirectly in favor of an **alternative hypothesis** (H_a). Usually H_o is a statement that no effect is present and H_a says the parameter differs from the null value in one direction **(one-sided test)** or both **(two-sided test)**.

The p-value is the probability assuming that H_o is true that the test statistic will take a value as extreme as that which is observed. Small p-values indicate strong evidence against H_0. However a statistically significant effect need not be **practically** important.

Remember: Tips for Success

1) Read the text.
2) Read the notes.
3) Try the assignment.
4) If needed, try the exercise questions.
5) Try the simulations and view the videos if you need more help with a concept.
6) Try the self tests for practice on each chapter of the text at www.whfreeman.com/ips.
7) Steady work = Success