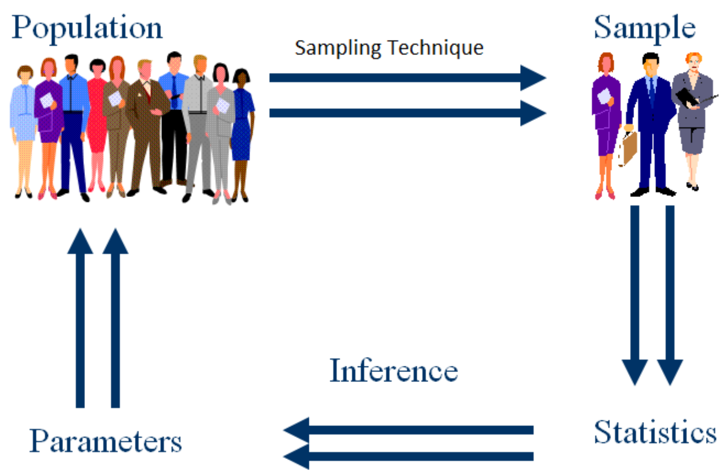
Chapter 5

ST 511 - Inferences About Population Central Values

Readings: Chapter 5 (for 5.8-5.9 read if interested)

Recall our overall idea:



Inference - refers to making mathematical claims about a parameter using sample data.

Two main methods for inference:

1. ______________________________ - Range of values we think contains the parameter.

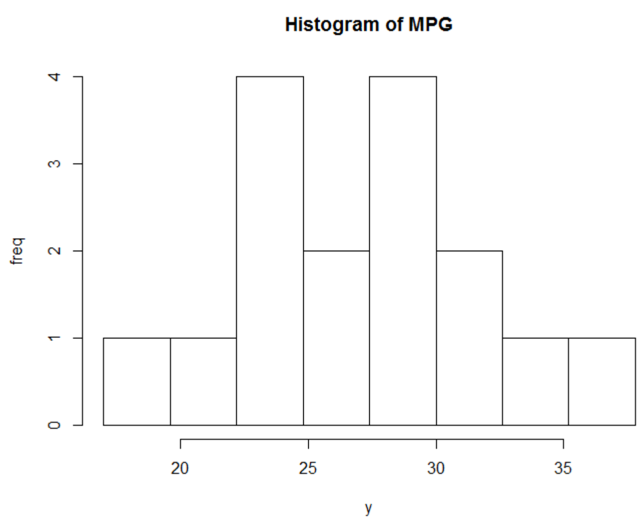
2. ______________________________ - Test of whether a specific parameter value is plausible.

Putting it all together: Engineers at Ford are attempting to improve the overall gas mileage of next year's model of one of their cars. From extensive testing on the previous year's model they know the gas mileage can be well modeled by a normal distribution with true mean gas mileage of 26.9 mpg and true standard deviation of 2.3 mpg.
To investigate if this year's model has improved the average gas mileage, data is collected on 16 automobiles. The average gas mileage of the sample is 28.25 mpg.

- Population -

- Parameter of Interest -

- Random Variables used to answer questions of interest -



- Sample -

- Statistic(s) -

- Inference using a ‘Hypothesis Test’ -

- Inference using a 'Confidence Interval' -

5.2 - Confidence Intervals

Confidence Intervals are better than point estimates, such as $\bar{y}$, as they ________________________________

Intervals that are created have a ____ = the probability **the procedure used** creates an interval that contains the parameter.

$\alpha =$ ____ - usually 0.01, 0.05, or 0.1.

An **observed interval**, such as the one in the example above, is interpreted in the following manner:

We say...

For a confidence level of $(1-\alpha)$, the general interpretation of an observed CI is...

What we mean by 'confident' for our example above is...

For a general $(1-\alpha)100\%$ CI, confidence is interpreted as...

Note that the interval is random and the parameter is fixed!

`http://bcs.whfreeman.com/ips4e/cat_010/applets/confidenceinterval.html`

Confidence Interval for μ_Y, the population mean

For a **random sample** of size n, where the population standard deviation, σ_Y, is known, a $(1-\alpha)100\%$ CI for μ_Y is given by

The interval is valid if **either**

1.

2.

Common z values used:

Ex: The length of music videos is of interest to advertisers. Assume we know the standard deviation of the length of music videos is 18 seconds (a dubious assumption we will learn to deal with later). In a random sample of 44 music videos, the average length was found to be 186 seconds. Find a 90% confidence interval for the mean length of all music videos. What does confidence mean here?

(Book examples 5.1 pg 228, 5.2 pg 229 for more practice.)

Factors that affect the width of confidence intervals:

1. Natural Variability -

2. Level of Confidence desired -

3. Sample size n

5.3 - Required sample size to have an interval of a given width:

Ex: Suppose that birth weights for boys are normally distributed. We want to estimate the population mean μ, the overall average birthweight for boys. Assuming that the population standard deviation, σ, is known to be 12 oz (based on past data), what minimum sample size is required if the width of a 90% CI is to be at most 6 ounces?

(Book examples 5.3/5.4 on page 231/232 for more practice.)

5.4/5.6 - Hypothesis Testing for μ_Y.

- Hypothesis testing and confidence interval estimation are related methods and are often both used to analyze the same situation.
- CI is a numerical answer to the question, 'What is the population value?'
- Hypothesis test is used to answer questions about particular values for a parameter.

Goal of hypothesis testing:

Step 1/2 - Determine Null and Alternative Hypotheses

______________________ - a statement or claim regarding parameter(s)

____ - Statement about parameter 'no effect' or 'no difference,' the default belief or status quo.

____ - Statement we hope to prove or give evidence for.

One-sided vs Two-Sided Hypotheses:

For the following examples, define the parameter of interest and determine the null and alternative hypotheses:

1. A certain type of light bulb is advertised as having an average lifetime of 750 hours. A potential customer likes the price and wants to purchase a large amount of them if it can be shown the average lifetime is higher than advertised. A random sample of 20 bulbs was selected and the lifetime of each bulb was determined. The mean was 766.4 hours. It is known that the lifetime of the light bulbs is normally distributed and the true standard deviation is 30.5 hours.

2. The average age of a person on facebook last year was 19.34 years. The standard deviation of the age of facebook users is 8.2 years. Suppose an advertising agency is interested in seeing if the average age is different this year. He randomly selects 150 profiles and finds that the sample mean is 18.99 years old.

There are two possible conclusions from a hypothesis test:

-
  - Observed value is 'significantly different' than hypothesized value (i.e. observed value is unlikely to have occurred simply due to random chance if the hypothesized value were true)
-

We never 'accept' H_0 as the book states. We believe the null hypothesis until we see significant evidence to the contrary. Thus, we either reject the null or we fail to reject (not enough evidence to the contrary). This does not imply that H_0 is true, just that we can't say it isn't true.

- The data collected will not 'prove H_0', but it may lead us to believe that it is pretty unlikely that H_0 is true.

Two types of errors we could make:

- ____________________________ - Reject H_0 when H_0 is 'true'
  - Probability of Type I error = P(Type I Error) =
- ____________________________ - Fail to reject H_0 when H_A is 'true'
  - Probability of Type II error = P(Type II Error) =

We consider the Type I error to be the most serious one. This is why we 'control' the type I error rate by setting it **prior to the experiment**.

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	no error	type II error
Reject H_0	type I error	no error

Idea follows US justice system. Consider the following example:

A person is on trial for a crime.

- The null hypothesis is H_0: Innocent
- The alternative is H_A: Guilty
  - A type I error would be

  - A type II error would be

For most crimes, a type I error is worse than a type II. This is the same in an experiment, usually making a type I error is worse, so we set *the type I error rate* at α.

For the light bulb example earlier, what would a type I error be in words? a type II error?

Step 3 - Check Assumptions and Find Test Statistic

To make inference, we will need a 'test statistic' (such as $\bar{Y}$ or $Z = \frac{\bar{Y}-\mu_Y}{\sigma_Y/\sqrt{n}}$) that we know the *sampling distribution* of.

Recall: If we are interested in a true mean, we can estimate it using the sample mean (a RV). We know the distribution of the sample mean is

$$\bar{Y} \sim N(\mu_Y, \sigma_Y^2/n)$$

if

We assume the null hypothesis is true and see if we find evidence to the contrary.

Assuming the null hypothesis is true, we can use the test statistic

For the two examples previously given,

- determine which assumptions are met,
- calculate the **observed** value of the test statistic,
- assuming the null is true, draw the sampling distribution and place the observed value of the test statistic on the distribution.

Step 4 - Find Rejection Region (RR) and/or find P-value - Make Decision

________________________________ is determined by our chosen α and the distribution of our test statistic under the null hypothesis.

RR - values of the test statistic for which the null hypothesis will be rejected.

For a '$>$' alternative and an $\alpha = 0.05$ let's find our RR

For a '$<$' alternative and an $\alpha = 0.01$ let's find our RR

For a '$\neq$' alternative and an $\alpha = 0.05$ let's find our RR

For the previous two examples, let's write down our RR and make our decision using $\alpha = 0.05$.

____________________________ - probability of getting a test statistic as extreme or more extreme than the observed value, assuming the null hypothesis is true.

- Measure of how likely it is to get this type of sample (or worse) if H_0 is true.
- A better measure than RR for evaluating the evidence for or against the null.

For a '$>$' alternative the p-value is

For a '$<$' alternative the p-value is

For a '$\neq$' alternative the p-value is

Decision determined by

- If p-value $\leq \alpha$, ____
- If p-value $> \alpha$, ____

For the previous two examples, let's draw the sampling distribution of the test statistic, shade the 'extreme' region, find the p-value, and make our decision using $\alpha = 0.05$.

Step 5 - Draw Conclusions (in the context of the problem)

Interpreting the result means that we say in a formal way what our conclusion means for this problem:

- Fail to reject H_0, we say... At the $\alpha 100\%$ significance level, there is not enough evidence to support the alternative hypothesis that (context).
- Reject H_0, we say... At the $\alpha 100\%$ significance level, there is enough evidence to reject the null hypothesis that (context) in favor of the alternative that (context).

For the previous two examples, let's interpret our results at the 0.05 significance level.

Overview of HT

1. Set up Alternative
2. Set up Null
3. Check Assumptions and Calculate Observed Test Stat
4. Find RR and/or P-value
5. Draw Conclusions in the context of the problem

Example: The average employee tenure (number of years workers have been with their current employer) in 2010 was 4.4 years with a standard deviation of 0.9 years. Tenure is believed to be higher in this year than it was in 2010. A sample of 90 employees produced a mean tenure period of 4.7 years.

1. Assuming the spread remained constant, conduct a 0.01 level (that means use $\alpha = 0.01$) test to determine if average tenure is greater than it was in 2010. (You need to do all 5 steps.)

2. Is the sample mean 4.7 'significantly different' from 4.4? Explain what we mean by significantly different in the context of the problem.

Relationship between CIs and Two-sided Tests

- If the null value μ_0 is contained in a $100(1-\alpha)\%$ CI for μ, then we fail to reject H_0 at level α.
- If the null value μ_0 is NOT contained in a $100(1-\alpha)\%$ CI for μ, then we reject H_0 at level α.

Let's calculate a confidence interval for μ in the employee tenure example.

5.5 - Power and Choosing a Sample Size

______________________ = 1 - P(Type II Error) = 1-β = 1 - P(failing to reject H_0 when H_A is true) = P(reject H_A when H_A is true)

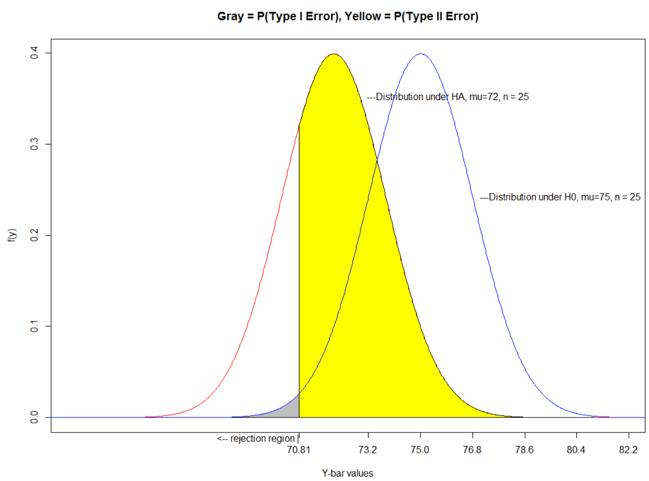
Ideally, we have small type I AND type II error rates (probabilities).
We 'control' the α = P(type I error) = type I error rate by setting this prior to the experiment.
The main way to deal with the type II error rate (or equivalently power) is by increasing the sample size.

Ex: The drying time of a certain type of paint (in minutes) under specified test conditions is known to follow a $N(75, 9^2)$. Chemists have designed a new additive to decrease average drying time. Let Y denote the drying time of this new additive. Lets assume the $Y \sim N(\mu, 9^2)$. We want to determine if there is strong evidence to suggest an improvement in average drying time. Suppose a random sample of 25 drying times is taken and the sample mean is 70.8.

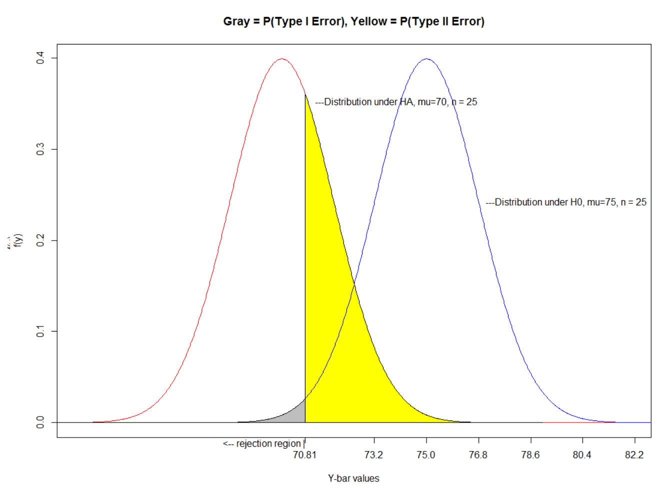
1. Conduct a hypothesis test using rejecting regions and $\alpha = 0.01$.

2. A HT is conducted assuming H_0 is true. Let the random sample of 25 drying times be denoted as $Y_1, Y_2, ..., Y_{25}$. What is the distribution of $\bar{Y}$ generally?

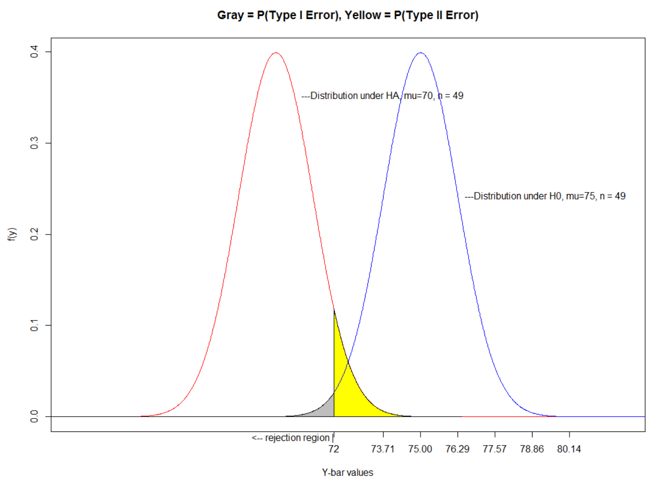
3. Assume now that in actuality the average drying time is truly 72 minutes. Find the probability of a type II error. (We denote this as $\beta(72)$.)



4. Assume now that in actuality the average drying time is 70 minutes. Find the probability of a type II error, $\beta(70)$.



5. Continuing with the assumption that $\mu = 70$, suppose now that a random sample of size n=49 is conducted. Find $\beta(70)$.



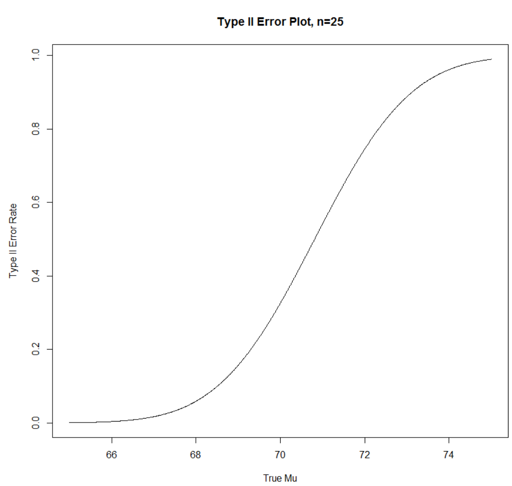
Generally, the type II error rate for a one-tailed test is given by

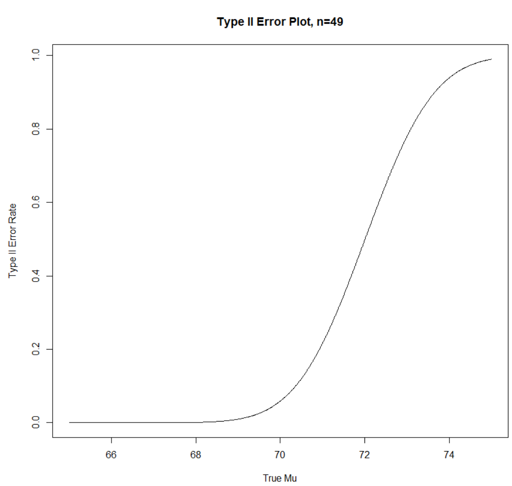
$$\beta(\mu_A) = P\left(Z \leq z_\alpha - \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}}\right)$$

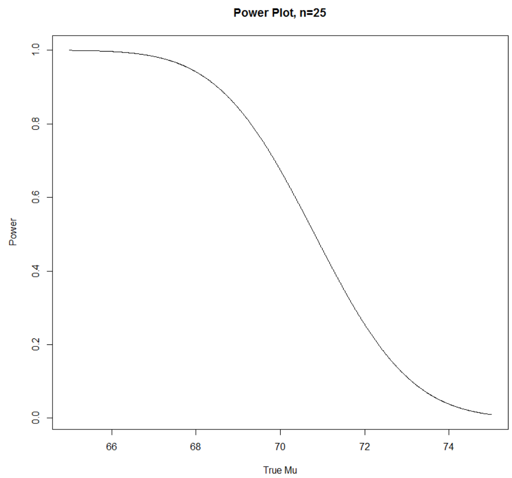
for a two-tailed test it is given by

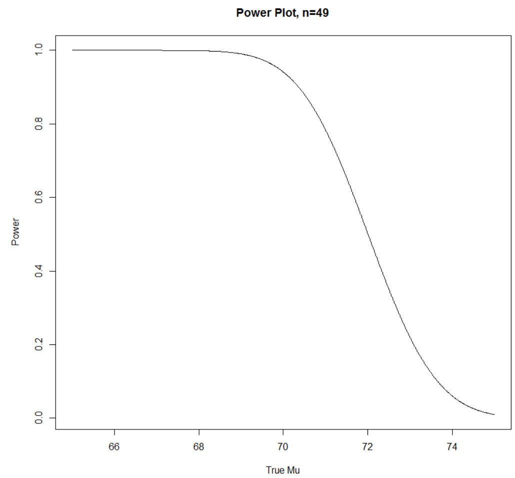
$$\beta(\mu_A) \approx P\left(Z \leq z_{\alpha/2} - \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}}\right)$$

Prior to an experiment, we would assume a value for σ_Y and plot the power (or type II error rate, β) as a function of μ_A.







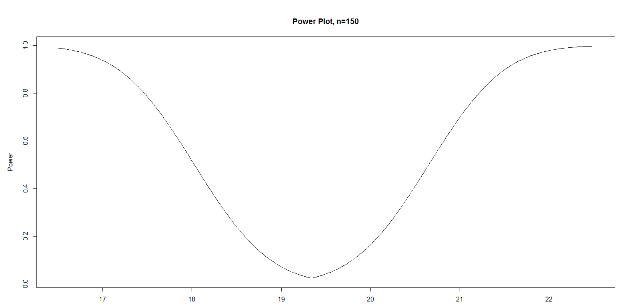


Recall Example: The average age of a person on facebook last year was 19.34 years. The standard deviation of the age of facebook users is 8.2 years. Suppose an advertising agency is interested in seeing if the average age is different this year. He randomly selects 150 profiles and finds that the sample mean is 18.99 years old.

Using $\alpha = 0.05$, our RR is $\{z_{obs} : |z_{obs}| > 1.96\}$.

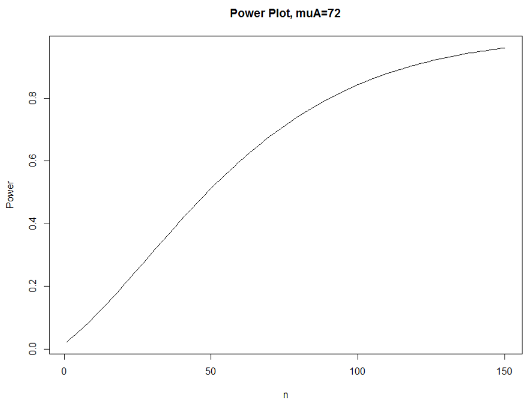
What is the power if $\mu_A = 18$ is the truth? if $\mu_A = 21$ is the truth?

Looking at the power curve below, what is the power when $\mu = 19.34$ is the truth? Why does this make sense?



Alternatively, we could fix the value of μ_A and let n vary to determine a given sample size for obtaining a certain power.

Drying time example with $\mu_A = 72$. Plot of power for varying n



Facebook age example with $\mu_A = 18$. Plot of power for varying n