Contents

[HW Summary 1](#_Toc512062069)

[Unit 1 1](#_Toc512062070)

[Unit 2 1](#_Toc512062071)

[Unit 3 2](#_Toc512062072)

[Unit 4 3](#_Toc512062073)

[Unit 5 3](#_Toc512062074)

[6Steps 4](#_Toc512062075)

[6 Step Process: 4](#_Toc512062076)

[Power 5](#_Toc512062077)

[Type 1 and Type 2 6](#_Toc512062078)

[Central Limit Theorem: 6](#_Toc512062079)

[P-vals and Confidence Intervals 6](#_Toc512062080)

[Rank Sum Test 7](#_Toc512062081)

[ANOVA 7](#_Toc512062082)

[Assumptions 8](#_Toc512062083)

[Sample T Test Assumptions: 8](#_Toc512062084)

[2 Sample T Test Assumptions: 8](#_Toc512062085)

[Assumptions of Linear Regression 8](#_Toc512062086)

# HW Summary

Unit 1

Randomized, Random Sample, Observational Studies

## Unit 2

6 Step Hypothesis

T –test to see if u equals a value

2 Sample t-test

Proc Power

### Unit 3

2 Sample T – Test with Assumptions

That ulog(16years) = ulog(12years) problem

## Unit 4

Rank Sum Test

Welch’s Two sample t-test

## Unit 5

One Way Anova

# 6Steps

### 6 Step Process:

Step 1: Hypothesis

Ho = 50, Ha <> 50

#### Step 2: Critical Value

1.97

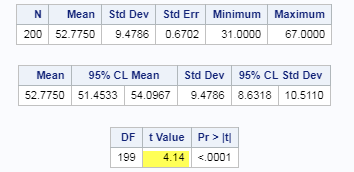
data quantile;

myquant = quantile('t',.975, 199); /\*199 is N - 1 for ttest AND .05 FOR TWO TAIL \*/

run;

#### Step 3: Test Statistic

4.14



Code:

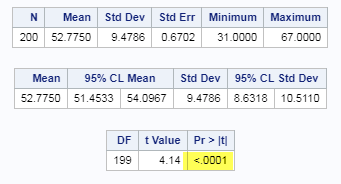
proc ttest data = WORK.hsb2 alpha= .05 h0=50 sides=2;

var write;

run;

Step 4: P-value

<.0001



#### Step 5:

Reject the null

#### Step 6:

There is strong evidence to suggest at the alpha = .05 level of significance (p-value <.0001) that the mean writing score is different than 50 points. A 95% confidence interval for the true mean writing score is (51.5 points, 54.1points).

# Power

Power – The probability of rejecting H0 when the null is false

As Power Increases…

* The sample size increases
* The standard deviation/standard error decreases
* The effect size increases

Effect size = True Effect = treatment effect = Theta

# Type 1 and Type 2

Type1 – Alpha – The probability of rejecting the H0 when the null is true

Type2 – Beta – The probability of FTR H0 when the null is false

As alpha increases, B decreases

As B increases, power decreases

As alpha increases, power increases

As B increases, the effect size decreases

# Central Limit Theorem:

<http://www.rossmanchance.com/applets/OneSample.html?population=gettysburg>

The Central Limit Theorem says that as n increases, the binomial distribution with n trials and probability p of success gets closer and closer to a normal distribution. That is, the binomial probability of any event gets closer and closer to the normal probability of the same event.

The normal distribution has the same mean μ = np and standard deviation as the binomial distribution.

SD gets smaller as the sample size increases

The more data you pick for each sample, the more normal the sample mean is

# Robustness and Resistant

Robust – A statistical procedure is robust to departures from a particular assumption if it is valid even when the assumption is not met.

CLT – Robust against normality

Resistant – A statistical procedure is resistant if it does not change very much when a small part of the data changes, perhaps drastically.

# P-vals and Confidence Intervals

* P < .05 Reject the Null (H0)
* P > .05 Fail to Reject the Null
* If the P value is less than your significance (alpha) level, the hypothesis test is statistically significant.
* If the confidence interval does not contain the null hypothesis value, the results are statistically significant.
* If the P value is less than alpha, the confidence interval will not contain the null hypothesis value.
* If 0 is found in your confidence interval

# Rank Sum Test

* No distributional assumptions
* Resistant to outliers
* Performs nearly as well as the t-test when the two populations are normal and considerably better when there are extreme outliers
* Works well with **ordinal** (as opposed to interval data)
* Works with censored values

Assumptions:

All observations are independent

The y values are ordinal

When you do separate comparisions, you need to multiply the p-vals by k.

# ANOVA

# Lack of Fit Test

See Print Outs

# Assumptions

## Sample T Test Assumptions:

1. Samples are drawn from a **normally** distributed population.
   1. If normal, move on if not
      1. Transform data or
      2. Central Limit Theorem if N>30 assume
      3. Sign Test or Wilcoxon Signed Rank Test
2. The observations in the sample are independent of one another.

## 2 Sample T Test Assumptions:

1. Samples are drawn from a **normally** distributed population.
2. If it is a two sample test, both populations are assumed to have the same standard deviation (same shape).
3. The observations in the sample are independent of one another.

## Assumptions of Linear Regression

1. There is normally distributed subpopulation of responses for each value of the explanatory variable.
2. The means of the subpopulations fall on a straight line function of the explanatory variable.
3. The subpopulation standard deviations are all equal (to
4. The selection of an observation from any of the subpopulations is independent of the selection of any other observations.

VIF > 10 is a red flag