Project
STA 2023: Statistics I
Spring 2023

Descriptive Analysis and Hypothesis Testing

Submitted to
Dr./Mr. XXXX
Florida Polytechnic University

Submitted by
Shriraj Mandulapalli
Computer Science: Software Engineering
Florida Polytechnic University

April 26, 2023

Table of Contents (followed by)

1. Introduction

Aim of this project. Include what is included in each of the sections 2 – 6.

2. Data

Include description of data. Also include the descriptive statistics of the data (Q1 and Q2).

3. Distribution of Trips by Purpose

Q3

4. Distribution of Trips by Mode of Transportation

Q4

5. Time of Day Distribution

Q5

6. Trip Length Distributions (Time and Distance)

Q6

7. Statistical Significance or Hypothesis Testing

Q7

8. Conclusion

Recap of the analysis done and highlight the conclusions.

Introduction:

The aim of this project includes the analysis of given data that corresponds with different groups / parameters regarding taking trips. The data includes how different trips are made, regarding transportation methods, from home or from other places, the age of the person making the trip, the trip distance, the trip time, the amount of trips, the amount of workers, and the gender of the person.

These different events are compared with one another and information is taken and displayed using multiple graphs or statistical outputs. Conclusions can be inferred due to the information displayed, regarding the difference between groups (age, sex, working or not, etc).

Data:

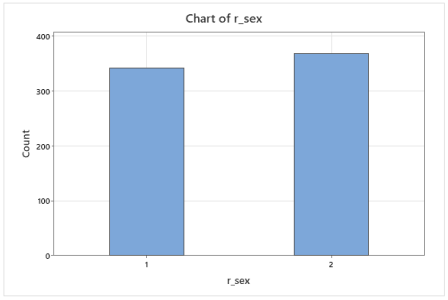
Statistics from Trip File:

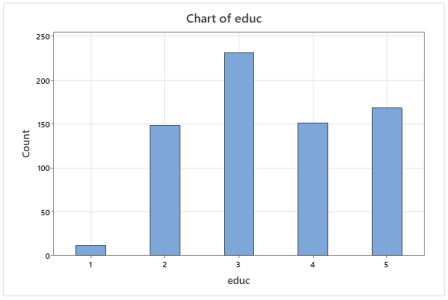
Auto-generated Statistics of specific variables (trip duration, trip distance, household size, household vehicle count, driver count, number of adults, number of workers and age)

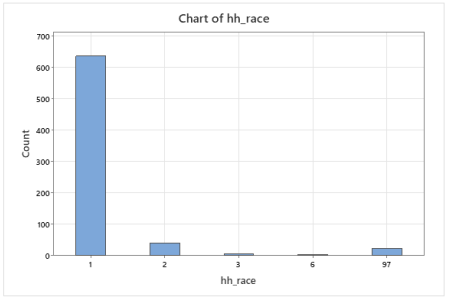
Statistics

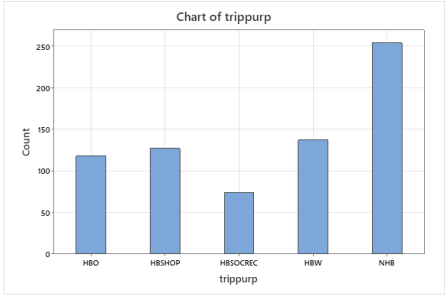
Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
strhr	13.111	4.525	1.000	9.000	13.000	17.000	23.000
trpmiles	8.246	11.240	0.038	1.500	4.196	10.140	92.325
hhsize	2.5338	1.2550	1.0000	2.0000	2.0000	3.0000	7.0000
hhvehcnt	2.2338	1.1186	0.0000	1.0000	2.0000	3.0000	6.0000
drvrcnt	1.8845	0.6724	0.0000	1.0000	2.0000	2.0000	4.0000
numadlt	1.9085	0.5782	1.0000	2.0000	2.0000	2.0000	3.0000
wrkcount	1.2803	0.8672	0.0000	1.0000	1.0000	2.0000	3.0000

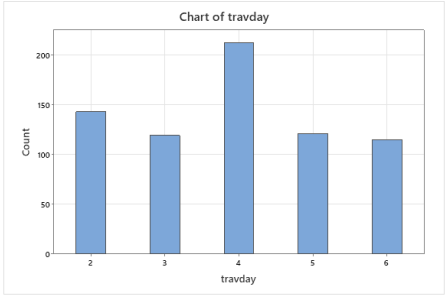
Auto-generated frequency distributions (bar charts) of categorical variables Gender, educational level, race, trip purpose, travel day and trip mode.

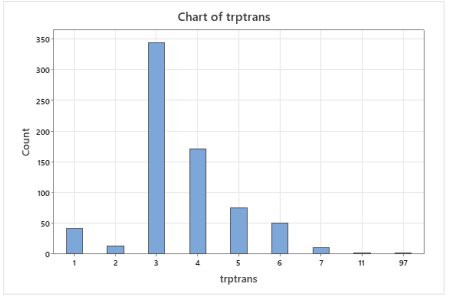






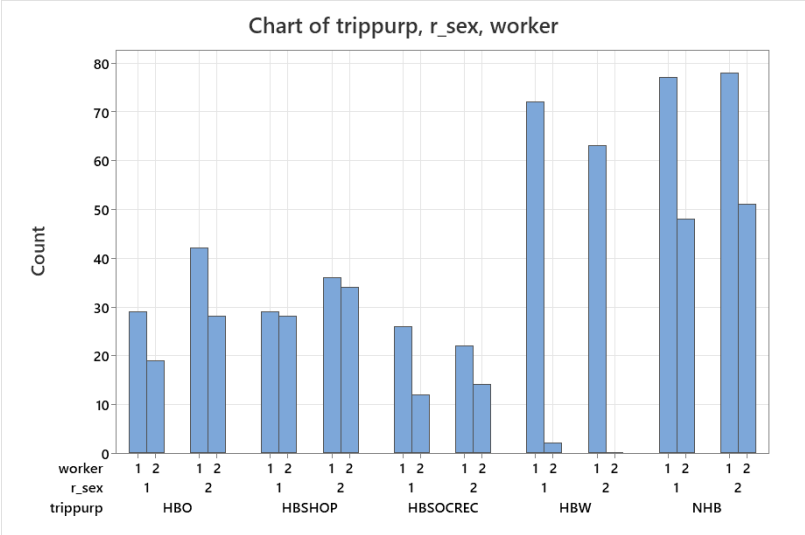






Distribution of Trips by Purpose

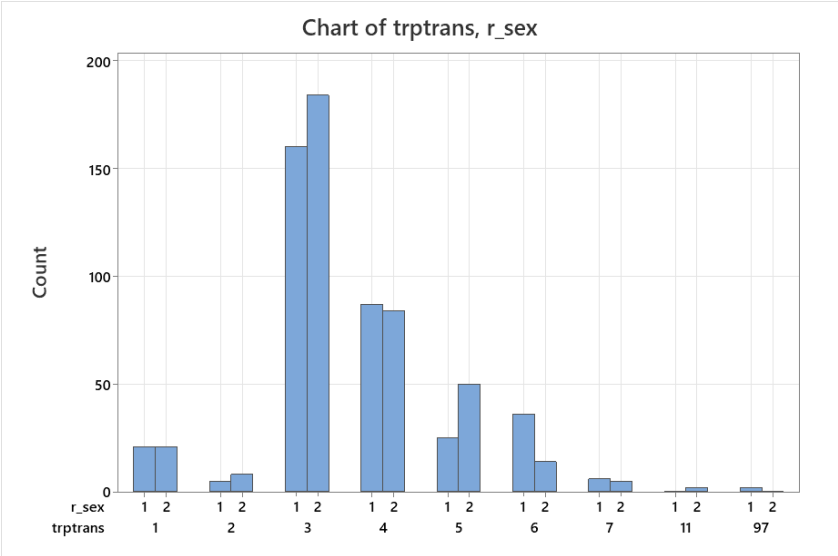
Chart of travel purposes depending on whether the person works and their gender:



Distribution of Trips by Mode of Transportation

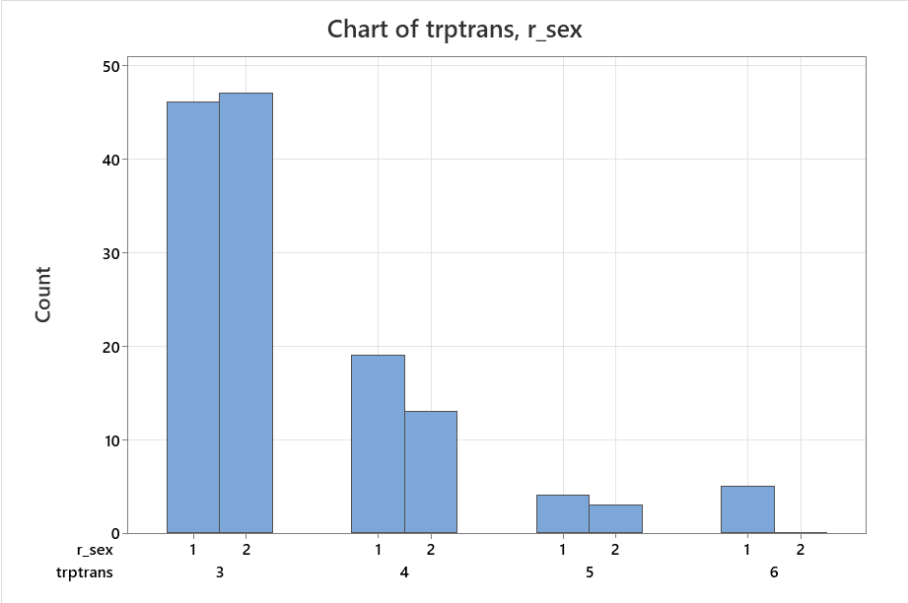
Distribution of all trips by mode of transportation regarding gender:

1 = Walking
2 = Bicycling
3 = Car
4 = SUV
5 = Van
6 = Truck
7 = Golf Cart / Segway
11 = Public/Commuter Bus
97 = Other



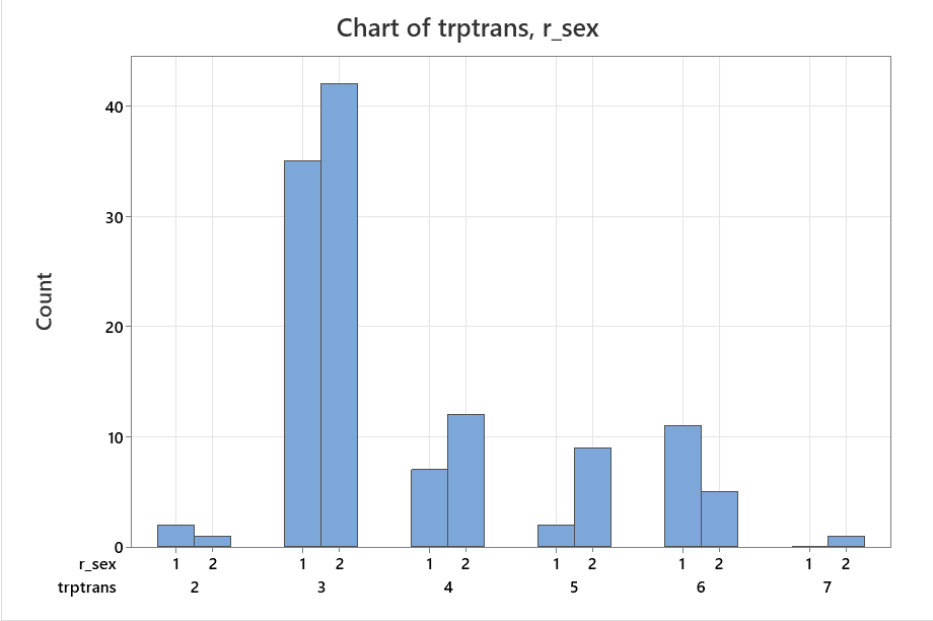
Distribution of home based work by specific mode of transportation regarding gender:

3 = Car
4 = SUV
5 = Van
6 = Truck



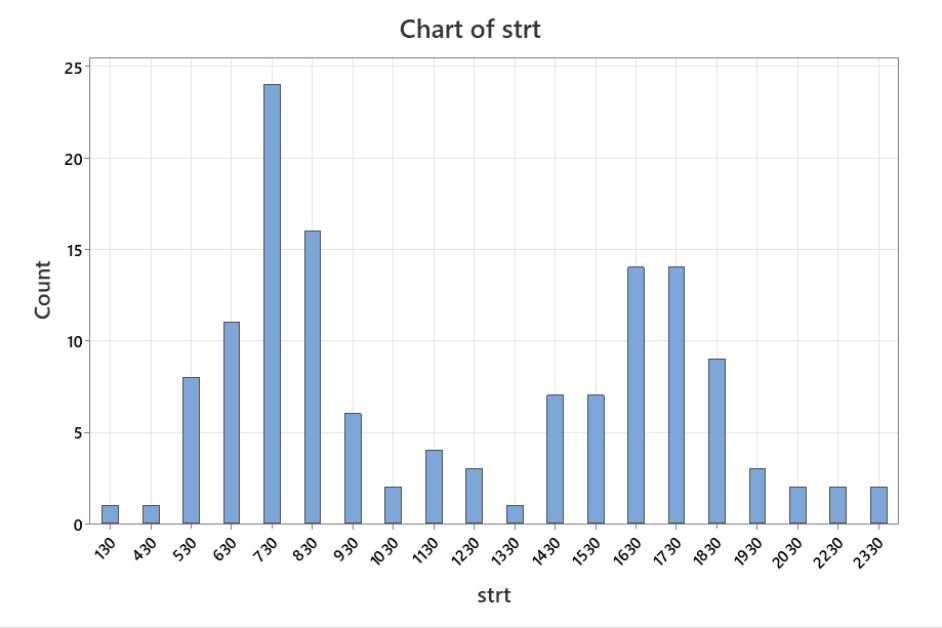
Distribution of home based shopping by specific mode of transportation regarding gender:

2 = Bicycling
3 = Car
4 = SUV
5 = Van
6 = Truck
7 = Golf Cart / Segway

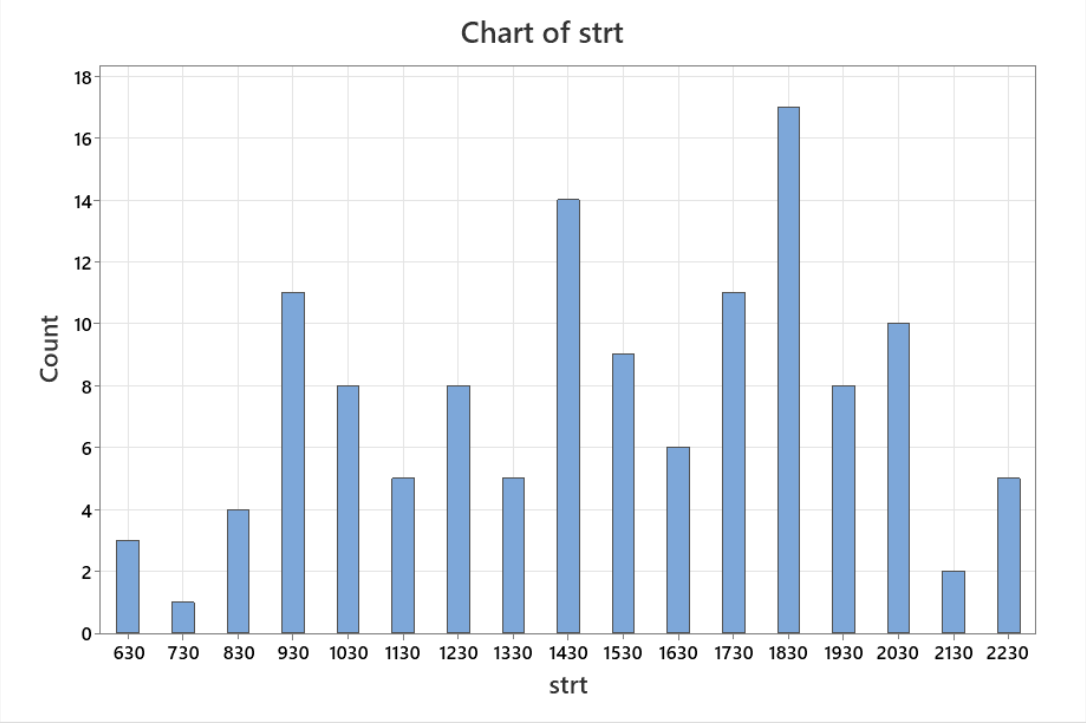


Time of Day Distribution

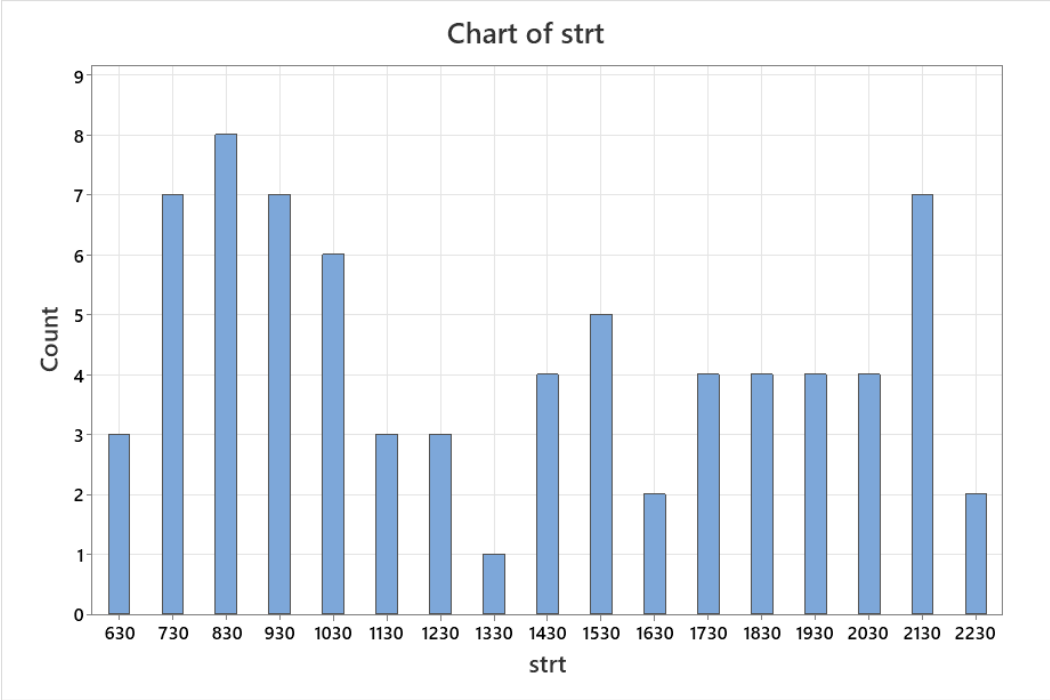
Time of day distribution regarding Home Based Work (In intervals of 30 Min):



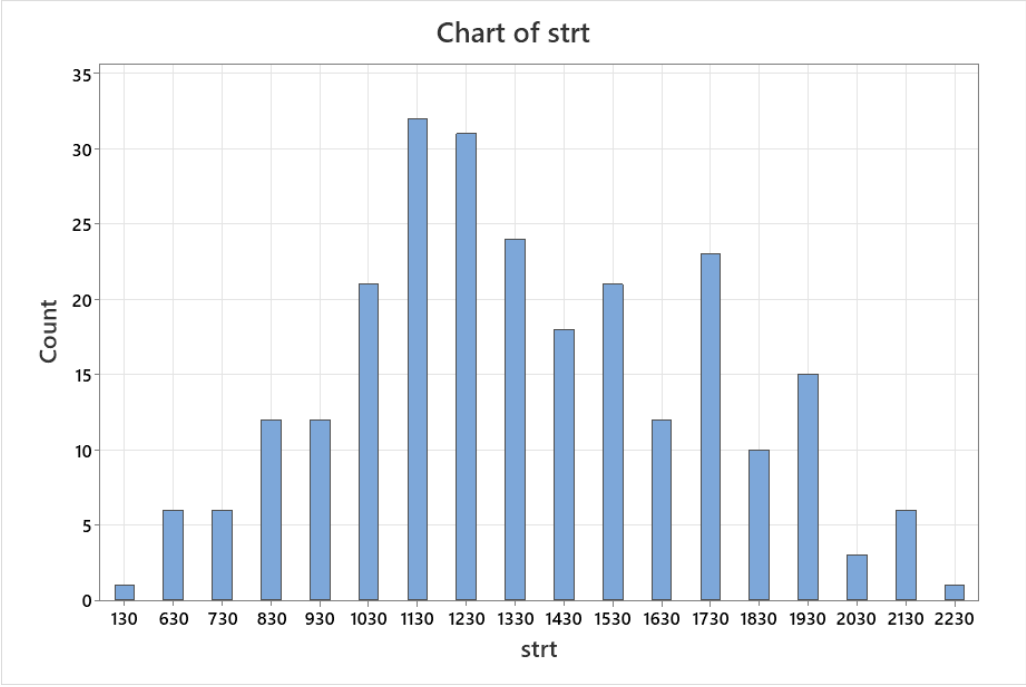
Time of day distribution regarding Home Based Shopping (In intervals of 30 Min):



Time of day distribution regarding Home Based Social Recreation (In intervals of 30 Min):



Time of day distribution regarding Non-Home Based (In intervals of 30 Min)::



Trip Length Distributions (Time and Distance)

Trip length (time and distance) for home based work trips:

Statistics

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
trpmiles	15.68	13.14	0.46	5.43	13.21	22.33	92.33
trvlcmin	34.49	20.82	5.00	20.00	30.00	45.00	120.00

Trip length (time and distance) for all work trips:

Statistics

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
trpmiles	8.246	11.240	0.038	1.500	4.196	10.140	92.325
trvlcmin	22.465	21.332	1.000	10.000	15.000	30.000	270.000

Trip length (time and distance) for non transit trips:

Statistics

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
trpmiles	8.384	11.316	0.038	1.540	4.257	10.198	92.325
trvlcmin	22.511	21.218	2.000	10.000	15.000	30.000	270.000

Trip length (time and distance) for transit trips:

Statistics

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
trpmiles	1.828	2.159	0.183	1.067	1.097	1.312	8.850
trvlcmin	20.33	26.89	1.00	5.00	10.00	15.00	74.00

Statistical Significance or Hypothesis Testing

Average daily trip frequencies (male vs female):

Method

μ_1: population mean of trpcnt_sum when r_sex_first_1 = 1
μ_2: population mean of trpcnt_sum when r_sex_first_1 = 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: trpcnt_sum

r_sex_first_1	N	Mean	StDev	SE Mean
1	80	4.28	2.39	0.27
2	83	4.43	2.60	0.29

Estimation for Difference

Difference	95% CI for Difference
-0.159	(-0.930, 0.613)

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.41	160	0.685

Average daily travel time (workers vs non-workers):

Method

μ_1: population mean of trvlcmin_sum when worker_first_1 = 1
μ_2: population mean of trvlcmin_sum when worker_first_1 = 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: trvlcmin_sum

worker_first_1	N	Mean	StDev	SE Mean
1	112	96.2	62.1	5.9
2	51	101.5	75.2	11

Estimation for Difference

Difference	95% CI for Difference
-5.2	(-29.2, 18.8)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.43	82	0.665

Average daily travel distance (persons age 50+ vs rest of sample):

Note: Since The Age group was split between Above 50 and Below 50. Everyone that was older than 50 had their age set to 2, and everyone under had their age set to 1.

Method

μ_1: population mean of trpmiles_sum when r_age_first = 1
μ_2: population mean of trpmiles_sum when r_age_first = 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: trpmiles_sum

r_age_first	N	Mean	StDev	SE Mean
1	73	34.9	27.0	3.2
2	90	36.7	32.2	3.4

Estimation for Difference

Difference	95% CI for Difference
-1.83	(-10.98, 7.33)

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-0.39	160	0.694

Conclusion:

In conclusion, the given data can be deciphered to show that the analysis of how trips change depending on specific groups like sex, age, working or not, etc, are all independent of trip time, distance, and method. That's it, there's my conclusion, nothing more to say. Enjoy the rest of your day / night (depending on when this is being graded). :)