Final Project — Milestone 2

ALY 6010 Probability Theory and Statistics

RAHUL AVINASH JADHAV

Northeastern University

College of Professional Studies, Northeastern University, Boston, MA 02115

Contact: jadhav.ra@northeastern.edu

Submitted to Professor: Prof. Roy Wada

Date of submission: 12/05/2021

Introduction

The dataset we are going to use for this assignment is the old dataset provided in Milestone 1. In this assignment we are going to perform t-test on net income of genders based on marital status and compare one sample and 2 sample t test of them.

The dataset we have consist of 8847 observations with 37 variables.

Analysis

Table 1: Summary table of the Variable (Marital status) used.

	MARSTAT_KEY_	GENDER	mean(INCOME_)	mean(NET_WORTH_)	count	proportion
1	WIDOWED	FEMALE	7157.605	25803.67	43	0.811
2	WIDOWED	MALE	16779.7	47570.5	10	0.189
3	Unspecified	MALE	0	0	2	1
4	SEPARATED	FEMALE	10351.79	13462.39	302	0.589
5	SEPARATED	MALE	15727.39	21860.66	211	0.411
6	NEVER MARRIED	FEMALE	14416.95	24001.4	994	0.428
7	NEVER MARRIED	MALE	16279.23	29787.06	1330	0.572
8	MARRIED	FEMALE	14004.01	76639.84	2528	0.518
9	MARRIED	MALE	29175.22	74743.08	2356	0.482
10	DIVORCED	FEMALE	15260.91	25556.46	596	0.556
11	DIVORCED	MALE	19168.32	30811.68	475	0.444

The above table provides mean Net worth and Income of genders based on Marital status. The unique marital Status we have in our dataset are Widowed, Unspecified, Separated, Never Married, Married and divorced. If it is to be observed from the above table, we can see that the mean income of the married Male is the highest whereas the mean net worth of the married female is highest.

- **T TEST BASED ON MARITAL STATUS**

Table 2: One sample t-test of mean net worth based on Marital status with Alternative as two sided

	Marital_status	Group	count	Mean(networth)	statistic	p_value	conf.low	conf.high
1	Separated	All	4	14010.90	4.51	0.01	5388.28	22633.53
2	Separated	Female	1	10523.49	2.04	0.29	-55042.61	76089.58
3	Separated	Male	2	16335.85	4.02	0.06	-1150.84	33822.54
4	MARRIED	All	5	60088.95	7.69	0.00	40007.30	80170.61
5	MARRIED	Female	2	61338.85	5.46	0.03	13007.90	109669.80
6	MARRIED	Male	2	58839.05	4.42	0.05	1530.33	116147.78
7	WIDOWED	All	3	39838.99	2.39	0.10	-13171.95	92849.92
8	WIDOWED	Female	1	25097.01	20.65	0.03	9653.16	40540.87
9	WIDOWED	Male	1	54580.96	1.56	0.36	-390800.64	499962.55
10	DIVORCED	All	5	26324.48	7.09	0.00	16784.88	35864.08
11	DIVORCED	Female	2	20744.02	6.05	0.03	5986.34	35501.71
12	DIVORCED	Male	2	31904.94	6.26	0.02	9984.91	53824.97
13	NEVER MARRIED	All	3	24958.00	7.05	0.01	13690.16	36225.84
14	NEVER MARRIED	Female	1	25836.98	10.17	0.06	-6451.74	58125.69
15	NEVER MARRIED	Male	1	24079.02	2.94	0.21	-80091.12	128249.17

From table 2 we can see one same t test performed based on marital status.

- Marital Status Separated - When performed T test on the whole dataset, we could see that the mean net worth was 14010.90 with t statistic as 4.51 and p value as 0.01.When performed t test on dataset of Gender female, the mean net worth was 10523.49 with t statistic as 2.04 and p value as 0.29.When performed T test on dataset of gender male, the mean net worth was 16.335.85 with t statistic as 4.02 and p value as 0.06.
- Marital Status Married - When performed T test on the whole dataset, we could see that the mean net worth was 60088.95 with t statistic as 7.69 and p value as 0.00.When performed t test on dataset of Gender female, the mean net worth was 61338.85 with t statistic as 5.46 and p value as 0.03.When performed T test on dataset of gender male, the mean net worth was 58839.05 with t statistic as 4.42 and p value as 0.05.
- Marital Status Widowed - When performed T test on the whole dataset, we could see that the mean net worth was 39838.99 with t statistic as 2.39 and p value as 0.10.When performed t test on dataset of Gender female, the mean net worth was 25097.01 with t statistic as 20.65 and p value as 0.03.When performed T test on dataset of gender male, the mean net worth was 54580.96 with t statistic as 1.56 and p value as 0.36.

- Marital Status Divorced - When performed T test on the whole dataset, we could see that the mean net worth was 26324.48 with t statistic as 7.09 and p value as 0.00.When performed t test on dataset of Gender female, the mean net worth was 20744.02 with t statistic as 6.05 and p value as 0.03.When performed T test on dataset of gender male, the mean net worth was 31904.94 with t statistic as 6.26 and p value as 0.02.
- Marital Status Never Married - When performed T test on the whole dataset, we could see that the mean net worth was 24958 with t statistic as 7.05 and p value as 0.01.When performed t test on dataset of Gender female, the mean net worth was 25836.98 with t statistic as 10.17 and p value as 0.06.When performed T test on dataset of gender male, the mean net worth was 24079.02 with t statistic as 2.94 and p value as 0.21.

Table 3: Welch Two Sample t-test on mean net worth based on Marital status with alternative and two sided

	Maritial_status	estimate	Mean net worth(Female)	Mean net worth(Male)	count	statistic	p.value	conf.low	conf.high
1.00	Separated	-5812.36	10523.49	16335.85	2.20	-0.88	0.46	-31731.36	20106.63
2.00	married	2499.79	61338.85	58839.05	3.89	0.14	0.89	-46424.11	51423.70
3.00	widowed	-29483.95	25097.01	54580.96	1.00	-0.84	0.55	-472614.70	413646.81
4.00	divorced	-11160.92	20744.02	31904.94	3.50	-1.82	0.15	-29207.87	6886.03
5.00	never married	1757.95	25836.98	24079.02	1.19	0.20	0.87	-73608.42	77124.32

From table 3 we can see Two same t test performed based on marital status

- Marital Status Separated – When performed two sample t test on mean net worth of female and male, we could see that the estimate was -5812.36. the mean net worth of female was 10523.49 and 16335.85 was of male. The count was 2.20 with Statistic of -0.88 and p value of 0.46. The confidence low was -31731.36 and confidence high was 20106.63

- Marital Status Married – When performed two sample t test on mean net worth of female and male, we could see that the estimate was 2499.79. the mean net worth of female was 61338.85 and 58839.05 was of male. The count was 3.89 with Statistic of 0.14 and p value of 0.89. The confidence low was -46424.11 and confidence high was 51423.70

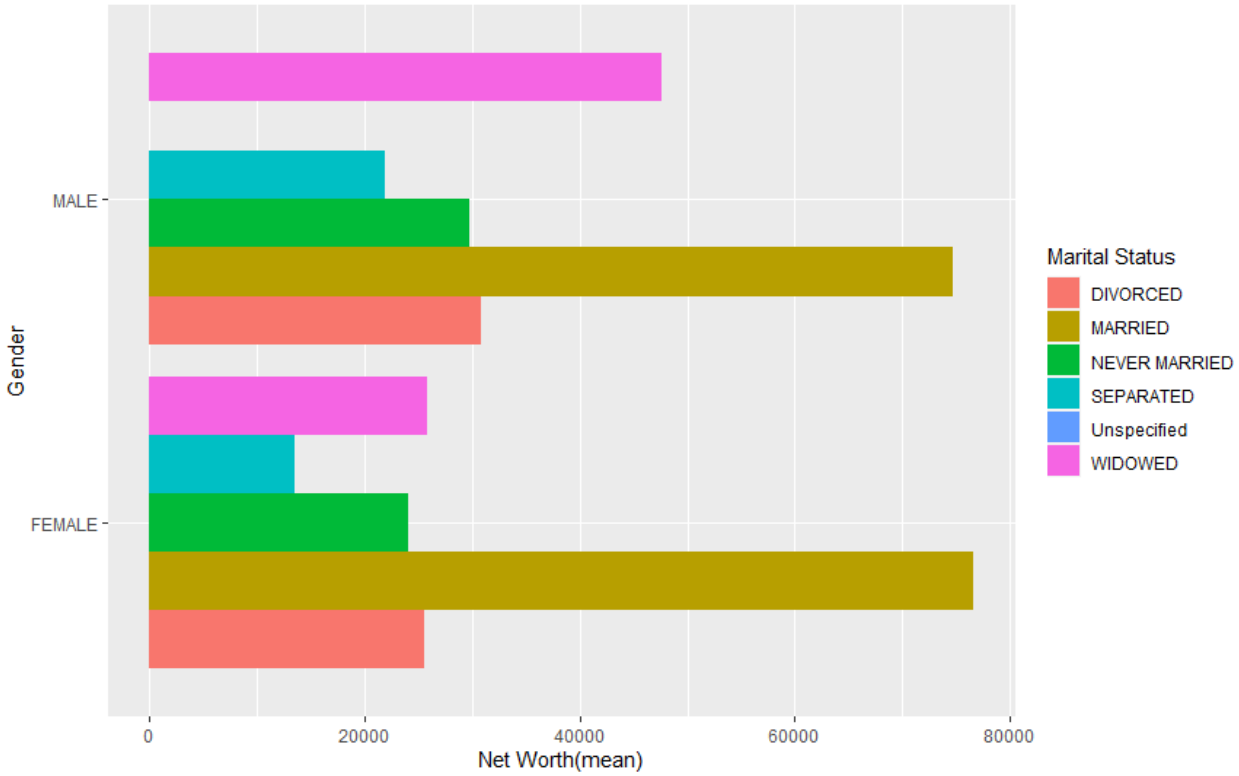
- Marital Status Widowed – When performed two sample t test on mean net worth of female and male, we could see that the estimate was -29483.95. the mean net worth of female was 25097.01 and 54580.96 was of male. The count was 1.00 with Statistic of -0.84 and p value of 0.55. The confidence low was -472614.70 and confidence high was 413646.81

- Marital Status Divorced – When performed two sample t test on mean net worth of female and male, we could see that the estimate was -11160.92. the mean net worth of female was 20744.02 and 31904.94 was of male. The count was 3.50 with Statistic of -1.82 and p value of 0.15. The confidence low was -29207.87 and confidence high was 6886.03

- Marital Status Never Married – When performed two sample t test on mean net worth of female and male, we could see that the estimate was 1757.95. the mean

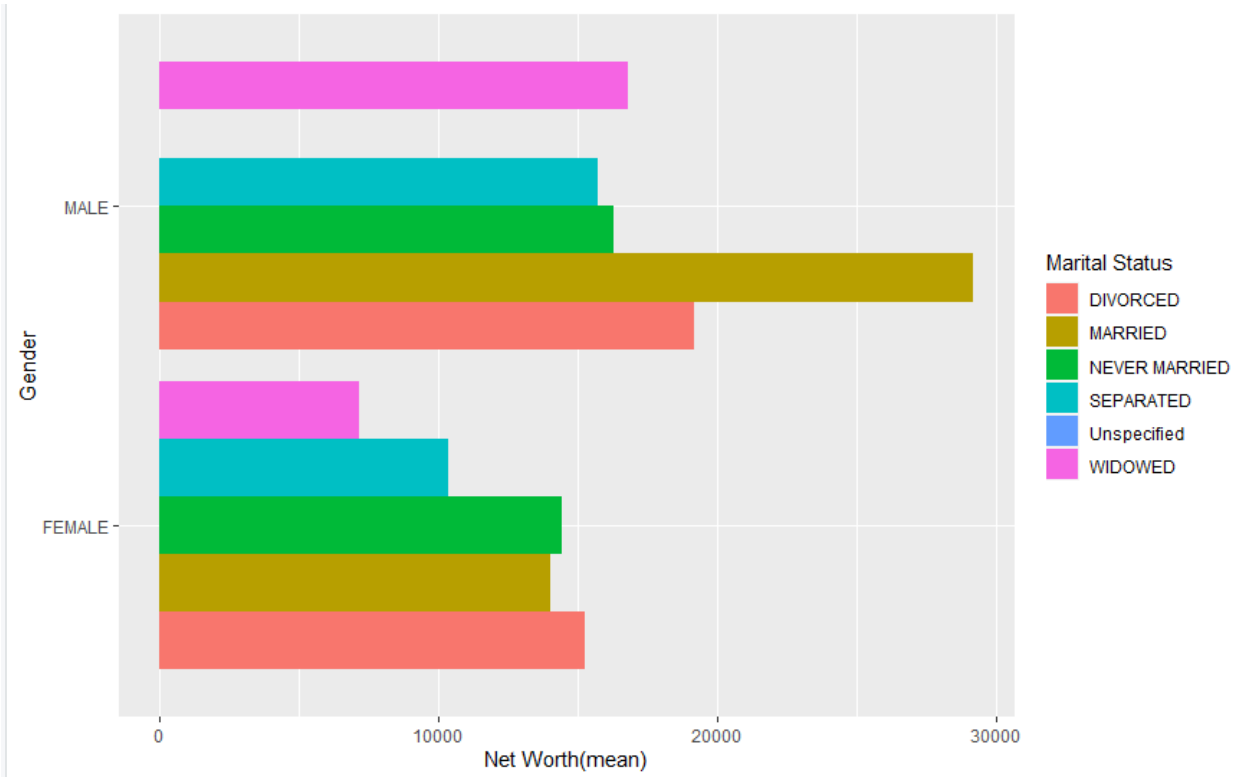
net worth of female was 25836.98 and 24079.02 was of male. The count was 1.19 with Statistic of 0.20 and p value of 0.87. The confidence low was -73608.42 and confidence high was 77124.32

Bar plot 1: Net worth based on Gender filled with Marital Status.



From Bar plot 1, we can see that the net worth of married female is more , where as net worth of separated female are the least.

Bar plot 2: Income worth based on Gender filled with Marital Status.



From bar plot 2 we can see that the income of married man is the highest where as the income of widowed female is the lowest.

Conclusion

- We have learned to do one sample and 2 sample t test on variable and got opportunity to compare them.
- Using one sample t test we did the testing on them, where as 2 sample t test is used to compare 2 groups and provide the t test for them.
- For our tests the hypothesis was the mean of the samples was 0 (default value) and the null hypothesis was the mean of the same was not 0 we did t test and rejected the null hypothesis and states that there was no evidence that the mean of the samples was 0.

Bibliography

- Kabacoff, R. (2011). *R in action: Data analysis and graphics with R*. Manning.