ABDULLAH GUL UNIVERSITY

A Statistical Analysis of The Relationship Between Spending Time on Social Media and Number of Followers

---

MATH301 – TERM PROJECT

AHMET YASİR CİLVEZ

ALİ NURETTİN DEMİR

---

1. Introduction

Social Media is starting become most important part of communication from early 2000's. Especially, smart phone technologies are very effective for spreading of using social media websites and applications. Social media using in Turkey also increased during this big spreading. Anadolu University SODIGEM's research shows that 52 million active social media users are in Turkey in 2017. It is 60% of total Turkish population. That's why researches on social media using is most popular in recent years.

Social media also have a very effective role of psychological situation of people. Sharings, texts, photos are shared for more taking attention of followers and increase followers numbers, generally. Most of people think that number of followers is an evidence of being an important people or not. Therefore, spending time on social media is increasing. In this research project, the relationship between spending time on social media and number of followers of male and female users is aimed to examine.

2. Methods and Materials

A. Data Collection

In this research, best way to collect data is making an online survey. Using Google forms, the form questions were that:

1) What is your gender?
2) Do you use social media?
3) How many hours do you spend in social media?
4) How many followers – friends do you have?

The participants are divided by male and female. 39 male and 54 female participants filled the survey. Graph of participants are shown in figure 1.

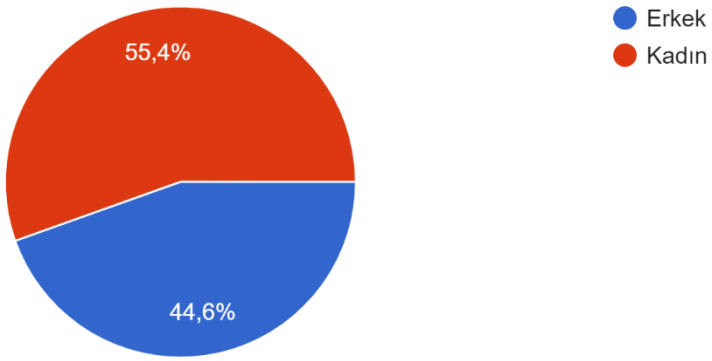


Figure 1 Graph shows the percentage of genders.

The distributions of answer of "How many hours do you spend in social media?" for two genders are shown in figure 2.

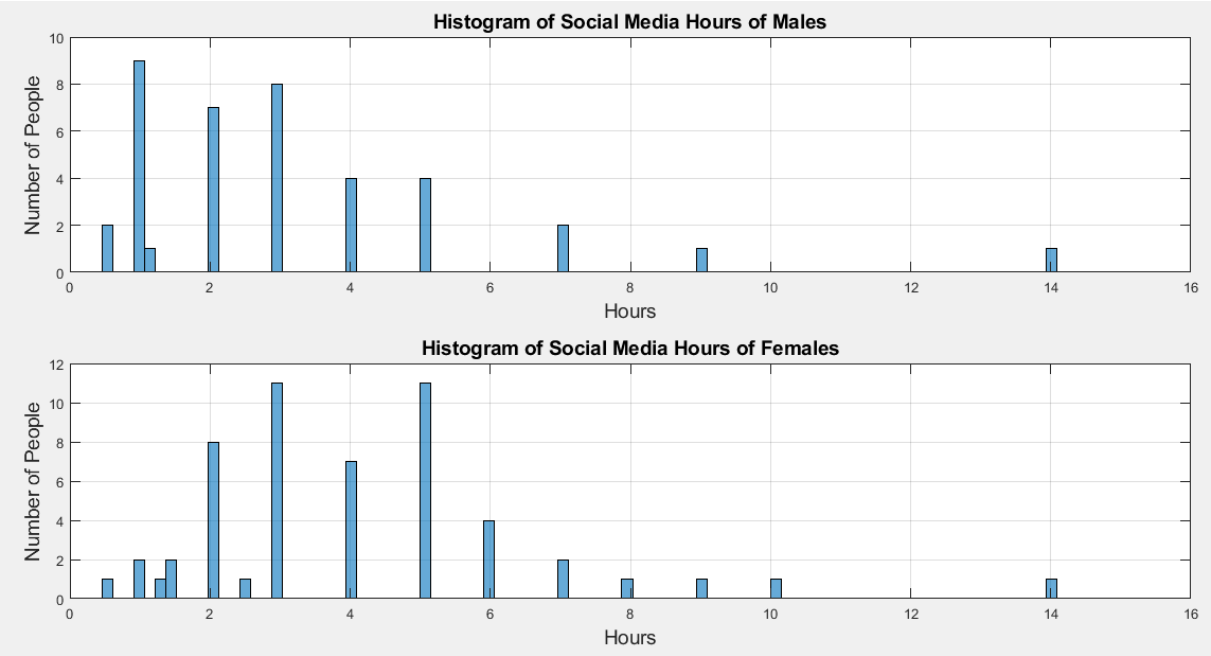


Figure 2 Distribution of spending hours

The distributions of answer of "How many followers – friends do you have?" for two genders are shown in figure 3.

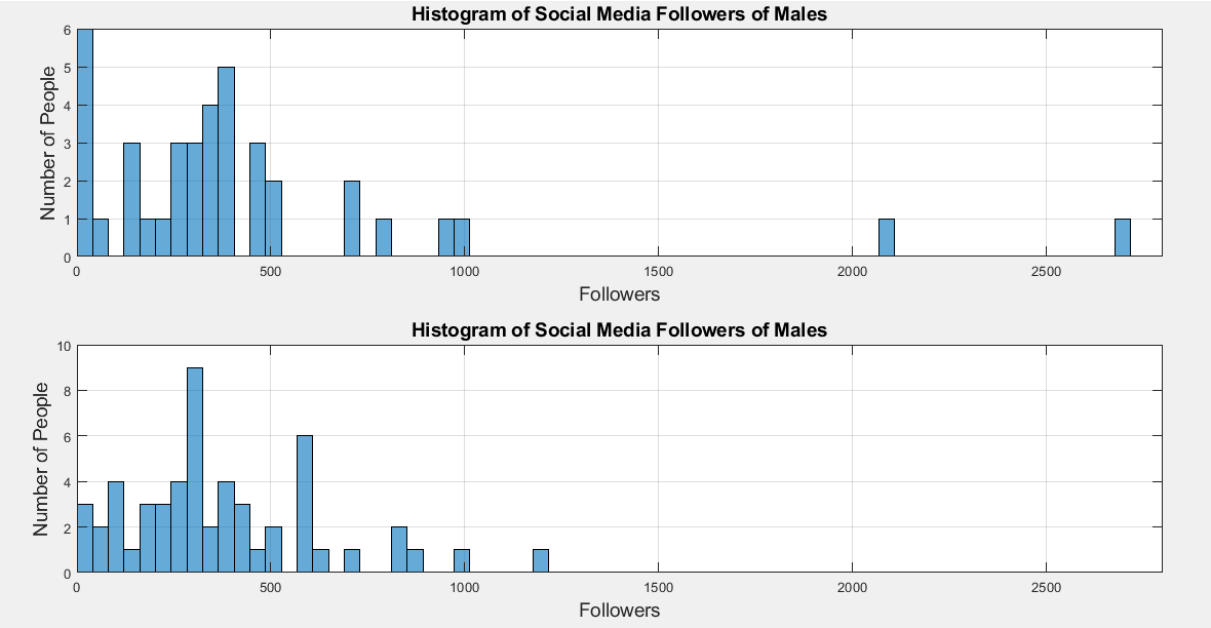


Figure 3 Distribution of number of followers

The Matlab code of histogram graphs are shown below.

```
edges = linspace(0, 15, 100);
subplot(2,1,1)
histogram(A, 'BinEdges',edges);
grid on;
xlim([0, 16]);
xlabel('Hours', 'FontSize', 14);
ylabel('Number of People', 'FontSize', 14);
title('Histogram of Social Media Hours of Males', 'FontSize', 14);
hold on;

subplot(2,1,2)
histogram(B, 'BinEdges',edges);
grid on;
xlim([0, 16]);
xlabel('Hours', 'FontSize', 14);
ylabel('Number of People', 'FontSize', 14);
title('Histogram of Social Media Hours of Females', 'FontSize', 14);
```

Figure 4 Distribution of spending hour code

```
edge = linspace(0, 3000, 75); % Create 20 bins.
subplot(2,1,1)
histogram(C, 'BinEdges',edge);
grid on;
xlim([0, 2800]);
xlabel('Followers', 'FontSize', 14);
ylabel('Number of People', 'FontSize', 14);
title('Histogram of Social Media Followers of Males', 'FontSize', 14);
hold on;

subplot(2,1,2)
histogram(D, 'BinEdges',edge);
grid on;
xlim([0, 2800]);
xlabel('Followers', 'FontSize', 14);
ylabel('Number of People', 'FontSize', 14);
title('Histogram of Social Media Followers of Males', 'FontSize', 14);
hold on;
```

Figure 5 Distribution of followers code

3. Results

A. Sample Variance and Mean

Calculation methods are used on collected data. These calculations and results are explained below. By calculations Matlab is used. The codes are shown below.

```
A = table2array(erkeklersorubir);
B = table2array(kadinlarbir);
meanA = mean(A)
meanB = mean(B)
varA  = var(A)
varB  = var(B)
C = table2array(erkeklersoruiki);
D = table2array(kadinlariki);
meanC = mean(C)
meanD = mean(D)
varC  = var(C)
varD  = var(D)
```

Sample mean and sample variance calculated by following formulas:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$$

Spending Hours	Mean	Variance
Males	3,1333	7,0275
Females	4,0796	5,9801

Number of Followers	Mean	Variance
Males	451,0769	274760
Females	378,2778	64901

B. T – Interval

A t interval should be used to estimate an unknown population mean. This occurs as a result of our sample standard deviation in calculating our margin of error. We have an average based on the data in this survey and we will use t-interval to estimate the true average in the population. T – interval is calculated by following formulas. The confidence level was 95%.

$$\overline{X} \pm t\frac{s}{\sqrt{n}}$$

$$Lower\ bound: (\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_{12}^2}{n_2}}$$

$$Upper\ bound: (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_{12}^2}{n_2}}$$

The Matlab code of calculating T – interval is shown below.

```
SEM = std(B)/sqrt(length(B))          % Standard Error
ts = tinv([0.025  0.975],length(B)-1) % T-Score
CI = mean(B) + ts*SEM                 % Confidence Intervals
```

Spending Hours	T - Interval		Confidence Interval		SEM (Standard Error)
Gender	**Lower**	**Upper**	**Lower**	**Upper**	
Males	-2,0244	2,0244	2,274	3,9927	0,4245
Females	-2,0057	2,0057	3,4122	4,7471	0,3328

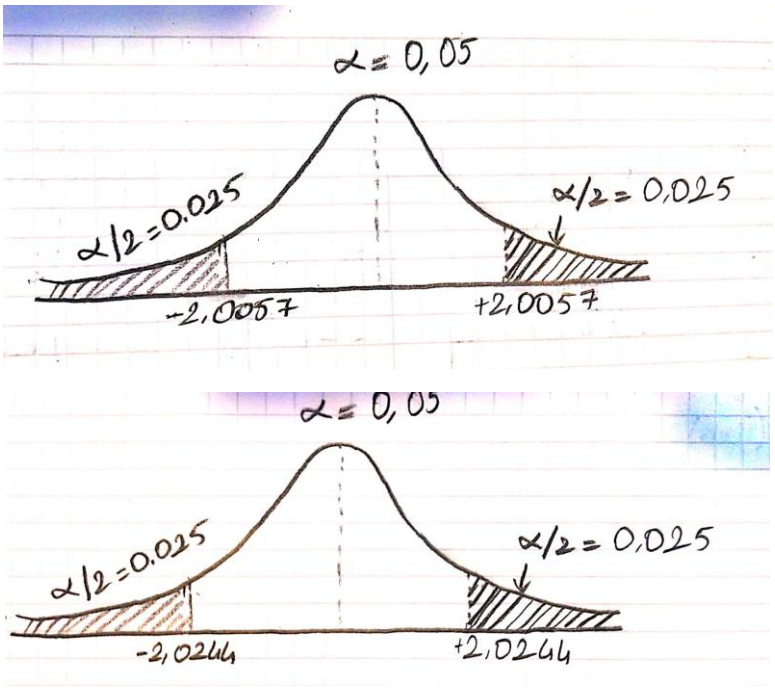
Number of Followers	T - Interval		Confidence Interval		SEM (Standard Error)
Gender	**Lower**	**Upper**	**Lower**	**Upper**	
Males	-2,0244	2,0244	281,1602	2,0244	83,9346
Females	-2,0057	2,0057	308,7424	447,8132	34,6681

C. Hypothesis Test

We will use the hypothesis testing to test the results of our survey to see if there are meaningful results. Basically, we will test whether our results are valid by examining the probability that a small portion of our results will occur by chance.

We conducted a hypothesis test to compare the impact of women and men social media usage times on their followers. You can see the use of hypothesis testing below for null and alternative hypothesis.

	2-Tailed Test	Right-Tailed	Left Tailed
Null hypothesis	$H_0: \mu = \mu_0$	$H_0: \mu \leq \mu_0$	$H_0: \mu \geq \mu_0$
Alternative hypothesis	$H_a: \mu \neq \mu_0$	$H_a: \mu > \mu_0$	$H_a: \mu < \mu_0$



D. Correlation Scores

Correlation values take values between +1 and -1. In the case of values leading up to the ends, the relationship between the two variables of the bivariate distribution is considered to be strong. Positive (+) values explain that the relationship between the two variables is directly proportional, Negative (-) values explain that the relationship is inversely proportional.

The correlation scores of this research is negative, therefore the relationship between both genders' spending hour and follower number have an opposite rate. A result of it, when female's spending our is increasing their number of followers are decreases and this changing has sharper rate than male's.

Gender	Correlation Scores
Males	-0,0319
Females	-0,0867

E. Regression Models

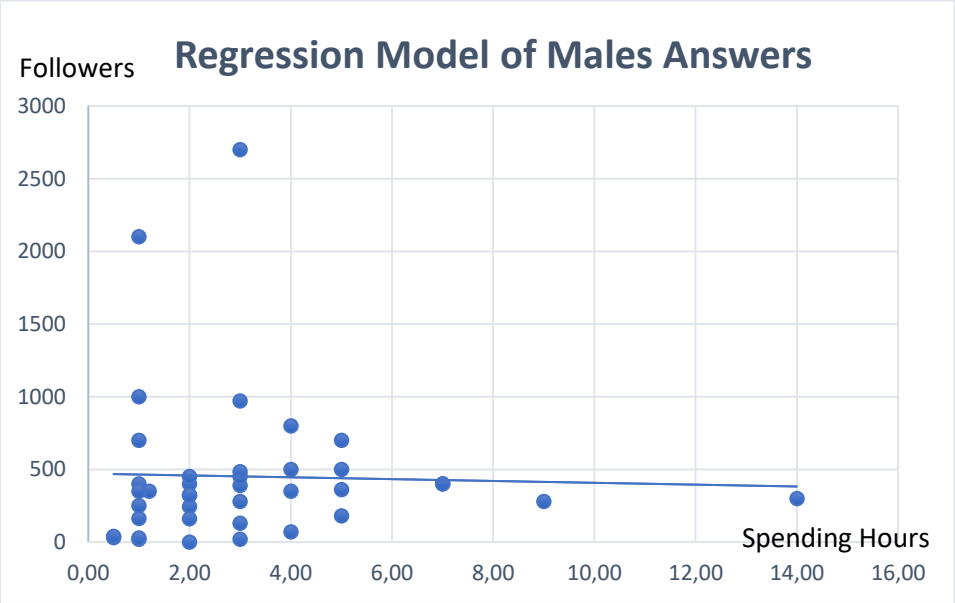
Regression is the target value modeling method. It is a statistical tool used to find the relationship between dependent variable and one or more variables. Regression reveals the line between the dependent variable and predictors with the least error and shows us.

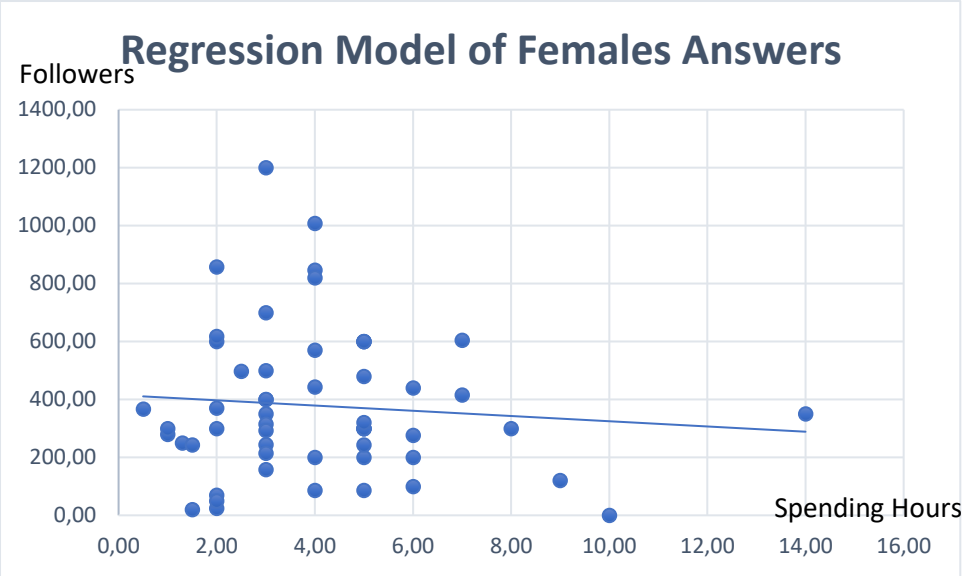
$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\,\bar{x},$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \frac{s_{x,y}}{s_x^2}$$

$$= r_{xy}\frac{s_y}{s_x}.$$





4. Discussion and Conclusion

In this research, the relationship between spending our in social media and the number of followers for male and female users is examined. The data is collected by a form. Some particular Matlab function is used in code design and also Microsoft Excel is used for storage data and making graphs. The research shows that, female users spend more time than male users, even though this fact, there is a negative correlation between spending hours and number of followers. It is estimating that females are spend more time, followers' numbers have a direct connection with spending hour time. At the end of the evaluation of this research it is proved that females spend more time than male but number of followers has an opposite relationship between spending hours. The hypothesis was that as a result of more spending hour is more follower, however, at the end of the research the hypothesis was not accurate.

Link: https://youtu.be/7pHkM_8osM4