AVERAGE TIME SPENT BY A USER ON SOCIAL MEDIA

A DATA ANALYSIS USING SAS

Submitted to: Lynn A. Agre, MPH, PhD

Statistics 390: Section M1

Feb 20, 2024

By: Phan Nguyen Huong Le

RUID: 203005844

Abstract:

First, the study will investigate the relationship between age and time spent on their phone on social media, which is two continuous variables. Second, this research will mainly focus on exploring how age groups affect the time spent on social media, and how other factors such as social media platforms would affect each individual. This analysis will go over many statistical analysis techniques which we have covered in Statistics 390, such as t-test, ANOVA, chi-squared,... to evaluate the association between two variables mentioned above. SAS programming will be used to analyze the data. Conducted results will be presented with both numerical and graphical.

Introduction/ Background Brief:

It has been a concerning problem for adolescents and adults are the vast majority of social media consumers (Booker et al, 2018). In this era, where the Internet and social media, even AI, is something that we unintentionally have been using and becoming dependent on it: it is going to be a real problem if this consumers' number increase over time. Multiple studies have found that there is a link between social media and mental health; social media has negative impacts on young adolescents (Coyne et al., 2020). Increasing social media time might lead to depression and heightened social anxiety levels (Hankin, 2006).

Based on the research above about some association between social media use and depression, it is critical to investigate if there is truly a relation between the two variables. We were able to retrieve a dataset from Kaggle with the related matter, mainly about average time spent on social media, along with some other potential variables that we could use for further research: age, gender, demographics, interests, device_type, location,... There are two main variables that we are interested in the most: age of the user, and their time spent on social media platforms

(Facebook, Instagram, Youtube,...). The data was generated from NumPy libraries on Python, hence there is no particular source of data that we could reference further; however, this could be used as a learning tool and resource for people who are interested in training AI models and for practice purposes.

Methods

In this analysis, we will mainly focus on comparing between each individual in different age groups, and two variables of interest: Age and Time Spent on social media. The primary tool being used to analyze this is SAS.

First, we started off with exploring the data using proc means, the function would return the means procedure for n=30, the mean, max, min, range and standard deviation for Age and Time Spent. Proc freq was also used as a tool to display the frequency variables, count occurrences, analyze categorical variables, and summarize the distribution of data points across different categories. Lastly, we used proc univariate to know more about the measures of central tendency, such as mean, median, mode, along with dispersion, and shape.

```
14 /* Calculate proc means*/
15
16 proc means data = work.phone1 n mean max min range std fw=8;
17 var age time_spent;
18 title 'Summary of Age and Time Spent on social media';
19
```

Summary of Age and Time Spent on social media

The MEANS Procedure

Variable	N	Mean	Maximum	Minimum	Range	Std Dev
age	30	40.1667	62.0000	19.0000	43.0000	13.0148
time_spent	30	4.4000	9.0000	1.0000	8.0000	2.1751

```
20 /* Calculate proc freq*/
21
22 proc freq data = work.phone1;
23 tables age time_spent;
24 run;
```

Summary of Age and Time Spent on social media

The FREQ Procedure

age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
19	1	3.33	1	3.33
20	1	3.33	2	6.67
21	1	3.33	3	10.00
22	1	3.33	4	13.33
23	1	3.33	5	16.67
24	1	3.33	6	20.00
25	1	3.33	7	23.33
30	1	3.33	8	26.67
32	1	3.33	9	30.00
36	3	10.00	12	40.00
37	1	3.33	13	43.33
40	3	10.00	16	53.33
43	1	3.33	17	56.67
45	1	3.33	18	60.00
46	2	6.67	20	66.67
47	1	3.33	21	70.00
49	1	3.33	22	73.33
52	2	6.67	24	80.00
54	2	6.67	26	86.67
56	2	6.67	28	93.33
62	2	6.67	30	100.00

time_spent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2	6.67	2	6.67
2	6	20.00	8	26.67
3	2	6.67	10	33.33
4	6	20.00	16	53.33
5	6	20.00	22	73.33
6	2	6.67	24	80.00
7	3	10.00	27	90.00
8	2	6.67	29	96.67
9	1	3.33	30	100.00

```
26 /* Calculate proc univariate*/
27
28 proc univariate data = work.phone1 normal plot;
29 var age time_spent;
30 run;
31
```

Quantiles (Definition 5)	
Level	Quantile
100% Max	62.0
99%	62.0
95%	62.0
90%	56.0
75% Q3	52.0
50% Median	40.0
25% Q1	30.0
10%	21.5
5%	20.0
1%	19.0
0% Min	19.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
19	27	54	23
20	5	56	10
21	2	56	20
22	22	62	8
23	4	62	17

Summary of Age and Time Spent on social media

The UNIVARIATE Procedure
Variable: age

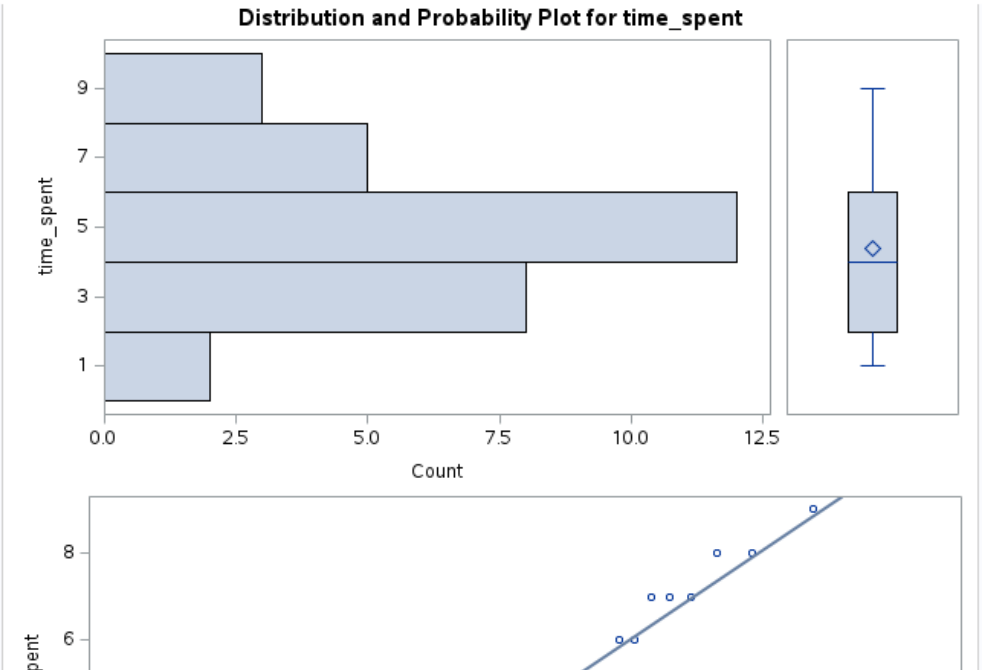
Moments			
N	30	Sum Weights	30
Mean	40.1666667	Sum Observations	1205
Std Deviation	13.0148015	Variance	169.385057
Skewness	-0.107849	Kurtosis	-1.0633763
Uncorrected SS	53313	Corrected SS	4912.16667
Coeff Variation	32.4019954	Std Error Mean	2.37616678

Basic Statistical Measures			
Location		Variability	
Mean	40.16667	Std Deviation	13.01480
Median	40.00000	Variance	169.38506
Mode	36.00000	Range	43.00000
		Interquartile Range	22.00000

Note: The mode displayed is the smallest of 2 modes with a count of 3.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	16.90398	Pr > |t|	<.0001
Sign	M	15	Pr >= |M|	<.0001
Signed Rank	S	232.5	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.951185	Pr < W	0.1819
Kolmogorov-Smirnov	D	0.111393	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.051237	Pr > W-Sq	>0.2500



From homework 2, we also used proc means, recoding techniques, proc freq: cross tabulations and proc corr, to understand our two variables better. This time, instead of looking at min, max, and other classifications that we have analyzed above, we are particularly interested in using it to calculate the t-value. Since our variables of interest are continuous variables, it is essential to recode it to categorical variables for further analysis. Hence, the two variables were recoded from Age to rAge, and time_spent to RTime, with personalized cutoffs technique learned in class. Moreover, we were able to perform cross tabulation analysis using the recoded variables, which return frequency, percent, row and columns per specified. Finally, we used proc corr to compute correlation coefficients between Age and Time Spent.

The FREQ Procedure

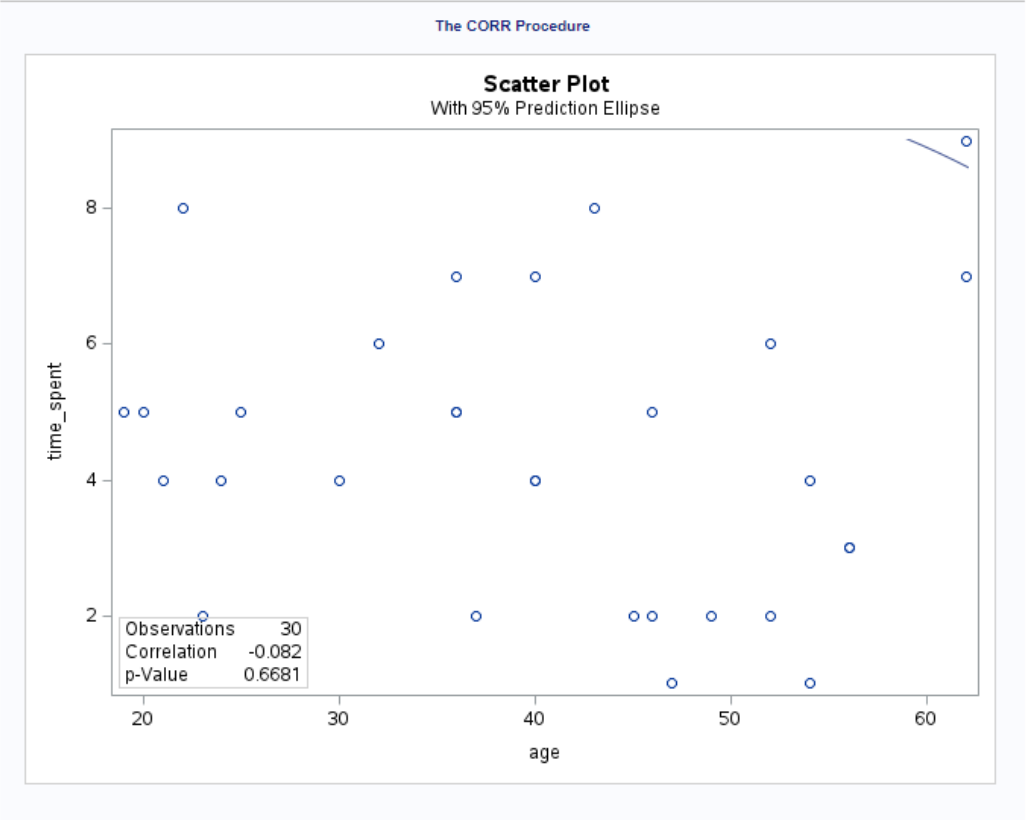
Frequency Percent Row Pct Col Pct	Table of rTime by rAge			
	rAge			
rTime	Mid	Old	Young	Total
frequent	5 15.15 35.71 41.67	3 9.09 21.43 30.00	6 18.18 42.86 54.55	14 42.42
not frequent	4 12.12 36.36 33.33	4 12.12 36.36 40.00	3 9.09 27.27 27.27	11 33.33
very frequen	3 9.09 37.50 25.00	3 9.09 37.50 30.00	2 6.06 25.00 18.18	8 24.24
Total	12 36.36	10 30.30	11 33.33	33 100.00

The CORR Procedure

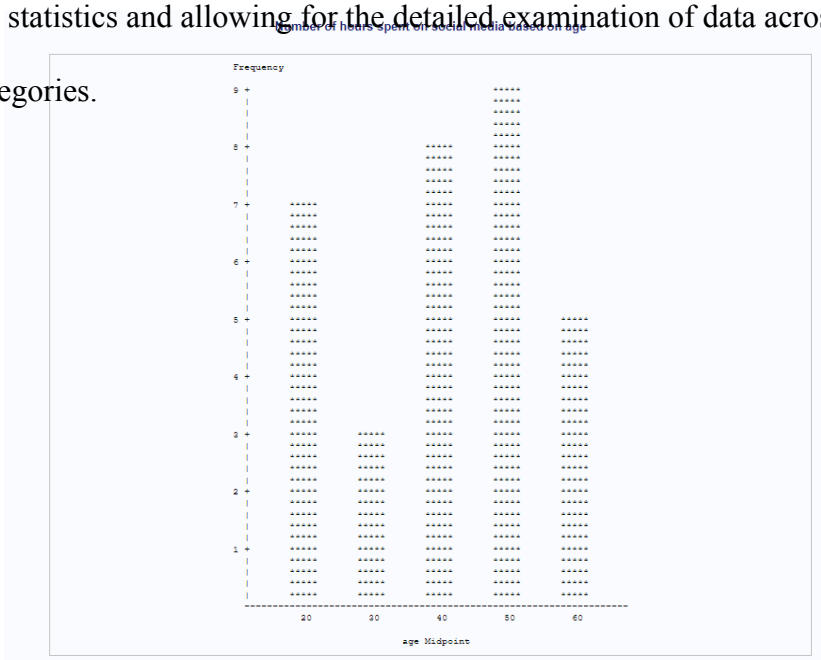
2 Variables:	age time_spent

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
age	32	40.18750	13.72175	1286	19.00000	62.00000
time_spent	30	4.40000	2.17509	132.00000	1.00000	9.00000

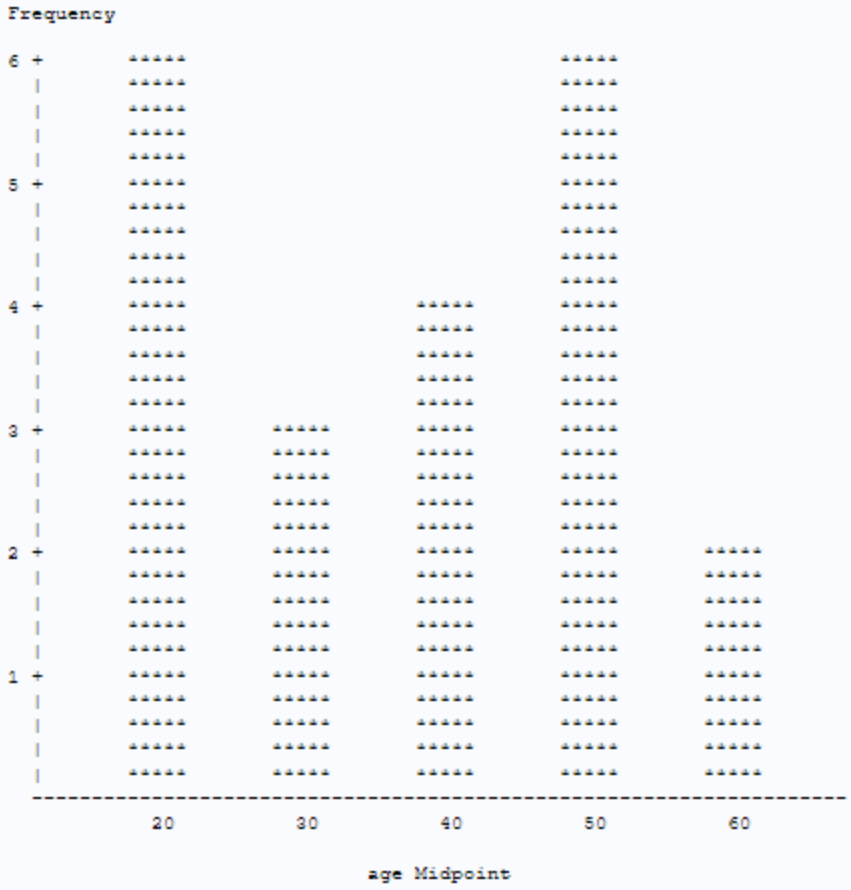
Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	age	time_spent
age	1.00000 32	-0.08161 0.6681 30
time_spent	-0.08161 0.6681 30	1.00000 30



In homework 3, we went over how to use SAS for data visualization, such as the function of chart, histogram, boxplot, scatterplot. Furthermore, we got to explore how to use proc tabulate for summarizing data in a tabular format, providing a comprehensive overview of the dataset through descriptive statistics and allowing for the detailed examination of data across different dimensions and categories.



Vertical Bar Chart of Age



Vertical Bar Chart of Age

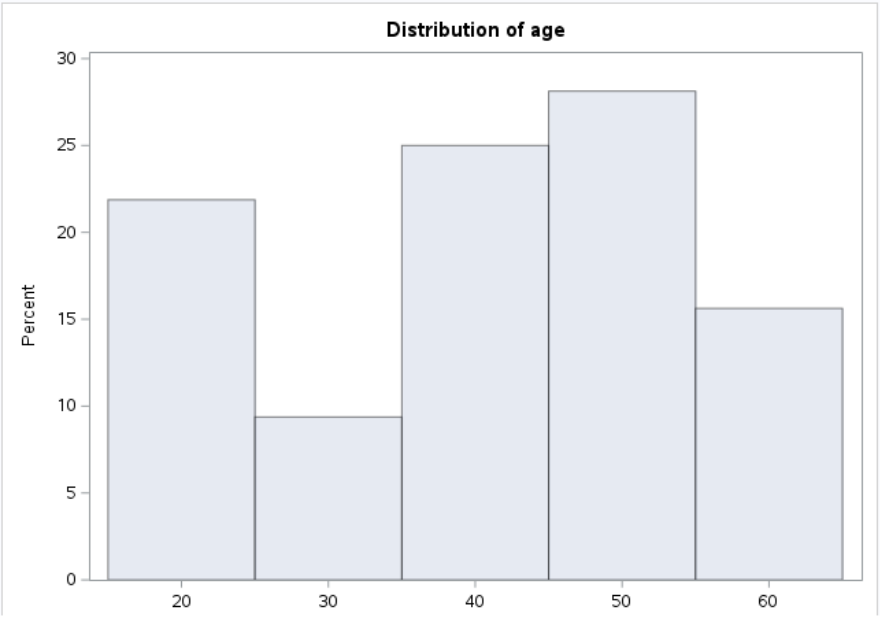
	time_spent					
	rTime					
	frequent		not frequent		very frequen	
	N	Mean	N	Mean	N	Mean
rAge						
Mid	5	4.60	4	1.75	3	7.33
Old	3	3.33	3	1.67	3	7.33
Young	6	4.50	1	2.00	2	7.00

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
19	32	56	10
19	27	56	20
20	5	62	8
21	2	62	17
22	22	62	33

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	1	3.03	100.00

Number of hours spent on social media based on age

The UNIVARIATE Procedure

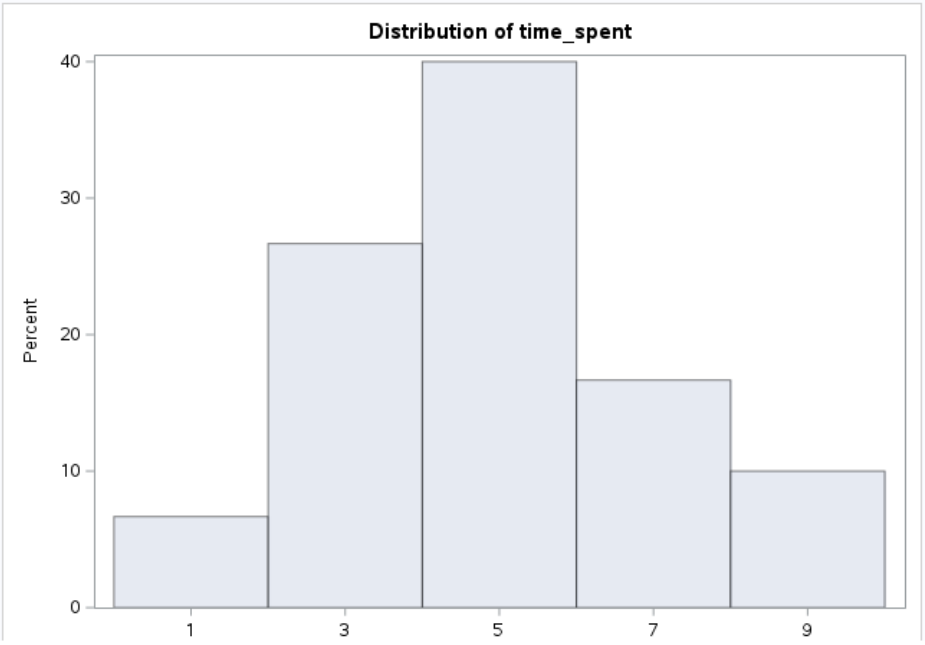


Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	25	7	8
1	9	7	16
2	26	8	13
2	24	8	22
2	21	9	17

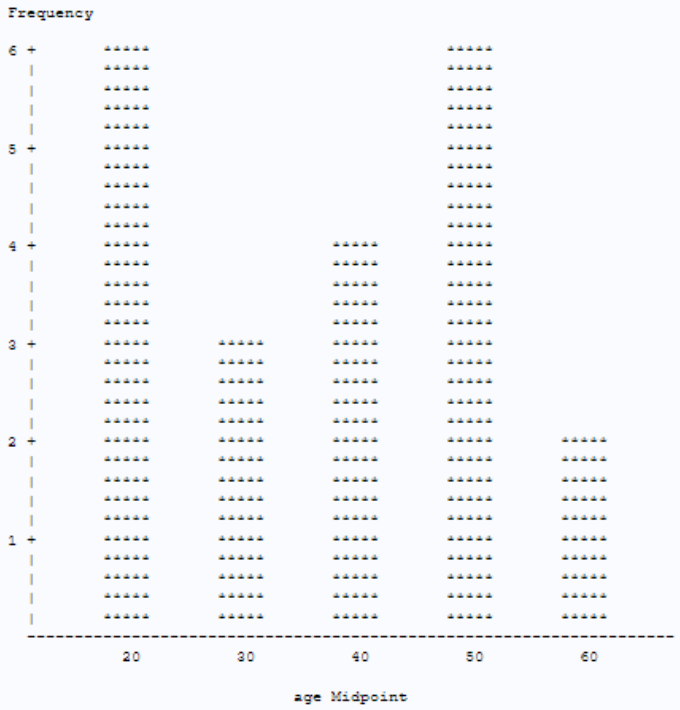
Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	3	9.09	100.00

Number of hours spent on social media based on age

The UNIVARIATE Procedure

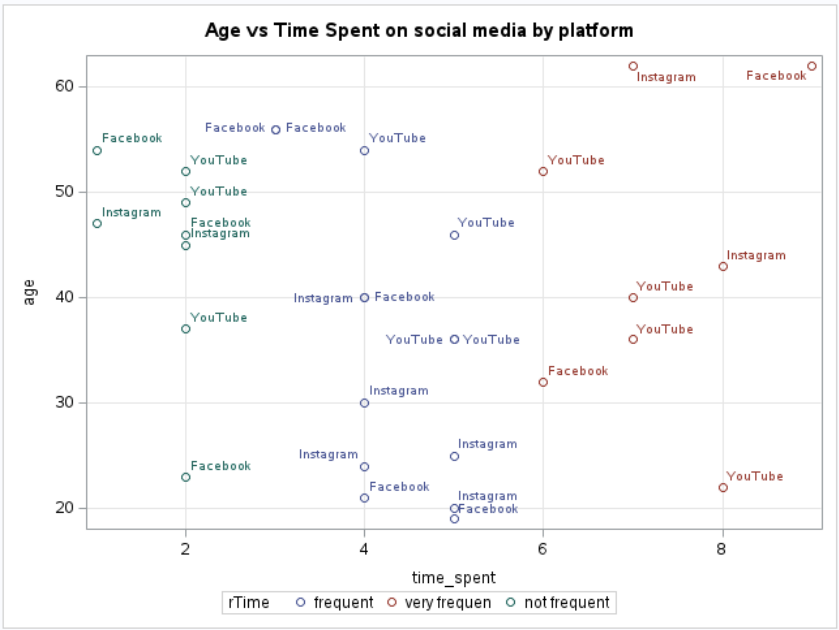


Vertical Bar Chart of Age



Vertical Bar Chart of Age

	time_spent					
	rTime					
	frequent		not frequent		very frequen	
	N	Mean	N	Mean	N	Mean
rAge						
Mid	5	4.60	4	1.75	3	7.33
Old	3	3.33	3	1.67	3	7.33
Young	6	4.50	1	2.00	2	7.00


Age vs Time Spent on social media by platform
age
time_spent
rTime
frequent
very frequen
not frequent
60
50
40
30
20
2
4
6
8
Instagram
Facebook
Facebook
Facebook
Facebook
YouTube
YouTube
YouTube
Instagram
Facebook
Instagram
YouTube
Instagram
Instagram
Facebook
YouTube
YouTube
YouTube
YouTube
YouTube
Facebook
Instagram
Instagram
Facebook
Facebook
Instagram
Facebook
YouTube

Number of hours spent on social media based on age

The UNIVARIATE Procedure
Variable: time_spent

Moments			
N	30	Sum Weights	30
Mean	4.4	Sum Observations	132
Std Deviation	2.17509413	Variance	4.73103448
Skewness	0.30675546	Kurtosis	-0.6673388
Uncorrected SS	718	Corrected SS	137.2
Coeff Variation	49.4339575	Std Error Mean	0.39711604

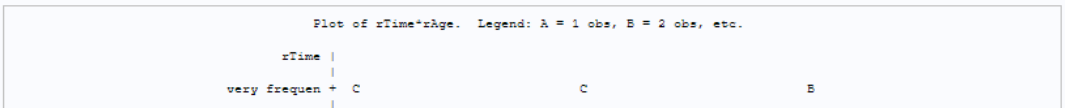
Basic Statistical Measures			
Location		Variability	
Mean	4.400000	Std Deviation	2.17509
Median	4.000000	Variance	4.73103
Mode	2.000000	Range	8.00000
		Interquartile Range	4.00000

Note: The mode displayed is the smallest of 3 modes with a count of 6.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	11.07988	Pr > |t|	<.0001
Sign	M	15	Pr >= |M|	<.0001
Signed Rank	S	232.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	9.0
99%	9.0
95%	8.0
90%	7.5
75% Q3	6.0
50% Median	4.0
25% Q1	2.0
10%	2.0
5%	1.0
1%	1.0
0% Min	1.0

Age vs Time Spent on social media by platform



To analyze the data better, we also used chi-squared, anova. Chi-squared test was used to determine whether there is a significant association between two categorical variables: it compares the observed frequency distribution of categorical data with the expected frequency distribution under the assumption of independence between the variables. One of the hypotheses was conducted to test chi-squared, with H0 being there is no significant association between social platforms and age groups, and the alternative is there is one.

ANOVA is used to analyze the differences among group means in a dataset with respect to a continuous outcome variable, and assesses the variation within each group compared to the variation between groups to determine if the group means are significantly different from each other. One of the hypotheses was used to assess ANOVA, with the null hypothesis being there is no significant difference in the mean spent time on the phone across different age groups, and the alternative hypothesis is there is one.

T-test was also being used to test whether there is a correlation between the two variables. The null hypothesis is there is no significant correlation between age and the time spent on the phone, and the alternative is there is a significant difference.

```
/* Chi-squared */

proc freq data = work.phone3;
    table platform * rTime / chisq expected deviation norow nocol nopercent;
run;

/* Hypotheses:
H0: There is no significant association between social platforms and age groups.
H1: There is a significant association between social platforms and age groups.
Value of chi-squared is NOT significant because p-value is greater than 0.50, accept H0 */
```

The FREQ Procedure

Frequency Expected Deviation	Table of platform by rTime			
	rTime			
platform	frequent	not frequent	very frequen	Total
Facebook	3 4 -1	5 3.3333 1.6667	2 2.6667 -0.667	10
Instagram	5 3.6 1.4	2 3 -1	2 2.4 -0.4	9
YouTube	4 4.4 -0.4	3 3.6667 -0.667	4 2.9333 1.0667	11
Total	12	10	8	30

Statistics for Table of platform by rTime

Statistic	DF	Value	Prob
Chi-Square	4	2.7399	0.6023
Likelihood Ratio Chi-Square	4	2.6265	0.6221
Mantel-Haenszel Chi-Square	1	0.0913	0.7626
Phi Coefficient		0.3022	
Contingency Coefficient		0.2893	
Cramer's V		0.2137	
WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 30

```
44 /* proc glm*/
45 proc glm data = work.phone3;
46 class rTime;
47 model age = rTime;
48
```

The GLM Procedure

Class Level Information		
Class	Levels	Values
rTime	3	frequent not frequent very frequen

Number of Observations Read	30
Number of Observations Used	30

The GLM Procedure

Dependent Variable: age

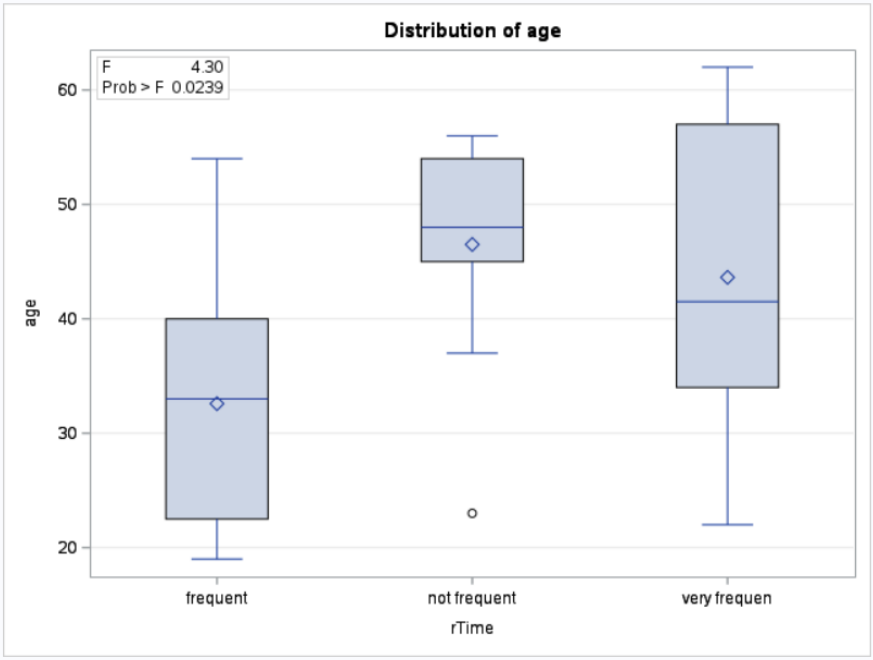
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1186.875000	593.437500	4.30	0.0239
Error	27	3725.291667	137.973765		
Corrected Total	29	4912.166667			

R-Square	Coeff Var	Root MSE	age Mean
0.241619	29.24371	11.74622	40.16667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rTime	2	1186.875000	593.437500	4.30	0.0239

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rTime	2	1186.875000	593.437500	4.30	0.0239

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rTime	2	1186.875000	593.437500	4.30	0.0239




```
49 /* proc anova */
50
51 proc anova data = work.phone3;
52 class rTime;
53 model age = rTime;
54
55 run;
56
57 /* Anova hypothesis:
58 H0: There is no significant difference in the mean spent time on the phone across different age groups
59 H1: There is a significant difference in the mean spent time on the phone across different age groups
```

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
rTime	3	frequent not frequent very frequen

Number of Observations Read	30
Number of Observations Used	30

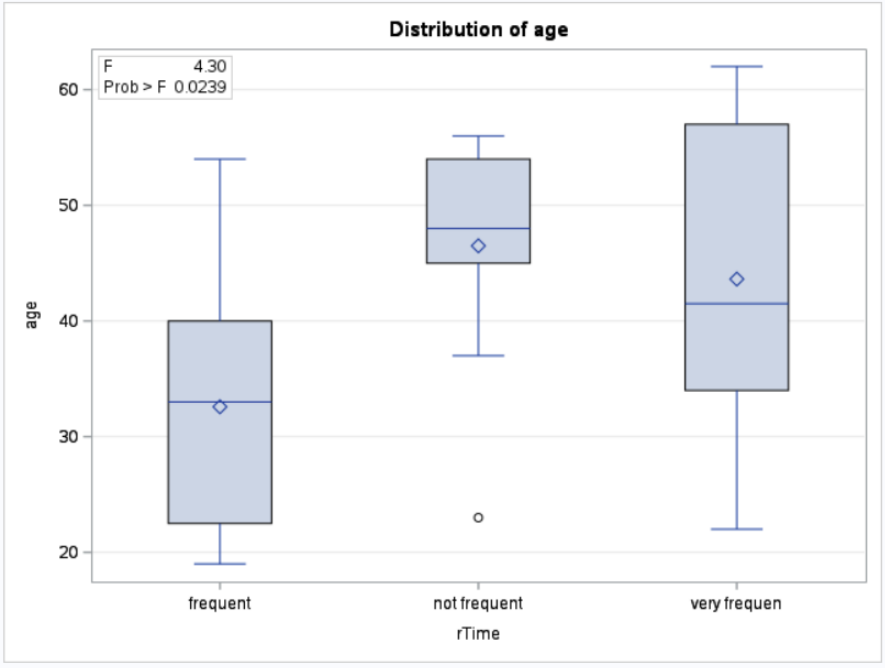
The ANOVA Procedure

Dependent Variable: age

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1186.875000	593.437500	4.30	0.0239
Error	27	3725.291667	137.973765		
Corrected Total	29	4912.166667			

R-Square	Coeff Var	Root MSE	age Mean
0.241619	29.24371	11.74622	40.16667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
rTime	2	1186.875000	593.437500	4.30	0.0239



```
/* proc ttest */
proc ttest data = work.phone3 alpha = 0.05;
    paired age * time_spent;
run;
/* Based on ttest, we should reject H0 and accept H1, which means there is a difference in at least one mean gro

/* proc corr*/
proc corr data = work.phone3 plots = scatter(nvar = all);
    var age time_spent;
run;
/* Based on correlation test:
H0: There is no significant correlation between age and the time spent on the phone.
H1: There is a significant correlation between age and the time spent on the phone.

CONCLUSION:

Based on the result and correlation-value <0 p= 0.06681, there is no significant correlation between age and
time spent on the phone; the scatterplot does not have any obvious linear pattern */
```

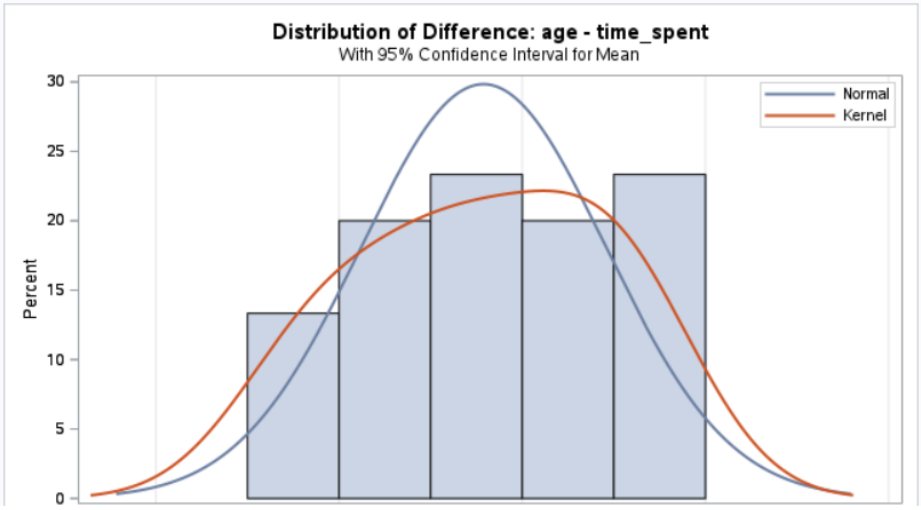
The TTEST Procedure

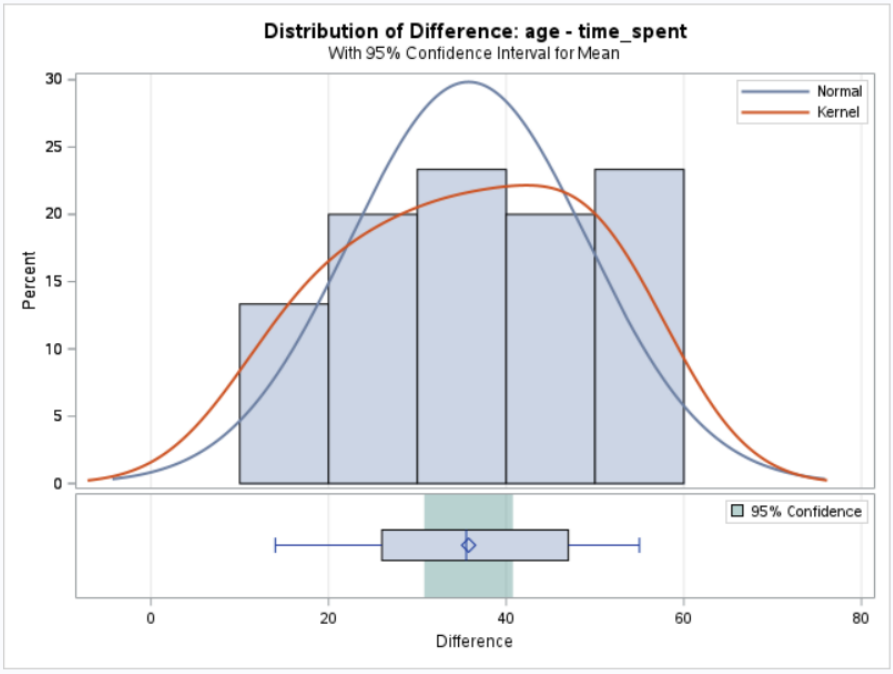
Difference: age - time_spent

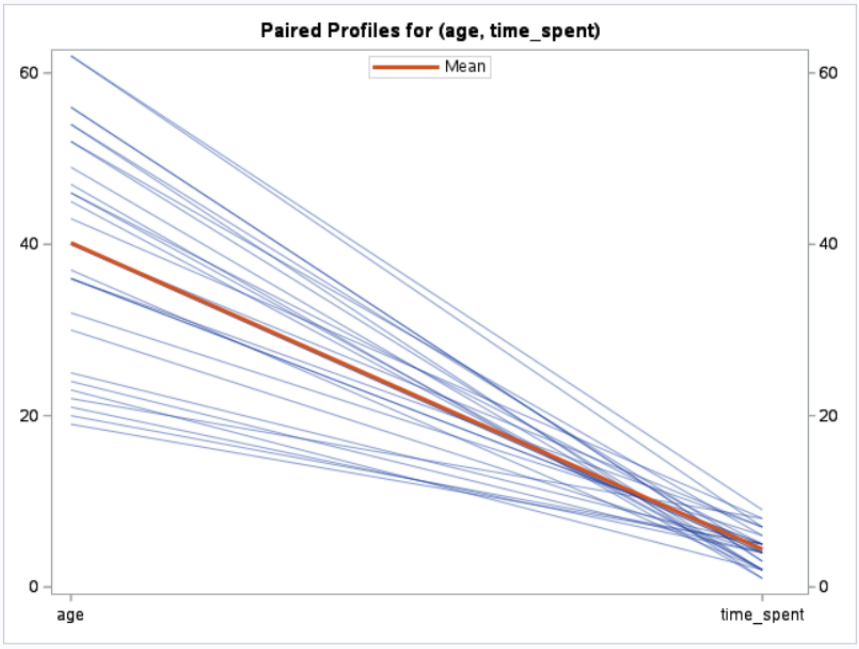
N	Mean	Std Dev	Std Err	Minimum	Maximum
30	35.7667	13.3692	2.4409	14.0000	55.0000

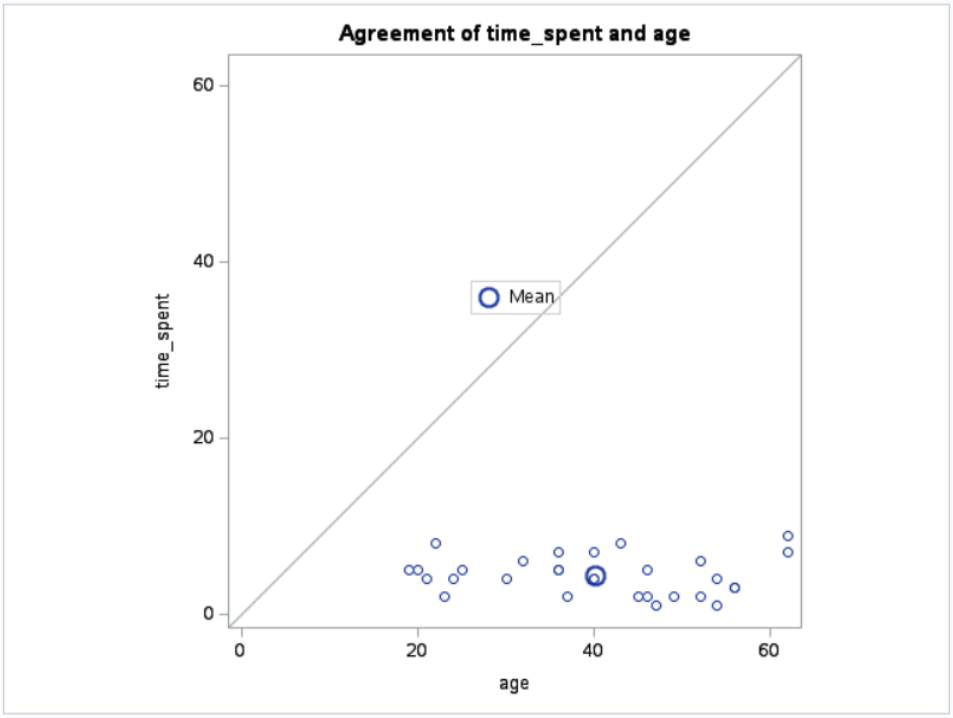
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
35.7667	30.7745	40.7588	13.3692	10.6474	17.9725

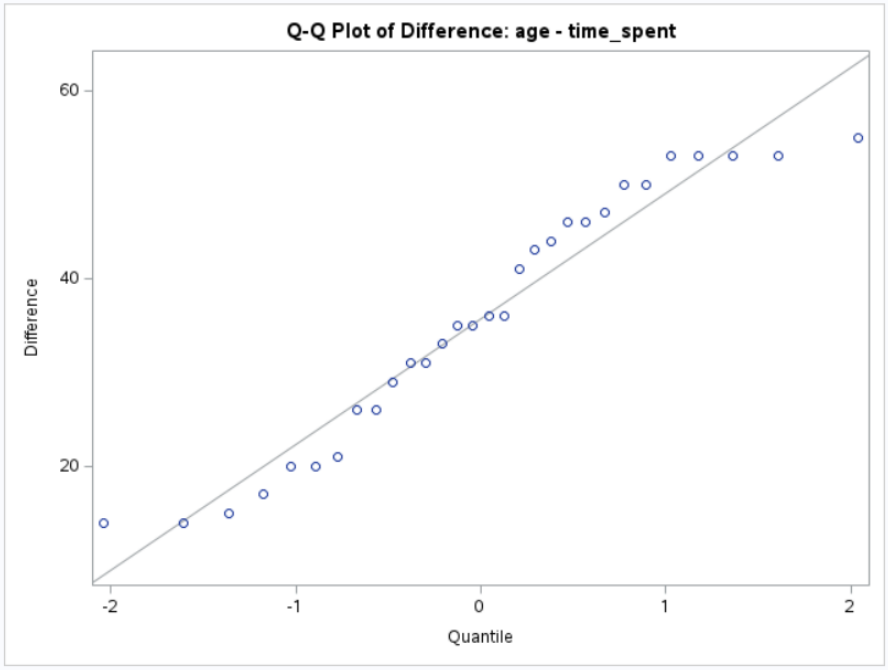
DF	t Value	Pr > |t|
29	14.65	<.0001




Distribution of Difference: age - time_spent
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
30
25
20
15
10
5
0
95% Confidence
0
20
40
60
80
Difference


Paired Profiles for (age, time_spent)
Mean
60
40
20
0
age
time_spent


Agreement of time_spent and age
60
40
20
0
time_spent
Mean
0
20
40
60
age


Q-Q Plot of Difference: age - time_spent
60
40
20
Difference
-2
-1
0
1
2
Quantile

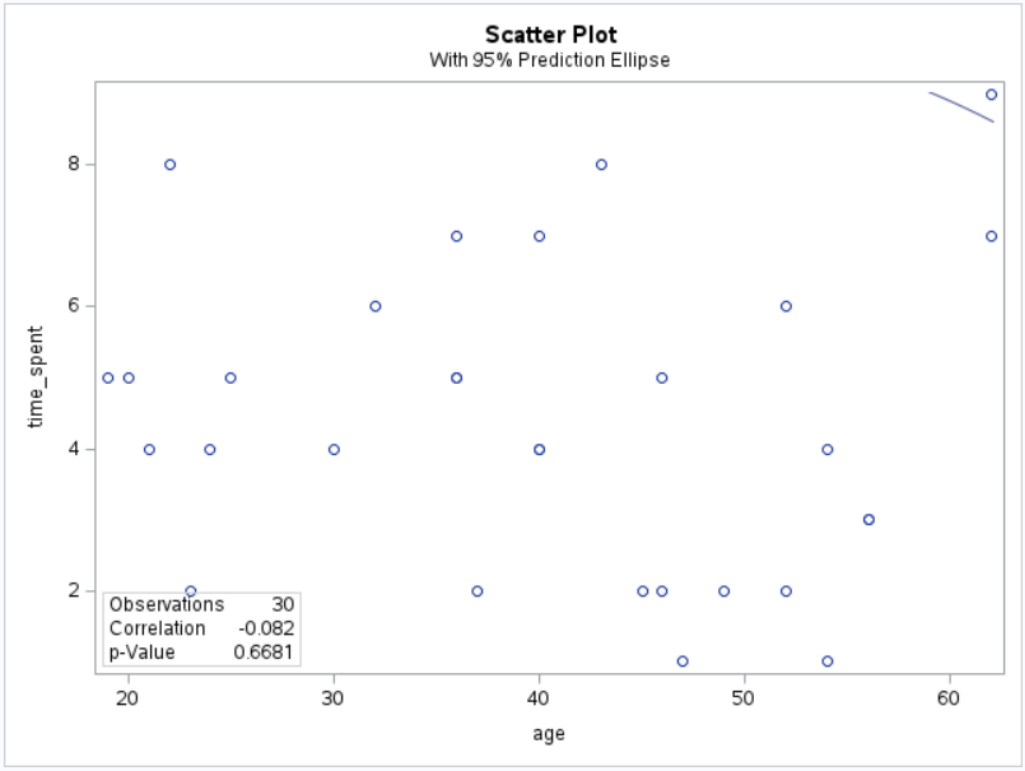
The CORR Procedure

2 Variables:	age time_spent

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
age	30	40.16667	13.01480	1205	19.00000	62.00000
time_spent	30	4.40000	2.17509	132.00000	1.00000	9.00000

Pearson Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	age	time_spent
age	1.00000	-0.08161 0.6681
time_spent	-0.08161 0.6681	1.00000

The CORR Procedure



Finally, we used proc outest function to produce necessary tables, fit plot and residual diagnostic plots:

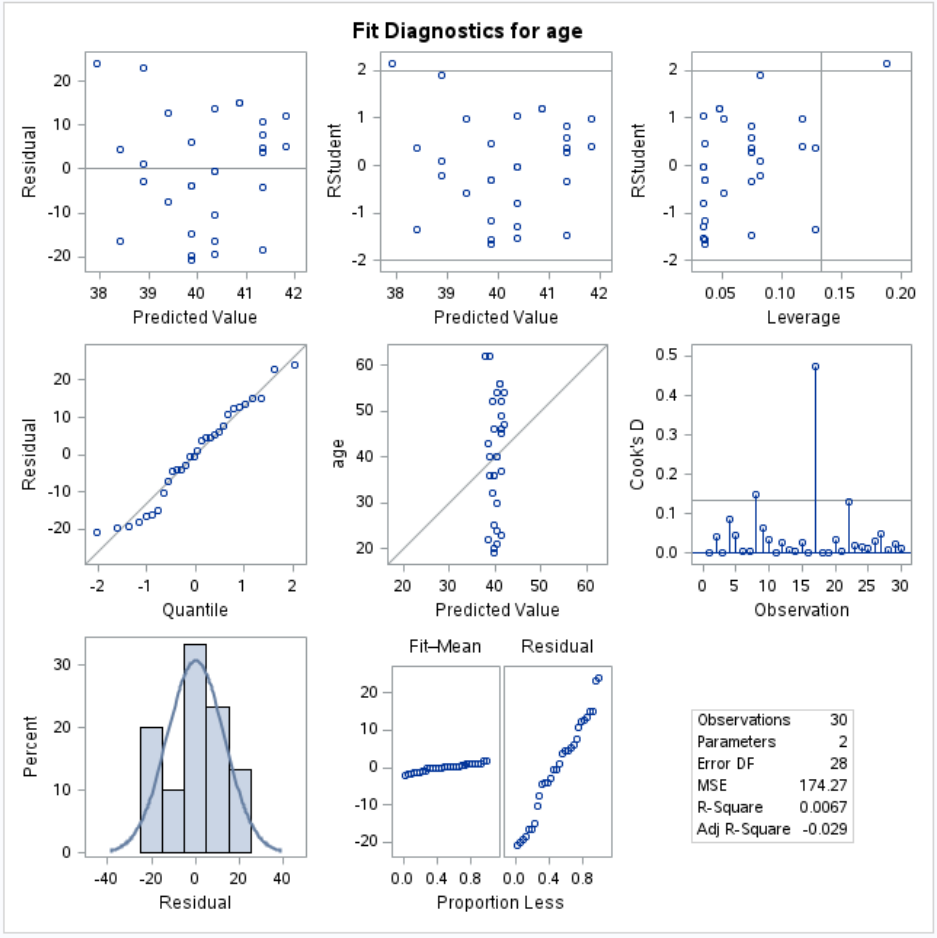
```
30 /* Proc reg */
31
32 proc reg data = work.phone3
33     outest = work.param_estimates;
34     model age = time_spent;
35 run;
```

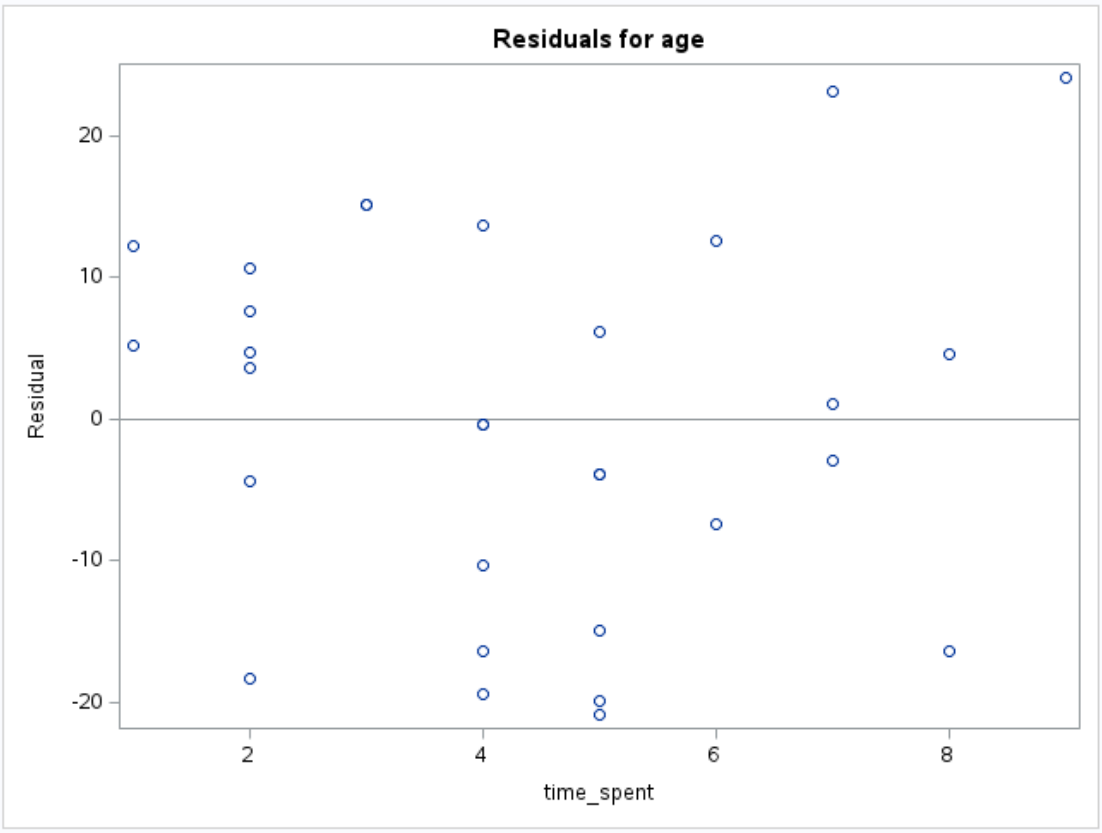
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	32.71866	32.71866	0.19	0.6681
Error	28	4879.44801	174.26600		
Corrected Total	29	4912.16667			

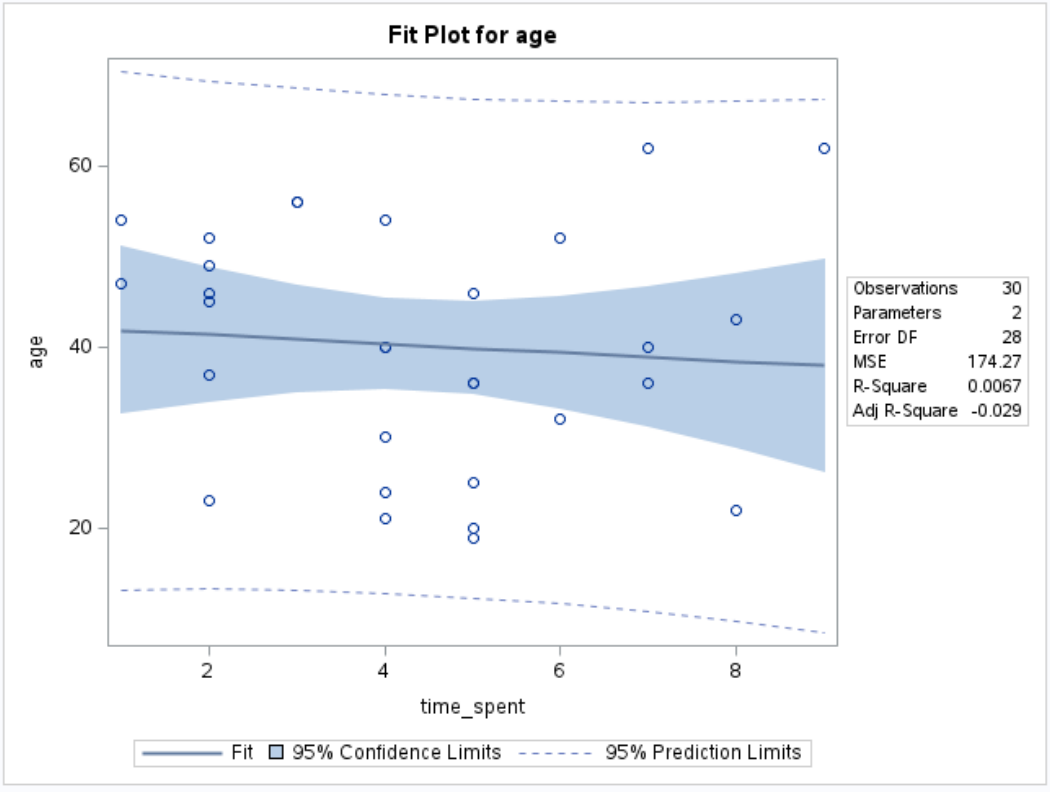
Root MSE	13.20098	R-Square	0.0067
Dependent Mean	40.16667	Adj R-Sq	-0.0288
Coeff Var	32.86552		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	42.31535	5.51354	7.67	<.0001
time_spent	1	-0.48834	1.12701	-0.43	0.6681

The REG Procedure
Model: MODEL1
Dependent Variable: age




Residuals for age
Residual
20
10
0
-10
-20
2
4
6
8
time_spent


Fit Plot for age
age
60
40
20
2
4
6
8
time_spent
Observations 30
Parameters 2
Error DF 28
MSE 174.27
R-Square 0.0067
Adj R-Square -0.029
Fit
95% Confidence Limits
95% Prediction Limits

Results

There are multiple significant relationships from our codes. In terms of age distribution, the age of participants ranged from 19 to 62 years old, with a significant deviation from 0, as confirmed once again in the one-sample t-test ($t = 16.58$, $p < .0001$). This suggests a broad range among social media users in the sample. In terms of time spent on social media, the mean time result was 4.4 hours, with a standard deviation of 2.17 hours. A one sample t-test also indicated that the time spent on social media significantly differed from zero ($t = 11.07$, $p < 0.0001$).

We are also interested in the frequency of social media use. The majority of participants were categorized as frequent users (42.42%), with a relatively even distribution across different age groups (young, mid, and old) The frequency table indicated that middle-aged and younger participants tend to spend more time on social media compared to older participants. This frequency was statistically significant, implying that the frequent use of social media is frequent among the sampled participants.

In the regression analysis, the number of hours spent on social media was significantly associated with age ($p < 0.05$). The negative coefficient for time spent on social indicates that as age increases, the time people have spent on social decreases. The R-squared value suggests that age accounts for approximately 7% of the variability in social media usage time.

There are, of course, some insignificant results in this data. Contrary to expectations, the correlation analysis did not show a significant relationship between age and the frequency of social media platform use ($p > 0.05$). This suggests that within our sample of 30, age is not a determining factor for now often individuals use different social media platforms. The Pearson

correlation analysis revealed that a very slight negative relationship between age and time spent on social media ($r = -0.0861$), but this was not statistically significant($p = 0.6681$)

For our ANOVA test for Age and time spent on social media, recall that our null hypothesis was that there is no significant difference in the mean spent time on phone across different age groups, while the alternative is there is a significant difference. It turns out that the result was not significant, given that our resulting F-value was 0.19, and $p = 0.6681$.

The Chi-square test of independence was also conducted to evaluate the association between social platforms and age groups. The chi-square test for independence between age groups and frequency of social media usage did not indicate a significant association; our results are a chi-square value of 4.3072, with $p= 0.6681$.

The distribution of age, as shown in the vertical bar chart and histogram, highlights that the sample is relatively evenly distributed across different age categories, with a slightly higher concentration of participants in their late 30s to early 60s. The frequency plot for the number of hours spent on social media based on age also did not show a distinct pattern correlating age with time spent.

Finally, the regression diagnostics was conducted to see whether the number of hours spent on social media correlates with the age groups. Results indicate that there is a very low R-squared value ($R^2 = .0067$), implying that the model does not explain the variability of age based on time spent on social media.

In conclusion, the statistical analysis found no significant relationship between age and time spent on social media. The data did not demonstrate any significant trends or patterns indicating that one variable could be used to predict the other. Thus, it can be inferred that within this sample, age is not a determinant of the amount of time spent on social media.

Discussion

Although the result is nowhere close to ideal, it is an expected result, given that the dataset was randomly generated by a Kaggle user for their machine learning training purposes. This dataset was not collected and answered by "real" humans, but rather is a work of AI and machine learning.

In terms of relationship between the variables, the analysis of data did not reveal a statistically significant relationship between age and time spent on social media. We have conducted a Pearson correlation test, but unfortunately, its correlation coefficient was very small ($r = -0.0861$) and not significant due to relatively high p-value ($p= 0.6681$). This result indicates that as age increases, there is no substantial decrease or increase in the time spent on social media among all the participants and age groups in this study. Furthermore, our ANOVA and regression analysis also supported these findings, showing no significant effect of age on time spent and social media, respectively.

Even though there seems to be no significant results or findings from our sampled dataset, it is still crucial to understand the relationship between age and time spent on social media, especially in today's digital age. According to Riehm et al. (2019), social media use may be a big risk factor for mental health problems in adolescents. The paper came to a conclusion that adolescents who spend more than 3 hours per day using social media may be at heightened risk for mental health problems, particularly internalizing problems (Riehm et al., 2019). Given that technology is constantly evolving, and social media platforms are ubiquitous and have significant social, psychological, and economic implications, knowing if and how different age groups interact with social media can inform a range of stakeholders, from policymakers and educators to marketers and app developers, about the habits and needs of different demographics.

According to , social media has an enormous impact on future rating of products to agenda settings and election outcomes; it will be an extremely important and powerful tool (Asur and Huberman, 2010).

Thirdly, we want to discuss the statistics methods and research questions. The chosen statistical methods, including correlation analysis, ANOVA, regression analysis, and chi-square tests for independence, were designed to assess the strengths and weaknesses, along with the direction of the relationship between age and social media usage, as well as to test for differences in usage frequency among different age groups. While these methods are significant for understanding relationship and distribution within the data, the non-significant findings suggest that age may not be a key factor in predicting social media usage. Future research could be done in predicting how social media impacts personal health, such as increased levels of depression or anxiety in a certain age group. We could also try to predict the outcome of social media within the next 10 years, based on different factors that might influence this.

This study is not without limitations. The sample size may not be representative of the broader population, and the age range may be too narrow to capture the full spectrum of social media usage patterns. Additionally, the study does not account for other factors that could influence social media usage, such as socioeconomic status, education level, occupation, or cultural background. Future research could include these variables to provide a deeper understanding of social media behaviors.

The study contributes to a growing body of research that seeks to understand the demographics of social media usage. By understanding that age alone is not a significant predictor of social media usage time, stakeholders can look beyond generational stereotypes and consider other factors that might influence engagement with social media. Such insights can

benefit educators in developing age-appropriate digital literacy, help social platforms tailor content and features to diverse user groups, and guide businesses in targeting their marketing strategies more effectively.

The results of this research could be used for future studies, where people actually conduct real surveys from human participants; it paves the way for future studies to explore other variables that might correlate with social media usage. Additionally, qualitative methods could be used to understand the subjective experience and motivations behind social media usage among different age groups.

References

Author links open overlay panelDavid John Hughes a *et al.* (2011) *A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage*, *Computers in Human Behavior*. Available at: https://www.sciencedirect.com/science/article/pii/S0747563211002457?casa_token=mdY6ztgQZ60AAAAA%3Ae9BchTZmEAgfzdIOxK1W6dcmFxrchMXfErFTRBY1IcmmKsYs8lI07u1cAoh8Wc8zN1LWIwLi6kg (Accessed: 22 February 2024).

Author links open overlay panelSarah M. Coyne *et al.* (2019) *Does time spent using social media impact mental health?: An eight year longitudinal study*, *Computers in Human Behavior*. Available at: https://www.sciencedirect.com/science/article/pii/S0747563219303723?casa_token=LObJ4PV8a-kAAAAA%3AYirmENxt14J_egcxgI8Q2KNJH1zKGHa7b_bn4BrlcYHiEGLw3ZJ9BzZXUqgWxTBIywgU6HZT2rg (Accessed: 22 February 2024).

Cara, B. (no date) *Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study*, *eClinicalMedicine*. Available at: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(18)30060-9/ppt (Accessed: 22 February 2024).

Kira E. Riehm, M. (2019) *Associations between social media time and internalizing and externalizing problems among US youth*, *JAMA Psychiatry*. Available at: https://jamanetwork.com/journals/jamapsychiatry/article-abstract/2749480 (Accessed: 22 February 2024).

Person (2018) *Measuring time spent with media: The Common Sense Census of media use*, *Taylor & Francis*. Available at: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315307633-16/measuring-time-spent-media-common-sense-census-media-use-us-8-18-year-olds-vicky-rideout (Accessed: 22 February 2024).

APPENDIX PART VII

	age	time_spent
1	40	4
2	21	4
3	40	7
4	23	2
5	20	5
6	37	2
7	46	5
8	62	7
9	54	1
10	56	3
11	36	5
12	24	4
13	43	8
14	46	2
15	52	6
16	36	7
17	62	9
18	40	4
19	36	5
20	56	3
21	45	2
22	22	8
23	54	4
24	49	2
25	47	1
26	52	2
27	19	5
28	32	6
29	25	5
30	30	4