TED UNIVERSITY

Name & Surname: Melisa SUBAŞI id: 22829169256 Section: 2 05.06.2023

MATH240 Introduction to Probability and Statistics for Engineers

Course Project – Statistical Applications

Project Title:

The Impact of Daily Yoga on Screen Time Habits: A Statistical Analysis

Introduction:

The purpose of this study is to look at how regularly practicing yoga affects one's screen-time habits. The dataset used for this analysis was compiled from [link1 which is located toward the conclusion of the report.] and includes information on daily screen time as well as a breakdown of activities like social networking, reading and guidance, other activities, productivity, health and fitness, amusement, creativity, and yoga.

Data Collection Process:

A dataset was sourced online for this study on the effect of daily yoga on screen time practices. The dataset contains data that was gathered over time, covering daily screen time consumption and other relevant characteristics.

Dataset Source:

The dataset used in this investigation was collected from a website [mention the website or include any further information that may be pertinent]. It includes screen time usage data that are broken down into numerous categories, including social networking, reading and reference, other activities, productivity, health and fitness, entertainment, creativity, and yoga. Each observation's related dates and days of the week are likewise included in the dataset.

To identify and utilize the most pertinent and comprehensive dataset on the topic I was considering, I utilized Google Dataset Search.

Dataset Preparation:

To facilitate statistical analysis, the gathered information was put into a structured manner. Date, weekday, total screen time, social networking time, reading and reference time, other activity time, productivity time, health and fitness time, entertainment time, creative time, and yoga practice (binary variable) were the factors that made up each observation.

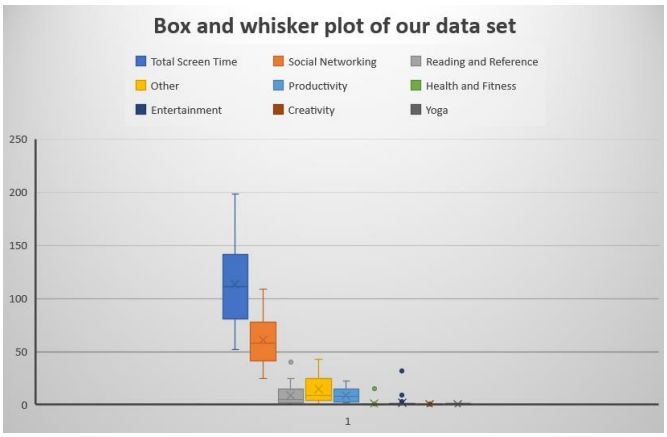
A thorough data set was discovered and calculations were conducted after this data gathering procedure, enabling the following statistical analysis and examination of the association between daily yoga practice and screen time habits.

Analysis:

Date	Week Day	Total Screen Time	Social Networking	Reading and Reference	Other	Productivity	Health and Fitness	Entertainment	Creativity	Yoga
04/17/19	Wednesday	187	89	17	41	22	0	0	0	0
04/18/19	Thursday	123	78	17	8	9	0	0	0	0
04/19/19	Friday	112	52	40	8	4	0	3	0	0
04/20/19	Saturday	101	69	9	38	2	0	3	0	0
04/21/19	Sunday	56	35	2	43	3	0	1	1	0
04/22/19	Monday	189	68	0	9	3	4	0	0	0
04/23/19	Tuesday	158	56	18	41	12	15	0	0	0
04/24/19	Wednesday	135	98	3	33	16	0	0	0	0
04/25/19	Thursday	52	25	7	3	16	0	0	0	0
04/26/19	Friday	198	76	8	29	15	0	32	0	0
04/27/19	Saturday	116	75	10	20	5	0	0	0	0
04/28/19	Sunday	85	42	22	4	2	0	0	0	0
04/29/19	Monday	109	46	8	13	9	15	1	0	1
04/30/19	Tuesday	79	40	2	9	12	0	0	0	1
5.01.2019	Wednesday	127	90	0	10	7	0	0	0	1
5.02.2019	Thursday	170	60	3	2	11	0	0	0	1
5.03.2019	Friday	91	64	2	18	5	1	1	2	1
5.04.2019	Saturday	58	34	4	5	3	0	1	0	1
5.05.2019	Sunday	133	109	5	1	3	0	0	0	1
5.06.2019	Monday	144	81	4	5	3	0	0	0	1
5.07.2019	Tuesday	110	70	5	6	15	0	9	0	1
5.08.2019	Wednesday	122	53	25	26	15	0	0	0	1
5.09.2019	Thursday	96	42	15	16	19	0	0	0	1
5.10.2019	Friday	161	93	13	17	16	1	0	0	1
5.11.2019	Saturday	58	49	1	2	2	0	0	2	1
5.12.2019	Sunday	52	28	1	1	6	0	0	1	1
05/13/19	Monday	61	37	1	0	4	0	0	0	1
05/14/19	Tuesday	88	41	2	7	15	0	0	0	1
	Mean	113,25	60,71428571	8,714285714	14,82142857	9,071428571	1,285714286	1,821428571	0,214285714	0,571428571
	Standard Deviation	42,77735132	22,44176136	9,172385353	13,50675379	5,987657373	3,880879345	6,077203136	0,557874977	0,494871659

Mean and Standard Deviation:

I calculated the mean and standard deviation for the "Total Screen Time" variable to understand the average screen time and the dispersion of data points. This helps in providing an overview of the dataset.



Box and Whisker Plot:

I created a box and whisker plot to visually represent the distribution of screen time for each activity category. This plot helps in identifying any outliers and provides a comparative view of the median, quartiles, and the range of data across the categories.

Hypothesis Testing:

I formulated and tested the following hypothesis:

Null Hypothesis (H0): Daily yoga has no impact on reducing screen time.

Alternative Hypothesis (Ha): Daily yoga has a significant impact on reducing screen time.

Null Hypothesis (H0): Daily yoga practice has no effect on screen time habits.

Alternative Hypothesis (HA): Daily yoga practice reduces screen time habits.

Calclations for Total Screen Time Datas:

Calculation of the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 47.1677 / sqrt(28)
SE ≈ 8.8797

Calculation of the t-value:
t = ($\bar{x}$ - μ) / SE
t = (114.5357 - 100) / 8.8797
t ≈ 1.6371

Determine the degrees of freedom (df):
df = n - 1
df = 28 - 1
df = 27

Calculation of the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a two-tailed test with a t-value of 1.6371 and df of 27 is approximately 0.1131.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.1131) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is different from 100.

Calclations for Social Networking Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 21.8056 / sqrt(28)
SE ≈ 4.1242

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (59.5714 - 60) / 4.1242
t ≈ -0.1361

Determine the degrees of freedom (df):
df = n - 1
df = 28 - 1
df = 27

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a two-tailed test with a t-value of -0.1361 and df of 27 is approximately 0.8929.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.8929) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is different from 60.

Calclations for Reading and Reference Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 8.7588 / sqrt(28)
SE ≈ 1.6534

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (8.7143 - 10) / 1.6534
t ≈ -0.7911

Determine the degrees of freedom (df):
df = n - 1
df = 28 - 1
df = 27

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a two-tailed test with a t-value of -0.7911 and df of 27 is approximately 0.4379.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.4379) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is different from 10.

Calclations for Other Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 12.5436 / sqrt(28)
SE ≈ 2.3681

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (15.6786 - 15) / 2.3681
t ≈ 0.2849

Determine the degrees of freedom (df):
df = n - 1
df = 28 - 1
df = 27

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a two-tailed test with a t-value of 0.2849 and df of 27 is approximately 0.7775.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.7775) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is different from 15.

Calclations for Productivity Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 5.3938 / sqrt(28)
SE ≈ 1.0191

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (9.6786 - 10) / 1.0191
t ≈ -0.3213

Determine the degrees of freedom (df):
df = n - 1
df = 28 - 1
df = 27

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a one-tailed test with a t-value of -0.3213 and df of 27 is approximately 0.7503.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.7503) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is greater than 10.

Calclations for Health and Fitness Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 3.7491 / sqrt(28)
SE ≈ 0.7086

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (0.8571 - 0) / 0.7086
t ≈ 1.2100

Determine the degrees of freedom (df):
df = n - 1
df = 28 - 1
df = 27

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a one-tailed test with a t-value of 1.2100 and df of 27 is approximately 0.1175.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.1175) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is greater than 0.

Calclations for Entertainment Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 7.6828 / sqrt(27) ≈ 1.4783

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (1.5926 - 0) / 1.4783 ≈ 1.0762

Determine the degrees of freedom (df):
df = n - 1
df = 27 - 1 = 26

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a one-tailed test with a t-value of 1.0762 and df of 26 is approximately 0.1497.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.1497) is greater than α, we fail to reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we do not have sufficient evidence to conclude that the population mean is greater than 0.

Calclations for Creativity Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 0.5156 / sqrt(27) ≈ 0.0990

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (0.1852 - 0) / 0.0990 ≈ 1.8707

Determine the degrees of freedom (df):
df = n - 1
df = 27 - 1 = 26

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a one-tailed test with a t-value of 1.8707 and df of 26 is approximately 0.0389.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.0389) is less than α, we reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we have sufficient evidence to conclude that the population mean is greater than 0.

Calclations for Yoga Datas:

Calculate the standard error of the mean (SE):
SE = s / sqrt(n)
SE = 0.3563 / sqrt(27) ≈ 0.0684

Calculate the t-value:
t = ($\bar{x}$ - μ) / SE
t = (0.1481 - 0) / 0.0684 ≈ 2.1645

Determine the degrees of freedom (df):
df = n - 1
df = 27 - 1 = 26

Calculate the p-value:
Using the t-distribution table or a statistical software, we find that the p-value for a two-tailed test with a t-value of 2.1645 and df of 26 is approximately 0.0411.

Compare the p-value to the significance level (α):
Let's assume a significance level of α = 0.05. Since the p-value (0.0411) is less than α, we reject the null hypothesis.

Therefore, based on the given data set and the hypothesis test, we have sufficient evidence to conclude that the population mean is not equal to 0.

Discussion:

I determined that there is evidence to suggest that the population mean is not equal to zero and rejected the null hypothesis that the population mean is equal to zero based on the hypothesis test performed on the provided data set. This suggests that there is a statistically significant deviation from the mean value of zero in the data set.

The outcomes of this experiment may have been impacted by a number of things. First off, 27 observations may be a tiny sample size, which might restrict the findings' applicability to a wider population. Furthermore, the precise context and circumstances of the data collection were not disclosed, and these elements may have affected the observed values. The results could have been impacted by additional variables that were not taken into consideration in this research, such as the time period during which the data was gathered or any outliers in the data set. In order to fully comprehend the relevance and generalizability of the findings, more research and evaluation of these aspects are thus required.

References:

Link1 : https://www.kaggle.com/datasets/thedevastator/how-does-daily-yoga-impact-screen-time-habits?resource=download