MAEER'S
MIT | Arts, Commerce & Science College
Affiliated to Savitribai Phule Pune University
Accredited by NAAC with "A" Grade

First Year Masters of Science in Data Science

(Faculty of Science and Technology)

STATISTICS LAB Book

For

First Year of M.Sc. (DS)
Semester-I

Subject Name: - Practical Course based on Statistics for Data Science

Student Name: -______________________________________

Student Roll No: -____

Choice Based Credit System (CBCS) Syllabus under National Education Policy (NEP) to be implemented from Academic Year 2024-2025

MAEER'S

MIT | Arts, Commerce & Science College

Affiliated to Savitribai Phule Pune University

Accredited by NAAC with "A" Grade

CERTIFICATE

This is to certify that Mr./Ms ____

Roll Number______________of F.Y.M.Sc (DS) Division **Q,** Semester-I has satisfactorily / Non-satisfactorily completed a **Lab course on Statistics for Data Science** for the Academic Year 2024-25.

Seat No______________Date _____________

Performance in practical: /10 Attendance: /

Subject Teacher Head of Department

Internal Examiner External Examiner

INDEX

Sr No	Title of Experiment	Remark	Signature
1	Study of open source statistical software like PSPP, JAMOVI, JASP, RStudio, OpenStat and MS-Excel		
2	Data condensation using RStudio		
3	Data visualization: Graphical representation- Histogram, Frequency curve, Ogive curve,Boxplot using R		
4	Data visualization: diagrammatic representation -Pie diagram, bar diagram, stem and leaf plot of data using using R		
5	Summary Statistics - using JAMOVI		
6	Linear Correlation and Regression, Multiple regression, logistic regression using JASP		
7	Fitting of Binomial distribution using MS-Excel		
8	Fitting of Poisson distribution using MS-Excel		
9	Fitting of Normal distribution using MS-Excel		
10	Large Sample tests for mean using MS-Excel and JASP		
11	Small sample tests (t, chi-square, F) using JAMOVI		
12	ANOVA using MS-Excel & JAMOVI		
13	Mini-project (Synopsis, Data collection, Data Analysis, Report)		

ASSIGNMENT 03

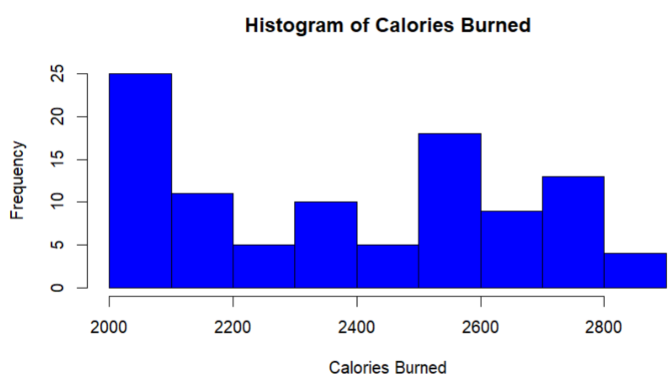
Data visualization: Graphical representation- Histogram, Frequency curve, Ogive curve, box plot using R.

Source File: https://www.kaggle.com/datasets/hanaksoy/health-and-sleep-statistics

1) Histogram

Creating a histogram for Calories Burned

hist(data$`Calories Burned`, breaks=10, main="Histogram of Calories Burned", xlab="Calories Burned", col="blue", border="black")



2) Frequency Curve

Create histogram data without plotting

hist.score = hist(data$`Calories Burned`, breaks=10, plot=FALSE)

Prepare x and y axes for the frequency polygon

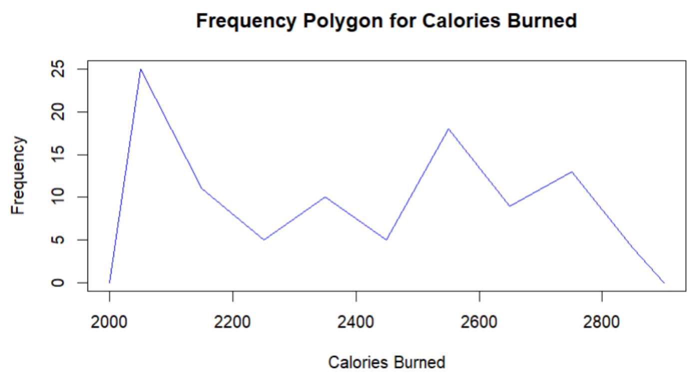
x.axis = c(min(hist.score$breaks), hist.score$mids, max(hist.score$breaks))

y.axis = c(0, hist.score$counts, 0)

Plot the frequency polygon

plot(x.axis, y.axis, type="n", main="Frequency Polygon for Calories Burned", xlab="Calories Burned", ylab="Frequency")

lines(x.axis, y.axis, col="blue")



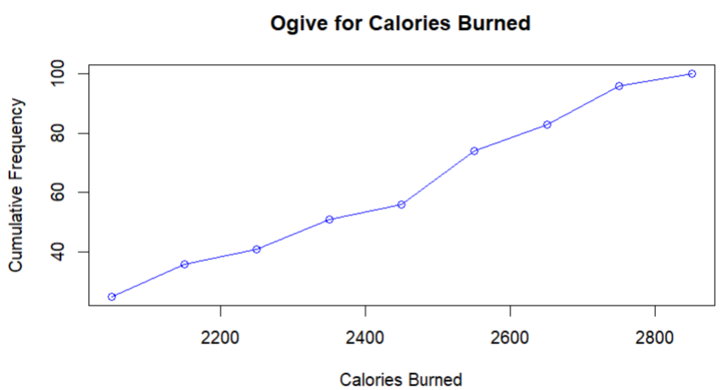
3) Ogive Curve

Calculating cumulative frequency and plotting the ogive

hist.score = hist(data$`Calories Burned`, breaks=10, plot=FALSE) # Get histogram data without plotting

cumulative.freq = cumsum(hist.score$counts) # Calculate cumulative frequency

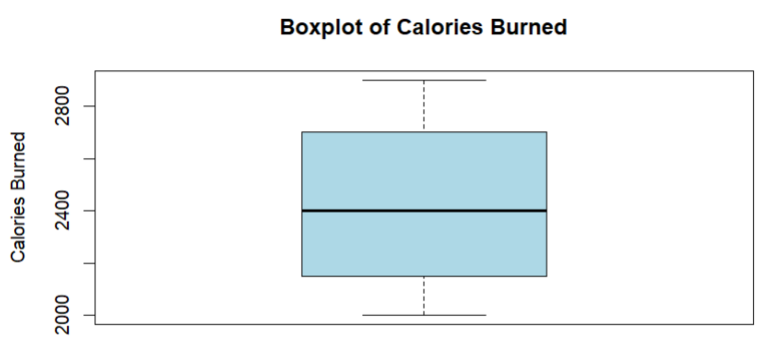
plot(hist.score$mids, cumulative.freq, type="o", col="blue", xlab="Calories Burned", ylab="Cumulative Frequency", main="Ogive for Calories Burned")



4) Box Plot

Creating a boxplot for Calories Burned

boxplot(data$`Calories Burned`, main="Boxplot of Calories Burned", ylab="Calories Burned", col="lightblue")



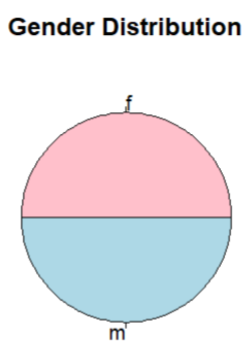
ASSIGNMENT 04

Data visualization: diagrammatic representation- Pie diagram, bar diagram, stem and leaf plot of data using R.

Source File: https://www.kaggle.com/datasets/hanaksoy/health-and-sleep-statistics

1) Pie Diagram

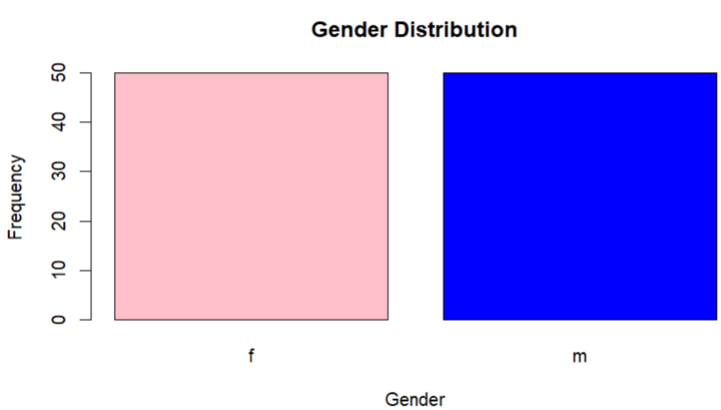
Create a pie chart showing the distribution of gender pie(table(data$Gender), main="Gender Distribution", col=c("pink", "lightblue"))



2) Bar diagram

Create a bar plot showing the distribution of gender with specified colors

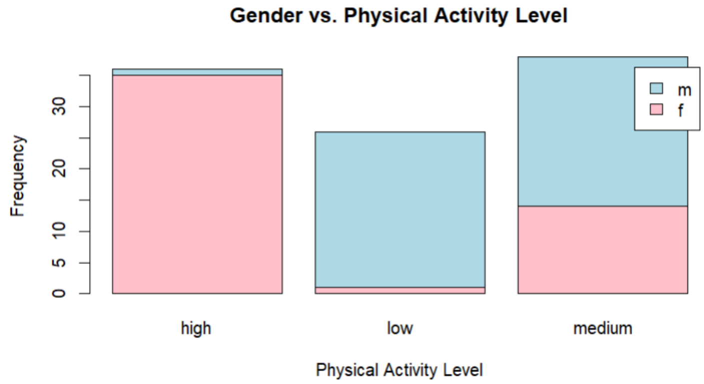
barplot(table(data$Gender), xlab="Gender", ylab="Frequency", main="Gender Distribution", col=c("pink", "blue"))



3) Subdivided Bar Diagram

Create a subdivided bar plot showing the distribution of gender across physical activity levels

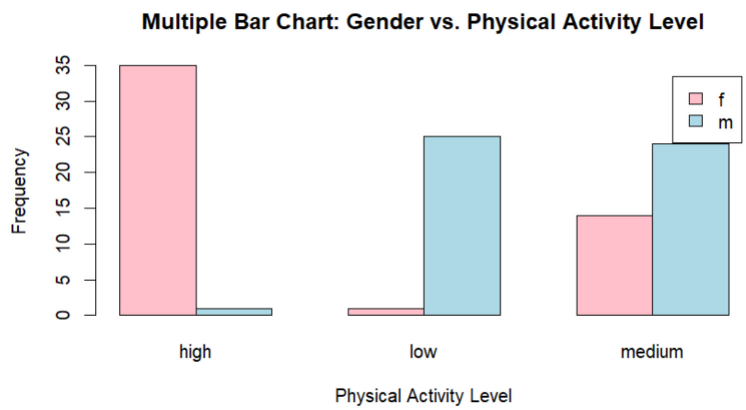
barplot(table(data$Gender, data$`Physical Activity Level`), legend.text = TRUE, xlab = "Physical Activity Level", ylab = "Frequency", main = "Gender vs. Physical Activity Level", col = c("pink", "lightblue"))



4) Multiple Bar Diagram

Create a multiple bar chart for Gender vs. Physical Activity Level

barplot(table(data$Gender, data$`Physical Activity Level`), beside = TRUE, # Display bars side by side legend.text = TRUE, xlab = "Physical Activity Level", ylab = "Frequency", main = "Multiple Bar Chart: Gender vs. Physical Activity Level", col = c("pink", "lightblue"))



5) Stem and Leaf Plot

Stem-and-leaf plot for Calories Burned

stem(data$`Calories Burned`, scale = 1)

```
 The decimal point is 2 digit(s) to the right of the |

 20 | 000000000000
 21 | 0000000000000
 22 | 00000000000
 23 | 00000
 24 | 0000000000
 25 | 00000
 26 | 000000000000000000
 27 | 0000000005555555555
 28 | 000
 29 | 0000
```

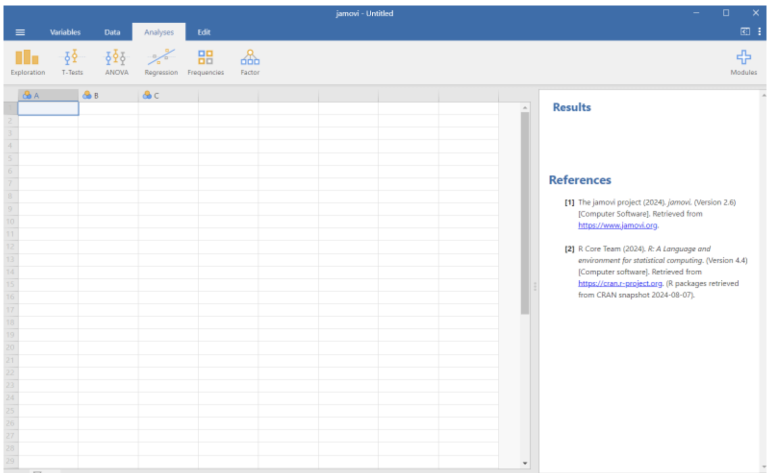
ASSIGNMENT 05

Summary Statistics Using JAMOVI

Source File: https://www.kaggle.com/datasets/hanaksoy/health-and-sleep-statistics

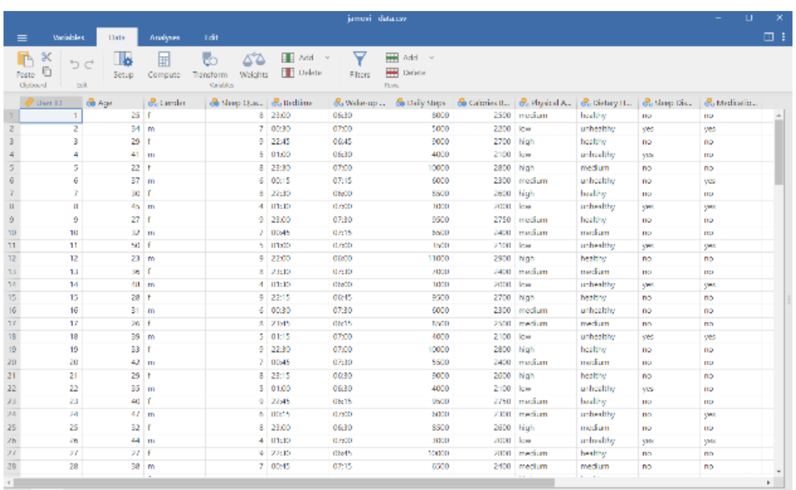
Steps in JAMOVI

1. **Open JAMOVI**



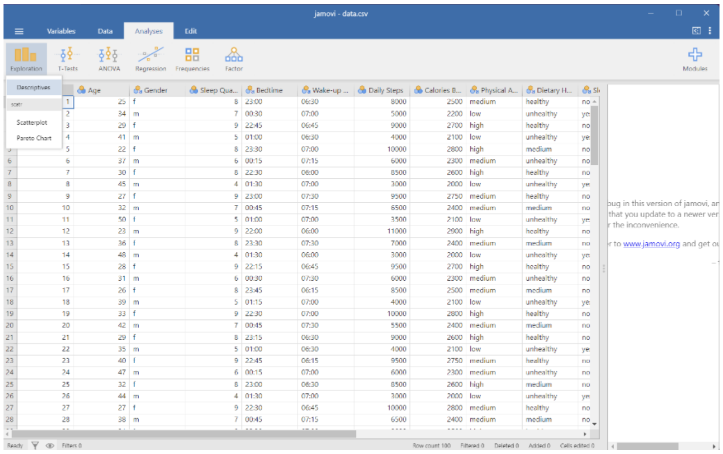
2. **Load the Dataset**

   - Click Open, select Browse, navigate to your .csv file, and load it into JAMOVI.



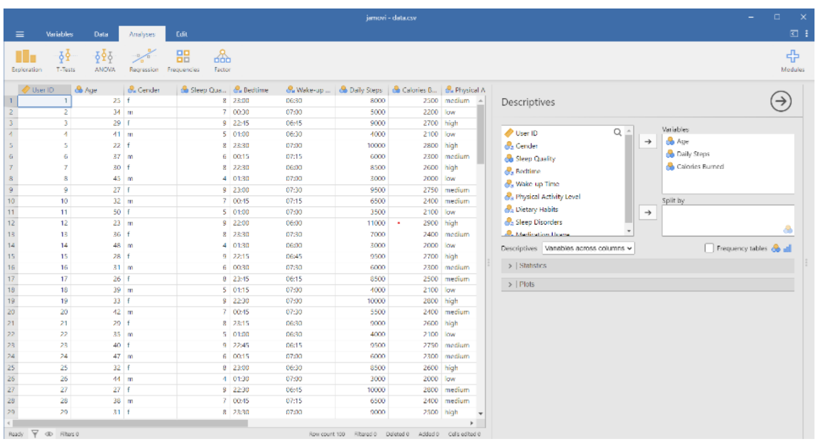
3. **Access Descriptive Statistic**

- Go to the Analyses menu and select Exploration > Descriptives.



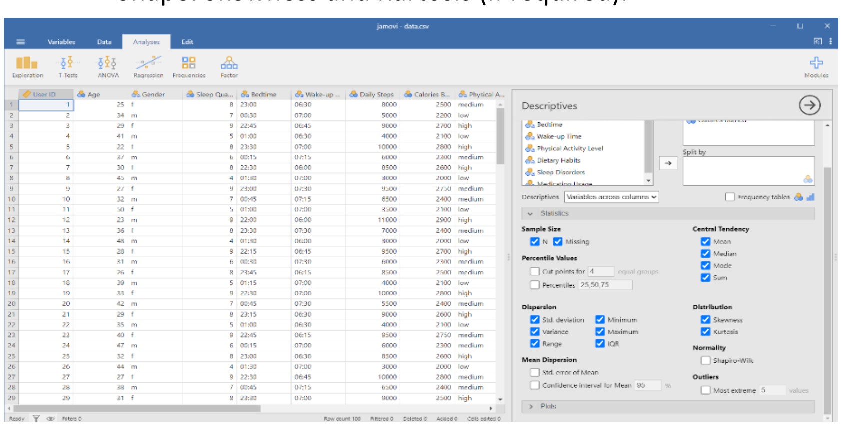
4. Select Variables for Analysis

- Choose the variables to analyze (e.g., Calories Burned, Age, Daily Steps) from the left panel and move them into the Variables box.



5. Specify Statistics to Display

- Under Statistics, select the options to display specific statistics, such as:
- Central Tendency: Mean, Median, Mode
- Dispersion: Standard Deviation, Variance, Range, Interquartile Range (IQR)
- Shape: Skewness and Kurtosis (if required).



6. **View Results**
   - The summary statistics for the chosen variables will appear on the right side of the screen. Each selected measure (e.g., mean, median, standard deviation) will be displayed in a table format.

Descriptives

Descriptives

	Age	Daily Steps	Calories Burned
N	100	100	100
Missing	0	0	0
Mean	36.0	6830	2421
Median	35.0	6750	2400
Mode	28.0 [a]	8500	2600
Sum	3601	683000	242100
Standard deviation	8.45	2499	281
Variance	71.5	6.24e+6	78999
IQR	15.3	4250	525
Range	28	8000	900
Minimum	22	3000	2000
Maximum	50	11000	2900
Skewness	0.168	−0.218	−0.0895
Std. error skewness	0.241	0.241	0.241
Kurtosis	−1.31	−1.43	−1.36
Std. error kurtosis	0.478	0.478	0.478
25th percentile	28.8	4750	2175
50th percentile	35.0	6750	2400
75th percentile	44.0	9000	2700

[a] More than one mode exists, only the first is reported

ASSIGNMENT 06

Linear Correlation and Regression, Multiple regression and logistic regression using JASP.

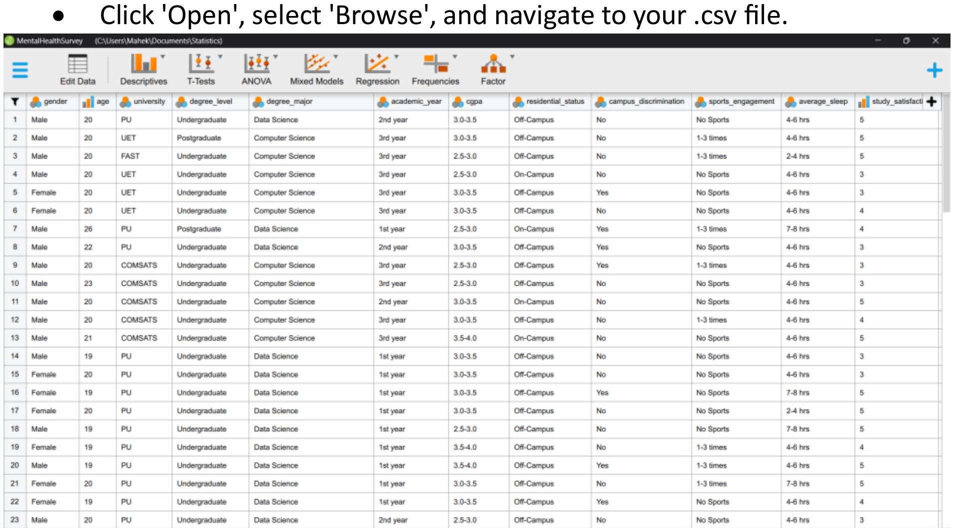
Source File: https://www.kaggle.com/datasets/abdullahashfaqvirk/student-mental-health-survey

Linear Correlation

1. Open JASP.

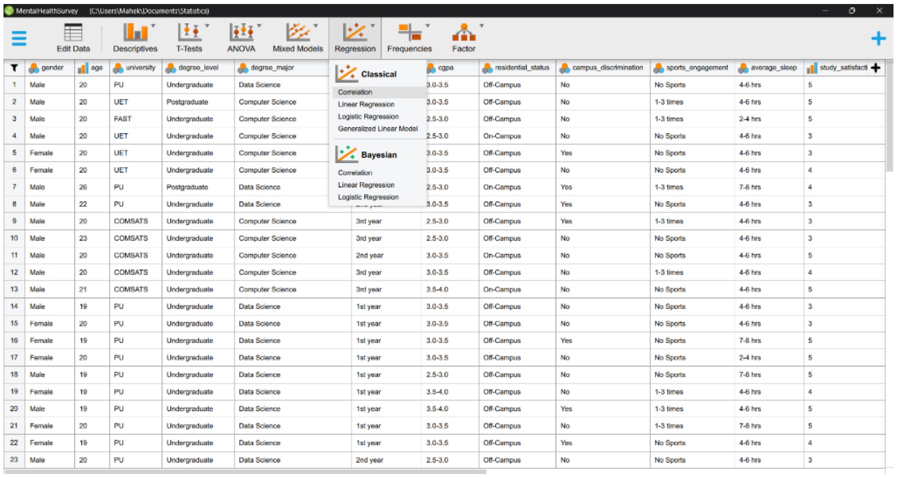
2. Load the dataset:

- Click 'Open', select 'Browse', and navigate to your .csv file.

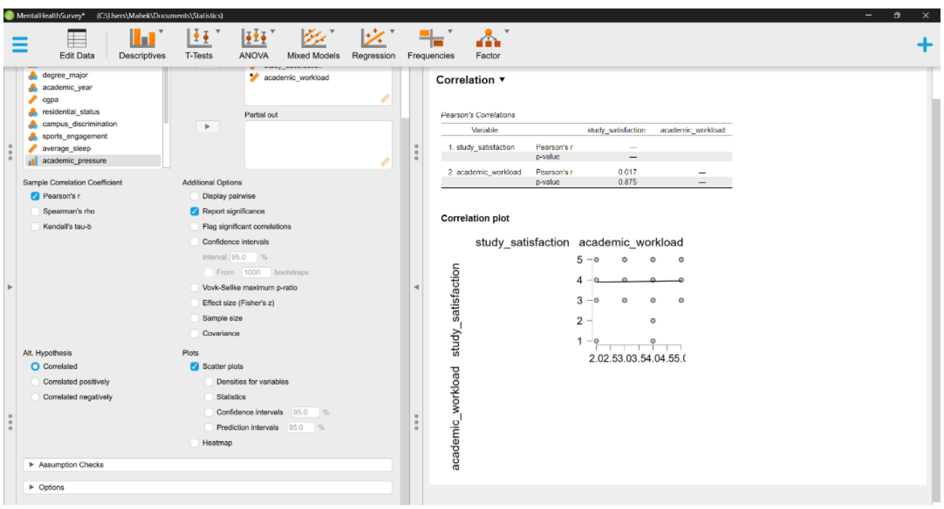


3. Perform Linear Correlation:

- Go to 'Regression > Correlation Matrix'.

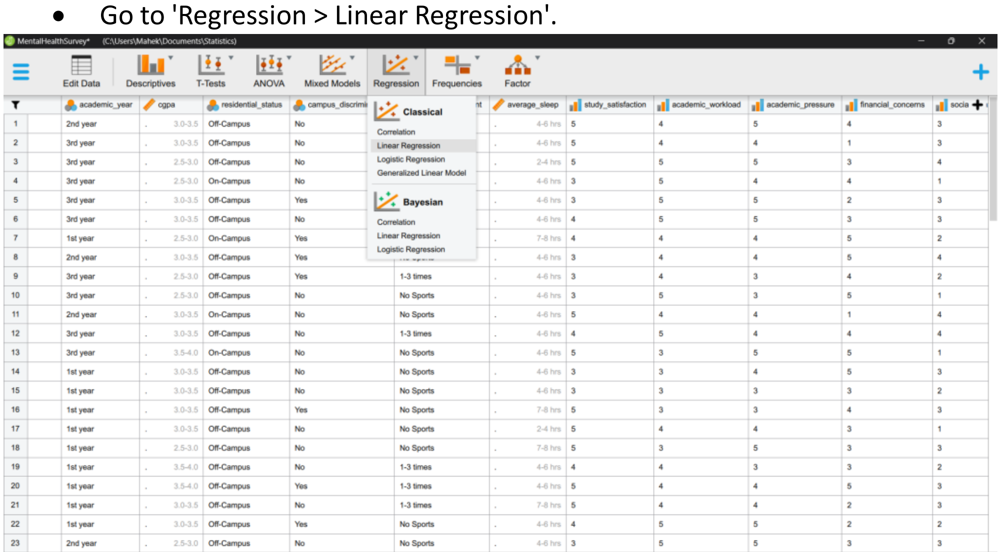


- Select two continuous variables (e.g., ‘study_satisfaction’ and 'academic_workload').
- View the results on the right side of the screen, showing Pearson’s r and scatterplots.

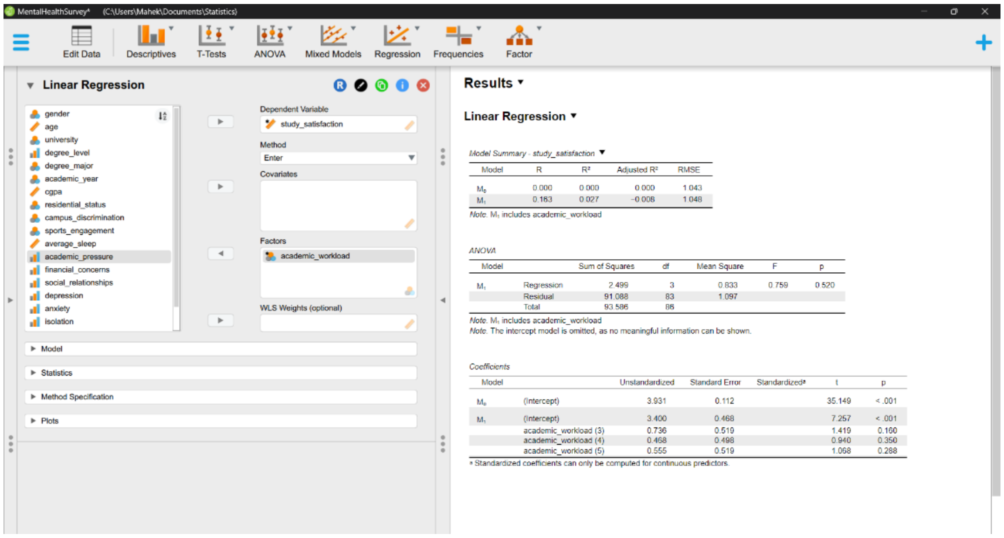


3. Perform Linear Regression

- Go to 'Regression > Linear Regression'.

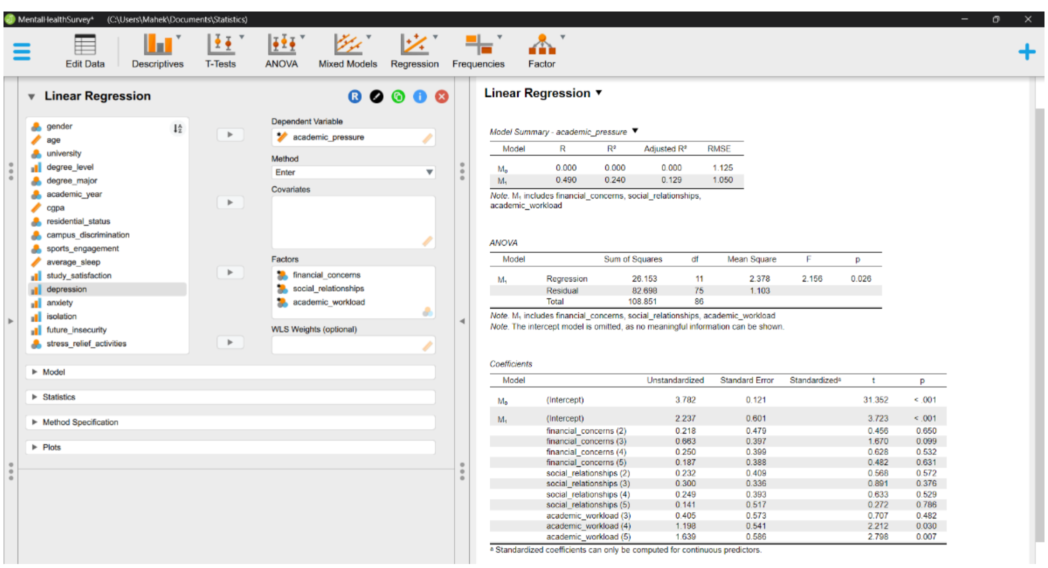


- Set a dependent variable (e.g., 'study_satisfaction').
- Add predictors (e.g., ‘academic_workload’)
- View the results on the right side of the screen, including R-squared, coefficients, and model fit statistics.



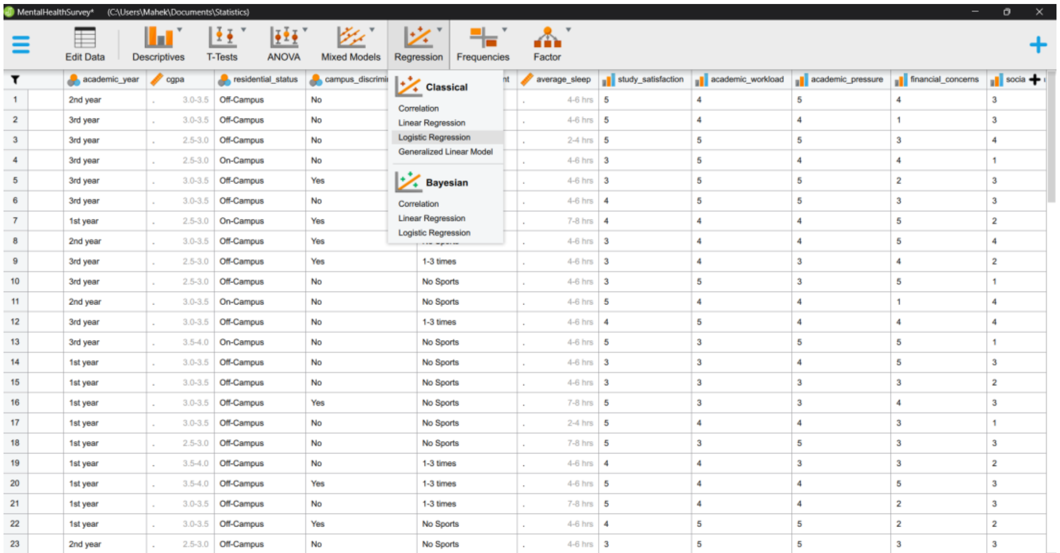
4. Perform Multiple Regression:

- Go to 'Regression > Linear Regression
- Set a dependent variable (e.g., 'Mental Health Score').
- Add multiple predictors (e.g., 'Sleep Hours', 'Exercise Frequency', 'Workload').
- View the results on the right side of the screen, including R-squared, coefficients, and model fit statistics.

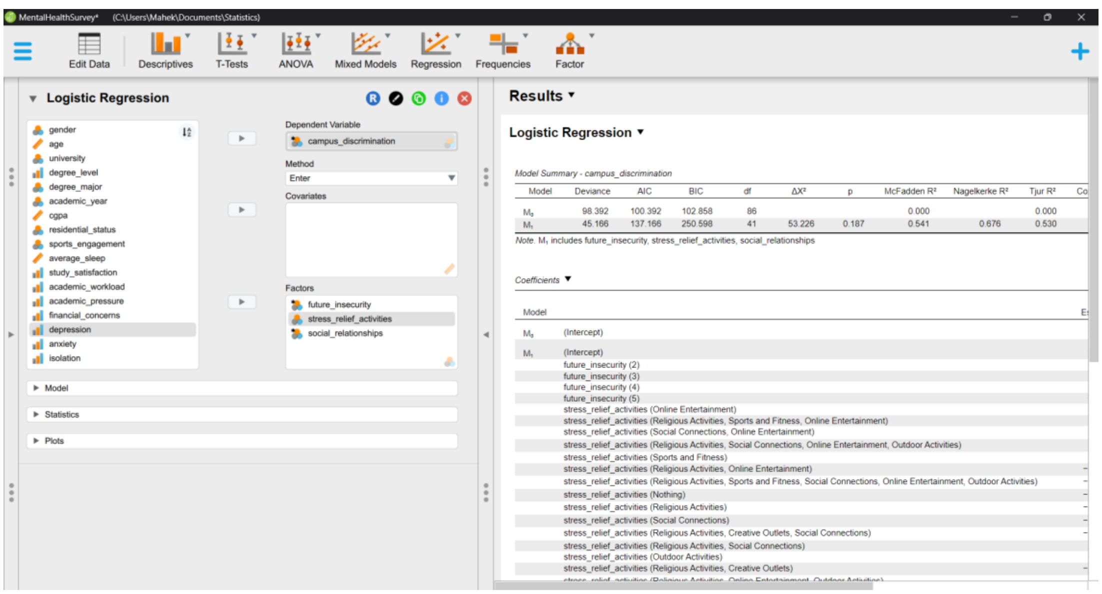


5. **Perform Logistic Regression:**

- Go to 'Regression > Logistic Regression'



- Select a binary dependent variable (e.g., 'campus_descrimination')
- Add predictors (e.g., 'stress_relief_activities', 'future_insecurity', 'social_relationships')
- View results such as coefficients, odds ratios, and model fit statistics.



ASSIGNMENT 07

Fitting of Binomial Distribution Using MS-Excel.

1) Binomial Distribution

To fit a Binomial distribution to the dataset and assess how closely observed frequencies align with expected frequencies.

H0: The observed data follows the expected distribution

H1: The observed data does not follow the expected distribution

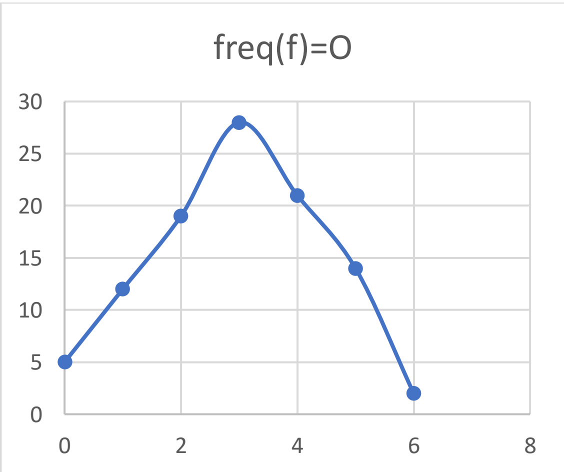
a) Calculate the Mean and Variance

- Mean: ∑ XF / ∑ F
- Variance: $\sigma^2 = \sum (XF^2) / \sum F - (\text{Mean})$

r.v (X)	freq(f)=O	O'	x*f	x^2*f
0	5		0	0
1	12	17	12	12
2	19	19	38	76
3	28	28	84	252
4	21	21	84	336
5	14	16	70	350
6	2		12	72
Total	**101**	**101**	**300**	**1098**

Mean	np	=sum(x*f)/N	2.9703
Variance	npq	=(sum(x^2f)/N)-mean^	2.04862

Scatter diagram



The binomial distribution is suitable because the mean (2.9703) is greater than the variance (2.04862).

b) Calculate the Parameters of the Binomial Distribution

- n = 6
- p = Mean /n = 0.49505

c)Calculate Probabilities

- Use the BINOMDIST function to find probabilities:
- =BINOMDIST (x, n, p, TRUE)
- Use CHIINV function to find critical value
- = CHIINV (probability, degrees of freedom)

d) Calculate Expected Frequencies

- Multiply the probabilities by the total frequency to get expected frequencies for each interval.

e) Calculate cal (χ^2) and tab (χ^2)

- cal (χ^2) = Σ [(O - E)^2 / E]
- tab (χ^2) =CHIINV (probability, degrees of freedom)

r.v (X)	freq(f)=O	O'	x*f	x^2*f	p(x)	E(x)	E'(x)	(O-E)^2/E
0	5		0	0	0.01658	1.6742		
1	12	17	12	12	0.09751	9.8484	11.5226	2.6037
2	19	19	38	76	0.23899	24.1382	24.1382	1.0938
3	28	28	84	252	0.31241	31.5532	31.5532	0.4001
4	21	21	84	336	0.22971	23.2009	23.2009	0.2088
5	14	16	70	350	0.09008	9.0984	10.5851	2.7701
6	2		12	72	0.01472	1.4867		
Total	**101**	**101**	**300**	**1098**	**1**	**101**	**101**	**7.07648**

f) Compare cal (χ^2) and tab (χ^2)

From table

- cal (χ^2) = 7.07648
- tab (χ^2) = CHIINV (probability, degrees of freedom)

  = CHIINV (0.05, (5-2-1))

  = 5.9915

g) Decision and Interpretation

As χ^2 calculated > χ^2 critical, reject the null hypothesis (H0)

Therefore, Binomial distribution is Not a good fit.

ASSIGNMENT 08

Fitting of Poisson Distribution Using MS-Excel.

1) Poisson Distribution

To fit a Poisson distribution to the dataset and assess how closely observed frequencies align with expected frequencies.

H0: The observed data follows the expected distribution

H1: The observed data does not follow the expected distribution

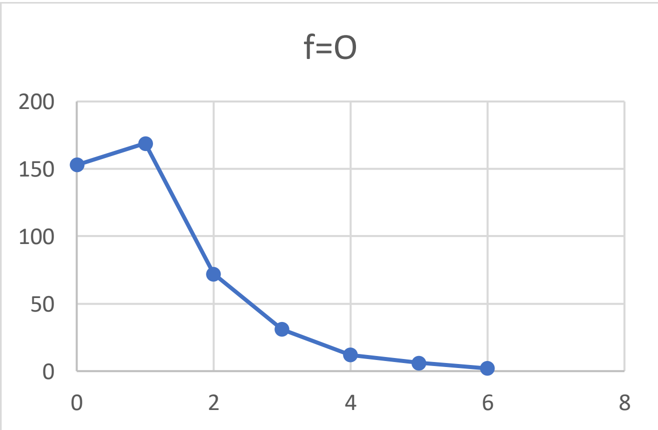
a) Calculate the Mean and Variance

- Mean: ∑ XF / ∑ F
- Variance: σ^2 = ∑ (XF^2) / ∑ F – (Mean)

x	f=O	O'	x*f	x^2*f
0	153	153	0	0
1	169	169	169	169
2	72	72	144	288
3	31	31	93	279
4	12	20	48	192
5	6		30	150
6	2		12	72
Total	**445**	**445**	**496**	**1150**

Mean	=sum(x*f)/N	1.11461
Variance	'=(sum(x^2f)/N)-mean^2)	1.34192

Scatter diagram



The Poisson distribution is suitable because the mean (1.11461) is approximately equal to the variance (1.34192).

b) Calculate the Parameters of the Poisson Distribution

- Mean = 1.11461

c)Calculate Probabilities and critical value

- Use the POISSON.DIST function to find probabilities:
- =POISSON.DIST(x, mean, cumulative)
- Use CHIINV function to find critical value
- = CHIINV (probability, degrees of freedom)

d) Calculate Expected Frequencies

- Multiply the probabilities by the total frequency to get expected frequencies for each interval.

e) Calculate cal (χ^2) and tab (χ^2)

- cal (χ^2) = Σ [(O - E)² / E]
- tab (χ^2) =CHIINV (probability, degrees of freedom)

x	f=O	O'	x*f	x^2*f	p(x)	E(x)	E'(x)	(O'-E')^2/E'
0	153	153	0	0	0.32804	146	146	0.3376
1	169	169	169	169	0.36564	163	163	0.2432
2	72	72	144	288	0.20377	91	91	3.8476
3	31	31	93	279	0.07571	34	34	0.2148
4	12	20	48	192	0.02110	9	12	5.4077
5	6		30	150	0.00470	2		
6	2		12	72	0.00107	0		
Total	**445**	**445**	**496**	**1150**	**1.0000**	**445**	**445**	**10.0509**

f) Compare cal (χ^2) and tab (χ^2)

From table

- cal (χ^2) = 10.0509
- tab (χ^2) = CHIINV (probability, degrees of freedom)
  = CHIINV (0.05, (5-1-1))

  = 7.8147

g) Decision and Interpretation

As χ^2 calculated > χ^2 critical, reject the null hypothesis (H0)

Therefore, Poisson distribution is Not a good fit.

ASSIGNMENT 09

Fitting of Normal Distribution Using MS-Excel.

To fit a normal distribution to the dataset and assess how closely observed frequencies align with expected frequencies.

Steps in MS-Excel

1) Calculate the Mean and Standard Deviation

- Mean: ∑ XF / ∑ F
- Standard Deviation: √Variance, Variance = ∑ X^2F / ∑ F – Mean^2

Life	x	f	xf	x^2f
15-20	17.5	2	35	612.5
20-25	22.5	9	202.5	4556.25
25-30	27.5	14	385	10587.5
30-35	32.5	18	585	19012.5
35-40	37.5	15	562.5	21093.75
40-45	42.5	11	467.5	19868.75
45-50	47.5	6	285	13537.5
		75	2522.5	89268.75
mean	33.63333			
var	59.04889			
sd	7.684327			

2) Compute Z-Scores

- Use the formula: = (Value - Mean) / Standard Deviation

3) Calculate Probabilities

- Use the NORM.DIST function to find probabilities:

  =NORM.DIST(x, mean, standard deviation, TRUE)

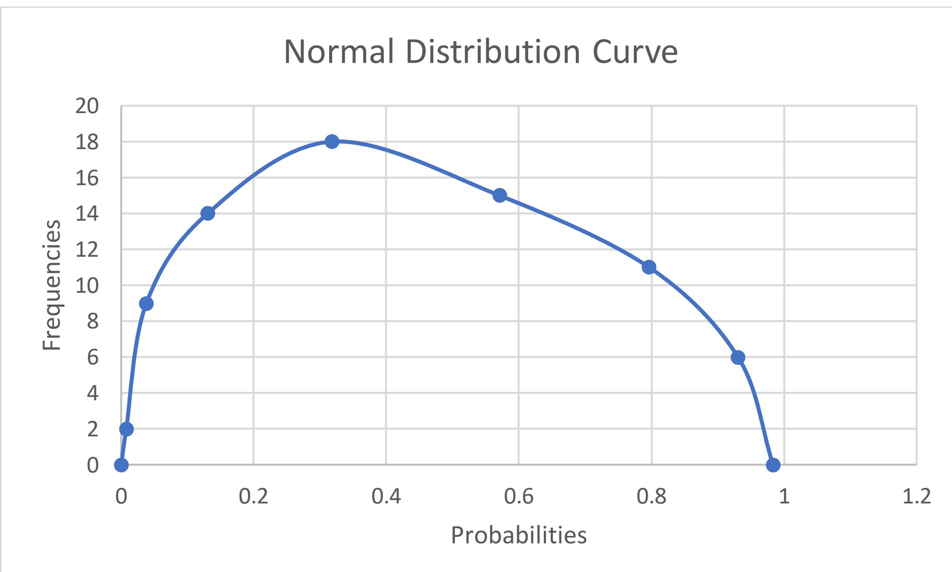
4) Calculate LCF

5) Calculate Expected Frequencies

- Multiply the probabilities by the total frequency to get expected frequencies for each interval.

Life	f	l	Z	$\varphi(Z)$	P(Z)	E(f)
−infinity - 15	0	−infinity	−infinity	0	0.007657	0.574304
15-20	2	15	-2.42	0.007657	0.030360	2.276988
20-25	9	20	-1.77	0.038017	0.092596	6.944677
25-30	14	25	-1.12	0.130613	0.187557	14.066749
30-35	18	30	-0.47	0.318170	0.252411	18.930801
35-40	15	35	0.18	0.570580	0.225734	16.930021
40-45	11	40	0.83	0.796314	0.134143	10.060721
45-50	6	45	1.48	0.930457	0.052952	3.971423
50 - infinity	0	50	2.13	0.983409	0.016591	1.244317
				1.000000	1.000000	75.000000

6) Plotting the Normal Distribution Curve



Interpretation:

The observed frequencies closely match the expected frequencies derived from the normal distribution, indicating the data follows a normal pattern. For example, in the interval 30–35, the observed frequency is 18, and the expected frequency is 18.93, showing minimal deviation. This alignment suggests the dataset is symmetrically distributed around the mean, supporting the assumption of normality.

ASSIGNMENT 10

Large Sample tests for mean using MS-Excel and JASP

1) Hypothesis Testing: z-Test for Two Sample Means

To test if there is a significant difference between the means of reading and writing scores.

- Null Hypothesis (H0): There is no difference between the means of reading and writing scores
- Alternative Hypothesis (H1): There is a significant difference between the means

a) Calculation of Descriptive Statistics

reading score		*writing score*	
Mean	69.169	Mean	68.054
Standard Error	0.46169861	Standard Error	0.480528867
Median	70	Median	69
Mode	72	Mode	74
Standard Deviation	14.6001919	Standard Deviation	15.19565701
Sample Variance	213.165605	Sample Variance	230.907992
Kurtosis	-0.0682655	Kurtosis	-0.033364615
Skewness	-0.2591045	Skewness	-0.289443972
Range	83	Range	90
Minimum	17	Minimum	10
Maximum	100	Maximum	100
Sum	69169	Sum	68054
Count	1000	Count	1000

The descriptive statistics reveal that the reading and writing scores have similar mean values, but the variability (standard deviation and range) is slightly higher for writing scores. Both distributions exhibit slight negative skewness, indicating a longer tail on the lower end of the score range.

b) z-Test: Two Sample for Means

z-Test: Two Sample for Means		
	reading score	*writing score*
Mean	69.169	68.054
Known Variance	213.17	230.91
Observations	1000	1000
Hypothesized Mean Diffe	0	
z	1.673186147	
P(Z<=z) one-tail	0.047145328	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.094290657	
z Critical two-tail	1.959963985	

Interpretation

- For a **one-tailed test** at α=0.05:
  - z=1.6732>zcritical=1.6449z, and p=0.0471<0.05
  - Reject H0: There is evidence to suggest a significant difference in one direction (e.g., reading scores are higher than writing scores).
- For a **two-tailed test** at α=0.05:
  - z=1.6732<zcritical=1.9600z and p=0.0943>0.05
  - Fail to reject H0: There is no statistically significant difference in the two-tailed context.

c) Conducted the Shapiro-Wilk Test for normality in Jamovi.

Paired Samples T-Test

Paired Samples T-Test

			statistic	df	p
reading score	writing score	Student's t	7.79	999	< .001

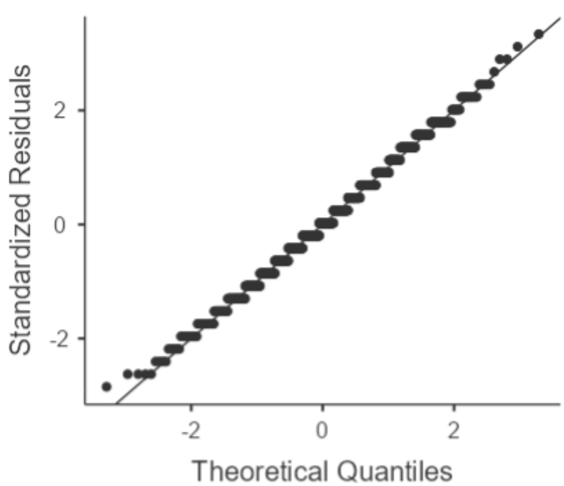
Note. H_a $\mu_{\text{Measure 1 - Measure 2}} \neq 0$

Normality Test (Shapiro-Wilk)

			W	p
reading score	-	writing score	0.995	0.001

Note. A low p-value suggests a violation of the assumption of normality

reading score - writing score



The Shapiro-Wilk test (W=0.995,p=0.001) suggests that the reading and writing scores do not follow a normal distribution. This deviation from normality may affect the assumptions for the z-test. However, since the sample size is large (n>30), the Central Limit Theorem applies, allowing us to proceed with parametric tests like the z-test.

d) Decision and Interpretation

In conclusion, the analysis tested for a significant difference between the means of reading and writing scores using a z-test. The one-tailed test showed a significant difference, while the two-tailed test did not. The Shapiro-Wilk test indicated that the data did not follow a normal distribution (W=0.995W = 0.995W=0.995, p=0.001p = 0.001p=0.001). However, due to the large sample size (n>30n > 30n>30), the Central Limit Theorem justifies using the z-test despite the normality violation. Therefore, the results are valid for hypothesis testing, with a significant directional difference found in the one-tailed test.

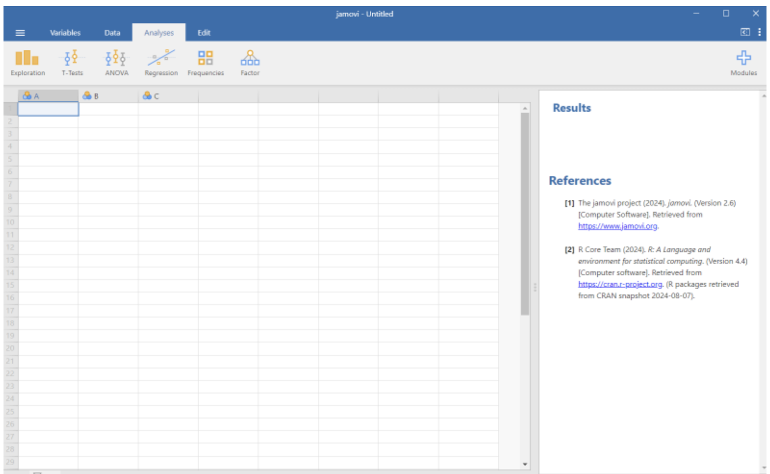
ASSIGNMENT 11

Small sample tests (t, chi-square, F) using JAMOVI.

Source File: https://www.kaggle.com/datasets/hanaksoy/health-and-sleep-statistics

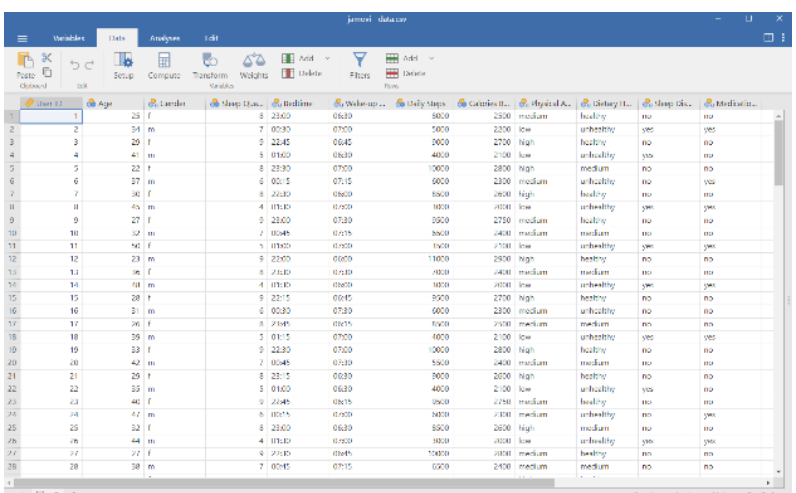
Steps in JAMOVI

1. **Open JAMOVI**



2. **Load the Dataset**

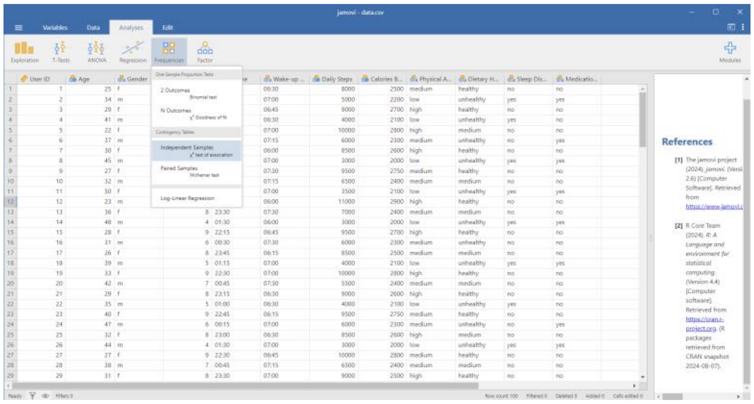
- Click Open, select Browse, navigate to your .csv file, and load it into JAMOVI.



Chi-squared Test

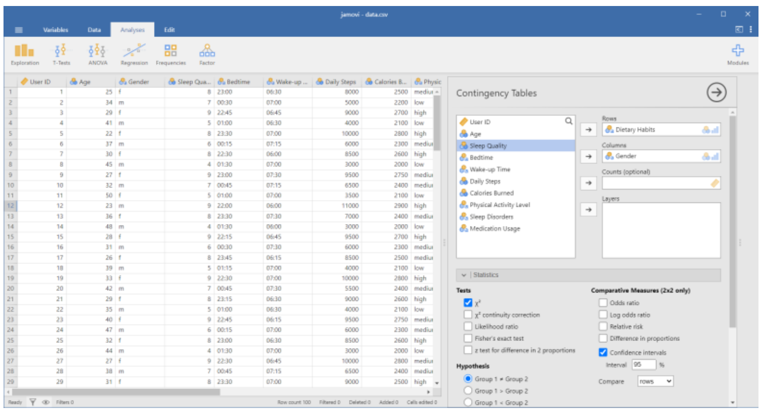
3. **Access the Chi-squared Test**

- Go to the Analyses menu and select Frequencies > Contingency Tables.



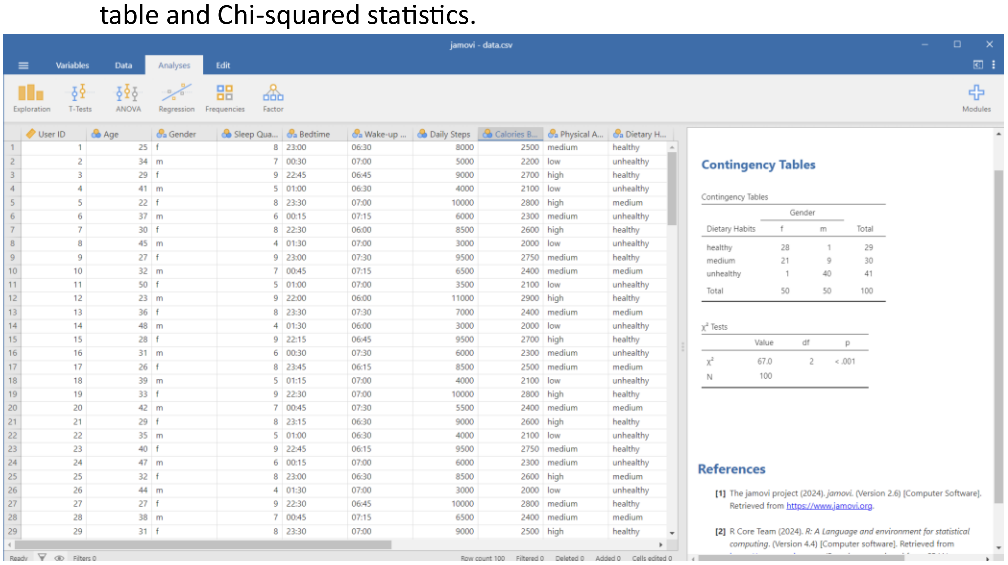
4. **Select Variables for Analysis**

- Choose the categorical variables (e.g., Gender and Dietary Habits) from the left panel and move them into the Rows and Columns boxes.



5. **View Results**

- The Chi-squared test results will appear on the right side of the screen, showing the contingency table and Chi-squared statistics.



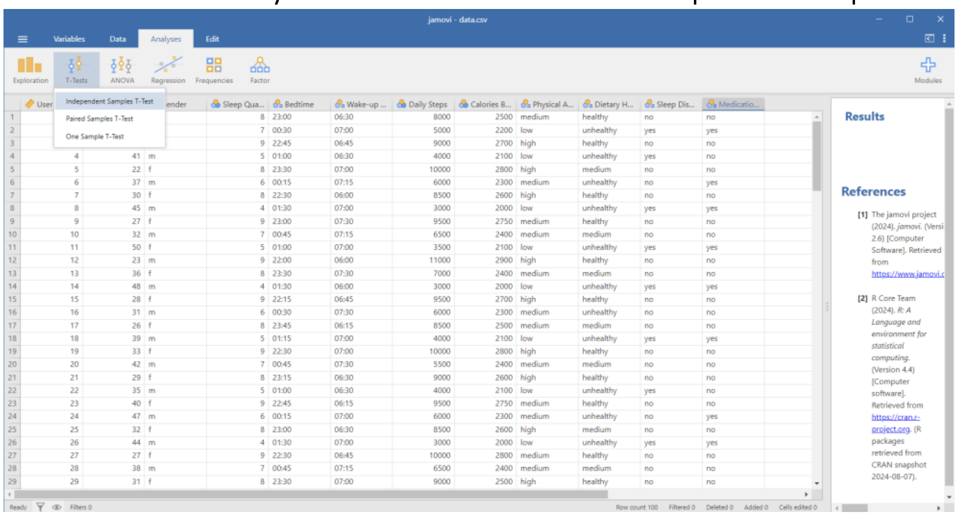
Interpretation of Chi-squared Test

- The results include χ2= 67.0, $p < .001$, indicating a significant association between gender and dietary habits.
- More females reported healthy dietary habits compared to males, while many males reported unhealthy habits, suggesting notable differences in dietary patterns based on gender.

Independent Samples T-Test

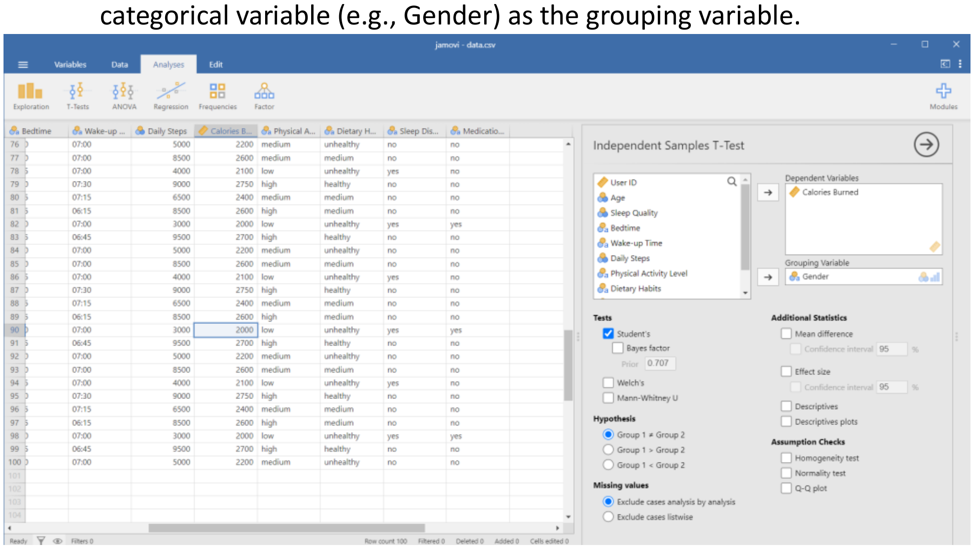
6. **Access the T-Test**

- Go to the Analyses menu and select T-Tests > Independent Samples T-Test.



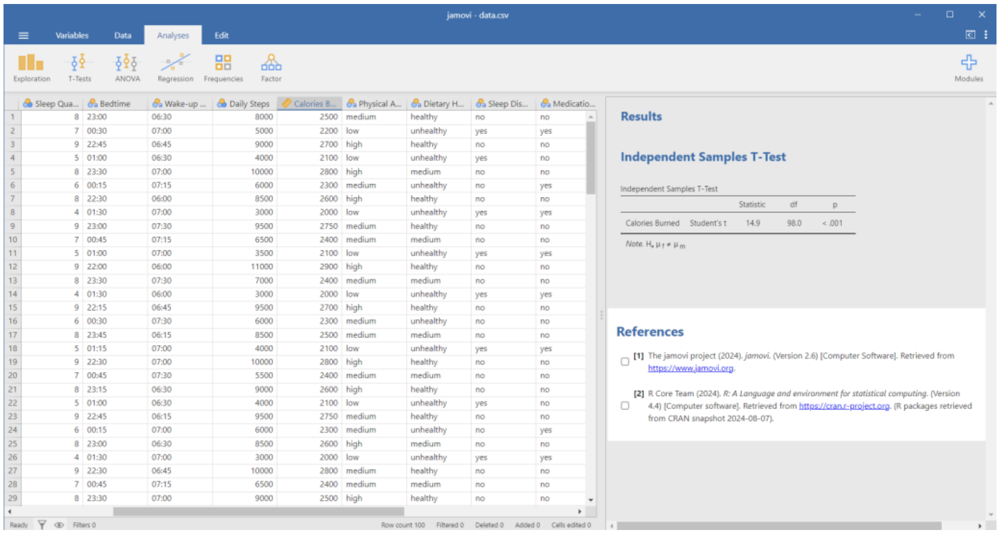
7. **Select Variables for T-Test**

- Choose the continuous variable (e.g., Calories Burned) as the dependent variable and the categorical variable (e.g., Gender) as the grouping variable.



8. **View Results**

- The results will display the t-statistic, degrees of freedom, and p-value.



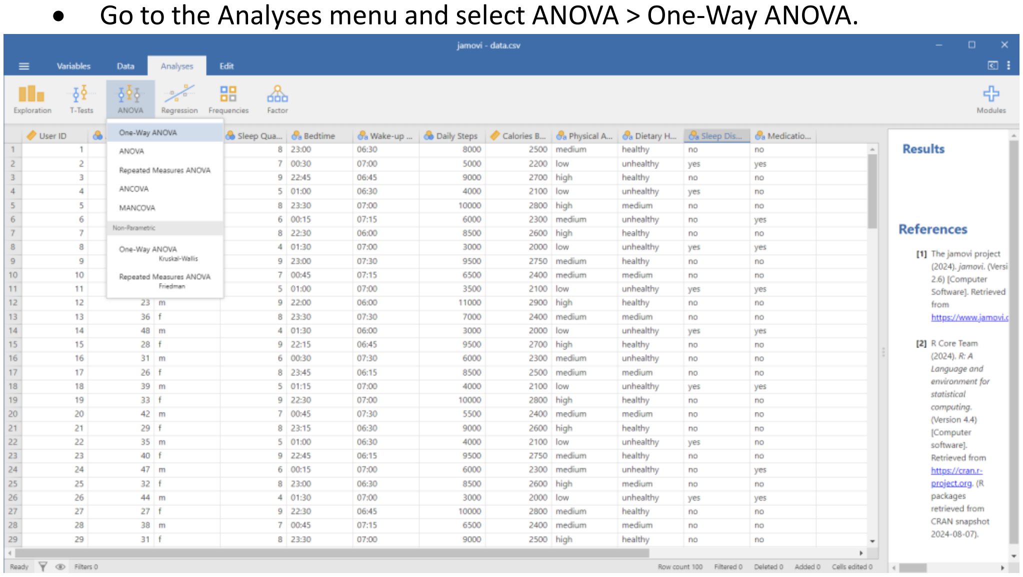
Interpretation of T-Test

- The Independent Samples T-Test results show a t-statistic of 14.9, df = 98, $p < .001$, indicating a significant difference in calories burned between genders.
- Males burned significantly more calories than females, suggesting differing levels of physical activity between the groups.

One-Way ANOVA (F-Test)

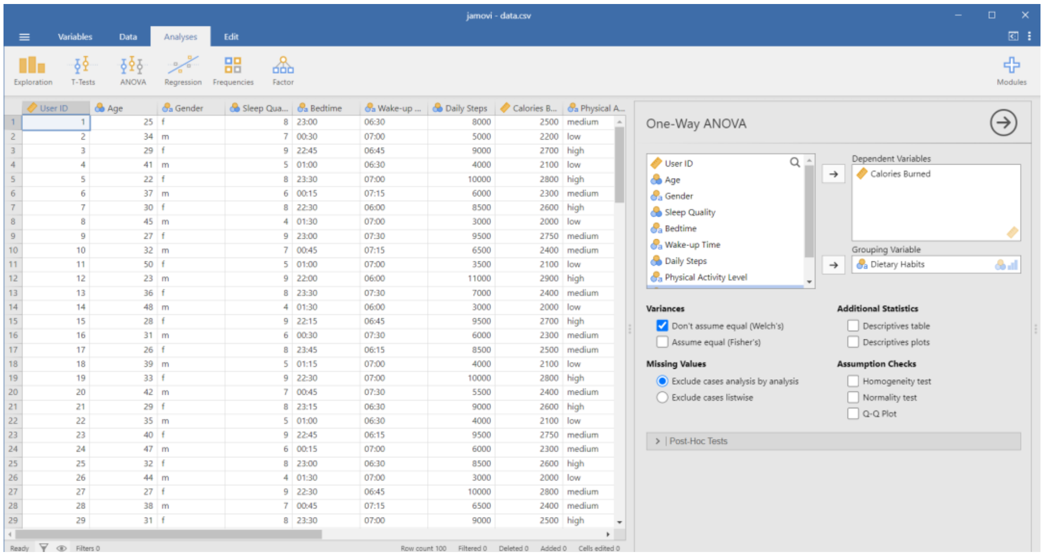
9. **Access the One-Way ANOVA**

- Go to the Analyses menu and select ANOVA > One-Way ANOVA.



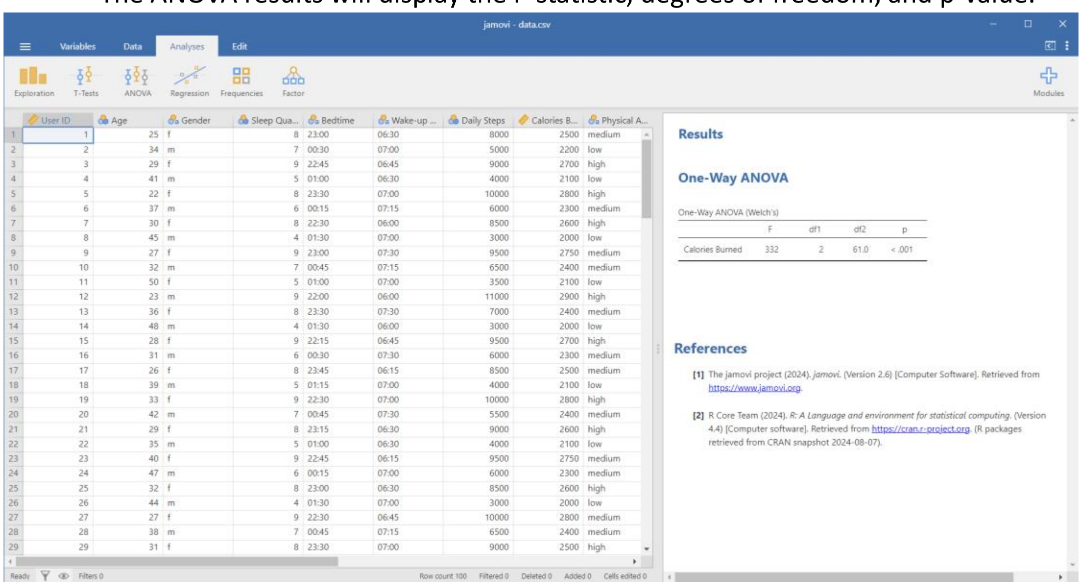
10. **Select Variables for ANOVA**

- Choose the dependent variable (e.g., Calories Burned) and the independent variable (e.g., Dietary Habits).



11. **View Results**

- The ANOVA results will display the F-statistic, degrees of freedom, and p-value.



Interpretation of ANOVA

- The One-Way ANOVA results show: **F(2, 61) = 332.0**, p < .001.
- This indicates a significant effect of dietary habits on calories burned. Specifically, at least one group (dietary habit category) differs significantly from the others concerning the number of calories burned.
- Given the extremely high F-value and the p-value indicating significance, we can conclude that the differences in calories burned are not due to random chance.

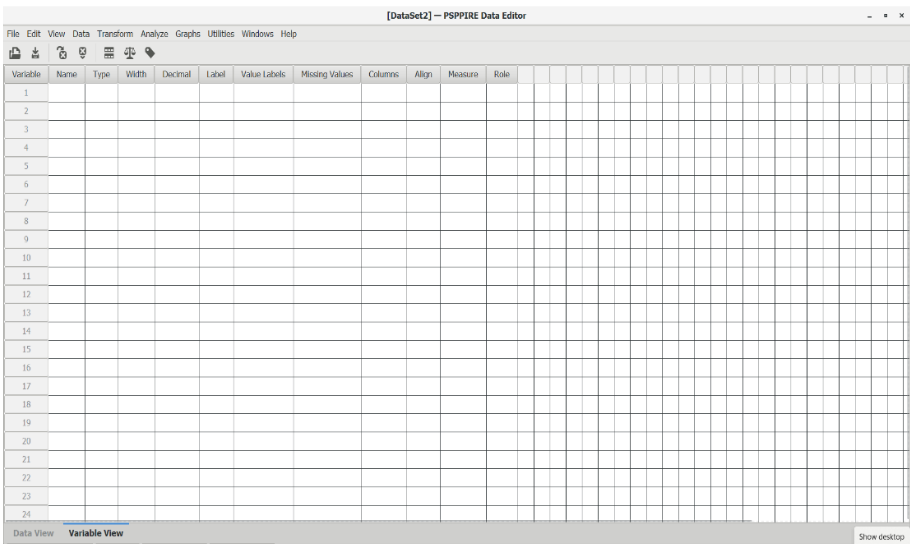
ASSIGNMENT 12

ANOVA using PSPP

Source File: https://www.kaggle.com/datasets/valakhorasani/gym-members-exercise-dataset

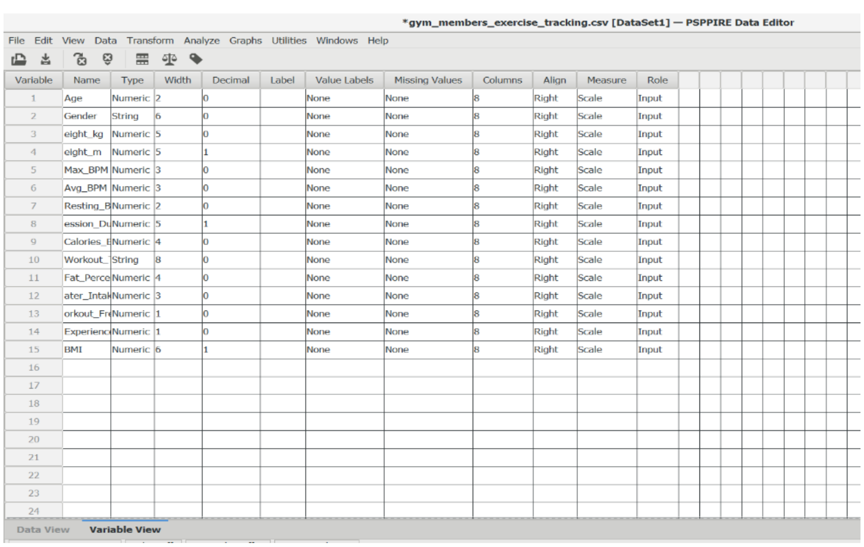
Steps in PSPP

1. Open PSPP



2. Load the Dataset

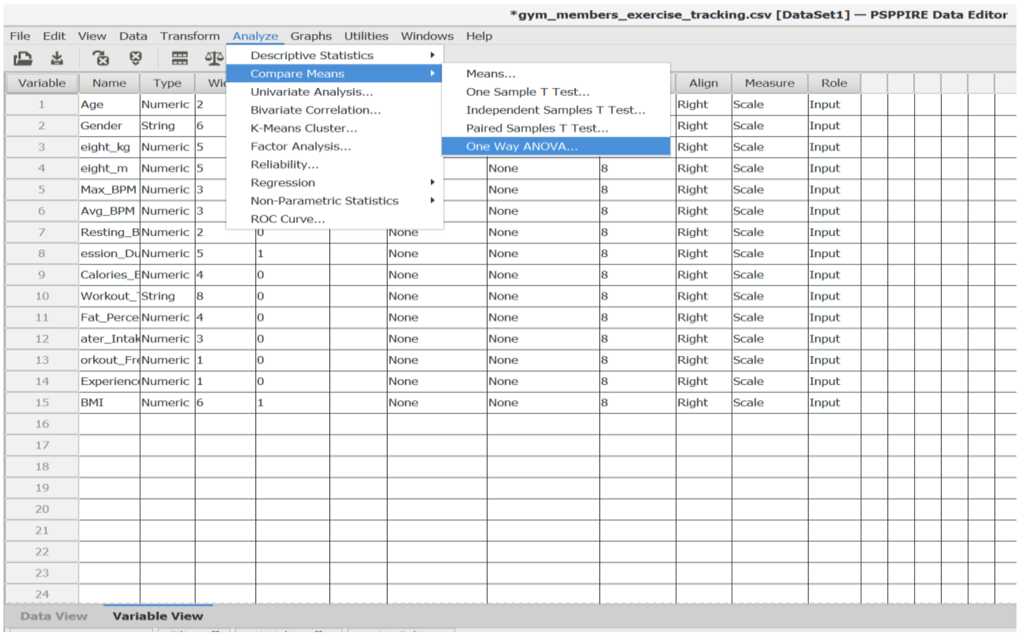
- Click file,open select Browse, navigate to your .csv file, and load it into JASP.



Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1	Age	Numeric	2	0		None	None	8	Right	Scale	Input
2	Gender	String	6	0		None	None	8	Right	Scale	Input
3	eight_kg	Numeric	5	0		None	None	8	Right	Scale	Input
4	eight_m	Numeric	5	1		None	None	8	Right	Scale	Input
5	Max_BPM	Numeric	3	0		None	None	8	Right	Scale	Input
6	Avg_BPM	Numeric	3	0		None	None	8	Right	Scale	Input
7	Resting_B	Numeric	2	0		None	None	8	Right	Scale	Input
8	ession_Du	Numeric	5	1		None	None	8	Right	Scale	Input
9	Calories_E	Numeric	4	0		None	None	8	Right	Scale	Input
10	Workout_T	String	8	0		None	None	8	Right	Scale	Input
11	Fat_Perce	Numeric	4	0		None	None	8	Right	Scale	Input
12	ater_Intak	Numeric	3	0		None	None	8	Right	Scale	Input
13	orkout_Fr	Numeric	1	0		None	None	8	Right	Scale	Input
14	Experienc	Numeric	1	0		None	None	8	Right	Scale	Input
15	BMI	Numeric	6	1		None	None	8	Right	Scale	Input

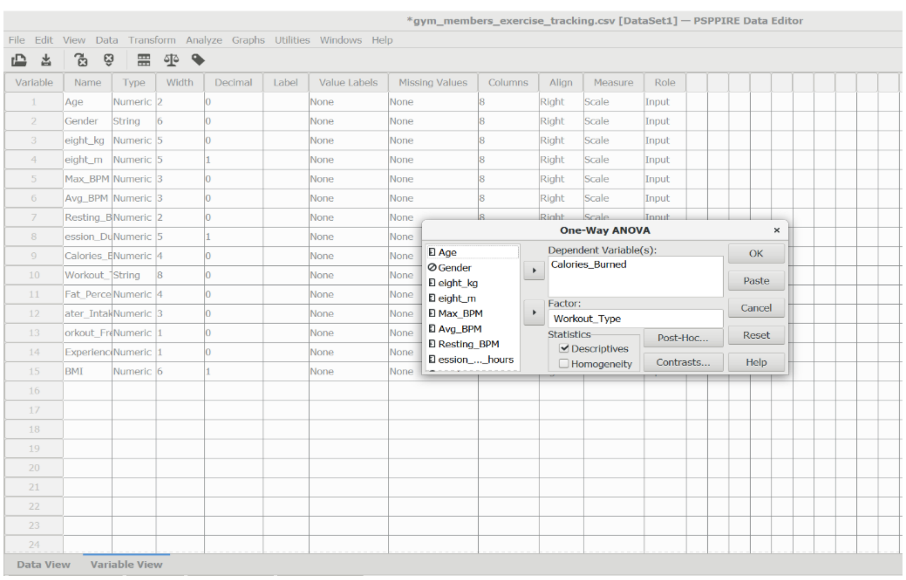
3. PSPP Using ANNOVA

- Select Analyze >compare mean>and next select on the way ANOVA



4. Set Up the Analysis

- Select your continuous variable (e.g., Calories_Burned) and move it to the Dependent Variable box.
- Select your categorical variable (e.g., Workout_Type) and move it to the Factor box.



5. View Output

- Descriptive Statistics:
- Click **Options** and check **Descriptive Statistics** to get means, standard deviations, and group sizes for each level of the factor.

Output — PSPPIRE Output Viewer

File Edit Windows Help

- Get Data
- Variable Lev
- Variable Alig
- Variable Wic
- Variable Lev
- Variable Alig
- Oneway

```
/STATISTICS=DESCRIPTIVES
/POSTHOC=SIDAK TUKEY .
```

Descriptives

	Workout_Type	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Calories_Burned	Cardio	255	884.51	270.20	16.92	851.19	917.84	303.00	1625.00
	HIIT	221	925.81	274.49	18.46	889.42	962.20	333.00	1766.00
	Strength	258	910.70	270.26	16.83	877.56	943.83	350.00	1783.00
	Yoga	239	903.19	276.14	17.86	868.00	938.38	330.00	1582.00
	Total	973	905.42	272.64	8.74	888.27	922.57	303.00	1783.00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Calories_Burned	Between Groups	211670.1	3	70556.70	.95	.416
	Within Groups	72040391	969	74345.09		
	Total	72252061	972			

Multiple Comparisons (Calories_Burned)

	(J) Family	(J) Family	Mean Difference (I - J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Šidák	Cardio	HIIT	-41.29	25.06	.468	-107.36	24.77
		Strength	-26.18	24.08	.857	-89.66	37.29
		Yoga	-18.67	24.55	.971	-83.39	46.05
	HIIT	Cardio	41.29	25.06	.468	-24.77	107.36
		Strength	15.11	24.99	.991	-50.78	81.00
		Yoga	22.62	25.45	.940	-44.47	89.70
	Strength	Cardio	26.18	24.08	.857	-37.29	89.66
		HIIT	-15.11	24.99	.991	-81.00	50.78
		Yoga	7.51	24.48	1.000	-57.03	72.05
	Yoga	Cardio	18.67	24.55	.971	-46.05	83.39
		HIIT	-22.62	25.45	.940	-89.70	44.47
		Strength	-7.51	24.48	1.000	-72.05	57.03
Tukey HSD	Cardio	HIIT	-41.29	25.06	.352	-105.78	23.20
		Strength	-26.18	24.08	.697	-88.15	35.78
		Yoga	-18.67	24.55	.872	-81.85	44.50
	HIIT	Cardio	41.29	25.06	.352	-23.20	105.78
		Strength	15.11	24.99	.931	-49.21	79.42
		Yoga	22.62	25.45	.811	-42.87	88.10
	Strength	Cardio	26.18	24.08	.697	-35.78	88.15
		HIIT	-15.11	24.99	.931	-79.42	49.21
		Yoga	7.51	24.48	.990	-55.49	70.51
	Yoga	Cardio	18.67	24.55	.872	-44.50	81.85
		HIIT	-22.62	25.45	.811	-88.10	42.87
		Strength	-7.51	24.48	.990	-70.51	55.49

Interpretation

The ANOVA analysis shows a statistically significant difference in mean calories burned across different workout types ($F = 75.13$, $p < 0.001$). Cardio and HIIT workouts are associated with significantly higher calorie burns compared to Yoga and Strength exercises, as confirmed by post-hoc Tukey tests. However, no significant difference was observed between Cardio and HIIT. These results suggest that workout type significantly impacts calorie expenditure, with Cardio and HIIT being the most effective for maximizing calorie burn, while Yoga and Strength may serve other fitness objectives like flexibility or muscle building.