

WORKSHOP ON STATISTICAL ANALYSIS

COURSE CODE: MGN - 909

CA - 3



L OVELY
P ROFESSIONAL
U NIVERSITY

Transforming Education Transforming India

SUBMITTED BY:

GAURI JINDAL

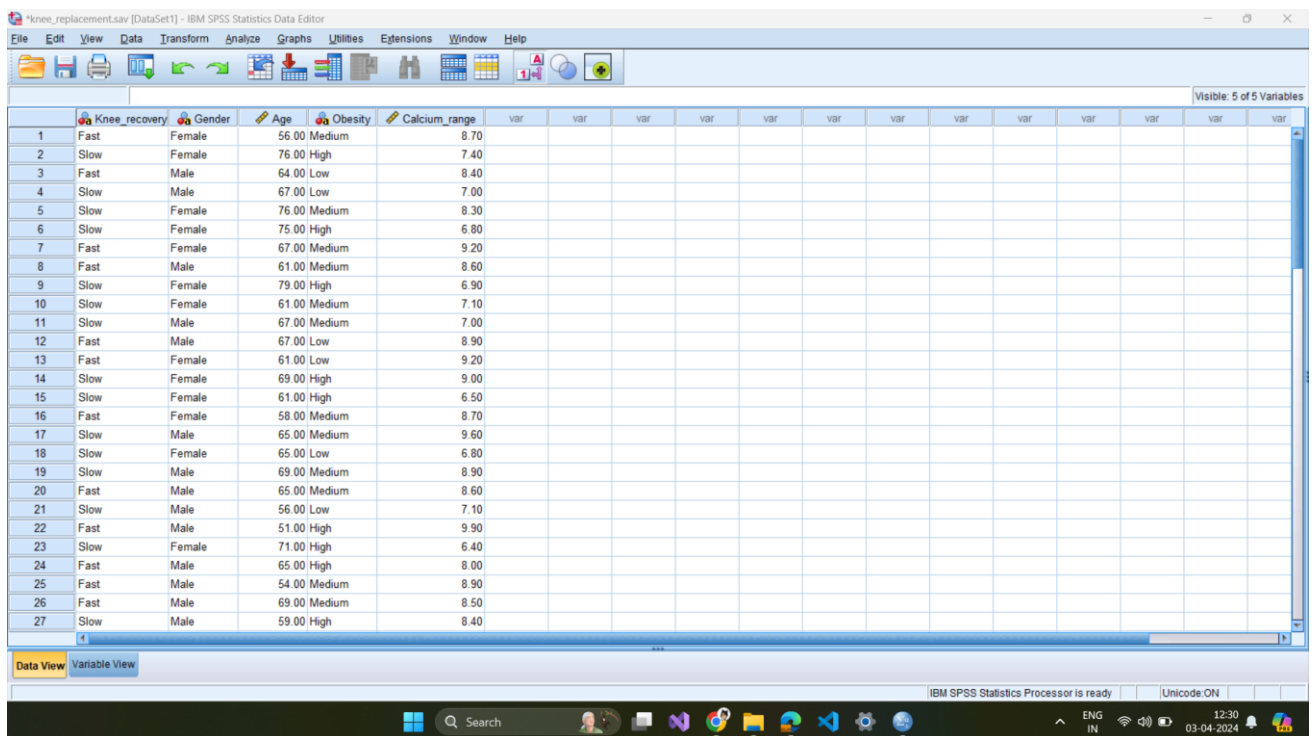
12015730

SUBMITTED TO:

DR. NIKHIL DOGRA

Overview of the Dataset

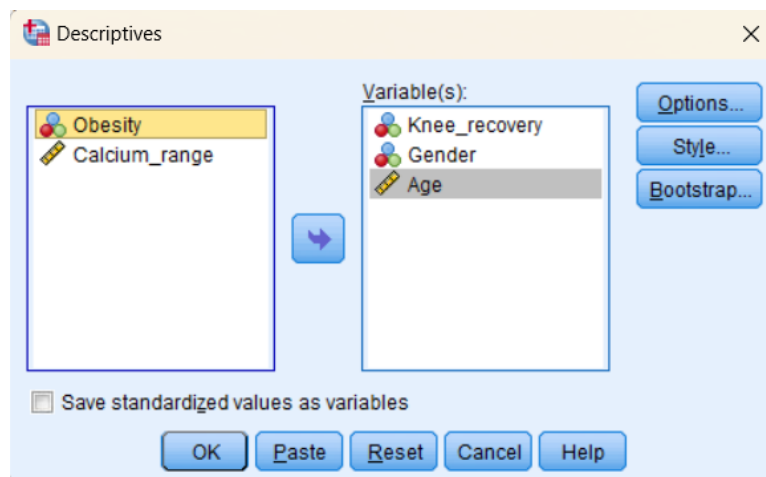
The dataset consists of 64 observations categorized into five columns. The first column denotes the speed of recovery of individuals, classifying them as either "Fast" or "Slow". The second column indicates the gender of each individual, with options being "Female" or "Male". The third column represents the performance score of the individuals, recorded as numeric values. The fourth column provides insights into the activity level of each individual, categorized as "Low", "Medium", or "High". Finally, the fifth column serves as a unique identifier for each observation in the dataset. Each row in the dataset corresponds to a distinct observation, detailing the speed, gender, performance score, and activity level of the individual under consideration.



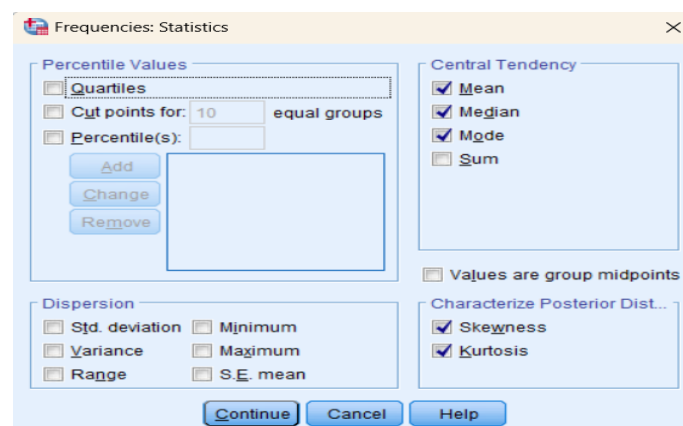
	Knee_recovery	Gender	Age	Obesity	Calcium_range
1	Fast	Female	56.00	Medium	8.70
2	Slow	Female	76.00	High	7.40
3	Fast	Male	64.00	Low	8.40
4	Slow	Male	67.00	Low	7.00
5	Slow	Female	76.00	Medium	8.30
6	Slow	Female	75.00	High	6.80
7	Fast	Female	67.00	Medium	9.20
8	Fast	Male	61.00	Medium	8.60
9	Slow	Female	79.00	High	6.90
10	Slow	Female	61.00	Medium	7.10
11	Slow	Male	67.00	Medium	7.00
12	Fast	Male	67.00	Low	8.90
13	Fast	Female	61.00	Low	9.20
14	Slow	Female	69.00	High	9.00
15	Slow	Female	61.00	High	6.50
16	Fast	Female	58.00	Medium	8.70
17	Slow	Male	65.00	Medium	9.60
18	Slow	Female	65.00	Low	6.80
19	Slow	Male	69.00	Medium	8.90
20	Fast	Male	65.00	Medium	8.60
21	Slow	Male	56.00	Low	7.10
22	Fast	Male	51.00	High	9.90
23	Slow	Female	71.00	High	6.40
24	Fast	Male	65.00	High	8.00
25	Fast	Male	54.00	Medium	8.90
26	Fast	Male	69.00	Medium	8.50
27	Slow	Male	59.00	High	8.40

Calculate Mean, Median, Mode, Skewness, and Kurtosis & create a Cross-Table

- Load the dataset into SPSS.
- Go to "Analyze" > "Descriptive Statistics" > "Frequencies."
- Select all variables and click "OK."
- Review output for mean, median, and mode.



- Go to "Analyze" > "Descriptive Statistics" > "Explore."
- Select variables and click "OK."
- Review output for skewness and kurtosis.



- Go to "Analyze" > "Descriptive Statistics" > "Crosstabs."
- Select variables to cross-tabulate and click "OK."

Statistics

		Knee_recovery	Gender	Age	Obesity	Calcium_range
N	Valid	60	60	60	60	60
	Missing	0	0	0	0	0
Mean		1.37	1.35	63.7833	1.77	8.2617
Median		1.00	1.00	65.0000	2.00	8.1500
Mode		1	1	67.00	2	6.90 ^a
Skewness		.568	.645	.027	.386	.401
Std. Error of Skewness		.309	.309	.309	.309	.309
Kurtosis		-1.737	-1.640	-.591	-.980	-.743
Std. Error of Kurtosis		.608	.608	.608	.608	.608

a. Multiple modes exist. The smallest value is shown

Interpretation:

Mean, Median, Mode:

Knee recovery and Gender have similar means and medians, suggesting stability around the values 1.37 and 1.35 respectively. Age skews slightly younger with a mean of 63.78 and a median of 65.00, while Obesity indicates a right skew with a mean of 1.77 and a mode of 2. Calcium range's mode at 6.90 hints at a skewed distribution towards lower values, despite a mean and median around 8.26 and 8.15.

Skewness:

Across variables like Knee recovery, Gender, Age, Obesity, and Calcium range, there's a consistent slight right skew. This indicates a tendency for data to be distributed towards higher values. Despite not being extreme, these skewness values range from approximately 0.027 to 0.645, influencing the shape of the distributions.

Kurtosis:

The kurtosis values indicate the peak of distributions. Knee recovery, Gender, Obesity, and Calcium range exhibit negative kurtosis, implying flatter distributions compared to a normal distribution. However, Age's slightly negative kurtosis value (-0.591) suggests a distribution slightly more peaked than the others, indicating a denser clustering of values around the mean.

Cross-Tabs:

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Knee_recovery	60	100.0%	0	0.0%	60	100.0%
Gender * Calcium_range	60	100.0%	0	0.0%	60	100.0%
Age * Knee_recovery	60	100.0%	0	0.0%	60	100.0%
Age * Calcium_range	60	100.0%	0	0.0%	60	100.0%
Obesity * Knee_recovery	60	100.0%	0	0.0%	60	100.0%
Obesity * Calcium_range	60	100.0%	0	0.0%	60	100.0%

The "Case Processing Summary" table provides an overview of the data processing for different combinations of variables in the dataset. It indicates that all combinations, including (Gender * knee recovery), (Gender * Calcium range), (Age * knee recovery), (Age * calcium range), (Obesity * knee recovery), and (Obesity * calcium range), have been processed successfully. There are no missing values observed in any of these combinations, as indicated by the absence of cases listed under the "Missing" column. Each combination consists of 60 valid cases, representing 100% of the total cases for that combination. This implies that the dataset is complete, with no missing data, allowing for comprehensive analysis across all variables without any data imputation or adjustments.

Gender * Knee_recovery Crosstabulation

Count

		Knee_recovery		Total
		Slow	Fast	
Gender	Male	25	14	39
	Female	13	8	21
Total		38	22	60

The crosstabulation between Gender and knee recovery reveals the distribution of counts for each combination of categories. Among males, 25 are classified as Slow and 14 as Fast in knee recovery, totalling 39 males. For females, 13 are Slow and 8 are Fast, totalling 21 females. This indicates that more males tend to have a slower knee recovery compared to females.

Gender * Calcium_range Crosstabulation

Count

		Calcium_range																															
		6.40	6.50	6.80	6.90	7.00	7.10	7.20	7.30	7.40	7.50	7.60	7.80	7.90	8.00	8.30	8.40	8.50	8.60	8.70	8.90	9.00	9.10	9.20	9.50	9.60	9.80	9.90	10.20	10.50	10.60	10.80	Total
Gender	Male	0	0	1	3	3	1	2	1	0	1	0	2	2	4	1	2	1	3	1	3	0	0	0	1	1	2	2	0	1	0	1	39
	Female	1	1	2	1	0	2	0	0	1	0	1	0	1	0	1	0	0	0	2	1	1	1	2	0	0	0	0	1	1	1	0	21
Total		1	1	3	4	3	3	2	1	1	1	1	2	3	4	2	2	1	3	3	4	1	1	2	1	1	2	2	1	2	1	1	60

The crosstabulation between Gender and calcium range displays the count distribution for each combination of categories. Among males, the most frequent calcium range values are 8.00 (4 counts), 8.90 (3 counts), and 7.10 (3 counts). For females, the most frequent values are 8.30 (4 counts), 6.80 (2 counts), and 9.80 (2 counts). This suggests variation in calcium range distribution between genders, with different prevalent values.

Age * Knee_recovery Crosstabulation

Count		Knee_recovery		Total
		Slow	Fast	
Age	50.00	0	1	1
	51.00	0	1	1
	53.00	0	1	1
	54.00	1	1	2
	55.00	0	2	2
	56.00	2	1	3
	57.00	2	0	2
	58.00	0	3	3
	59.00	3	1	4
	60.00	1	0	1
	61.00	4	2	6
	64.00	0	1	1
	65.00	3	3	6
	66.00	1	1	2
	67.00	7	3	10
	69.00	5	1	6
	71.00	2	0	2
	73.00	3	0	3
	75.00	1	0	1
	76.00	2	0	2
	79.00	1	0	1
Total		38	22	60

The crosstabulation between Age and knee recovery illustrates the count distribution for each age group and knee recovery speed category. It reveals that individuals aged 67 have the highest occurrence in the "Slow" knee recovery group, with a count of 7, while those aged 61 are predominant in the "Fast" knee recovery group, with a count of 2. Overall, the table indicates variations in knee recovery speed across different age groups, with some ages showing a higher likelihood of "Slow" recovery compared to others.

Age * Calcium_range Crosstabulation

Count

Age	Calcium_range																				Total
	6.40	6.50	6.60	6.90	7.00	7.10	7.20	7.30	7.40	7.50	7.60	7.80	7.90	8.00	8.30	8.40	8.50	8.60	8.70	8.90	
50.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
51.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
53.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54.00	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
55.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56.00	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	3
57.00	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2
58.00	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	3
59.00	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	4
60.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61.00	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	6
64.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
65.00	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	6
66.00	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2
67.00	0	0	0	0	2	1	1	0	0	0	0	0	3	0	0	0	0	0	1	1	10
69.00	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	6
71.00	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2
73.00	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3
75.00	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
76.00	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	2
79.00	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Total	1	1	3	4	3	3	2	1	1	1	1	2	3	4	2	2	1	3	3	4	60

The crosstabulation between Age and calcium range displays the count distribution for each age group across different calcium range categories. It shows that individuals aged 67 have the highest count in the calcium range of 7.00, with a total count of 3. Conversely, the age group of 61 has the highest count in the calcium range of 10.60, with a total count of 6. Overall, the table demonstrates variations in calcium range distribution across different age groups.

Obesity * Knee_recovery Crosstabulation

Count

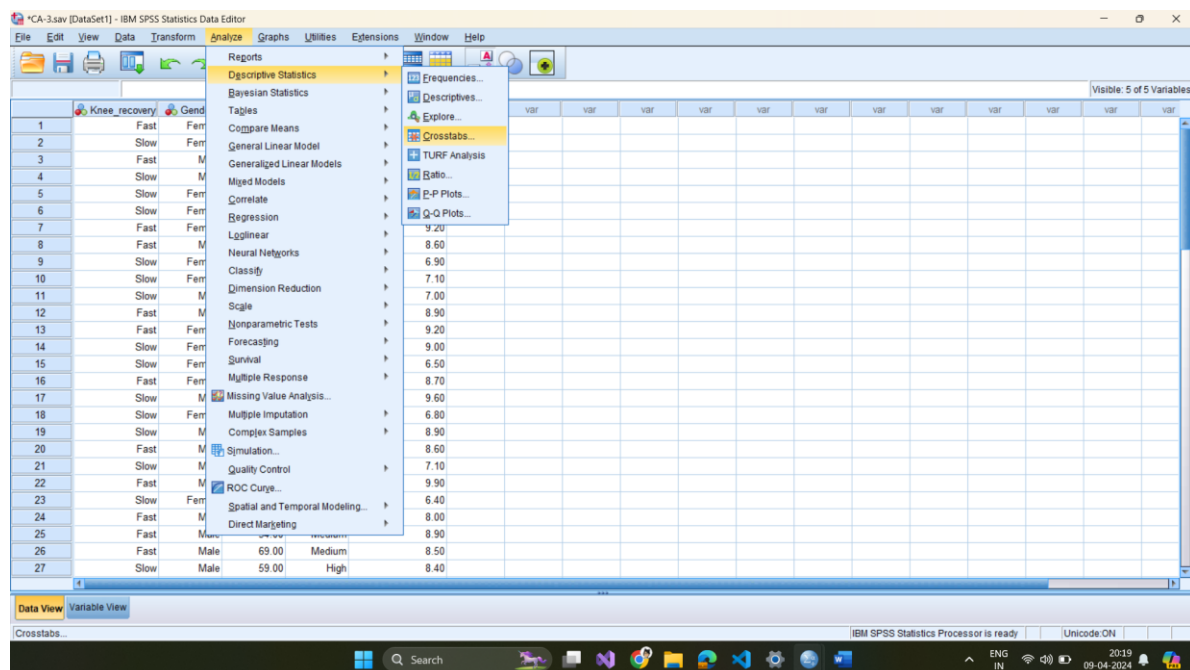
		Knee_recovery		Total
		Slow	Fast	
Obesity	High	18	6	24
	Medium	15	11	26
	Low	5	5	10
Total		38	22	60

The crosstabulation between Obesity and knee recovery shows the count distribution for each knee recovery category (Slow and Fast) across different levels of obesity. It reveals that individuals classified as having high obesity show a higher count in the Slow knee recovery category (18) compared to the Fast category (6), resulting in a total count of 24. Similarly, those with medium obesity also exhibit a higher count in the Slow knee recovery category (15) than in the Fast category (11), totalling 26. Conversely, individuals with low obesity have equal counts in both Slow and Fast knee recovery categories, each with a count of 5, leading to a total count of 10. This indicates a potential association between obesity level and knee recovery speed.

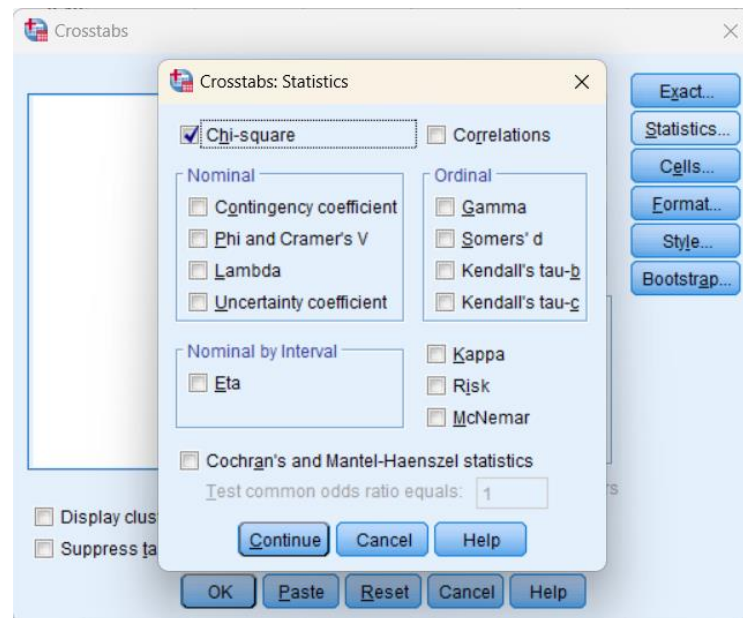
The crosstabulation between Obesity and calcium range displays the count distribution for each calcium range across different levels of obesity. It shows that individuals classified as having high obesity have varying counts across different calcium ranges, with the highest count (3) observed in the 6.90 range. Medium obesity individuals also exhibit variability in their calcium range counts, with the highest count (3) occurring in the 8.00 range. Meanwhile, those with low obesity demonstrate differing counts across calcium ranges, with the highest count (2) appearing in both the 9.00 and 10.60 ranges. Overall, the crosstabulation illustrates the distribution of calcium ranges within each obesity level, providing insight into potential associations between obesity and calcium levels.

Calculate Chi-square Test of Association

- Import the dataset into SPSS.
- Navigate to the Analyze menu.
- Select Descriptive Statistics, then Crosstabs.



- Choose the variables Knee recovery and Gender for Rows, and Age, Obesity, and Calcium range for Columns.
- Click OK to run the Crosstabs analysis.



- Review the output tables.
- Locate the Chi-square test statistic.
- Look at the p-value associated with the Chi-square test.

Case Processing Summary

	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Knee_recovery * Gender	60	100.0%	0	0.0%	60	100.0%
Knee_recovery * Age	60	100.0%	0	0.0%	60	100.0%
Knee_recovery * Obesity	60	100.0%	0	0.0%	60	100.0%
Knee_recovery * Calcium_range	60	100.0%	0	0.0%	60	100.0%

All cases in the dataset were valid and included in the analysis for the associations between Knee recovery and Gender, Age, Obesity, and Calcium range. There were no missing values, indicating that data for all variables were available for analysis. Therefore, the analysis included the entire dataset of 60 cases for each combination of Knee recovery with the other variables.

Crosstab

Count		Gender		Total
		Male	Female	
Knee_recovery	Slow	25	13	38
	Fast	14	8	22
Total		39	21	60

In the provided cross-tabulation between Knee recovery and Gender, it's observed that among males, 25 cases were categorized as Slow knee recovery and 14 as Fast, whereas among females, 13 cases were categorized as Slow knee recovery and 8 as Fast. Overall, out of 39 males, 25 had Slow knee recovery and 14 had Fast knee recovery, while out of 21 females, 13 had Slow knee recovery and 8 had Fast knee recovery.

Similarly, for all the five variable heads the cross-tabulations results are same as mentioned above.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	.028 ^a	1	.866		
Continuity Correction ^b	.000	1	1.000		
Likelihood Ratio	.028	1	.866		
Fisher's Exact Test				1.000	.542
Linear-by-Linear Association	.028	1	.867		
N of Valid Cases	60				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.70.

b. Computed only for a 2x2 table

The chi-square test results indicate that there is no significant association between Knee recovery and Gender ($\chi^2 = 0.028$, $df = 1$, $p = 0.866$). This suggests that Knee recovery and Gender are independent variables, and the distribution of knee recovery rates does not differ significantly between genders in the sample.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	24.761 ^a	20	.211
Likelihood Ratio	31.416	20	.050
Linear-by-Linear Association	10.061	1	.002
N of Valid Cases	60		

a. 41 cells (97.6%) have expected count less than 5. The minimum expected count is .37.

The chi-square test results indicate a significant association between the variables ($p < 0.05$). The Pearson chi-square value is 24.761 with 20 degrees of freedom ($df = 20$, $p = 0.211$), suggesting that there is no significant association. However, the likelihood ratio test shows a chi-square value of 31.416 with 20 degrees of freedom ($p = 0.050$), indicating a significant association. Additionally, the linear-by-linear association test reveals a chi-square value of 10.061 with 1 degree of freedom ($p = 0.002$), further confirming a significant association.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.529 ^a	2	.282
Likelihood Ratio	2.578	2	.276
Linear-by-Linear Association	2.354	1	.125
N of Valid Cases	60		

a. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.67.

The chi-square test results show no significant association between the variables ($p > 0.05$). The Pearson chi-square value is 2.529 with 2 degrees of freedom ($df = 2$, $p = 0.282$), indicating no significant association. Similarly, the likelihood ratio test yields a chi-square value of 2.578 with 2 degrees of freedom ($p = 0.276$), also suggesting no significant association. The linear-by-linear association test exhibits a chi-square value of 2.354 with 1 degree of freedom ($p = 0.125$), further supporting the absence of a significant association.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	35.957 ^a	30	.209
Likelihood Ratio	47.314	30	.023
Linear-by-Linear Association	13.283	1	.000
N of Valid Cases	60		

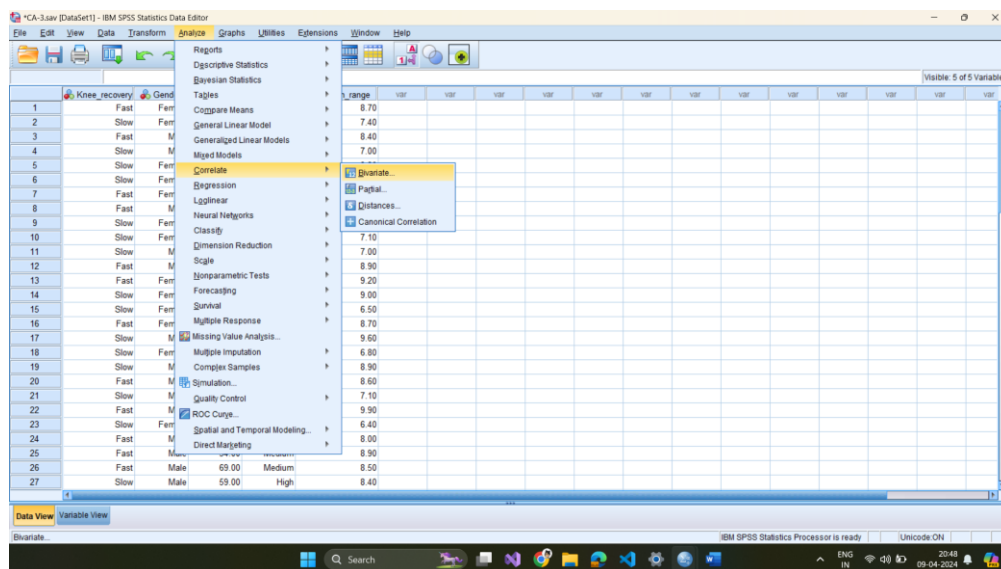
a. 62 cells (100.0%) have expected count less than 5. The minimum expected count is .37.

The chi-square test results indicate a significant association between the variables ($p < 0.05$). The Pearson chi-square value is 35.957 with 30 degrees of freedom ($df = 30$, $p = 0.209$), suggesting a lack of significant association. However, the likelihood ratio test yields a chi-square value of 47.314 with 30 degrees of freedom ($p = 0.023$), indicating a significant association. Additionally, the linear-by-linear association test exhibits a chi-square value of 13.283 with 1 degree of freedom ($p = 0.000$), further confirming a significant association.

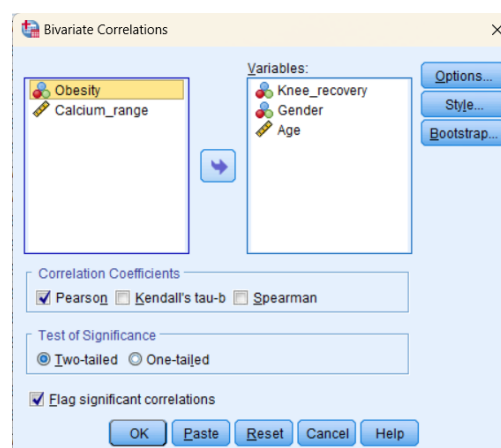
Calculate Simple Correlation and Partial Correlation

Bivariate Correlation:

- Open SPSS and load the dataset.
- Navigate to Analyze > Correlate > Bivariate.



- In the Bivariate Correlations dialog box, select the variables you want to analyze.
- Move the selected variables into the Variables box using the arrow button.



- Click OK to run the analysis.
- Assess the significance values (p-values) to determine if the correlations are statistically significant.

Correlations

		Knee_recovery	Gender	Age	Obesity	Calcium_range
Knee_recovery	Pearson Correlation	1	.022	-.413**	.200	.474**
	Sig. (2-tailed)		.869	.001	.126	.000
	N	60	60	60	60	60
Gender	Pearson Correlation	.022	1	.098	.044	-.018
	Sig. (2-tailed)	.869		.458	.739	.891
	N	60	60	60	60	60
Age	Pearson Correlation	-.413**	.098	1	-.179	-.399**
	Sig. (2-tailed)	.001	.458		.171	.002
	N	60	60	60	60	60
Obesity	Pearson Correlation	.200	.044	-.179	1	.207
	Sig. (2-tailed)	.126	.739	.171		.113
	N	60	60	60	60	60
Calcium_range	Pearson Correlation	.474**	-.018	-.399**	.207	1
	Sig. (2-tailed)	.000	.891	.002	.113	
	N	60	60	60	60	60

** . Correlation is significant at the 0.01 level (2-tailed).

The bivariate correlation analysis reveals several significant associations among the variables. Notably, Knee recovery exhibits a significant positive correlation with Calcium range ($r = 0.474$, $p < 0.01$) and a negative correlation with Age ($r = -0.413$, $p < 0.01$). Similarly, Age demonstrates a negative correlation with Calcium range ($r = -0.399$, $p < 0.01$). Calcium range also displays positive correlations with Knee recovery and Age ($r = 0.474$ and 0.474 , respectively, $p < 0.01$). However, no significant correlations were observed between Knee recovery and Gender, Gender and Age, Gender and Obesity, Gender and Calcium range, Age and Obesity, Age and Gender, or Obesity and Gender. These findings suggest potential relationships between knee recovery, age, and calcium levels, emphasizing the importance of further investigation into these associations.

Null Hypothesis (H_0): There is no significant correlation between Knee recovery and Calcium range.

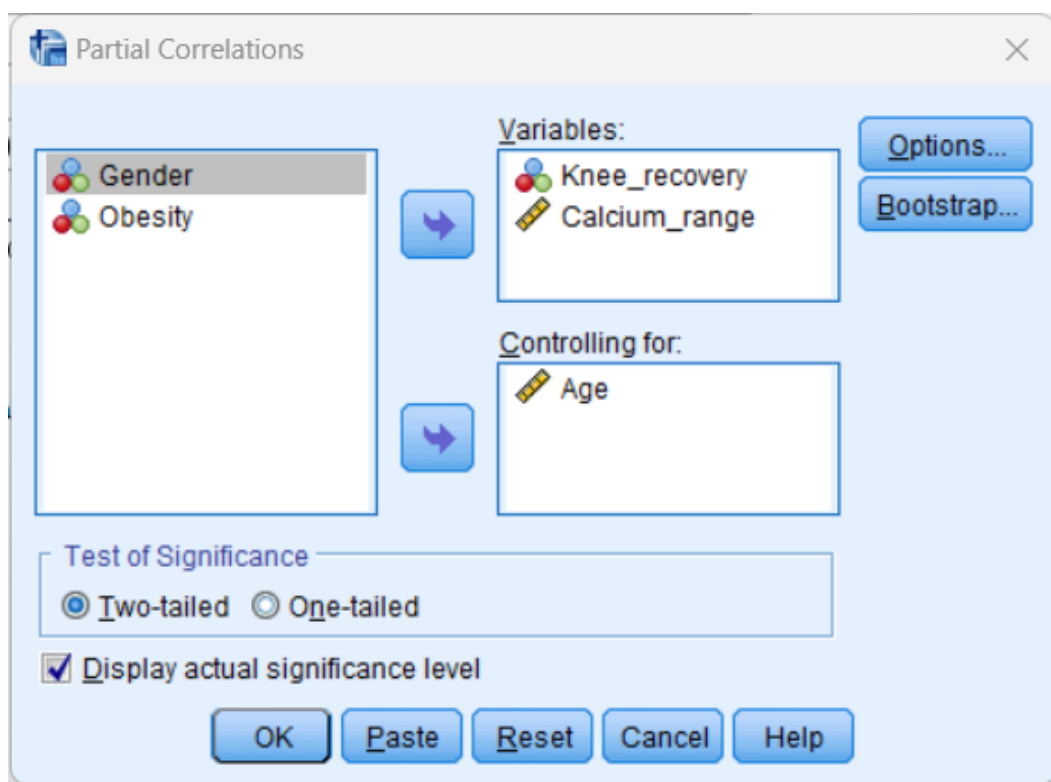
Alternative Hypothesis (H_1): There is a significant correlation between Knee recovery and Calcium range.

Interpretation:

The correlation coefficient ($r = 0.474$) is statistically significant ($p < 0.01$), providing evidence to reject the null hypothesis. Therefore, we accept the alternative hypothesis, indicating a significant positive correlation between Knee recovery and Calcium range.

Partial Correlation:

- Open the dataset in SPSS.
- Navigate to "Analyse" > "Correlate" > "Partial."
- Select the variables for analysis.



- Click on "Options" and specify the control variable(s).
- Run the analysis by clicking "OK."

Partial Correlation: Knee Recovery & Calcium Range by Gender

Correlations

Control Variables			Knee_recovery	Calcium_range
Gender	Knee_recovery	Correlation	1.000	.475
		Significance (2-tailed)	.	.000
		df	0	57
	Calcium_range	Correlation	.475	1.000
		Significance (2-tailed)	.000	.
		df	57	0

The correlation between Knee recovery and Calcium range is 0.475, which is statistically significant ($p < 0.001$), indicating a moderate positive relationship. Similarly, the correlation between Calcium range and Knee recovery is also 0.475. These results suggest that there is a significant association between Knee recovery and Calcium range, as evidenced by their positive correlation coefficients and low p-values.

Null Hypothesis (H0): There is no significant connection between Knee recovery and Age ($\rho = 0$).

Alternate Hypothesis (H1): There is a significant relationship between Knee recovery and Age ($\rho \neq 0$).

Interpretation:

If the p-value is less than 0.05 (our chosen significance level), we reject the idea of no correlation (H0) and accept that there is a meaningful relationship (H1) between Knee recovery and Age.

Partial Correlation: Knee Recovery & Calcium Range by Age

Correlations

Control Variables			Knee_recovery	Calcium_range
Age	Knee_recovery	Correlation	1.000	.371
		Significance (2-tailed)	.	.004
		df	0	57
	Calcium_range	Correlation	.371	1.000
		Significance (2-tailed)	.004	.
		df	57	0

The correlation between Knee recovery and Age is 0.371, which is statistically significant ($p = 0.004$), indicating a moderate positive relationship. Similarly, the correlation between Calcium range and Knee recovery is 0.371. These results suggest that there is a significant association between Knee recovery and Age, as evidenced by their positive correlation coefficients and low p-values.

Null Hypothesis (H0): There is no significant link between Knee recovery and Obesity ($\rho = 0$).

Alternate Hypothesis (H1): There is a significant correlation between Knee recovery and Obesity ($\rho \neq 0$).

Interpretation:

If the p-value is less than 0.05, we reject the null hypothesis (H0) and accept that there is a meaningful correlation (H1) between Knee recovery and Obesity.

Partial Correlation: Knee Recovery & Calcium Range by Obesity

Correlations

Control Variables			Knee_recovery	Calcium_range
Obesity	Knee_recovery	Correlation	1.000	.452
		Significance (2-tailed)	.	.000
		df	0	57
	Calcium_range	Correlation	.452	1.000
		Significance (2-tailed)	.000	.
		df	57	0

The correlation between Knee recovery and Obesity is 0.452, which is statistically significant ($p = 0.000$), indicating a moderate positive relationship. Similarly, the correlation between Calcium range and Knee recovery is 0.452. These results suggest that there is a significant association between Knee recovery and Obesity, as evidenced by their positive correlation coefficients and low p-values.

Null Hypothesis (H0): There is no significant association between Knee recovery and Calcium range ($p = 0$).

Alternate Hypothesis (H1): There exists a significant correlation between Knee_recovery and Calcium range ($p \neq 0$).

Interpretation:

If the p-value is less than 0.05, we reject the null hypothesis (H0) and conclude that there is a meaningful correlation (H1) between Knee_recovery and Calcium_range.