Section 1

T-Test for one sample

Introduction

Here we use t distribution as we need compare sample mean with unknown population mean and also, we have to evaluate the population standard deviation.

Step 1: Idea

One of the economists at University of Manitoba estimates that the average amount that students spend on monthly groceries is 120 dollars. However, you believe that floods in British Columbia and its impact will significantly increase the price of monthly groceries. You randomly select 30 students at the University of Manitoba and ask them about the monthly expenditure on groceries.

Students	Monthly Expenditure on groceries ($)
1	100
2	90
3	145
4	113
5	150
6	120
7	135
8	95
9	50
10	190
11	200
12	155
13	125
14	75
15	127

16	**116**
17	**86**
18	**128**
19	**156**
20	**178**
21	**148**
22	**180**
23	**117**
24	**95**
25	**131**
26	**97**
27	**106**
28	**178**
29	**109**
30	**97**

Step 2: Hypothesis and Alpha

Null Hypothesis → Ho: $\mu = 120$

Research (Alternative) Hypothesis → Ha: $\mu > 120$

Alpha (α) = 0.05

Step 3: SPSS Output

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
monthly_expenditure	30	126.4000	36.10005	6.59094

Table 1

One-Sample Test

	Test Value = 120					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
monthly_expenditure	.971	29	.340	6.40000	-7.0800	19.8800

Table 2

Step 4: Explanation of SPSS Output

Table 1: One-Sample Statistics show the following characteristics:

- Column 1 identifies the variable name which is monthly_expenditure.
- Column 2 shows total number of participants (N) in our case total number of students was 30.
- The Mean of the given sample is shown in column 3 which is 126.4 dollars.
- The Std. Deviation of the sample is displayed in column 4; in our case our standard deviation is 36.10005.
- Column 5, Std. Error Mean, is the Std. Deviation divided by the square root of total number of participants (N). In this case, the value is 6.59094.

Table 2: One-Sample Test exhibits the following attributes:

- Column 1 identifies the variable name which is monthly_expenditure.
- Column 2, t, is the t- score which can be obtained by $t = (M-\mu)/S$ and the t score of the above sample is 0.971.
- Column 3, df, is the degree of freedom which can be obtained by N-1 formula. In this case df is 29.
- Column 4, Sig. (2-tailed), is the 2-tail probability (p value) which is 0.340. We need to divide this p value by half because SPSS by default gives 2-tailed probability. So, the result after dividing by half is 0.17.
- Column 5, Mean Difference, is the difference between sample mean and population mean which is (126.40 – 120 = 6.40000).

- Column 6, 95% Confidence Interval of the Difference, which means SPSS generate 95% of the confidence interval around the Mean Difference. Here, the upper bound of the sample is 19.8800 whereas the lower bound is -7.0800.

Step 5: Conclusion

The p value which we obtained in column 4, table 2 for one tail is 0.17 which is greater than our alpha value i.e., 0.05. Thus, we fail to retain research hypothesis (Ha), and thus we retained the null hypothesis (Ho). Hence, the floods in British Columbia and its impact does not the increase the prices of the groceries.

Section 2

T-Test for dependent means

Introduction

When you don't know the population mean as well as standard deviation then we use dependent mean t-test. In this case, we usually have two samples one before experiment and one after.

Step 1: Idea

During physical education class, the instructor asked the class of 30 students to throw a javelin as far as they could, and the instructor recorded the distance of the javelin throw for each student. However, for the second trail, the teacher asked students to try to throw a

javelin at a 35-degree angle, and then he recorded the distance of the javelin throw. The results are depicted below:

Students	Trail 1 Distance (m)	Trail 2 Distance (m)
1	40.0	45.7
2	50.3	55.9
3	34.5	45.9
4	20.3	30.0
5	45.8	55.7
6	34.6	43.9
7	45.0	55.0
8	50.7	59.2
9	60.8	67.5
10	43.4	54.9
11	60.9	75.9
12	63.6	77.0
13	70.1	80.9
14	55.6	65.9
15	34.4	47.8
16	70.8	83.9
17	44.7	57.9
18	57.5	70.8
19	49.9	52.0
20	72.3	87.9
21	45.7	58.9
22	56.8	65.8
23	60.3	73.5
24	57.8	69.8

25	**67.4**	**79.8**
26	**64.9**	**76.0**
27	**46.4**	**57.8**
28	**49.0**	**59.0**
29	**50.4**	**62.9**
30	**59.0**	**72.0**

Step 2: Hypothesis and Alpha

Research (Alternative) Hypothesis (μ1) is that throwing javelin with 35-degree angle will cover more distance than throwing at different angle (μ2).

Null Hypothesis (H0): μ1 = μ2.

Research (Alternative) Hypothesis (H1): μ1 > μ2.

Alpha (α) = 0.05

Step 3: SPSS Output

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Trail_1	52.0967	30	12.28872	2.24360
	Trail_2	62.9733	30	13.46369	2.45812

Table 1

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Trail_1 & Trail_2	30	.978	.000

Table 2

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Trail_1 - Trail_2	-10.87667	2.96970	.54219	-11.98557	-9.76776	-20.061	29	.000

Table 3

Step 4: Explanation of SPSS Output

Table 1 shows the following characteristics:

Column 1 shows the name of the variable which is Trail_1 and Trail_2. Mean of each trial is shown in column 2 which is 52.0967 and 62.9733 for Trail_1 and Trail_2 respectively. Column 3 shows the total number of participants which is 30 in our case and standard deviation for both the trails are shown in column 4.

Table 2 shows the following characteristics:

Column 1 illustrate the name of the variable which is Trail_1 and Trail_2.
Column 2 depicts total number of athletes whereas column 3 shows the correlation between the variables which is 0.978.

Table 3 depicts the following attributes:

In table 3 we are most concerned with t scores, degree of freedom and Sig. (2-tailed).

As in column 6, the t score is -20.061 which can be calculated by $t = (M-\mu)/S$ and column 7 illustrate the df (N-1) which is 29. Column 8, Sig. (2-tailed), is the 2-tail probability (p value) which is 0.000. We need to divide this p value by half because SPSS by default gives 2-tailed probability. So, the result after dividing by half is 0.000.

Step 5: Conclusion

The p value which we obtained in column 7, table 3 for one tail is 0.00 which is smaller than our alpha value i.e., 0.05. Thus, we reject the null hypothesis and support research hypothesis. Hence, throwing javelin at 35-degree angle will cover more distance than throwing at some different angle.

Section 3

T-Test for independent means

Introduction

When you don't know the population mean as well as standard deviation then we use independent mean t-test. Moreover, you have two samples but these sample independently related to each other.

Step 1: Idea

The Muscle Pharm Company has launched a new whey protein with some additional nutritional components. To test the effect of this product, the company selected 48 athletes. Testers gave 24 participants new whey protein, while the other 24 participants used Muscle Pharm ordinary protein powder an hour before the workout. Then each athlete was asked to run 1000 meters and the timing was recorded by the researchers. The results are shown below:

Group 1 (Ordinary powder)	Timing (minute's)	Group 2 (New powder)	Timing (minute's)
1.	7.2	25.	3.1
2.	5.5	26.	3.14
3.	8.9	27.	5.4
4.	7.0	28.	4.4
5.	10	29.	5.7
6.	6.5	30.	3.9
7.	6.9	31.	3.0
8.	9.0	32.	4.9
9.	5.5	33.	5.8
10.	7.3	34.	6.4
11.	5.3	35.	2.3
12.	6.4	36.	3.9
13.	8.9	37.	4.5
14.	4.5	38.	5.9
15.	6.0	39.	4.8
16.	5.4	40.	3.9
17.	7.8	41.	2.7
18.	8.4	42.	3.5
19.	6.9	43.	4.9
20.	7.5	44.	2.70
21.	8.9	45.	3.7
22.	10.9	46.	5.1
23.	7.6	47.	4.7
24.	8.9	48.	2.9

Step 2: Hypothesis and Alpha

The research (alternative) hypothesis (μ1) is that the new protein powder will make athletes stronger and allow them to cover distance in less time than athletes who use ordinary protein powder (μ2)

Null hypothesis: H0: μ1 = μ2

Research hypothesis: H1: μ1 ≠ μ2

Alpha (α) = 0.05

Step 3: SPSS Output

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Timing	1	24	7.3833	1.61909	.33050
	2.00	24	4.2183	1.16256	.23731

Table 1

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Timing	Equal variances assumed	1.920	.173	7.779	46	.000	3.16500	.40687	2.34602	3.98398
	Equal variances not assumed			7.779	41.736	.000	3.16500	.40687	2.34375	3.98625

Table 2

Step 4: Explanation of SPSS Output

Table 1 shows the following characteristics:

Timing is the dependent variable shown in column 1. Column 2 depicts the total of participants in each group that is 24 each so total is 48. Mean of each group 1 and group 2 is shown in column 3 which is 1.61909 and 1.16256 respectively. Furthermore, Standard deviation and Std. Error mean is shown in column 4 and 5 sequentially.

Table 2 shows the following characteristics:

As we can see in table 2 timing is dependent variable two independent t-test were run on our data one is “Equal variances assumed” and other is “Equal variances not assumed”. Here we will use “Equal variance assumed” as we can see in Levene’s Test for Equality of Variances our Sig. value is 0.173 which is greater than 0.05. Moreover, “t” and “df” shows the t score and degree of freedom which is 7.779 and 46 respectively. Sig. (2-tailed) column is the 2-tail probability (p value) which is 0.000.

Step 5: Conclusion

The Sig. (2-tailed) which we obtained is 0.00 which is smaller than our alpha value i.e. 0.05. Thus, we reject the null hypothesis and support research hypothesis. Hence, new protein powder will make athletes stronger and allow them to cover distance in less time than athletes who use ordinary protein powder.

Section 4

One- Way ANOVA

Introduction

When researcher wants to determine the mean difference between three or more samples then he can use ANOVA (analysis of variance).

Step 1: Idea

In a science exhibition, there are four best projects: the water cycle, the human body, genes, and solar power. The head of department decided to call 40 students who are doing masters and randomly assigned 10 participants to each project and asked them to give the rating of these projects based on idea, design, and presentation. The student needs to give a rating based on a scale of 0 to 10, where 0 is incompetent and 10 is fantastic.

Water cyclist	Cycle Scores	Human Bodies	Bodies Scores	Genes model	Genes scores	Solar model	Solar scores
1.	7	1.	10	1.	5	1.	10
2.	6	2.	9	2.	6	2.	7
3.	8	3.	6	3.	3	3.	6
4.	9	4.	8	4.	2	4.	8
5.	6	5.	5	5.	7	5.	9
6.	5	6.	9	6.	4	6.	7
7.	7	7.	10	7.	6	7.	5
8.	9	8.	8	8.	5	8.	4
9.	8	9.	9	9.	8	9.	6
10.	7	10.	7	10.	5	10.	8

Step 2: Hypothesis and Alpha

Here we considered: Cycle Scores = μ1, Bodies Scores = μ2, Genes Scores = μ3, Solar Scores = μ4.

Research hypothesis: $\mu 1 \neq \mu 2 \neq \mu 3 \neq \mu 4$.

Null hypothesis: $\mu 1 = \mu 2 = \mu 3 = \mu 4$.

Alpha (α) = 0.05

Step 3: SPSS Output

Descriptives

Scores

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Cycle Scores	10	7.20	1.317	.416	6.26	8.14	5	9
Bodies Scores	10	8.10	1.663	.526	6.91	9.29	5	10
Genes Scores	10	5.10	1.792	.567	3.82	6.38	2	8
Solar Scores	10	7.00	1.826	.577	5.69	8.31	4	10
Total	40	6.85	1.942	.307	6.23	7.47	2	10

Table 1

ANOVA

Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	47.700	3	15.900	5.759	.003
Within Groups	99.400	36	2.761		
Total	147.100	39			

Table 2

Multiple Comparisons

Dependent Variable: Scores

	(I) Model	(J) Model	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Cycle Scores	Bodies Scores	-.900	.743	.624	-2.90	1.10
		Genes Scores	2.100*	.743	.037	.10	4.10
		Solar Scores	.200	.743	.993	-1.80	2.20
	Bodies Scores	Cycle Scores	.900	.743	.624	-1.10	2.90
		Genes Scores	3.000*	.743	.001	1.00	5.00
		Solar Scores	1.100	.743	.460	-.90	3.10
	Genes Scores	Cycle Scores	-2.100*	.743	.037	-4.10	-.10
		Bodies Scores	-3.000*	.743	.001	-5.00	-1.00
		Solar Scores	-1.900	.743	.068	-3.90	.10
	Solar Scores	Cycle Scores	-.200	.743	.993	-2.20	1.80
		Bodies Scores	-1.100	.743	.460	-3.10	.90
		Genes Scores	1.900	.743	.068	-.10	3.90
Scheffe	Cycle Scores	Bodies Scores	-.900	.743	.692	-3.08	1.28
		Genes Scores	2.100	.743	.063	-.08	4.28
		Solar Scores	.200	.743	.995	-1.98	2.38
	Bodies Scores	Cycle Scores	.900	.743	.692	-1.28	3.08
		Genes Scores	3.000*	.743	.003	.82	5.18
		Solar Scores	1.100	.743	.541	-1.08	3.28
	Genes Scores	Cycle Scores	-2.100	.743	.063	-4.28	.08
		Bodies Scores	-3.000*	.743	.003	-5.18	-.82
		Solar Scores	-1.900	.743	.107	-4.08	.28
	Solar Scores	Cycle Scores	-.200	.743	.995	-2.38	1.98
		Bodies Scores	-1.100	.743	.541	-3.28	1.08
		Genes Scores	1.900	.743	.107	-.28	4.08

*. The mean difference is significant at the 0.05 level.

Table 3

Scores

	Model	N	Subset for alpha = 0.05	
			1	2
Tukey HSD[a]	Genes Scores	10	5.10	
	Solar Scores	10	7.00	7.00
	Cycle Scores	10		7.20
	Bodies Scores	10		8.10
	Sig.		.068	.460
Scheffe[a]	Genes Scores	10	5.10	
	Solar Scores	10	7.00	7.00
	Cycle Scores	10	7.20	7.20
	Bodies Scores	10		8.10
	Sig.		.063	.541

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.000.

Table 4

Step 4: Explanation of SPSS Output

Table 1, “Descriptive” shows the following characteristics:

Column 1 and 2 shows all the scores that we have in our research and total number of participants. Columns 3 and 4 give the mean and standard deviation of all each score. Column 5 and 6 shows the Std. Error and confidence interval (upper and lower bound) whereas column 7 and 8 shows the minimum and maximum.

Table 2, ANOVA shows the following attributes:

Column 1 divides the variability into three components: “Between Groups”, “Within Groups and “Total”. Column 2 represents “Sum of Squares”. Here the “Sum of Squares” for between groups is 47.700 and for within groups is 99.400 and total is 147.100. Column 3 represent the “df” degree of freedom for between groups and within groups and gives total degree of freedom. The df for between group is 3 whereas df for within group is 36 which can be obtained by 9+9+9+9 = 36. However, column 5 represent

Mean Square which are 15.900 and 2.761 for between groups and within groups respectively. Moving forward, column 5 shows the F value, which is 5.759 in our case, this value is the ratio of between-groups and within-groups variability. Column 6 represents the Sig. value or p value which is 0.003 which is less than our alpha level that is 0.005.

Table 3, “Multiple Comparisons” shows the following characteristics:

Column 1 (I) Model (J) Model shows the comparison with other model scores. For instance, Cycle Scores are compared with all other three models scores. Similarly, all other model scores are compared with each other. However, column 2 represent the Mean Difference (I-J) which can be calculated by subtracting the model Under (J) from Model under (I). Moving forward, column 4, Sig. shows the significance level. Values under 0.05 are considered significant. Moving forward to table 4,” Homogenous Subsets” assist to distinguish between the similarities and differences the groups possess.

Step 5: Conclusion

The p value which we obtained is 0.003 which is smaller than our alpha value i.e., 0.05.

Thus, we reject the null hypothesis and support research hypothesis.

Section 5

2 x 2 Factorial Design

Introduction

In factorial ANOVA there can two or three independent variables and these variables can have their two or more levels. Factorial ANOVA can be used when the complexity of the samples increased.

Step 1: Idea

Charles Pfizer has finally discovered the COVID-19 vaccine. To give approval to this vaccine, he did an experiment on 40 participants. He assigned two levels of dose to participants: Moderate dose and High dose. Moreover, he decided to give these to COVID-19 patients and non-COVID-19 patients. After that, he took each participant's blood test and graded them on a scale of 0 to 10. Where 0 indicates ineffectiveness and 10 indicates maximum effectiveness.

		DOSE Moderate dose	High dose
Patients	Covid-19 Patients	4,5,3,7,2,5,6,1,4,1 Mean =3.8	10,9,8,7,10,7,8 9,6,8 Mean = 8.2
	Non Covid-19 Patients	6,8,7,4,3,4,5,2,7,4 Mean = 5	10,8,9,6,8,10,9,7,5,3 Mean = 7.5

Step 2: Hypothesis and Alpha

1. The null hypothesis for patients states that no interaction will take place whereas research hypothesis states that there will be interaction

2. The null hypothesis for drug main effect states that there will be no difference between moderate dose and high dose whereas research hypothesis states that there will be difference.
3. Alpha (α) = 0.05

Step 3: SPSS Output

Between-Subjects Factors

		Value Label	N
Patients	1.00	Covid-19 Patients	20
	2.00	Non Covid-19 Patients	20
Dose	1.00	Moderate Dose	20
	2.00	High Dose	20

Table 1

Tests of Between-Subjects Effects

Dependent Variable: Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	128.675[a]	3	42.892	11.549	.000
Intercept	1500.625	1	1500.625	404.058	.000
Patients	.625	1	.625	.168	.684
Dose	119.025	1	119.025	32.049	.000
Patients * Dose	9.025	1	9.025	2.430	.128
Error	133.700	36	3.714		
Total	1763.000	40			
Corrected Total	262.375	39			

a. R Squared = .490 (Adjusted R Squared = .448)

Table 2

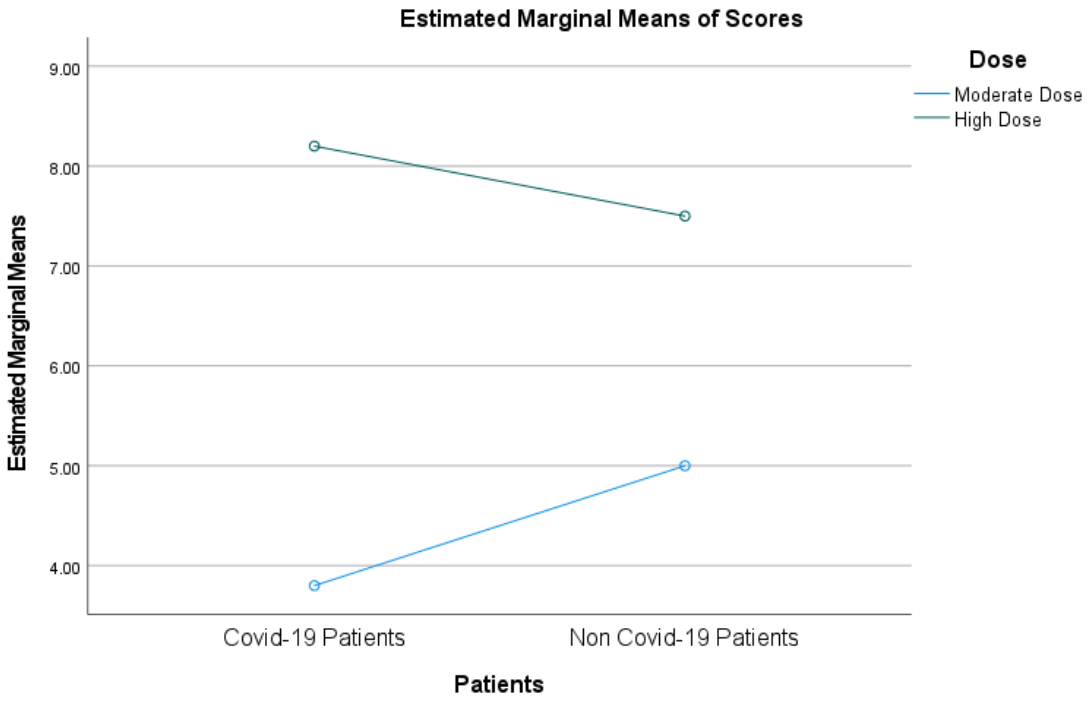

Table 3

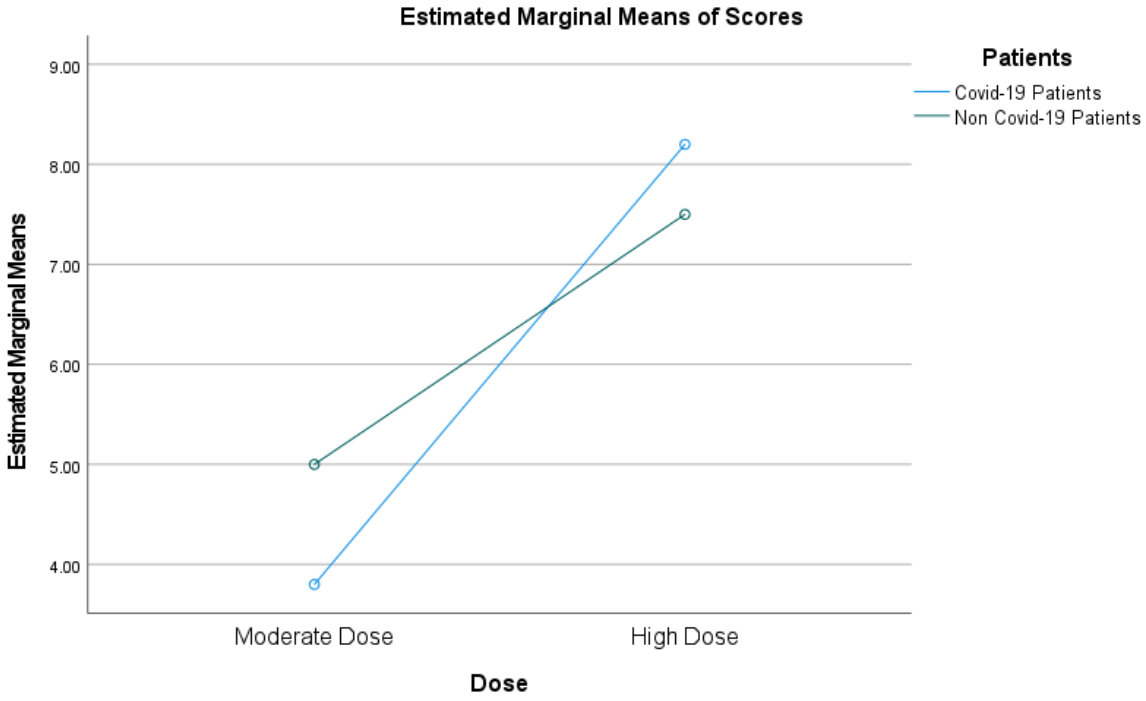

Table 4

Step 4 SPSS Explanation

Interaction

As we can see the graphs are depicted in table 3 and 4 which shows that there is an interaction between the independent variable because in graphs shown above the lines are not parallel. Here in table 3 independent variable, patients are on x axis with each level represented on horizontal axis whereas in table 4 another independent variable, dose is x axis along with its levels.

Table 1 Between-subjects factors gives the general information about the research data. Here N = 20 means total number of patients for each independent variable.

Table 2 Tests of Between-Subjects Effects Dependent Variable: Altered show the following characteristics:

Column 1 represent Source, here we are going to focus mainly on Patients, Dose and Patients*Dose.

Firstly, we have look at interaction, here Patients*Dose row depicts the interaction effect. The values associated with this row are “Type III Sum of Squares” which is 9.025, “df” = 1, “Mean Square” = 9.025, “F” = 2.430 and “Sig.” = 0.128. Here sig value is 0.128 which is greater than 0.05 thus, we retain the null hypothesis associated with interaction.

Now its turn to consider the main effects. Firstly, we will consider the main effect of the dose. In table 2, Dose row associated with “Type III Sum of Squares” which is 119.025, “df” = 1, “Mean Square” = 119.025, “F” = 32.049 and “Sig.” = 0.000. Since, p value is less than alpha value 0.005 thus we reject the null hypothesis.

Now, we will consider the main effects of patients. In table 2, Patient’s row associated with “Type III Sum of Squares” which is 0.625, “df” = 1, “Mean Square” = 0.625, “F” = 0.168 and “Sig.” = 0.684. Since, p value is greater than alpha value 0.005 thus we retain the null hypothesis.

Step 5 Conclusion

Here we retain the null hypothesis which states there is no interaction among the patients because p value 0.625 is greater than alpha value.

Here we reject the null hypothesis which states that there no interaction between the doses because p value 0.000 is less than alpha value.

Section 6

Chi-square test for goodness of fit

Introduction

It is non-parametric test that is performed on categorial data. Chi-Square test for goodness of fit mainly focus on population proportions.

Step 1 Idea

In 2008, John distributes candies on his birthday in party of 30 members. His father arranged the candies for him, and he helped John to distribute the candies. However, in 2009 John again distributes the candies to his same 30 buddies during his birthday party. His father, on the other hand, remembered the number candies each children took last time. However, he again observed the number of candies each child took in a party. Following are the results according to children's preferences.

Gender (Boys)	No. of Candies	Gender (Girls)	No. of Candies
1	5	1	3
2	6	2	4
3	3	3	5
4	1	4	6

5	**3**	**5**	**5**
6	**7**	**6**	**8**
7	**10**	**7**	**3**
8	**4**	**8**	**4**
9	**6**	**9**	**6**
10	**4**	**10**	**8**
11	**7**	**11**	**6**
12	**8**	**12**	**3**
13	**3**	**13**	**4**
14	**2**	**14**	**6**
15	**7**	**15**	**4**

Step 2 Hypothesis and Alpha level

Research hypothesis (H1): there is a difference between observed and expected values.

Null Hypothesis (H0): there is no difference between observed and expected values

Alpha (α) = 0.05

Step 3 SPSS Output

No_of_Candies

	Observed N	Expected N	Residual
1.00	1	3.3	-2.3
2.00	1	3.3	-2.3
3.00	6	3.3	2.7
4.00	6	3.3	2.7
5.00	3	3.3	-.3
6.00	6	3.3	2.7
7.00	3	3.3	-.3
8.00	3	3.3	-.3
10.00	1	3.3	-2.3
Total	30		

Table 1

Test Statistics

	No_of_Candies
Chi-Square	11.400[a]
df	8
Asymp. Sig.	.180

a. 9 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 3.3.

Table 2

Step 4 SPSS Explanation

Table 1 show the following characteristics:

Column 1 shows number of ranges of the candies taken by children, here minimum number is 1 and maximum 10. Column 2 depicted observed frequency for instance,6 different children took 6 candies each. Column 3 represent the expected result whereas column 4 is residual which is calculated by observed n – expected n.

Table 2, “Test statistics gives you the value of Chi-Square, which is 11.400, df which is 8 and Sig. value which 0.180

Step 5 Conclusion

The p value is 0.180 which is greater than alpha level 0.005. Thus, we retain the null hypothesis which states there is no difference between the expected and observed value.

Section 7

Chi-square test for independence

Introduction

It is non-parametric test that is performed on categorial data. Chi-Square test for independence shows the relationship between two variables.

Step 1: Idea

The researcher wants to determine the relationship between the gender (boys or girls) and type of games (outdoor games or indoor games). During his class, he asked a total of 30 students (15 boys and 15 girls) about their preferred games, with 1 indicating outdoor games and 2 indicating indoor games.

Gender (Boys)	Games (Outdoor or indoor)	Gender (Girls)	Games (Outdoor or indoor)
1	1	1	2
2	2	2	1
3	1	3	1
4	1	4	2
5	1	5	2
6	1	6	2
7	2	7	2
8	1	8	1
9	1	9	1
10	2	10	1
11	2	11	1
12	1	12	2
13	1	13	2

14	**1**	**14**	**2**
15	**1**	**15**	**1**

Step 2: Hypothesis and Alpha level

Research hypothesis (H1): gender and type of games are related.

Null Hypothesis (H0): gender and type of game are not related.

Alpha (α) = 0.05

Step 3: SPSS Output

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Game_perf * Gender	30	100.0%	0	0.0%	30	100.0%

Table 1

Game_perf * Gender Crosstabulation

Count

		Gender		
		Boys	Girls	Total
Game_perf	Outdoor games	11	7	18
	Indoor Games	4	8	12
Total		15	15	30

Table 2

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.222[a]	1	.136		
Continuity Correction[b]	1.250	1	.264		
Likelihood Ratio	2.256	1	.133		
Fisher's Exact Test				.264	.132
Linear-by-Linear Association	2.148	1	.143		
N of Valid Cases	30				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.00.

b. Computed only for a 2x2 table

Table 3

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.272	.136
	Cramer's V	.272	.136
N of Valid Cases		30	

Table 4

Step 4: SPSS Explanation

Here we are most likely to concerned with table 3, “Chi-Square Tests” as it will assist to reject or retain the null hypothesis. Looking at row 1 and column 1 which gives the “Pearson Chi-Square” and value for this is 2.222. Moving forward in same row gives you the df which is 1. Column 4, “Asymptotic Significance (2-sided)” gives the p value which is 0.136.

Step 5: Conclusion

The p value which we got from the analysis is 0.136 which is greater than alpha level 0.005. Thus, we support null hypothesis. Hence, gender and type of game are not related.

Section 8

Mann-Whitney U

Introduction

Mann-Whitney U test is used when we want to compare two independent groups which possess ordinal (ranked) data.

Step 1: Idea

You're now pursuing a bachelor's degree in computer science at the University of Manitoba, and you're undecided about whether to pursue a career in networking or development. So, you enlist the help of 30 experts (15 developers and 15 networkers) and ask them about their annual packages. Below is the data depicted.

Developers	Salaries ($)	networker	Salaries ($)
1	20,000	1	78,000
2	30,000	2	45,000
3	100,000	3	89,000
4	250,000	4	78,000
5	70,000	5	56,000
6	65,000	6	76,000
7	90,000	7	45,000
8	80,000	8	67,000
9	45,000	9	89,000
10	78,000	10	90,000
11	76,000	11	100,000
12	56,000	12	150,000
13	90,000	13	56,000
14	110,000	14	45,000
15	98,000	15	90,000

Step 2: Hypothesis and Alpha level

Research hypothesis (H1): there is a difference in their salaries.

Null Hypothesis (H0): there is no difference in their salaries.

Alpha (α) = 0.05.

Step 3: SPSS Output

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Salaries	30	80400.0000	41375.98837	20000.00	250000.00
professionals	30	1.5000	.50855	1.00	2.00

Table 1

Ranks

	professionals	N	Mean Rank	Sum of Ranks
Salaries	Developers	15	15.97	239.50
	Networkers	15	15.03	225.50
	Total	30		

Table 2

Test Statistics[a]

	Salaries
Mann-Whitney U	105.500
Wilcoxon W	225.500
Z	-.291
Asymp. Sig. (2-tailed)	.771
Exact Sig. [2*(1-tailed Sig.)]	.775[b]

a. Grouping Variable: professionals

b. Not corrected for ties.

Table 3

Step 4: Explanation of SPSS Output

Table 1, Descriptive statistics shows the summary of the data: total number of participants, mean of given sample, standard deviation, minimum and maximum

Here we are mostly concerned with table 3 as it will help to reject or retain the null hypothesis. Looking at row 3 and 4 which gives you the z and p value which is -0.291 and 0.771 respectively.

Step 5: Conclusion

The p value which we obtained from the analysis is 0.771 which is greater than alpha level 0.005. Thus, we support null hypothesis. Hence, there is no difference in their salaries of networkers and developers.

References:

Class Notes: Unit 6, 7 and 8

Aron, A., Coups E. J., Aron, E. (2019). Statistics for the Behavioral and Social Sciences: A Brief Course (6th ed.). Pearson