Module 4 Task

Name: Obubelebara George

PART 1.

1. ** *Public Perception Dataset*

The Public Perception dataset includes data from 1,034 telephone interviews among residents of Orange County, Florida. The survey instrument has seven sections encompassing 96 separate items. The first section asks the importance of various issues to residents, from controlling government spending to building a light rail transportation system. The rationale for asking these questions is that their ranking may inform county officials in their strategic planning and allocation decisions.

Among the first section questions, students will use three variables that ask the importance of "helping public schools," "fighting illegal drug use," and "dealing with problems of gangs."

How important are the following issues for you? Please state whether you consider each issue Very Important (=3), Important (=2), Somewhat Important (=1), or Unimportant (=0).

helping public schools (Pubschl)	[ ]	[ ]	[ ]	[ ]
fighting illegal drug use (Figtdrug)	[ ]	[ ]	[ ]	[ ]
dealing with the problems of gangs (Gangs)	[ ]	[ ]	[ ]	[ ]

* Race (ETHNIC)

Do you describe yourself as White (=1), Black/African-American (=2), Hispanic (=3), Asian/Pacific Islander (=4), Native American (=5), or some other ethnic group (=6)?

Use the Public Perceptions data set. The three most important issues were helping public schools (Pubschl), fighting illegal drug use (Figtdrug), and dealing with the problems of gangs (Gangs). Do whites and nonwhites agree on the importance of these priorities? On which issue is there a difference? Discuss the practical importance of any significant differences.

(1) Pubschl – race (Whites vs. Non-Whites)

- What is the Null hypothesis (H_0) and Alternative hypothesis (H_a)
- Create the contingency table
- Create a table with expected frequencies
- Calculate the -chi-square values for each cell

- Report the hypothesis test results using chi-square analysis

Helping Public Schools Hypotheses

H0: There is no significant difference in opinion between Whites and non-Whites regarding the importance of helping public schools.
H1: There is a significant difference in opinion between Whites and non-Whites regarding the importance of helping public schools.

Observed

Helping Public Schools	White (=1)	Non-White (=2)	Grand Total
Very Important (=3)	486	177	663
Important (=2)	199	56	255
Somewhat Important (=1)	43	7	50
Unimportant (=0)	22	3	25
Grand Total	750	243	993

Expected Frequency

Helping Public Schools	White (=1)	Non-White (=2)	Grand Total
Very Important (=3)	500.755	162.245	663
Important (=2)	192.598	62.402	255
Somewhat Important (=1)	37.764	12.236	50
Unimportant (=0)	18.882	6.118	25
Grand Total	750	243	993

Chi-square value

Helping Public Schools	White (=1)	Non-White (=2)	Grand Total
Very Important (=3)	0.435	1.342	1.777
Important (=2)	0.213	0.657	0.870
Somewhat Important (=1)	0.726	2.240	2.966
Unimportant (=0)	0.515	1.589	2.104
Grand Total	**1.888**	**5.828**	**7.716**

Critical value=	7.8147	
Chi-square=	7.716	**= CHISQ.INV.RT(p,df)**

Interpretation:

degree of freedom (df)=	3	**= (row-1)*(column-1) = (4-1)*(2-1)**
p=	0.05225572	**= CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Because the p <0.05, we fail to reject the null hypothesis.

The calculated chi-square value (7.716) is less than the critical value (7.8147) and the p-value (0.05225572) is slightly greater than the chosen significance level (0.05), we fail to reject the null hypothesis.

Therefore, we do not have sufficient evidence to conclude that there is a significant difference between the perceptions of whites and non-whites regarding the importance of helping public schools. It is also important to note that the p-value is very close to the significance level, suggesting a borderline significance. Therefore, further analysis may be warranted to explore potential differences more thoroughly.

(2) Figtdrug – race (Whites vs. Non-Whites)

- What is the Null hypothesis (H_0) and Alternative hypothesis (H_a)
- Create the contingency table
- Create a table with expected frequencies
- Calculate the -chi-square values for each cell
- Report the hypothesis test results using chi-square analysis

Fighting Drugs Hypotheses:

H_0: There is no significant difference in opinion between Whites and non-Whites regarding the importance of fighting drugs.
H_a: There is a significant difference in opinion between Whites and non-Whites regarding the importance of fighting drugs.

Observed

Fighting Drugs	White (=1)	Non-White (=2)	**Grand Total**
Very Important (=3)	522	175	697
Important (=2)	141	52	193
Somewhat Important (=1)	51	7	58
Unimportant (=0)	39	11	50
Grand Total	**753**	**245**	**998**

Expected Frequency

Fighting Drugs	White (=1)	Non-White (=2)	**Grand Total**
Very Important (=3)	525.893	171.107	697
Important (=2)	145.620	47.380	193
Somewhat Important (=1)	43.762	14.238	58
Unimportant (=0)	37.725	12.275	50
Grand Total	**753**	**245**	**998**

Chi-square value

Fighting Drugs	White (=1)	Non-White (=2)	**Grand Total**
Very Important (=3)	0.029	0.089	0.117
Important (=2)	0.147	0.451	0.597
Somewhat Important (=1)	1.197	3.680	4.877
Unimportant (=0)	0.043	0.132	0.175
Grand Total	**1.416**	**4.351**	**5.767**

Critical value	7.8147	

Interpretation:

Chi-square=	5.767	**=CHISQ.INV.RT(p,df)**
degree of freedom (df)=	3	**=(row-1)*(column-1) = (4-1)*(2-1)**
p=	0.123509304	**=CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Since the calculated chi-square value (5.767) is less than the critical value (7.8147) and the p-value (0.123509304) is greater than the significance level (0.05), we fail to reject the null hypothesis. This suggests that there is no significant difference in opinion between Whites and non-Whites regarding the importance of fighting drugs. Based on the analysis, we do not have enough evidence to conclude that there is a significant difference in opinion between Whites and non-Whites regarding the importance of fighting drugs.

(3) Gangs – race (Whites vs. Non-Whites)

- What is Null hypothesis (H_0) and Alternative hypothesis (H_a)
- Create the contingency table
- Create a table with expected frequencies
- Calculate the -chi-square values for each cell
- Report the hypothesis test results using chi-square analysis

Dealing with Gangs Hypotheses:

H_0: There is no significant difference in opinion between Whites and non-Whites regarding the importance of dealing with gangs.
H_a: There is a significant difference in opinion between Whites and non-Whites regarding the importance of dealing with gangs.

Observed

Dealing with Gangs	White (=1)	Non-White (=2)	**Grand Total**
Very Important (=3)	481	169	650
Important (=2)	203	62	265
Somewhat Important (=1)	54	8	62
Unimportant (=0)	16	5	21
Grand Total	754	244	998

Expected Frequency

Dealing with Gangs	White (=1)	Non-White (=2)	**Grand Total**
Very Important (=3)	491.082	158.918	650
Important (=2)	200.210	64.790	265
Somewhat Important (=1)	46.842	15.158	62
Unimportant (=0)	15.866	5.134	21
Grand Total	754	244	998

Chi-square value

Dealing with Gangs	White (=1)	Non-White (=2)	**Grand Total**
Very Important (=3)	0.207	0.640	0.847
Important (=2)	0.039	0.120	0.159
Somewhat Important (=1)	1.094	3.380	4.474
Unimportant (=0)	0.001	0.004	0.005

Interpretation:

Grand Total	1.341	4.144	**5.485**

Critical value=	7.8147	
Chi-square=	5.485	**=CHISQ.INV.RT(p,df)**
degree of freedom (df)=	3	**=(row-1)*(column-1) = (4-1)*(2-1)**
p=	0.139562301	**=CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Since the calculated chi-square value (5.485) is less than the critical value (7.8147) and the p-value (0.139562301) is greater than the significance level (0.05), we fail to reject the null hypothesis. This suggests that there is no significant difference in opinion between Whites and non-Whites regarding the importance of dealing with gangs.

(4) Do Whites and non-Whites agree on the importance of these priorities? On which issues is there a difference? Discuss the practical importance of any significant difference.

Overall, Whites and non-Whites generally agree on the importance of these priorities, as indicated by the lack of statistically significant differences in opinion across all three issues involving helping public schools, fighting gangs, and dealing with gangs.

It is essential to note that while statistical significance was not found, practical importance may still exist. Even though the differences were not significant at the chosen significance level of 0.05, there could still be practical implications to consider.

In short, the data doesn't say there are zero differences, just that they're not statistically big. However, looking for patterns in the data can still be useful in creating solutions and promoting fairness. So, when it comes to research, it's important to consider both big statistical differences and the real-world meaning of the results. However, while there are no significant differences in opinion between racial groups, it's still important to recognize and address any disparities or challenges faced by certain communities in accessing and benefiting from initiatives related to public schools, drug control, and gang prevention. Understanding the needs and perspectives within each community can help policymakers develop more effective and equitable strategies to address these critical issues.

2. Among the survey instrument, the first section addresses a few wide-ranging questions on some matters of current interest to the county. Of particular interest was how often residents watched the local county cable TV station, called Orange TV.

The following questions van be answered with a simple yes or no.

(Trust): Do you trust Orange County Government to do what is right most of the time?
Yes (=1) No (=2). Can't Say (=3)

(Watch): Do you watch Orange TV?
Yes (=1) No (=2). Can't Say (=3)

Use the Public Perception dataset. Is the relationship between watching Orange TV (watch), the county's cable television station, and trusting the county government to do what is right most of the time (Trust) statistically significant? Do you consider this a causal relationship or an association? Does the analysis satisfy the assumptions of the chi-square test? If not, how might you address this problem? *Please provide your solution after addressing the assumption violation problem.*

Hypotheses

H_0: There is no significant relationship between watching Orange TV and trusting the county government to do what is right most of the time.

H_a**:** There is a significant relationship between watching Orange TV and trusting the county government to do what is right most of the time.

Observed

Trust v. Watch	Yes (=1)	No (=2)	Don't Say (=3)	Grand Total
Yes (=1)	338	361	8	707
No (=2)	137	144		281
Don't Say (=3)	22	17	2	41
Grand Total	497	522	10	1029

Expected Frequency

Trust v. Watch	Yes (=1)	No (=2)	Don't Say (=3)	Grand Total
Yes (=1)	341.476	358.653	6.871	707
No (=2)	135.721	142.548	2.731	281
Don't Say (=3)	19.803	20.799	0.398	41
Grand Total	497	522	10	1029

Chi-Square Value

Trust v. Watch	Yes (=1)	No (=2)	Don't Say (=3)	Grand Total
Yes (=1)	0.035	0.015	0.186	0.236
No (=2)	0.012	0.015	2.731	2.758
Don't Say (=3)	0.244	0.694	6.437	7.375
Grand Total	0.291	0.724	9.354	**10.369**

Critical Value	9.4877	
Chi-square=	10.369	**=CHISQ.INV.RT(p,df)**
degree of freedom (df)=	4	**=(row-1)*(column-1) = (3-1)*(3-1)**
p=	0.105763061	**= CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Before addressing the assumption violation issue, I'll still go ahead to interpret the above. Since the chi-square statistic (10.369) exceeds the critical value (9.4877) for a significance level of 0.05, we reject the null hypothesis. Additionally, the p-value (0.105763061) is greater than 0.05, which further supports the conclusion of not rejecting the null hypothesis.

Therefore, based on these results, we can conclude that there is no statistically significant association between watching Orange TV and trusting the county government to do what is right most of the time.

Do you consider this a causal relationship or an association?

Regarding whether there is a casual relationship vs. association, the analysis conducted using the chi-square test determines an association between the variables but does not establish causation. The test only tells us whether there is a relationship between watching Orange TV and trusting the county government; it does not tell us the direction or cause of this relationship.

Addressing the Assumption Violation Issue:

Regarding the assumptions of the chi-square test, It is essential to ensure that the assumptions of independence of observations are met. If these assumptions are violated, the results of the chi-square test may not be reliable.

Each cell should have an expected frequency of at least 5. Violation of this assumption can affect the validity of the chi-square test. If you check the Expected table above, the highlighted data in the expected frequency were less than 5. This shows that it can affect the validity of the chi-square test.

In this case, since the assumptions are violated, we can consider merging categories with low frequencies. Since the expected cell frequencies are too low, one possible solution would be to combine categories with low frequencies to create larger expected cell counts. For example, merging the "Can't say" category with "No" since the frequencies are less than 5. The calculations after merging were done as shown below:

Observed

Trust v. Watch	Yes (=1)	No (=2&3)	**Grand Total**
Yes (=1)	338	369	707
No (=2)	137	144	281
Don't Say (=3)	22	19	41
Grand Total	497	532	1029

Expected Frequency

Trust v. Watch	Yes (=1)	No (=2&3)	Grand Total
Yes (=1)	341.48	365.52	707
No (=2)	135.72	145.28	281
Don't Say (=3)	19.80	21.20	41
Grand Total	497	532	1029

Chi-Square Value

Trust v. Watch	Yes (=1)	No (=2&3)	Grand Total
Yes (=1)	0.04	0.03	0.07
No (=2)	0.01	0.01	0.02
Don't Say (=3)	0.24	0.23	0.47
Grand Total	0.29	0.27	0.56

Critical Value	5.9915	
Chi-square=	0.563329457	**=CHISQ.INV.RT(p,df)**
degree of freedom (df)=	2	**=(row-1)*(column-1) = (3-1)*(2-1)**
p=	0.754526613	**=CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Interpretation:

Since the chi-square statistic (0.56) is less than the critical value (5.9915) for a significance level of 0.05, and the p-value (0.754526613) is greater than 0.05, we fail to reject the null hypothesis.

Therefore, even after merging categories to address the violation of assumptions, the conclusion remains the same: there is no statistically significant association between watching Orange TV and trusting the county government to do what is right most of the time.

PART 2. Performance Report Project –Citizen Satisfaction Data

Using your variables in your research proposal,

1) Develop chi-square hypotheses (relationship between two categorical variables)
2) Create a contingency table
3) Run a chi-square analysis *(* Remember that you should address the assumption violation issue.)* 4) Report the result

** Suggested table for the report:

Table # shows the contingency table and the chi-square test shows that there is a significant relationship between them (^{2}= 799.681, p= 0.00).

IV (Snow Removal on Major City Streets)	**DV (Resident Satisfaction with Snow Removal Services)**					
	Very dissatisfied	Dissatisfied	Neutral	Satisfied	Very satisfied	Total
Very dissatisfied	56.00%	0.64%	0.25%	0.00%	0.00%	1.15%
Dissatisfied	28.00%	22.93%	0.25%	0.00%	0.00%	3.15%
Neutral	16.00%	34.39%	21.32%	0.30%	0.00%	10.54%
Satisfied	0.00%	39.49%	66.42%	56.61%	0.00%	51.18%
Very satisfied	0.00%	2.55%	11.76%	43.09%	100.00%	33.98%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Chi-square	^{2}= 799.681, p= 0.00					

- Hint: N(%) can be calculated using PivotTable (count – number and percentage)
- Hint: You can create this table before addressing the assumption violation issue. But, your chi-square value and p-value should be calculated after addressing the assumption violation issue.

1) Develop chi-square hypotheses (relationship between two categorical variables)

H_0: There is no significant relationship between residents' satisfaction with snow removal services and snow removal on major city streets.

H_a: There is a significant relationship between residents' satisfaction with snow removal services and snow removal on major city streets.

2) Create a contingency table

Observed

IV v. DV	Very Dissatisfied (=1)	Dissatisfied (=2)	Neutral (=3)	Satisfied (=4)	Very Satisfied (=5)	Grand Total
Very Dissatisfied (=1)	14	1	1			16
Dissatisfied (=2)	7	36	1			44
Neutral (=3)	4	54	87	2		147
Satisfied (=4)		62	271	381		714
Very Satisfied (=5)		4	48	290	132	474
Grand Total	25	157	408	673	132	1395

Expected Frequency

IV v. DV	Very Dissatisfied (=1)	Dissatisfied (=2)	Neutral (=3)	Satisfied (=4)	Very Satisfied (=5)	Grand Total
Very Dissatisfied (=1)	0.287	1.801	4.680	7.719	1.514	16
Dissatisfied (=2)	0.789	4.952	12.869	21.227	4.163	44
Neutral (=3)	2.634	16.544	42.994	70.918	13.910	147
Satisfied (=4)	12.796	80.357	208.826	344.460	67.561	714
Very Satisfied (=5)	8.495	53.346	138.632	228.675	44.852	474
Grand Total	25	157	408	673	132	1395

Chi Square Value

IV v. DV	Very Dissatisfied (=1)	Dissatisfied (=2)	Neutral (=3)	Satisfied (=4)	Very Satisfied (=5)	Grand Total
Very Dissatisfied (=1)	655.837	0.356	2.893	7.719	1.514	668.319

Dissatisfied (=2)	48.929	194.666	10.947	21.227	4.163	279.933
Neutral (=3)	0.708	84.800	45.043	66.975	13.910	211.436
Satisfied (=4)	12.796	4.194	18.511	3.876	67.561	106.938
Very Satisfied (=5)	8.495	45.646	59.252	16.446	169.333	299.171
Grand Total	726.764	329.662	136.646	116.243	256.481	1565.796

Critical value	26.2962	
Chi-square=	1565.796	**=CHISQ.INV.RT(p,df)**
degree of freedom (df)=	16	**=(row-1)*(column-1) = (3-1)*(3-1)**
p=	0.00	**=CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Addressing the Assumption Violation Issue:

Each cell should have an expected frequency of at least 5. Violation of this assumption can affect the validity of the chi-square test. If you check the Expected table above, the highlighted data in the expected frequency were less than 5. This shows that it can affect the validity of the chi-square test.

In this case, since the assumptions are violated, we can consider merging categories with low frequencies. Since the expected cell frequencies are too low, one possible solution would be to combine categories with low frequencies. For example, merging the "Very Dissatisfied" category with "Dissatisfied" and the "Very Satisfied " category with "Satisfied " since the frequencies are less than 5. The contingency table after addressing the assumption violation is shown below:

Observed

Row Labels	Dissatisfied (=1&2)	Neutral (=3)	Satisfied (=4&5)	**Grand Total**
Dissatisfied (=1&2)	58	2		60
Neutral (=3)	58	87	2	147
Satisfied (=4)	62	271	381	714
Very Satisfied (=5)	4	48	422	474
Grand Total	182	408	805	1395

Expected Frequency

Row Labels	Dissatisfied (=1&2)	Neutral (=3)	Satisfied (=4&5)	**Grand Total**
Dissatisfied (=1&2)	7.828	17.548	34.624	60
Neutral (=3)	19.178	42.994	84.828	147

Satisfied (=4)	93.153	208.826	412.022	714
Very Satisfied (=5)	61.841	138.632	273.527	474
Grand Total	182	408	805	1395

Chi-Square Value

Row Labels	Dissatisfied (=1&2)	Neutral (=3)	Satisfied (=4&5)	**Grand Total**
Dissatisfied (=1&2)	321.570	13.776	34.624	369.970
Neutral (=3)	78.583	45.043	80.875	204.502
Satisfied (=4)	10.418	18.511	2.336	31.265
Very Satisfied (=5)	54.100	59.252	80.593	193.944
Grand Total	464.671	136.583	198.427	**799.681**

Critical value	12.5916	
Chi-square=	**799.681**	**=CHISQ.INV.RT(p,df)**
degree of freedom (df)=	6	**=(row-1)*(column-1) = (4-1)*(3-1)**
p=	0.00	**=CHISQ.TEST(observed array, expected array)**
alpha=	0.05	

Interpretation:

Since the computed chi-square value (799.681) is much larger than the critical value (12.5916), and the p-value (0.00) is less than the significance level (0.05), we reject the null hypothesis.

Therefore, we have sufficient evidence to conclude that there is a significant relationship between residents' satisfaction with snow removal services and snow removal on major city streets at the 0.05 significance level.

Additional Comment:

After addressing the assumption violation issue, I revised the table for the report by incorporating the merged categories, which led to slight adjustments in the percentages. While calculating the N% using the pivot table, I used the row column percentage.

Table # below shows the contingency table and the chi-square test shows that there is a significant relationship between them (^{2}= 799.681, p= 0.00).

IV (Snow Removal on Major City Streets)	DV (Resident Satisfaction with Snow Removal Services)			
	Very Dissatisfied/ Dissatisfied	Neutral	Very Satisfied/ Satisfied	**Total**
Very Dissatisfied/Dissatisfied	31.87%	0.49%	0.00%	4.30%
Neutral	31.87%	21.32%	0.25%	10.54%
Satisfied	34.07%	66.42%	47.33%	51.18%
Very Satisfied	2.20%	11.76%	52.42%	33.98%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Chi-square	^{2}= 799.681, p= 0.00			