Categorical Data Problems

Q.1. A study is conducted to compare whether incidence of muscle aches differs among athletes exposed to 5 types of pain medication. A total of 500 people who are members of a large fitness center are randomly assigned to one of the medications. After a lengthy workout, each is given a survey to determine presence/absence of muscle pain. For the 5 groups: 25, 38, 32, 40, and 35 are classified as having muscle pain, respectively. The following output gives the results for the Pearson chi-square statistic for testing (α=0.05):

H_0: True incidence rate of muscle pain doesn't differ among medications

H_A: Incidence rates are not all equal

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.150[a]	4	.188

Test Statistic: X_{obs}^2 **= 6.150**

Reject H_0 if the test statistic falls in the range(s) $X_{obs}^2 \geq \chi_{.05,5-1}^2 = 9.488$

P-value **.188**

Conclude (Circle One):   Medication effects not all equal   **No differences in effects**

Give the expected number of incidences of muscle pain for each medication under H_0:

  (25+38+32+40+35)(100)/500 = 170(100)/500 = 34

Q.2. A study is conducted to compare 3 treatments for curing insomnia. A sample of 300 insomniacs are obtained, and 100 are assigned to treatment A, 100 to treatment B, and 100 to treatment C at random. The following table gives the observed counts of subjects who are able to fall asleep within 30 minutes and unable to fall asleep within 30 minutes**.**

Treatment\Sleep	Yes	No	Total
A	60	40	100
B	50	50	100
C	40	60	100
Total	150	150	300

Complete the following table for expected cell counts under the hypothesis of no association between treatment and sleep status.

Treatment\Sleep	Yes	No	Total
A	50	50	100
B	50	50	100
C	50	50	100
Total	150	150	300

Compute the chi-square statistic

$$X^2_{obs} = \frac{(60-50)^2}{50} + \frac{(40-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} = 2+2+0+0+2+2 = 8$$

How large will the Pearson chi-square statistic need to be to conclude that the treatment effects differ if we conduct the test at the $\alpha = 0.05$ significance level?

$$RR: X^2_{obs} \geq \chi^2_{.05,(3-1)(2-1)} = \chi^2_{.05,2} = 5.991$$

Q.3. Among a group of 100 children exposed to a petting zoo, 20 contracted a particular symptom. Among a second group of 100 children not exposed to the petting zoo, 12 contracted the symptom. Give the estimated odds ratio (exposed group divided by not exposed group), and the corresponding 95% confidence interval for the population odds ratio. What do you conclude at the $\alpha=0.05$ significance level?   **Zoo increases odds   /   Zoo decreases odds   /   No association exists**

Group \ Outcome	Symptom Present	Symptom Absent
Exposed to Petting Zoo	20	80
Not Exposed to Petting Zoo	12	88

$$\hat{OR} = \frac{20/80}{12/88} = \frac{20(88)}{12(80)} = 1.833 \quad v_{OR} = \frac{1}{20} + \frac{1}{80} + \frac{1}{12} + \frac{1}{88} = 0.157 \quad \Rightarrow \quad z_{.025}\sqrt{v_{OR}} = 1.96\sqrt{0.157} = 0.777$$

95% CI for OR: $\left(1.833e^{-0.777}, 1.833e^{0.777}\right) \equiv (0.843, 3.987)$

Q.4. A study reports a 95% Confidence Interval for the Relative Risk of a disease for Group A relative to Group B to be (0.50 , 0.70). We conclude that Group A has a higher probability of getting the disease than Group B.

FALSE (Entire interval is BELOW 1)

Q.5. A study considered the effectiveness of wearing helmets among skiers and snowboarders in Canada. The researchers considered 2 groups in a Case/Control Study:

- Group 1 (Cases) – Skiers/Snowboarders suffering head or neck injuries (219 out of 824 wore helmets)
- Group 2 (Controls) – Skiers/Snowboarders suffering other injuries (929 out of 3295 wore helmets)

p.5.a. Fill in the following contingency table

Helmet\Injury	Head/Neck	Other	Total
Yes	219	929	1148
No	605	2366	2971
Total	824	3295	4119

p.5.b. Odds of wearing a helmet:

p.5.b.i. Among Head/Neck Injury **219 / 605 = 0.362**

p.5.b.ii. Among Other **929 / 2366 = 0.393**

p.5.c. Odds Ratio (Head/Neck relative to Other) **OR = 0.362 / 0.393 = 0.921**

p.5.d. 95% Confidence Interval for (population) Odds Ratio:

v_{OR} = (1/219) + (1/929) + (1/605) + (1/2366) = 0.00772   1.96(.00772)$^{1/2}$ = 0.172

Lower Bound **0.921e$^{-0.172}$ = 0.775**   Upper Bound **0.921e$^{0.172}$ = 1.094**

p.5.e. Conclusion based on your confidence interval

- Higher odds of wearing helmet among those with head/neck injury
- Higher odds of wearing helmet among those with other injury
- **Cannot conclude odds of wearing helmet differ by injury type**

Q.6. A manufacturer wishes to compare defective rates of two suppliers of a component used in their product. Let π_1 be the true defective rate for supplier 1, and π_2 be the true defective rate for supplier 2. The engineering department samples 500 parts from each supplier (n_1 = n_2 = 500). For supplier 1, X_1 = 60 parts were defective, while for supplier 2, X_2 = 25 parts were defective.

p.6.a. Compute a 95% Confidence Interval for the difference in true probabilities of defective parts for the 2 suppliers: π_1 - π_2:

$$\hat{\pi}_1 = \frac{60}{500} = 0.12 \qquad \hat{\pi}_2 = \frac{25}{500} = 0.05 \qquad \hat{\pi}_1 - \hat{\pi}_2 = 0.07 \qquad \hat{SE}\left\{\hat{\pi}_1 - \hat{\pi}_2\right\} = \sqrt{\frac{0.12(0.88)}{500} + \frac{0.05(0.95)}{500}} = 0.0175$$

$$95\% \text{ CI for } \pi_1 - \pi_2: \quad 0.070 \pm 1.96(.0175) \equiv 0.070 \pm 0.034 \equiv (.036, 0.104)$$

p.6.b. Based on your interval, which is the best statement?

i.   **Higher defect rate for supplier 1**   ii. Lower defect rate for supplier 1   iii. No significant difference

Q.7. The following contingency table gives the results of a ballot order experiment on voters' choice.

Observed				
Order\Party	Dem	Rep	Ind	Total
Dem/Rep/Ind	63	310	5	378
Rep/Ind/Dem	69	302	5	376
Ind/Dem/Rep	61	308	5	374
Total	193	920	15	1128
Expected				
Order\Party	Dem	Rep	Ind	
Dem/Rep/Ind	64.7	308.3	5.0	
Rep/Ind/Dem	64.3	306.7	5.0	
Ind/Dem/Rep	64.0	305.0	5.0	
Chi-Square				
Order\Party	Dem	Rep	Ind	
Dem/Rep/Ind	0.04	0.01	0.00	
Rep/Ind/Dem	0.34	0.07	0.00	
Ind/Dem/Rep	0.14	0.03	0.00	

p.7.a. Complete the Expected and Chi-Square tables (note you just need to compute the upper left-hand corners of the tables).

p.7.b. Give the test statistic for testing whether vote cast (party) is independent of order of party on the ballot.

$$X^2_{obs} = 0.04 + 0.01 + 0.00 + 0.34 + 0.07 + 0.00 + 0.14 + 0.03 + 0.00 = 0.63$$

p.7.c. Rejection Region: $X^2_{obs} \geq \chi^2_{.05,(3-1)(3-1)} = \chi^2_{.05,4} = 9.488$

p.7.d. The researchers claim that this study is evidence that ballot order "influences" candidate choice. Based on this study, the researchers are: Correct   or   **Incorrect** ?

Q.8. In a survey to analyze the coffee drinking habits of mean and women, a random sample of 400 males and 500 females was collected. Of the males, 200 said they drink decaffeinated coffee. Of the females, 300 said they drink decaffeinated coffee.

p.8.a. Give a 95% confidence interval for the population proportion of males who would say they drink decaffeinated coffee.

$$\hat{\pi}_M = \frac{200}{400} = 0.500 \qquad \hat{SE}\left\{\hat{\pi}_M\right\} = \sqrt{\frac{0.5(0.5)}{400}} = 0.025$$

95% CI: $0.500 \pm 1.96(0.025) \equiv 0.500 \pm 0.049 \equiv (0.451, 0.549)$

p.8.b. Is there significant evidence to indicate the population of males and the population proportion of females who would say they drink decaffeinated coffee are different? $\sqrt{\left(\frac{500}{900}\right)\left(\frac{400}{900}\right)\left(\frac{1}{400} + \frac{1}{500}\right)} = 0.0333$

$$\hat{\pi}_M = \frac{200}{400} = 0.500 \quad \hat{\pi}_F = \frac{300}{500} = 0.600 \quad \hat{\pi}_M - \hat{\pi}_F = -0.100 \quad TS: z_{obs} = \frac{-0.100}{0.0333} = -3.003$$

Yes, since |-3.003| > z.025 = 1.96

Q.9. An insurance company wished to determine if having a fire extinguisher in the kitchen had any influence on the dollar amount of damage in a kitchen fire. They selected a random sample of 300 claims where an extinguisher was present and 500 claims where there was no extinguisher present. The results were classified by the dollar amount of damage. A summary of the results in $ amount of damage:

Extinguisher\Damage	Low(<250)	Med(250-499)	High (>500)	Total
Yes	145	120	35	300
No	215	160	125	500
Total	360	280	160	800

Expected	Low(<250)	Med(250-499)	High (>500)	Total
Yes	135	105	60	300
No	225	175	100	500
Total	360	280	160	800

p.9.a. Complete the table of expected values

p.9.b. Compute the Chi-square statistic and give the rejection region for testing whether the distribution of damages differs by extinguisher presence/absence (H_0: No association):

$$TS: \quad X^2_{obs} = \frac{100}{135} + \frac{225}{105} + \frac{625}{60} + \frac{100}{225} + \frac{225}{175} + \frac{625}{100} = 0.74 + 2.14 + 10.42 + 0.44 + 1.29 + 6.25 = 21.28$$

$$RR: \quad X^2_{obs} \geq \chi^2_{.05,2} = 5.991$$

Q.10. A 95% Confidence Interval for the odds ratio of group1 relative to group 2 is reported as (0.80 , 0.90). The appropriate conclusion is, where p_i= probability of success for population i (circle correct answer):

p.10.i. $\pi_1 = \pi_2$ p.10.ii. $\pi_1 > \pi_2$ p.10.iii. $\pi_1 < \pi_2$

Q.11 A study was conducted to compare adverse events between two dose levels of Escitalopram (10 mg, and 20 mg). Among $n_{10} = 535$ subjects receiving the 10 mg dose, $y_{10} = 105$ reported symptoms of nausea. Among $n_{20} = 542$ subjects receiving the 20 mg dose, $y_{20} = 125$ reported symptoms of nausea.

p.11.a. Complete the following contingency table:

Dose\Event	Nausea Present	Nausea Absent	Total
10mg	105	430	535
20mg	125	417	542
Total	230	847	1077

p.11.b. Compute the following quantities:

$$\hat{\pi}_{10} = \frac{105}{535} = .1963 \quad \hat{\pi}_{20} = \frac{125}{542} = .2306 \quad \bar{\pi} = \frac{230}{1077} = .2136$$

p.11.c. Given the following quantities $\sqrt{\bar{\pi}(1-\bar{\pi})\left[\dfrac{1}{n_{10}}+\dfrac{1}{n_{20}}\right]} = .0250 \quad \sqrt{\dfrac{\hat{\pi}_{10}\left(1-\hat{\pi}_{10}\right)}{n_{10}}+\dfrac{\hat{\pi}_{20}\left(1-\hat{\pi}_{20}\right)}{n_{20}}} = .0249$

p.11.c.i) Test $H_0:\pi_{10} = \pi_{20}$ versus $H_A:\pi_{10} \neq \pi_{20}$

Test Statistic: $\mathbf{z_{obs} = (.1963-.2306) / .0250 = -.0343 / .0250 = -1.37}$ Rejection Region $|z_{obs}| \geq 1.96$ P-value $> .05$

p.11.c.ii) Obtain a 95% Confidence Interval for $\pi_{10} - \pi_{20}$: $-.0343 \pm 1.96(.0249) \equiv -.0343 \pm .0488 \equiv (-.0831 , .0145)$

Q.12. An ergonomic study was conducted to compare computer keyboard preferences among 5 keyboards by gender. A sample of 66 females, and a sample of 36 males were obtained, and each participant selected their preferred keyboard.:

Gender\Keyboard	Kybrd1	Kybrd2	Kybrd3	Kybrd4	Kybrd5	Total
Female	11	15	22	10	8	66
Male	4	8	10	6	8	36
Total	15	23	32	16	16	102
Expected	Kybrd1	Kybrd2	Kybrd3	Kybrd4	Kybrd5	Total
Female	9.71	14.88	20.71	10.35	10.35	
Male	5.29	8.12	11.29	5.65	5.65	
Total						

p.12.a. Complete the table of expected values

p.12.b. Compute the Chi-square statistic by completing the following table and give the rejection region for testing whether the distribution of damages differs by extinguisher presence/absence (H_0: No association):

Chi-Square	Kybrd1	Kybrd2	Kybrd3	Kybrd4	Kybrd5	Total
Female	0.171	0.001	0.081	0.012	0.535	#N/A
Male	0.315	0.002	0.148	0.022	0.980	#N/A
Total	0.315	0.003	0.229	0.034	1.515	

Test Statistic $\mathbf{X_{obs}^2 = 2.096}$ Reject H_0 if test statistic $\geq \chi_{.05,4}^2 = \mathbf{9.477}$

Q.13. In a random sample of n = 100 adult fish from a large lake, y = 63 have a particular genetic trait. Test whether a majority of all adult fish in the lake have the trait. H_0: $\pi \le 0.50$ versus H_A: $\pi > 0.50$

$$\hat{\pi} = 0.63 \quad \hat{SE}\left\{\hat{\pi}\right\}_0 = \sqrt{\frac{0.50(0.50)}{100}} = 0.050 \quad TS: z_{obs} = \frac{0.63 - 0.50}{.050} = 2.600 \quad RR: z_{obs} \ge z_{.05} = 1.645$$

Q.14. A study of the largest ski resorts in Scotland looked at whether event of ski injury was related to whether or not it was the participants first day at the resort (FDP). The following table gives the results. Obtain the estimated Odds Ratio, and its 95% Confidence Interval, where: $OR = \dfrac{\text{odds(Injury|FDP)}}{\text{odds(Injury|non-FDP)}}$

Group\Outcome	Injured	Uninjured
First Day Participant	381	144
Not First Day	1743	1638

Hint: $(1/381) + (1/144) + (1/1743) + (1/1638) = .01075$

$$\hat{OR} = \frac{381(1638)}{144(1743)} = 2.4864 \quad 1.96\sqrt{.01075} = .2032 \quad e^{.2032} = 1.2253$$

95% CI: $\left(\dfrac{2.4864}{1.2253}, 2.4864(1.2253)\right) \equiv \left(2.0292, 3.0466\right)$

Does this study provide evidence of an association between first day participation and injury? **Yes** or No

Q.15. A study was conducted to test whether the proportions of males and females who prefer one brand of coffee (A) to another brand (B). Samples of 50 male and 50 female coffee drinkers and their preferences were obtained, with 30 males preferring Brand A, and 20 females preferring A. Let π_M and π_F represent the population proportions of males and females preferring brand A.

p.15.a. Obtain a point estimate of $\pi_M - \pi_A$:
p.15.b. Suppose we want to test H_0: $\pi_M - \pi_F = 0$ versus H_A: $\pi_M - \pi_F \ne 0$, give the estimated standard error under the null hypothesis.
p.15.c. Conduct the test, by giving the test statistic, rejection region, and conclusion.

Test Statistic: z_{obs} = **2.00** Rejection Region: $|z_{obs}| \ge z_{.025}$ = **1.96** Reject H_0? **Yes** / No

p.15.d. Give the observed and expected cell counts for the Chi-Square test

Observed	Brand A	Brand B	Total
Male	30	20	50
Female	20	30	50
Total	50	50	50

Expected	Brand A	Brand B	Total
Male	25	25	50
Female	25	25	50
Total	50	50	100

Q.16. Interchanging the rows (levels of the independent variable) in a contingency table will:

   a) Effect the chi-square statistic, but not the numbers of concordant/discordant pairs
   b) Effect the numbers of concordant/discordant pairs, but not the chi-square statistic
   c) Effect both the chi-square statistic and the numbers of concordant/discordant pairs
   d) Effect neither the chi-square statistic or numbers of concordant/discordant pairs

Q.17. A study considered the effect of executive succession circumstances in small companies. Of 19 companies with planned successions (e.g. long-planned retirement of chief executive), 12 had undiminished profitability and 7 had diminished profitability. Of 17 companies without planned successions (e.g. untimely death or legal problems of chief executive) 3 had undiminished profitability and 14 had diminished profitability. Which method of analysis would be most appropriate for this study.

   a) McNemar's Test
   b) Linear Regression
   c) Gamma / Kendall's Tau
   d) Fisher's Exact Test

Q.18. Two researchers conduct a chi-square test with contingency tables of the exact same dimensions (say $r=2$ and $c=3$). Jack sampled 100 individuals from each of the 2 populations and Jill sampled 200 individuals from each population. They obtained the exact same conditional distributions for each of the two populations (although the conditional distribution for population 1 is not identical to the conditional distribution for population 2).

   a) Their chi-square statistics will be the same
   b) Jack's chi-square statistic will be twice as large as Jill's
   c) Jill's chi-square statistic will be twice as large as Jack's
   d) Jill's chi-square statistic will be 4 times as large as Jack's
   e) Jack's chi-square statistic will be 4 times as large as Jill's

Q.19. A researcher is interested in comparing married men's and women's attitudes toward same-sex marriage. A random sample of 1000 married couples is obtained. Of these couples, both the male and female opposed it for 200 of the couples, both the male and female favored it for 300 of the couples. The male favored and the female opposed it for 260 of the couples. For the remaining couples the male opposed and the female favored.

p.19.a. Give the test statistic for testing whether the proportions of males and females favoring same sex marriage are the same (H_0) or differ (H_A)

Male \ Female	Oppose	Favor
Oppose	200	240
Favor	260	300

$$TS: z_{obs} = \frac{240 - 260}{\sqrt{240 + 260}} = \frac{-20}{22.36} = -0.89 \qquad RR: |z_{obs}| \geq 1.96$$

p.19.b. What do we conclude at the $\alpha=0.05$ significance level?

- **Do not conclude the proportions differ**
- Conclude a higher proportion of males favor same-sex marriage
- Conclude a higher proportion of females favor same-sex marriage

Q.20. A study compared St. John's Wort (SJW), Sertraline, and placebo in patients with major depressive disorder. Patients were assigned at random to one of the three treatments and were classified as having any response or no response. The contingency table is given below.

Trt \ Outcome	Any Response	No Response	Total
SJW	43	70	113
Sertraline	53	56	109
Placebo	50	66	116
Total	146	192	338

p.20.a. Give the conditional distributions for each treatment and overall.

Trt \ Outcome	Any Response	No Response	Total
SJW	100(43/113)=38.1	61.9	100%
Sertraline	48.6	51.4	100%
Placebo	43.1	56.9	100%
Total	43.2	56.8	100%

p.20.b. Give the expected count for Any Response among SJW patients under the hypothesis of no association between response and treatment.

E_{11} = 113(146) / 338 = 48.81 = .381(113)

Q.21. A study considered the association between governors' strength and their control over state agencies. State administrators were classified by the strength of their state governor (very weak, weak, moderately strong, strong) and were asked who had control over their agency (legislature, about equal between legislature and governor, governor). The following table contains the measures gamma and Kendall's tau based on this sample.

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.177	.028	6.250	.000
	Gamma	.255	.040	6.250	.000
N of Valid Cases		865			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Obtain a 95% confidence interval for the true (population) value of Kendall's tau.

.177 ± 1.96(.028) ≡ .177 ± .055 ≡ (.122 , .232)

Q.22. A study considered the occurrence of upholstery in households in Philadelphia during 4 time periods. The following table gives a cross-tabulation of period by upholstery for samples of households within the periods.

PERIOD * UPHOLSTE Crosstabulation

Count

		UPHOLSTE		Total
		No	Yes	
PERIOD	1	61	19	80
	2	68	14	82
	3	64	18	82
	4	53	27	80
Total		246	78	324

p.22.a. Give the expected number of Yes in Period 1 under the hypothesis that upholstery occurrence is independent of period.

$E_{1,Yes}$ = 80(78) / 324 = 19.26

p.22.b. Give the contribution for that cell to the chi-square statistic.

$(19 - 19.26)^2 / 19.26 = .004$

Q.23. A test for independence is conducted where the nominal explanatory variable has 4 levels and the nominal response variables has 3 levels. If we conduct the chi-square test at the $\alpha=0.05$ significance level, we will conclude the variables are not independent (dependent) if the test statistic:

a) < 12.59 b) < 21.03 c) **> 12.59**   d) > 21.03 e) < 16.92 f) > 16.92

Q.24. A researcher is interested in the distribution of college professors party affiliations. She samples professors at public universities from each of the following colleges: medicine, law, engineering, and liberal arts/sciences. The following table gives the conditional distributions (percentages) for party affiliations within college.

College\Party	Independent	Republican	Democrat	Total % (cases)
Medicine	5%	50%	45%	100% (n=200)
Law	10%	40%	50%	100% (n=150)
Engineering	2%	58%	40%	100% (n=100)
Arts/Sciences	15%	30%	55%	100% (n=200)
Marginal Distribution (%)	8.77%	42.77%	48.46%	100% (n=650)

The researcher wishes to test whether the distribution of party affiliation is independent of college (H_0) or that it is dependent of college (H_A).

p.24.a. Give the observed cell count (f_o) for independent party affiliation among medicine professors.

$.05(200) = 10$

p.24.b. Give the expected cell count under H_0 (f_e) for independent party affiliation among medicine professors.

$.0877(200) = 17.54$

p.24.c. Give the contribution to the chi-square statistic for the cell corresponding to independent party affiliation among medicine professors.

$(10 - 17.54)^2 / 17.54 = 3.24$

Q.25. A sample of Gator fans are classified on the frequency they listen to Sports talk radio (never, occasionally, frequently) and their attitude toward coach Ron Zook (unfavorable, neutral, favorable). The following SPSS output gives gives the following results for the measure of association gamma.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T^b	Approx. Sig.
Ordinal by Ordinal	Gamma	-.432	.052	-7.803	.000
N of Valid Cases		450			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

p.25.a. Give a 95% Confidence interval for Gamma based on the population of all Gator fans

-.432 ± 1.96(.052) ≡ (-.534 , -330)

p.25.b. How would you describe the results? **Circle one**.

i)   **The more people listen to talk radio, the less favorable they are to Zook.**

ii)   The more people listen to talk radio, the more favorable they are.

iii)   People's attitudes toward Zook are independent of how much they listen to talk radio.

Q.26. You are conducting a literature view in your field of study. The author of an article has conducted a chi-squared test for independence (Nominal variables). He has 2 groups he's comparing and 3 possible outcomes. He reports a chi-square statistic of 5.50 and a P-value of 0.025. What do you conclude?

a) **He has understated P-value**   b) He has overstated P-value   c) Cannot determine

Q.27. Gamma and Kendall's τ_B will be appropriate association measures when two categorical variables are measured on nominal scales. TRUE or **FALSE**

Q.28. A chi-square test is conducted in a two-by-two contingency table. Each cell's $(f_o-f_e)^2$ value is 100. What can be said about the chi-square statistic?

a) **It will be larger if each f_e = 100 than if each f_e = 1000**
b) It will be smaller if each f_e = 100 than if each f_e = 1000
c) It will be the same whether each f_e = 100 or each f_e = 1000

Q.29. In a contingency table, where the conditional distribution of responses is the same for each level of the independent variable, what can be said about the chi-square statistic and the measures Gamma and Kendall's τ_B:

   a) **They will each be 0**
   b) They will each be 1
   c) We need more information

Q.30. A small study is conducted with 5 individuals from one political party (Party A) and 5 individuals from the other major political party (Party B). Of the people from Party A, 4 favor a political action. Of the people from Party B, 1 favors the political action. Write out the contingency table for this data, as well as the contingency table with the same margins, but stronger evidence that the true population proportion of Party A favoring the political action is higher than the proportion of Party B.

Observed Contingency Table   **Table with Stronger Evidence**

Party\Opinion	Favor	Oppose
A	4	1
B	1	4

Party\Opinion	Favor	Oppose
A	5	0
B	0	5

Q.31. A survey is conducted among random samples of Democrats, Republicans, and Independents. Each respondent is asked whether they favor a new social initiative. The counts are given in the following table.

Construct the conditional distributions of opinion for each party.

Party/Opinion	Favor	Neutral	Oppose	Total
Democrat	60	80	60	200
Republican	75	100	75	250
Independent	50	50	50	150
Total	185	230	185	600

P\O	F	N	O	Total
D	30	40	30	100
R	30	40	30	100
I	33.33	33.33	33.33	100

Q.32. A studio marketer is interested in whether there is an association between length of movie preview (short, medium, long) and individual's stated intentions to see the movie. Students are exposed to a 25 minute program, where the preview is embedded in the program. At the end of the task, a series of questions, including intention to go to the movie (no, possibly, certainly). The following table gives the results (there were 20 students exposed to each preview length). Give the number of concordant and discordant pairs.

Ad Length/Intention	No	Possibly	Certainly
Short	4	8	8
Medium	8	6	6
Long	12	6	2

$$C = 2(4+8+8+6)+6(4+8)+6(4+8)+6(4) = 52+72+72+24 = 220$$
$$D = 8(12+6+8+6)+6(12+6)+8(12+8)+6(12) = 256+108+160+72 = 596$$

Q.33. A sample of UFO reports were cross-classified by whether there was presence/absence of physical effects (rows) and presence/absence of multimedia evidence (columns). Since these are paired data in the sense that these are two types of evidence that can be present or absent on the same sampling units (reports), use McNemar's test to test the hypotheses:

PHYSICAL * MEDIA Crosstabulation

Count

		MEDIA		
		Absent	Present	Total
PHYSICAL	Absent	42	32	74
	Present	99	30	129
Total		141	62	203

H_0: Proportions of UFO sightings leaving physical and multimedia evidence are equal.

H_A: The proportions are unequal

Test Statistic: z_{obs} = (32-99) / (32+99)$^{1/2}$ = -67 / 11.45 = -5.85

P-value: ≈ 0

Conclusion (α=0.05 significance level):

    i)   Do not conclude proportions differ

    ii)   **Conclude Proportion with physical evidence higher**

    iii)   Conclude Proportion with media evidence higher

Q.34. The following computer output shows the results of computing gamma and Kendall's τ_B for a pair of ordinal variables.

Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	-.398	.036	-11.115	.000
	Gamma	-.546	.046	-11.115	.000
N of Valid Cases		385			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Which statement best describes the result (α=0.05 significance level)?

i)   The distribution of outcomes is not associated with the independent variable

ii)   There is a tendency for people who score high on one variable to score high on the other variable

iii)   **There is a tendency for people who score high on one variable to score low on the other variable**

Q.35. The reason we use measures like gamma or Kendall's τ_B for ordinal independent and dependent variables as opposed to chi-square test is that often we specifically hypothesize positive or negative association as opposed to any possible association. **TRUE**   FALSE

Q.36. You obtain a chi-square statistic of 0, when there were 3 groups (rows) and 2 outcomes (columns), with 100 people per row, and 60% of people in column 1 and 40% in column 2. Give the table of observed counts (f_o).

Row\Column	1	2
1	60	40
2	60	40
3	60	40

Q.37. A study is conducted to determine whether there is an association between packaging and quality assessment. Cereal was packaged in one of three type boxes (plain, medium, and fancy imagery). Samples of 100 subjects tasted the cereal from the 3 box types (each person just made one assessment, there were 300 total people). The ratings were: Low quality, average quality, high quality. The following table gives the results.

Rating

		Low	Average	High
	Plain	50	30	20
Box	Medium	30	40	30
	Fancy	20	30	50

p.37.a. Give the number of Concordant Pairs: **7500+2400+2400+2000 = 14300**

p.37.b. Give the number of Discordant Pairs: **2400+1500+1500+800 = 6200**

p.37.c. Compute the estimated Gamma: **(14300-6200) / (14300+6200) = 8100 / 20500 = .395**

p.37.d. The estimated standard error of $\hat{\gamma}$ is .069. Can we conclude there is an association between packaging and quality assessment at the 0.05 significance level? Why? **Yes   .395 ± 2(.069) does not contain 0**

Q.38. A researcher sampled n=300 registered voters who had been registered to vote over a certain period of time in his state. The voters classified themselves with respect to religiosity (Low, Medium, High frequency of church attendance) and past presidential voting frequency (Low, Medium, High). The results are given below:

Religiosity\Voting	Low	Medium	High
Low	50	30	20
Medium	30	40	30
High	20	30	50

Same answers as Q.37.

p.38.a. How many concordant pairs are there?

p.38.b. How many discordant pairs are there?

p.38.c. Compute the estimate of γ.

Q.39. Two movie reviewers are being compared with respect to their rates of giving positive reviews. A sample of 200 movies that each reviewer has rated is obtained, giving the following cross-tabulation. Give the test statistic and P-value for testing

H_0: $\pi_A = \pi_B$ against H_A: $\pi_A \neq \pi_B$ where π_A and π_B are the true proportions of positive reviews for reviewers A and B.

Reviewer A\B	Positive	Not Positive
Positive	60	30
Not Positive	40	70

Q.40. A content analysis was conducted to compare the reporting of 3 news organizations. Random samples of 50 stories were obtained from each organization, and each was classified by their attitude toward the current government: Positive, Neutral, or Negative. The table of observed frequencies is given below.

fo	Positive	Neutral	Negative	Total
A	30	10	10	50
B	10	30	10	50
C	10	10	30	50
Total	50	50	50	150

p.40.a. Obtain the expected count for each cell under the null hypothesis of no association between news organization and attitude toward the current government. Hint: It is the same for each of the 9 cells.   **16.67**

p.40.b. The chi-square statistic is $X_{obs}^2 = 48$. Can we conclude that there is an association between news organization and attitude toward the current government at the 0.05 level? Why?   **Yes, critical Chi-square with df=4 is 9.477**

Q.41. A researcher is interested in determining whether people's attitudes toward a political candidate is effected by exposure to an advertisement for the candidate. Subjects were asked if they had a positive attitude toward the candidate prior to seeing the video (Yes or No). The subjects then watched an hour-long television show, with advertisement being played during the TV show. Afterwards, the subjects were again asked whether they had a positive attitude toward the candidate (Yes or No). Which method of analysis would be appropriate?

Pearson's Chi-Square Test   Fisher's Exact Test   **McNemar's Test**   Test for Ordinal Association

Q.42. A study was conducted with large samples of British, Belgian, and Norwegian food consumers being classified into one of 5 classifications (naïve, sensible, skeptical, denying, or other). If we wished to test whether there was an association between classification and nationality, we would reject the null hypothesis of no association, if the test statistic falls in the range:

> 3.841   > 5.991   > 7.815   > 9.488   > 11.070   **> 15.507**   Need to know sample sizes

Q.43. If in a sample of individuals, classified on 2 ordinal scales, 75% of all combined Concordant and Discordant pairs are concordant, then the estimate of Gamma will be:

0.75   -0.75   **0.50**   -0.50   0.25   -0.25   0   None of the above

Q.44. A study was conducted, with samples of 2 brands of cars, and whether or not the car could reach a given speed in a given amount of time (Pass or Fail). The researchers sampled 10 cars from each brand, and found that 3 of Brand A did pass the test, and 1 from Brand B did. Fill in the contingency table for this outcome, and the table for the more extreme outcome in favor of Brand A, given the marginal totals.

Observed	Pass	Fail	Total
Brand A	3	7	10
Brand B	1	9	10
Total	4	16	20

More Extreme	Pass	Fail	Total
Brand A	4	6	10
Brand B	0	10	10
Total	4	16	20

Q.45. In a test of the endowment effect, the following experiment was conducted with data given below.

- 3 Gift Conditions (Explanatory Variable): Subject given a mug (1), given candy bar (2), given nothing (3)
- 2 Possible Outcomes: Preference between Mug (1) or Candy (2)

We wish to test H_0: No association between Gift Condition and Product preference

	Observed				Expected		
Condition \ Preference	Mug	Candy	Total		Mug	Candy	Total
Given Mug	68	8	76		38	38	76
Given Candy Bar	10	78	88		44	44	88
Given Nothing	32	24	56		28	28	56
Total	110	110	220		110	110	220

p.45.a. Complete the table of Expected Counts under the null hypothesis.

p.45.b. Give the rejection region for the chi-square test: Reject H_0 if the test statistic $\geq$ **5.991**

p.45.c. Compute the contribution to the chi-square statistic for the upper left cell of the table (Given Mug / Prefer Mug)

$(68-38)^2 / 38 = 23.68$

p.45.d. Based on p.45.b. and p.45.c., the P-value for the fully computed chi-square statistic is   > 0.05   or   **< 0.05**

p.45.e. Let π_{NM} be the population proportion who had been given nothing who prefer the mug to the candy bar. Compute a 95% Confidence Interval for π_{NM} Note: Use only information from observed values of Condition 3

$$\hat{\pi}_{NM} = \frac{32}{56} = .571 \quad \hat{SE}\left\{\hat{\pi}_{NM}\right\} = \sqrt{\frac{.571(.429)}{56}} = .066 \qquad .571 \pm 1.96(.066) \equiv .571 \pm .130 \equiv (.441, .701)$$

Q.46. A study considered the effect of juvenile criminal offenders being tried as adults. A sample of 2097 youths tried as adults were matched pairwise with 2097 youths tried as juveniles, with matching based on prior criminal behavior. The response was whether or not the youth was re-arrested during a subsequent follow-up period. The data are given below where the counts are the number of the 2097 pairs that fall in that cell of the table. Let π_J be the true probability of a youth tried as a Juvenile to be re-arrested, and π_A be similarly defined for a youth tried as an Adult. Use McNemar's test to test: $H_0 : \pi_J = \pi_A \quad H_A : \pi_J \neq \pi_A$

Adult \ Juvenile	Rearrest	No Rearrest
Rearrest	158	515
No Rearrest	290	1134

p.46.a. Compute the test statistic and rejection region for this test.

Test Statistic: z_{obs} = **7.93**   Rejection Region: $|z_{obs}| \geq$ **1.96**

p.46.b. Which statement best describes the conclusion?

   Tried as Adult has higher probability   Tried as Juvenile has higher Probability   No Difference

Q.47. A study considered mortality among fire-setters and matched controls (non-fire-setters) in Finland. Among a sample of 435 fire-setters, 91died of unnatural causes during the study period. Among 1740 controls, 66 died of unnatural causes during the study period. Compute the sample odds ratio of death by unnatural causes (fire-setters relative to non-fire-setters), and compute a 95% Confidence Interval for the population odds ratio).

$\hat{OR} = 6.710$ 95% CI for OR: **(4.788 , 9.403)**

Q.48. If 2 populations have the same probability of a success, then the Odds Ratio and the Relative Risk for the 2 populations will both be what value?

OR = RR = 1