A screenshot of a computer code

Description automatically generatedA close up of numbers

Description automatically generatedThe Women’s Health Initiative conducted a randomized experiment to see if hormone therapy was helpful for post-menopausal women. The women were randomly assigned to receive the estrogen plus progestin hormone therapy or a placebo. After 5 years, the number of women who developed cancer in each group was determined. *Proportion of women in the Hormone group with Cancer* = 107/(107+8399) = 0.01257936. *The proportion of women in the Placebo group with Cancer =* 88/(88+8014) = 0.1086152. *H0*: p1 – p2 = 0, *Ha*: p1 – p2 ≠ 0. *Test stat X2* = 1.0553. Not enough evidence to conclude the proportion of women in the population who get cancer is different between the two groups. *90% CI for the difference in proportion of women with cancer between the two groups* is (-0.0010, 0.0045). We are 90% confidence the proportion of women in the population taking hormone replacement therapy who get cancer is between 0.0010 less than to 0.0045 more than the proportion of women in the population taking a placebo who get cancer. *90% CI for relative risk of developing cancer when taking hormone therapy*: We are 90% confident the proportion of women taking hormone therapy and who get cancer is between 0.9153 and 1.4654 times the proportion of women taking a placebo and who get cancer. *90% CI for the odds ratio of developing cancer when taking hormone therapy:* We are 90% confident the odds of a women from this population who takes hormone therapy developing cancer is between 0.9144 to 1.4721 times the odds of a woman from this population who takes a placebo developing cancer.

A white paper with black text

Description automatically generatedA screenshot of a computer code

Description automatically generatedA number of numbers on a white background

Description automatically generatedIn the popular movie from 1997 about the Titanic, first class passengers appeared to be able to get to the life boats, while third class passengers were kept away. Is there truth to this appearance? Was the proportion of passengers rescued different for each class of ticket? *H0:* p1 = p2 = p3 = p4, *Ha*: at least 1 pi is different. Test statistic X2 = 187.79. Extremely strong evidence at least one of the ticket classes has a different proportion of rescued passengers. *Pairwise hypothesis tests have following p-values:* all pairs of proportions have very small p-values except for the pair of third class and crew. This indicates strong evidence the proportion of rescued is different between 1st and 2nd class, between 1st and 3rd class, between 2nd and 3rd class, and between the crew and 1st class, and the crew and 2nd class. Yes, the proportion of passengers rescued differed among the ticket classes.

A math equation with black text

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generatedA white background with black text

Description automatically generatedIn 1996, in the General Social Survey of 1,895 adults in the United States conducted by the National Opinion Research Center, respondents were asked about their attitudes towards premarital sex. The question asked was When is premarital sex wrong? and the possible answers were Always Wrong, Almost Always Wrong, Sometimes Wrong, Not Wrong at All. People’s attitudes about social behaviors tend to be related to other more general background variables about the individual. Among other questions, respondents were asked about one such variable, their religious affiliation. *Conditional distribution of attitude towards premarital sex given religious affiliation is Catholic:* 1st row. *Conditional distribution of attitude towards premarital sex given religious affiliation is Protestant*: 2nd row. *H0:* the distribution of attitudes towards premarital sex are the same for each religious affiliation. *Ha*: at least one of the religious affiliations has a different distribution of attitudes towards premarital sex. *Test statistic X2* = 157.02. Strong evidence at least one of the religious affiliations has a different distribution of attitudes towards premarital sex.

A black background with white text

Description automatically generatedThe following table of counts was obtained from a random sample of 1397 respondents from the population of adults in USA, 1982. Each respondent was cross-classified with respect to opinions regarding gun registration as a part of comprehensive gun control legislation (Favor or Oppose) and imposing the death penalty on adults convicted of certain violent acts (Favor or Oppose). *Mosaic plot interpretation:* There is a difference in the height of the vertical segments within each combination of opinions on gun registration and the death penalty. This indicates a variation in the proportion of respondents favoring or opposing gun registration across different opinions on the death penalty. *Conduct a test of independence for the two variables. Use α = 0.*05 *H0*: There is no association between opinions on gun registration and the death penalty. *Ha*: There is an association between opinions on gun registration and the death penalty. *Test Statistic*: 5.1503. *p-value*: 0.02324. We have moderate evidence that opinion about gun registration and death penalty are not independent in the population in 1982. *Calculate an estimate of the φ correlation for the two variables:* -0.0607179

**90% z = 1.645**

**95% z = 1.96**

A black background with white text

Description automatically generatedThe operations manager of a company that manufactures tires wants to determine whether there are any differences in the quality of workmanship among the three daily shifts. She randomly selects 496 tires and carefully inspects them. Each tire is either classified as perfect, satisfactory, or defective, and the shift that produced it is also recorded. The two categorical variables of interest are: shift and condition of the tire produced. *Mosaic plot interpretation:* The amount of perfect condition tires decreases as the shift increases. The amount of defective tires increases as shift increases. *Why is a test of independence OK to use in this case?* This test is ok because the tires are selected randomly and we don’t know how many will be in each group beforehand. So we should use test of independence. *How would we need to change the data collection to use test of multinomial equality?* We would need to adjust group size before sampling. We could sample 150 tires from each shift. *Conduct a test of independence for the two variables. Use α = 0.05* *H0*: Tire condition and shift are independent. *Ha*: Tire condition and shift are not independent. *Test Statistic*: 8.646696. *P-value*: 0.07056327. There is no significant evidence of an association between shift and the condition of the tire. *Calculate an estimate of Cramer’s V for this contingency table:* 0.09336181.

A number on a black background

Description automatically generatedA screen shot of a number

Description automatically generatedIn a study of disparities between mother and child perceptions of ability, 6th graders were asked to rate their own academic ability as Below Average, Average, or Above Average. The mother of each child was also asked to rate the childs academic ability as either Below Average, Average, or Above Average. *Mosaic plot interpretation:* The vertical segments differ noticeably among the classes, suggesting a potential association between child and mother perceptions, indicating that the two variables are not independent. *Calculate a test of independence for the two variables. Use α = 0.05* *H0*: There is no association between child and mother perceptions of academic ability. *Ha*: There is an association between child and mother perceptions of academic ability. *Test Statistic*: 17.21778. *P-value*: 0.001753395. There is significant evidence of an association between child and mother perceptions of academic ability. *Goodman-Krustal Gamma: More than Random Agreement?* Given that the confidence interval includes zero, the Goodman-Krustal Gamma suggests that there is no conclusive evidence of more than random agreement between mother and child perceptions of ability.

In 1990, a random sample of 1,600 adults in the US asked whether or not they approved of the job George H.W. Bush was doing as president. One month later, the same people were asked the same question. *Calculate an estimate of the proportion of adults in the U.S. who approved of the job President Bush was doing for each time period:* Proportion of Approval for First Rating = 944/1600 = 0.59 Proportion of Approval for Second Rating = 880/1600 = 0.55. *Why are the two ratings not independent from each other?* The estimates aren’t independent because they are based on the same group of individuals surveyed twice. The responses in the second rating are likely influenced by the responses in the first rating, indicating a relationship between the two time periods’ approval estimates. *Use McNemar’s test to determine if the proportion of adults in the U.S. who approved of the job President Bush was doing is different between the two time periods. H0:* The proportion of adults in the US who approved of the job President Bush was doing is the same between the two time periods. *H­a­:* The proportion of adults in the US who approved of the job President Bush was doing is different between the two time periods. *Test Statistic*: 17.356. *P-value*: 3.099e-05. There is sufficient evidence to conclude that the proportion of adults in the US who approved of the job President Bush was doing differs between the two time periods. We are *95% confident* that the true difference in the proportion of adults in the US who approved of the job President Bush was doing between the two time periods lies between 0.0213 and 0.0587. This indicates that there is a statistically significant difference in the approval rates between the two time periods, with a higher proportion of approval during the first time period compared to the second time period.

A black background with white text

Description automatically generatedDuring the 1970’s, 80’s and 90’s, Gene Siskel and Roger Ebert worked as film critics for the Chicago Tribune and Chicago Sun Times, respectively. In their syndicated TV show, At the Movies, they presented their reviews of recently released movies using a Thumbs Up/Thumbs Down system. This system also allowed them to give a mixed review. *Calculate the distribution of reviews for Siskel and the distribution of reviews for Ebert on these 160 movies*. Distribution of Ratings for Siskel = Down: 0.51875, Mixed: 0.20000, Up: 0.28125. Distribution of Ratings for Ebert = Down: 0.5500, Mixed: 0.1875, Up: 0.2625. *Who had more Thumbs Up reviews?* = Siskel. *Who had more Thumbs Down reviews?* = Ebert. *Use the extension to McNemar’s test to determine if there was a difference in the distribution of reviews between the two critics. H0*: There is no difference in the distribution of reviews between the two critics. *Ha:* There is a difference in the distribution of reviews between the two critics. *Test Statistic*: 0.585. *P-value*: 0.746. There is not enough evidence to conclude that there is a difference in the distribution of reviews between the two critics. *On what proportion of the movies did Siskel and Ebert agree on their reviews?* 0.63125. *Calculate Cohen’s kappa*: 0.389. *Calculate weighted kappa using the squared weight function:* 0.458. Based on the calculated Cohen’s Kappa values, both unweighted (0.389) and weighted (0.458), it appears that Siskel and Ebert had a moderate level of agreement on their movie reviews. This suggests that while there was some level of consensus between the two critics, there were also instances of disagreement in their assessments of the movies.

A number on a black background

Description automatically generatedResearchers at a large university are studying the accuracy of computer scored open-ended responses by comparing them to person scored responses. In this example, the computer and a person scored the same 1,011 student responses to the same open-ended question. Each question was scored as either a 1 = minimal understanding of concept, 2 = moderate understanding of concept, or 3 = full understanding of concept. *On what proportion of the scores do the computer and person agree?* 0.8081108. *Cohen’s kappa* = 0.688. *Weighted Cohen’s kappa* = 0.77. Based on the calculated Cohen’s kappa values, both unweighted (0.688) and weighted (0.77), it appears that there is substantial agreement on scores between the computer and the person. These kappa values indicate a strong level of agreement beyond what would be expected by chance alone. So, it can be concluded that there is agreement on scores between the computer and the person.