Chi-Square Test of Independence

Example 8.1 Consider data from the Minnesota Study Survey regarding the opinions of marijuana use for 12th grade students in Fillmore County.



The following data was obtained from the Minnesota Department of Education website for Fillmore County.



*Source*: Minnesota Department of Education; <http://education.state.mn.us/MDE/Learning_Support/Safe_and_Healthy_Learners/Minnesota_Student_Survey/index.html>

Consider the following investigation comparing Males to Females for Grade 12.

Research Question: Do differences exist between Genders in the opinions regarding marijuana use for 12th graders in Fillmore County?

|  |  |
| --- | --- |
| Data in JMP | Analyze > Fit Y by X |

Initial JMP output – again, fix the order of the response variable using Value Ordering



Observed Outcomes (Observed)





Compare Against No Difference Situation (Expected)





Computing the Expected Counts



Test of Significance: Chi –Square Test

|  |  |
| --- | --- |
| Observed Graph | Expected Graph (No Difference) |



|  |  |
| --- | --- |
| JMP Testing Output | <http://www.di-mgt.com.au/chisquare-calculator.html>  |

Example 8.2: Next, we will investigate the possible association between opinions of marijuana use and grade level. Certainly, young children have a belief, typically a strong belief, that drugs are “bad”. The question here centers around the possible shift in opinions from Grade 9 and Grade 12 students from Fillmore County.



Consider the following investigation which compares across grade levels 9 and 12.

Research Question: Are there differences in the opinions regarding marijuana use between Grade 9 and Grade 12 students from Fillmore County?

Writing the appropriate null and alternative hypothesis for your investigation.

Ho: Opinions regarding marijuana are independent of Age (Grade 9 & 12)

HA: Opinions regarding marijuana depend on Age (Grade 9 & 12)

The *observed* data from the study.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level | No Risk | Slight Risk | Moderate Risk | Great Risk | Total |
| Grade 9 | 14 | 2 | 29 | 127 | 172 |
| Grade 12 | 12 | 16 | 29 | 63 | 120 |
| Total | 26 | 18 | 58 | 190 | 292 |

The analysis in JMP

|  |  |
| --- | --- |
| Data in JMP | Mosaic Plot |

Questions:

1. How do the trends differ between Grade 9 and Grade 12? Why might these trends be expected? Discuss.
2. What is the likelihood of a No Risk outcome from a Grade 9 student? How about a Grade 12 student? What might this imply about the age in which a student makes the decision to believe that smoking marijuana once or twice a week is of No Risk? Explain.

|  |  |
| --- | --- |
| Observed Outcomes | Expected Outcomes |

Testing Output





Questions:

1. Make the appropirate statistical decision.

Decision: If the p-value < 0.05, then data is said to support the research question.

* Data supports research question
* Data does not support research question
1. Write a final conclusion that would be appropriate for the DARE administrator from Fillmore County.

A Test for Trend: Cochran Armitage Trend Test



Comment: A **Cochran Armitage Trend Test** can be used to verify whether or not the trend see in the above mosaic plot is likely to appear on repeated sampling from the population. Search Wikipedia for additional information regarding this test.



The one-sided p-value for the Cochran Armitage Trend test is less than 0.0005 (i.e. Prob > Z value).

Question:

1. What is the practical interpretation of this p-value? Write a statement regarding this test that would be useful for a DARE officer in Fillmore County.

Example 8.3 Consider the following data presented in the HIV / AIDS Surveillance Report 2013 on the Minnesota Department of Health web site. The information presented here is comparing the mode of exposure for diagnosis across race/ethnicity.

*Source*: Minnesota Department Health; Link: <http://www.health.state.mn.us/divs/idepc/diseases/hiv/stats/index.html>

|  |  |
| --- | --- |
|  | How to read the White Males graph* A total of 3447 white males are living with HIV or AIDS in MN
* 86% (about 2960) were exposed due to MSM (men having sex with men)
* 2% were exposed through heterosexual contact
* About 11% were exposed from IDU (Injecting Drug Use) or MSM/IDU combination
 |
|  |  |
|  |  |

Questions

1. Most of the above pie charts contain a large portion of red. What does this mean?
2. The African-born pie chart is strikingly different. How might the HIV / AIDS crisis in several African countries contribute to this difference in trends?
3. What would be the likely outcome of a Chi-Square Test of Independence for testing the relationship between mode of exposure and race? Explain your answer.

I obtained the counts based on the percentages presented above for White, African-American, Hispanic, African-born, and Asian and put them into the following table.



A mosaic plot that shows the trends across race.



Questions

1. Is it easier or more difficult to make comparisons across race using a mosaic plot or individual pie charts? Explain.
2. Does this plot suggest that a relationship exists between the mode of exposure and race? If so, briefly discuss this relationship.
3. The p-value for the Chi-Square Test of Independence is *very* small -- much smaller than 0.05. What is the practical interpretation of this p-value?
4. Suppose you are in charge of creating a pamphlet for the Minnesota Department of Health. Our investigation here suggests that an HIV/AIDS pamphlet appropriate for a clinic in certain area (say Winona) may need to be changed for a clinic in a different area (say a Somalia Community Health Clinic). Carefully explain why the p-value from our analysis supports this conclusion. Also, what is *gained* by using a p-value to help us make this decision? Why should we even look at a p-value in an investigation like this? Discuss.
5. Is it unethical to investigate data in the manner in which we have here? That is, should HIV / AIDS surveillance reports include information regarding race? Discuss.

Example 8.4 Consider a study done by Timonen, et al. (2001) that investigating the risk of suicide and it possible relationship with income level. This study was done in Finland because it has one of the highest death rates from suicide and treatment methods are similar across income levels (i.e. most of the population is treated similarly in public hospitals, regardless of socioeconomic status). Consider the following data from their study.

|  |  |
| --- | --- |
| Violent\* vs. Non Violent Across Occupations\*Violent Includes: Hanging, drowning, shooting, wrist cutting, jumping from a height | Reason for Admission into TreatmentAcross Occupations |

Source: Timonen, M., Viilo, K., Hakko, H., Väisänen, E., Räsänen, P., and Särkioja, T. (2001). “*Risk of suicide related to income level in mental illness: Psychiatric disorders are more severe among suicide victims of higher occupational level*.” British Medical Journal, Vol. 323 (7306) : 232.

Consider the following research questions.

1. Is the likelihood of a violent suicide dependent on type of occupation?
2. Students tend to be younger and the method of suicide for this age group is likely to be different than other age groups. Remove the student data from the dataset and redo the first research question.

Is the likelihood of a violent suicide dependent on type of occupation (with students removed)?

1. Is the likelihood for the reason of admission vary across occupations?

Complete the following for each research question.

* Obtain a p-value from a Chi-Square Test of Independence.
* Answer each research question using your p-value.
* Obtain the appropriate mosaic plot for each research question that can be used to make visual comparisons across the occupations.
* Write two or three statements, appropriate for a mental health clinic, that communicate the practical interpretation of the outcomes from each research question.

Example 8.5 The Minnesota Department of Corrections often creates reports for the Minnesota State Legislature to identify the importance of funding programs that are known to improve public safety. The following data was presented in one such report. The following question and supporting data were provided on p17 of their report.

Source: <http://www.doc.state.mn.us>



Research Question: Is completion of sex offender treatment while on probation associated with a decrease in recidivism?

Complete the following for your investigation.

* Put this data into JMP and complete a Chi-Square Test of Independence.
* Obtain a mosaic plot for this data.

Questions

1. What is the p-value from your test? In the context of this problem, what is the practical interpretation of this p-value?
2. What is *gained* by running a Chi-Square Test of Independence? Why would somebody actually do this test in an investigation of this type?
3. Notice, their research question is actually one-sided (i.e. interest in a decrease in recidivism), but the Chi-Square Test of Independence is a two-sided test. Explain why a Cochran Armitage Trendtype test (see page 8) may actually be a better test for this investigation.