Matt Isaac

A01515095

Biostatistics

Homework 3 – Due Monday, 26 Feb

1. Address the research questions below.
   1. Are physicians detecting depression any diﬀerently than surveys? (For this question, ignore how physicians detected depression, and assume each physician assessed each patient for depression.)
      1. Two-way table of category counts:

| **Table of SurveyDep by PhysicianDep** | | | |
| --- | --- | --- | --- |
| **SurveyDep** | **PhysicianDep** | | |
| **Frequency Row Pct** | **1** | **2** | **Total** |
| **1** | 9 14.06 | 55 85.94 | 64 |
| **2** | 0 0.00 | 47 100.00 | 47 |
| **Total** | 9 | 102 | 111 |

* + 1. Name of the appropriate test: Two-sample Binomial Test
    2. Null and alternative hypothesis:

H0: pphysician = psurvey

HA: pphysican ≠ psurvey

* + 1. P-value from the test: 0.0073

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 7.1926 | 0.0073 |

* + 1. Conclusion in context of research: Since the p-value is < 0.05, we reject H0. This implies that there is a significant difference between the physicians detecting depression and the survey detecting depression.
  1. Are there diﬀerences between the survey-based rates of depression (SurveyDep) in the seven diﬀerent diagnostic groups?
     1. Two-way table of category counts:

| **Table of Cardiac\_dx by SurveyDep** | | | |
| --- | --- | --- | --- |
| **Cardiac\_dx** | **SurveyDep** | | |
| **Frequency Col Pct** | **1** | **2** | **Total** |
| **1** | 19 29.69 | 18 38.30 | 37 |
| **2** | 15 23.44 | 5 10.64 | 20 |
| **3** | 14 21.88 | 13 27.66 | 27 |
| **4** | 5 7.81 | 4 8.51 | 9 |
| **5** | 6 9.38 | 1 2.13 | 7 |
| **6** | 3 4.69 | 4 8.51 | 7 |
| **7** | 2 3.13 | 2 4.26 | 4 |
| **Total** | 64 | 47 | 111 |

* + 1. Name of the appropriate test: Chi-square General Association test (Could also use Row Mean Scores Differ test – they will be equivalent).
    2. Null and alternative hypothesis:

H0: There is no association between Cardiac\_dx group and survey-based rates of depression.

HA: The distribution over Survey-based rates of depression differs between the different levels of Cardiac\_dx. (i.e. some association/dependence exists).

* + 1. P-value from the test: 0.3821 (as seen in table below)

| **Cochran-Mantel-Haenszel Statistics (Based on Table Scores)** | | | | |
| --- | --- | --- | --- | --- |
| **Statistic** | **Alternative Hypothesis** | **DF** | **Value** | **Prob** |
| **1** | **Nonzero Correlation** | 1 | 0.0046 | 0.9459 |
| **2** | **Row Mean Scores Differ** | 6 | 6.3788 | 0.3821 |
| **3** | **General Association** | 6 | 6.3788 | 0.3821 |

* + 1. Conclusion in context of research: Since our p-value is > 0.05, we fail to reject H0. We do not have reason to believe that there is a linear association between Cardiac\_dx group and the survey-based rates of depression.
  1. Are there diﬀerences between the survey-based rates of depression in male and female patients?
     1. Two-way table of category counts:

| 1. **Table of Gender by SurveyDep** | | | |
| --- | --- | --- | --- |
| **Gender** | **SurveyDep** | | |
| **Frequency Row Pct** | **1** | **2** | **Total** |
| **1** | 34 55.74 | 27 44.26 | 61 |
| **2** | 30 60.00 | 20 40.00 | 50 |
| **Total** | 64 | 47 | 111 |

* + 1. Name of the appropriate test: 2-sample Binomial test
    2. Null and alternative hypothesis:

H0: pmale = pfemale

HA: pmale ≠ pfemale

* + 1. P-value from the test: 0.6511 (see table below)

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 0.2045 | 0.6511 |

* + 1. Conclusion in context of research: Since our p-value is > 0.05, we fail to reject H0. That is, we have no evidence that depression rates are dependent on gender.

1. Are there diﬀerences between the survey-based rates of depression in patients who are married, single, widowed, or divorced?
   * 1. Two-way table of category counts:

| **Table of MaritalStatus by SurveyDep** | | | |
| --- | --- | --- | --- |
| **MaritalStatus** | **SurveyDep** | | |
| **Frequency Col Pct** | **1** | **2** | **Total** |
| **1** | 40 62.50 | 36 76.60 | 76 |
| **2** | 4 6.25 | 1 2.13 | 5 |
| **3** | 12 18.75 | 9 19.15 | 21 |
| **4** | 8 12.50 | 1 2.13 | 9 |
| **Total** | 64 | 47 | 111 |

* + 1. Name of the appropriate test: Chi-square Row Mean Scores Differ test (assuming MaritalStatus to be nominal and the 2-category Survey-based depression variable to be ordinal).
    2. Null and alternative hypothesis:

H0: There is no association between marital status and survey-based rates of depression.

HA: Mean scores differ among different marital status groups.

* + 1. P-value from the test: 0.1474 (see table below)

| **Cochran-Mantel-Haenszel Statistics (Based on Table Scores)** | | | | |
| --- | --- | --- | --- | --- |
| **Statistic** | **Alternative Hypothesis** | **DF** | **Value** | **Prob** |
| **1** | **Nonzero Correlation** | 1 | 2.9304 | 0.0869 |
| **2** | **Row Mean Scores Differ** | 3 | 5.3580 | 0.1474 |
| **3** | **General Association** | 3 | 5.3580 | 0.1474 |

* + 1. Conclusion in context of research: Since our p-value is greater than 0.05, we fail to reject H0. We have no evidence that there is any association between marital status and survey-based rates of depression.

1. Are there diﬀerences between the survey-based rates of depression in patients that had a prior cardiac history versus those with a new diagnosis?
   * 1. Two-way table of category counts:

| **Table of PriorCardHx by SurveyDep** | | | |
| --- | --- | --- | --- |
| **PriorCardHx** | **SurveyDep** | | |
| **Frequency Row Pct** | **1** | **2** | **Total** |
| **1** | 41 62.12 | 25 37.88 | 66 |
| **2** | 23 51.11 | 22 48.89 | 45 |
| **Total** | 64 | 47 | 111 |

* + 1. Name of the appropriate test: 2-sample Binomial test
    2. Null and alternative hypothesis:

H0: pPriorCardHx = pNoPriorCardHx

HA: pPriorCardHx ≠ pNoPriorCardHx

* + 1. P-value from the test: 0.2491

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 1.3286 | 0.2491 |

* + 1. Conclusion in context of research: We fail to reject H0 because our p-value is > 0.05. This indicates that we have found evidence that Depression diagnosis from the survey is independent of PriorCardHx.

2. In class (Handout #7) we discussed how the two-sample binomial test statistic Z and the chi-square test statistic χ2 for testing independence in a two-by-two table are related. Clearly demonstrate this relationship in general (Z2 = χ2), using the generic category counts a, b, c, and d as in the following table:

* 1. Report the test statistic Z (in terms of category counts a, b, c, and d only).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Response 1 | Response 2 |  |
| Group 1 | a  (or X1) | b  (or X1 – n1) | a + b  (or n1) |
| Group 2 | c  (or X2) | d  (or X2 – n2) | c + d  (or n2) |

From handout 7,

Let , estimated by .

Also, from handout 7, ,

where p is estimated by

Thus,

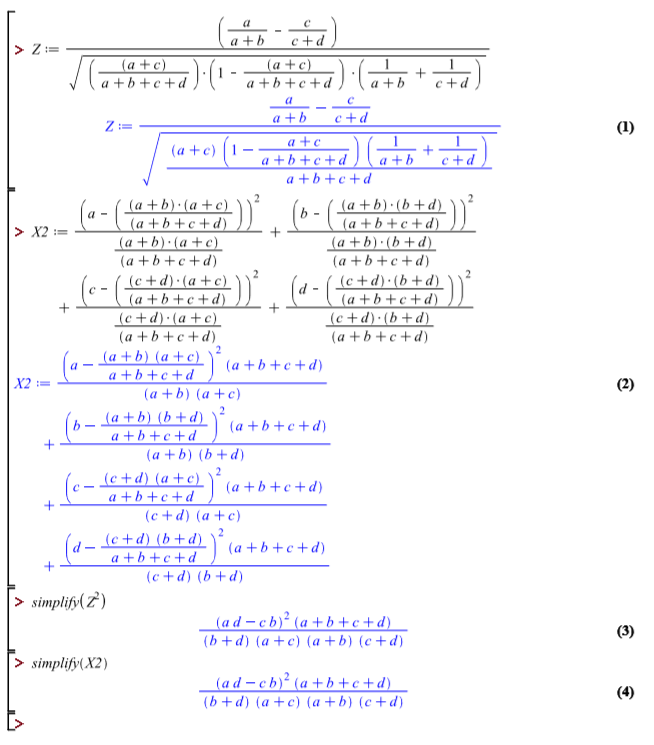
* 1. Report the test statistic χ2 (in terms of category counts a, b, c, and d only).

Where .

Thus,

* 1. Show that Z2 = χ2.

The Maple code on the page below shows the equality of and .



Appendix: SAS code

/\* Generated Code (IMPORT) \*/

/\* Source File: depression.csv \*/

/\* Source Path: /home/mattisaac0/BioStatistics \*/

/\* Code generated on: 2/21/18, 9:42 PM \*/

%web\_drop\_table(WORK.depression);

FILENAME REFFILE '/home/mattisaac0/BioStatistics/depression.csv';

PROC IMPORT DATAFILE=REFFILE

DBMS=CSV

OUT=WORK.depression;

GETNAMES=YES;

RUN;

\* proc print data = depression;

%web\_open\_table(WORK.depression);

proc print data = depression;

run;

/\* Code for problem 1(a)\*/

\* Frequency Table;

proc freq data = depression;

tables SurveyDep \* PhysicianDep / chisq nopercent nocol;

run;

/\* Code for problem 1(b)\*/

proc freq data = depression;

tables Cardiac\_dx \* SurveyDep / chisq nopercent norow cmh;

run;

/\* Code for problem 1(c)\*/

proc freq data = depression;

tables gender \* SurveyDep / chisq nopercent nocol;

run;

/\* Code for problem 1(d)\*/

proc freq data = depression;

tables MaritalStatus \* SurveyDep / chisq nopercent norow cmh;

run;

/\* Code for problem 1(e)\*/

proc freq data = depression;

tables PriorCardHx \* SurveyDep / chisq nopercent nocol;

run;