**Lab 11. Categorical Data**

**MSDS 6370**

**Objective:**

* For the student to learn more about examining relationships with survey data.

1. In a recent Gallup Poll of 1,011 randomly selected investors who have more than $10,000 of funds to invest, it was found that 51% thought that the Federal Government was going to raise interest rates this month.
2. Find and interpret a 95% confidence interval for the true proportion of these investors who feel that interest rates will increase this month.

(47.9%, 54.1%)

1. Find the margin of error for the above interval.

MOE = 0.03082

1. Let’s say that Gallop wanted to obtain a margin of error of only 1% (with 95% confidence) … how many more people would they need to sample?

N = 9604

1. Consider the HRS SAS data set.
2. Get the summary table (4 by 4) for the education level vs the race counts. (The variable name for the education level is *edcat* and for the race category it is *racecat*) You can use the PROC TABULATE.

|  | **Race Categories 1=Hispanic 2=White 3=Black 4=Other** | | | |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** |
| **N** | **N** | **N** | **N** |
| **Education 1=0-11 2=12 3=13-15 4=16+** | 1017 | 2362 | 991 | 100 |
| **1** |
| **2** | 338 | 4945 | 743 | 108 |
| **3** | 221 | 3044 | 484 | 93 |
| **4** | 117 | 3330 | 333 | 130 |

1. Carry out a Chi-square test of the following hypothesis:

*H*0: Education level is independent of race category.

*H*a: Education level is not independent of race category.

|  |  |
| --- | --- |
| **Pearson Chi-Square** | 1532.2855 |

1. Write a sentence interpreting your result.

In this case, the results of the test of independence are the same; we conclude education level and race are not independent.

1. Now reanalyze the data above, using the Rao-Scott Chi-square test to determine if education level and race category are independent or not. Include your SAS code and output, as well as a sentence interpreting your output.

|  |  |
| --- | --- |
| **Rao-Scott Chi-Square** | 706.3498 |
| **DF** | 9 |
| **Pr > ChiSq** | <.0001 |

In this case, the results of the test of independence are the same; we conclude education level and race are not independent.

proc surveyfreq data=hisp;

weight KWGTR;

strata stratum;

cluster SECU;

table edcat\*racecat/ chisq;

run;