CHRISTOPHER PRENER, PH.D.
FALL, 2016

WEEK 16
LECTURE 17

QUANTITATIVE ANALYSIS

CHI-SQUARED

AGENDA

1. Follow-up
2. Chi-squared Logic
3. Cross-tabulations
4. Chi-squared Test in Stata

1 FOLLOW-UP

2 CHI-SQUARED LOGIC

PROBLEMS WITH THE MEAN

THE MEAN OF A NOMINAL VARIABLE IS A "USELESS" STATISTIC

value	label	frequency
1	White	86
2	African American	43
3	Hispanic	24
4	Asian	18
mean		79

PROBLEMS WITH THE MEAN

THE MEAN OF A NOMINAL VARIABLE IS A "USELESS" STATISTIC

value	label	frequency
4	White	86
1	African American	43
2	Hispanic	24
3	Asian	18
mean		122.25

PROBLEMS WITH THE MEAN

THE MEAN OF A NOMINAL VARIABLE IS A "USELESS" STATISTIC

value	label	frequency
3	White	86
4	African American	43
1	Hispanic	24
2	Asian	18
mean		122.5

PROBLEMS WITH THE MEAN

THE MEAN OF A NOMINAL VARIABLE IS A "USELESS" STATISTIC

value	label	frequency
2	White	86
3	African American	43
4	Hispanic	24
1	Asian	18
mean		103.75

PROBLEMS WITH THE MEAN

THE MEAN OF A NOMINAL VARIABLE IS A "USELESS" STATISTIC

ALL OF THE STATISTICS WE'VE COVERED (S^2, S, COV, R, T) RELY ON DEVIANCE (X_I – X–BAR) AS A FOUNDATIONAL ELEMENT

PROBLEMS WITH THE MEAN

CHI^2 (X^2) ALLOWS FOR THE COMPARISON OF NOMINAL DATA WITHOUT RELYING ON MEANS TO UNDERSTAND THE RELATIONSHIP

CHI^2 COMPARES THE OBSERVED FREQUENCIES (F_O) TO THE EXPECTED FREQUENCIES (F_E) OF THE CATEGORIES IN BOTH THE X AND Y VARIABLES

3 CROSS-TABULATIONS

CROSS TABS IN STATA

```
tabulate xvar yvar [, row col cell]
```

```
. tabulate foreign rep78
```

```
           |                Repair Record 1978
  Car type |         1          2          3          4          5 |     Total
-----------+-------------------------------------------------------+----------
  Domestic |         2          8         27          9          2 |        48
   Foreign |         0          0          3          9          9 |        21
-----------+-------------------------------------------------------+----------
     Total |         2          8         30         18         11 |        69
```

CROSS TABS IN STATA

```
tabulate xvar yvar [, row col cell]
```

```
. tabulate rep78 foreign
```

Repair Record 1978	Car type	
	Domestic	Foreign	Total
1	2	0	2
2	8	0	8
3	27	3	30
4	9	9	18
5	2	9	11
Total	48	21	69

CROSS TABS IN STATA

```
tabulate xvar yvar [, row col cell]
```

```
. tabulate rep78 foreign
```

Repair Record 1978	Car type	
	Domestic	Foreign	Total
1	2	0	2
2	8	0	8
3	27	3	30
4	9	9	18
5	2	9	11
Total	48	21	69

CROSS TABS IN STATA

```
tabulate xvar yvar [, row col cell]
```

```
. tabulate rep78 foreign, row
```

```
+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+
```

Repair Record 1978	Car type: Domestic	Foreign	Total
1	2	0	2
	100.00	0.00	100.00
2	8	0	8
	100.00	0.00	100.00
3	27	3	30
	90.00	10.00	100.00
4	9	9	18
	50.00	50.00	100.00
5	2	9	11
	18.18	81.82	100.00
Total	48	21	69
	69.57	30.43	100.00

CROSS TABS IN STATA

```
tabulate xvar yvar [, row col cell]
```

```
. tabulate rep78 foreign, col
```

```
+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+
```

Repair Record 1978	Car type: Domestic	Foreign	Total
1	2	0	2
	4.17	0.00	2.90
2	8	0	8
	16.67	0.00	11.59
3	27	3	30
	56.25	14.29	43.48
4	9	9	18
	18.75	42.86	26.09
5	2	9	11
	4.17	42.86	15.94
Total	48	21	69
	100.00	100.00	100.00

CROSS TABS IN STATA

```
tabulate xvar yvar [, row col cell]
```

```
. tabulate rep78 foreign, cell
```

```
+----------------+
| Key            |
|----------------|
|    frequency   |
| cell percentage |
+----------------+
```

Repair Record 1978	Car type: Domestic	Foreign	Total
1	2	0	2
	2.90	0.00	2.90
2	8	0	8
	11.59	0.00	11.59
3	27	3	30
	39.13	4.35	43.48
4	9	9	18
	13.04	13.04	26.09
5	2	9	11
	2.90	13.04	15.94
Total	48	21	69
	69.57	30.43	100.00

4 CHI-SQUARED TEST IN STATA

HYPOTHESES

- H0 = there is no difference in the variation between *x* and *y*

- H1 = there is a difference in the variation between *x* and *y*

ASSUMPTIONS

- nominal data
- independence
- sufficient sample size
- less than 20% of cells can have an expected count of less than 5 cases

EXPECTED COUNTS

```
tabulate xvar yvar [, expected chi2 exact V]
```

```
. tabulate rep78 foreign, expected
```

```
+--------------------+
| Key                |
|--------------------|
|     frequency      |
| expected frequency |
+--------------------+
```

Repair Record 1978	Car type: Domestic	Foreign	Total
1	2	0	2
	1.4	0.6	2.0
2	8	0	8
	5.6	2.4	8.0
3	27	3	30
	20.9	9.1	30.0
4	9	9	18
	12.5	5.5	18.0
5	2	9	11
	7.7	3.3	11.0
Total	48	21	69
	48.0	21.0	69.0

CALCULATING CHI^2

```
. tabulate rep78 foreign, chi2

    Repair |
    Record |       Car type
      1978 |  Domestic    Foreign |     Total
-----------+----------------------+----------
         1 |         2           0 |         2
         2 |         8           0 |         8
         3 |        27           3 |        30
         4 |         9           9 |        18
         5 |         2           9 |        11
-----------+----------------------+----------
     Total |        48          21 |        69

          Pearson chi2(4) =  27.2640   Pr = 0.000
```

CALCULATING CHI^2

```
. tabulate rep78 foreign, chi2

   Repair |
   Record |       Car type
     1978 |  Domestic    Foreign |     Total
----------+----------------------+----------
        1 |         2          0 |         2
        2 |         8          0 |         8
        3 |        27          3 |        30
        4 |         9          9 |        18
        5 |         2          9 |        11
----------+----------------------+----------
    Total |        48         21 |        69

          Pearson chi2(4) =  27.2640   Pr = 0.000
```

The results of the Chi2 test ($chi^2 = 27.264$, $p < .001$) suggest that there is a relationship between origin and repair record. Foreign cars are expected to have higher repair record values than domestic cars.

EXPECTED COUNTS

```
tabulate xvar yvar [, expected chi2 exact V]
```

```
. tabulate rep78 foreign, expected
```

```
+--------------------+
| Key                |
|--------------------|
|     frequency      |
| expected frequency |
+--------------------+
```

Repair Record 1978	Car type: Domestic	Foreign	Total
1	2	0	2
	1.4	0.6	2.0
2	8	0	8
	5.6	2.4	8.0
3	27	3	30
	20.9	9.1	30.0
4	9	9	18
	12.5	5.5	18.0
5	2	9	11
	7.7	3.3	11.0
Total	48	21	69
	48.0	21.0	69.0

FISCHER'S EXACT TEST

```
. tabulate rep78 foreign, exact

Enumerating sample-space combinations:
stage 5:  enumerations = 1
stage 4:  enumerations = 3
stage 3:  enumerations = 24
stage 2:  enumerations = 203
stage 1:  enumerations = 0
```

Repair Record 1978	Car type: Domestic	Foreign	Total
1	2	0	2
2	8	0	8
3	27	3	30
4	9	9	18
5	2	9	11
Total	48	21	69

Fisher's exact = 0.000

CRAMER'S V

```
. tabulate rep78 foreign, chi2 exact V

[output omitted]

    Repair |
    Record |       Car type
      1978 |  Domestic    Foreign |     Total
-----------+----------------------+----------
         1 |         2           0 |         2
         2 |         8           0 |         8
         3 |        27           3 |        30
         4 |         9           9 |        18
         5 |         2           9 |        11
-----------+----------------------+----------
     Total |        48          21 |        69

          Pearson chi2(4) =  27.2640   Pr = 0.000
               Cramér's V =   0.6286
           Fisher's exact =                 0.000
```

SMALL = 0.1
MEDIUM = 0.3
LARGE = 0.5

DOCUMENT DETAILS

Document produced by Christopher Prener, Ph.D for the Saint Louis University course SOC 5050: QUANTITATIVE ANALYSIS - APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository README.md file for additional details.