

Running a Chi-Square Test of Independence

(Data Analysis Tools Week 2 Assignment)

Expected Activities

- Run a Chi-Square Test of Independence.
- Need to analyze and interpret post hoc paired comparisons in instances where original statistical test was significant, and were examining more than two groups (i.e. more than two levels of a categorical, explanatory variable).
- Submit syntax used to run a Chi-Square Test (copied and pasted from your program) along with corresponding output and a few sentences of interpretation.

SAS Program

```
LIBNAME mydata "/courses/d1406ae5ba27fe300 " ACCESS=readonly;

DATA new;
    SET mydata.gapminder;
    KEEP country lifeexpectancy urbanrate urban le;
    LABEL lifeexpectancy="Life Expectancy";
    LABEL urbanrate="Urbanisation Rate";
    LABEL urban="Urbanisation Groups";
    LABEL le="Life Expectancy Groups";

    /* Delete records with missing data */
    IF urbanrate=. THEN
        delete;
    IF lifeexpectancy=. THEN
        delete;

    /* Data Management for variable urbanrate */
    IF urbanrate < 50 THEN
        urban=1;
    IF urbanrate >=50 AND urbanrate < 75 THEN
        urban=2;
    IF urbanrate >=75 THEN
        urban=3;

    /* Data Management for variable lifeexpectancy */
    IF lifeexpectancy < 70 THEN
        le="Less than 70 Yrs";
    IF lifeexpectancy >=70 THEN
        le="70Yrs or more";

PROC SORT;
    BY country;
```

```
PROC FREQ;  
    TABLES le*urban/CHISQ;  
  
DATA COMPARISON1;  
    SET NEW;  
    IF (urban=1) OR (urban=2);  
  
PROC SORT;  
    BY country;  
  
PROC FREQ;  
    TABLES le*urban/CHISQ;  
    Title "Comparison 1";  
  
DATA COMPARISON2;  
    SET NEW;  
    IF (urban=1) OR (urban=3);  
  
PROC SORT;  
    BY country;  
  
PROC FREQ;  
    TABLES le*urban/CHISQ;  
    Title "Comparison 2";  
  
DATA COMPARISON3;  
    SET NEW;  
    IF (urban=2) OR (urban=3);  
  
PROC SORT;  
    BY country;  
  
PROC FREQ;  
    TABLES le*urban/CHISQ;  
    Title "Comparison 3";  
  
RUN;
```

Output

Chi-Square Tests

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of le by urban				
	le(Life Expectancy Groups)	urban(Urbanisation Groups)			
		1	2	3	Total
70Yrs or more	21	52	39	112	
	11.17	27.66	20.74	59.57	
	18.75	46.43	34.82		
	27.63	73.24	95.12		
Less than 70 Yrs	55	19	2	76	
	29.26	10.11	1.06	40.43	
	72.37	25.00	2.63		
	72.37	26.76	4.88		
Total	76	71	41	188	
	40.43	37.77	21.81	100.00	

Statistics for Table of le by urban

Statistic	DF	Value	Prob
Chi-Square	2	59.2165	<.0001
Likelihood Ratio Chi-Square	2	65.6271	<.0001
Mantel-Haenszel Chi-Square	1	56.4253	<.0001
Phi Coefficient		0.5612	
Contingency Coefficient		0.4894	
Cramer's V		0.5612	

Sample Size = 188

Post hoc Chi-Square Tests

Comparison 1

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of le by urban			
le(Life Expectancy Groups)	urban(Urbanisation Groups)		
	1	2	Total
70Yrs or more	21 14.29 28.77 27.63	52 35.37 71.23 73.24	73 49.66
Less than 70 Yrs	55 37.41 74.32 72.37	19 12.93 25.68 26.76	74 50.34
Total	76 51.70	71 48.30	147 100.00

Statistics for Table of le by urban

Statistic	DF	Value	Prob
Chi-Square	1	30.5432	<.0001
Likelihood Ratio Chi-Square	1	31.7011	<.0001
Continuity Adj. Chi-Square	1	28.7460	<.0001
Mantel-Haenszel Chi-Square	1	30.3354	<.0001
Phi Coefficient		-0.4558	
Contingency Coefficient		0.4148	
Cramer's V		-0.4558	

Fisher's Exact Test

Cell (1,1) Frequency (F)	21
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1.0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 147

Comparison 2

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of le by urban			
le(Life Expectancy Groups)	urban(Urbanisation Groups)		
	1	3	Total
70Yrs or more	21 17.95 35.00 27.63	39 33.33 65.00 95.12	60 51.28
Less than 70 Yrs	55 47.01 96.49 72.37	2 1.71 3.51 4.88	57 48.72
Total	76 64.96	41 35.04	117 100.00

Statistics for Table of le by urban

Statistic	DF	Value	Prob
Chi-Square	1	48.5558	<.0001
Likelihood Ratio Chi-Square	1	56.5421	<.0001
Continuity Adj. Chi-Square	1	45.8920	<.0001
Mantel-Haenszel Chi-Square	1	48.1408	<.0001
Phi Coefficient		-0.6442	
Contingency Coefficient		0.5416	
Cramer's V		-0.6442	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	21
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1.0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 117

Comparison 3

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of le by urban			
le(Life Expectancy Groups)	urban(Urbanisation Groups)		
	2	3	Total
70Yrs or more	52 46.43 57.14 73.24	39 34.82 42.86 95.12	91 81.25
Less than 70 Yrs	19 16.96 90.48 26.76	2 1.79 9.52 4.88	21 18.75
Total	71 63.39	41 36.61	112 100.00

Statistics for Table of le by urban

Statistic	DF	Value	Prob
Chi-Square	1	8.1695	0.0043
Likelihood Ratio Chi-Square	1	9.6323	0.0019
Continuity Adj. Chi-Square	1	6.7962	0.0091
Mantel-Haenszel Chi-Square	1	8.0965	0.0044
Phi Coefficient		-0.2701	
Contingency Coefficient		0.2607	
Cramer's V		-0.2701	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	52
Left-sided Pr <= F	0.0028
Right-sided Pr >= F	0.9997
Table Probability (P)	0.0025
Two-sided Pr <= P	0.0048

Sample Size = 112

Model Interpretation for Chi-Square Tests

While examining the association between the Life Expectancy (categorical response) and the Urbanization Rate of the countries (categorical explanatory) from the data set, a chi-square test of independence revealed that those with High Urbanization Rate tend to have high value results with Life Expectancy equal to or over 70 years (95.12 %), compared to countries whose Urbanization Rate is in the mid-range (73.24%) or lower ranges (27.63%).

In this analysis, $X^2 = 59.2165$, $df=2$ and $P < 0.0001$.

Model Interpretation for post hoc Chi-Square Tests

A Chi Square test of independence revealed that among the countries present in the dataset, the ranges of the Urbanization Rate ("3" high, "2" mid, and "1" lower") and Life Expectancy were significantly associated, $X^2 = 59.2165$, $df=2$ and $P < 0.0001$.

Post hoc comparisons of urbanization rates, by pairs of its ranges, revealed that countries with higher urbanization rate reported higher value of life expectancy equal to/over 70 years. The relationship between Life Expectancy and Urbanization Rate was proven in each of these comparisons tests.