Testing a Potential Moderator

(Data Analysis Tools Week 4 Assignment)

Expected Activities

- Run an ANOVA, Chi-Square Test or correlation coefficient that includes a moderator.
- Submit syntax used to test moderation (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 urbanrate lifeexpectancy alcconsumption urban ac le;
      LABEL lifeexpectancy="Life Expectancy";
      LABEL urbanrate="Urbanisation Rate";
      LABEL alcconsumption="Alcohol Consumption";
      LABEL urban="Urbanisation Groups";
      LABEL ac="Alcohol Consumption Category";
      LABEL le="Life Expectancy Groups";

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

```
        /* Data Management for variable alcconsumption */
        IF alcconsumption <=10 THEN
                ac=1;
        IF alcconsumption > 10 THEN
                ac=2;

PROC SORT;
        BY ac;

PROC ANOVA;
        CLASS urban;
        MODEL lifeexpectancy=urban;
        MEANS urban;
        BY ac;

PROC FREQ;
        TABLES le*urban/CHISQ;
        by ac;

PROC CORR;
        VAR lifeexpectancy urbanrate;
        BY ac;
        Title 'Pearson Correlation';

RUN;
```

Output

The ANOVA Procedure

Alcohol Consumption Category=1

Class Level Information

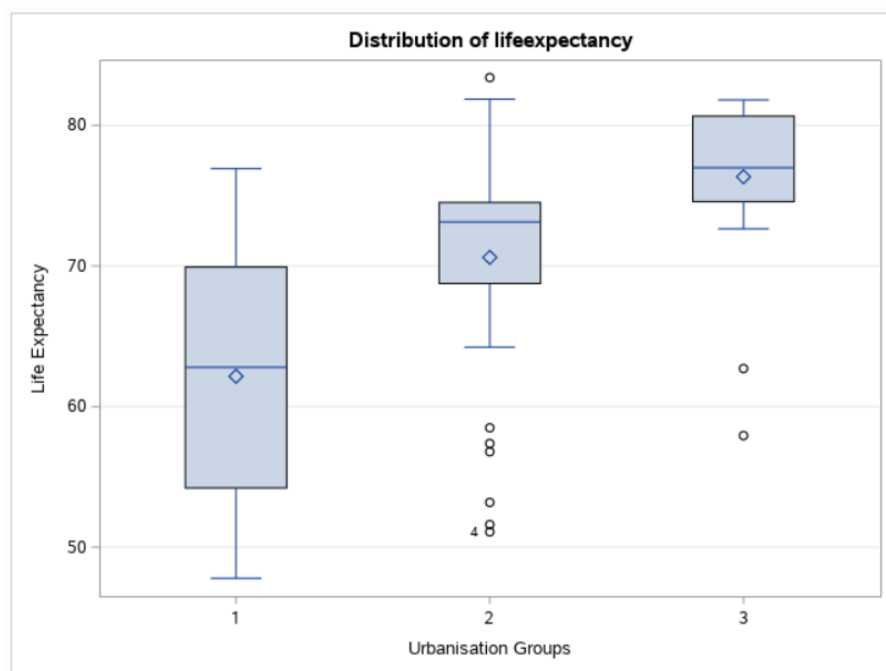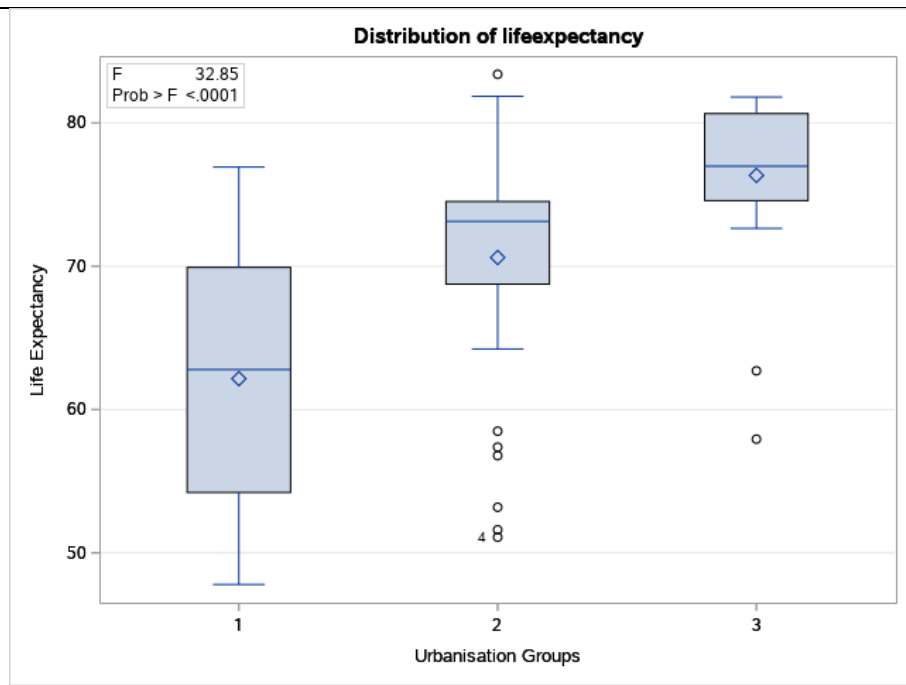| Class	Levels	Values
urban	3	1 2 3

Number of Observations Read	134
Number of Observations Used	134

The ANOVA Procedure

Dependent Variable: lifeexpectancy Life Expectancy

Alcohol Consumption Category=1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4340.53037	2170.26518	32.85	<.0001
Error	131	8653.57456	66.05782		
Corrected Total	133	12994.10493			

R-Square	Coeff Var	Root MSE	lifeexpectancy Mean
0.334038	12.04219	8.127596	67.49269

Source	DF	Anova SS	Mean Square	F Value	Pr > F
urban	2	4340.530366	2170.265183	32.85	<.0001

Distribution of lifeexpectancy

F	32.85
Prob > F	<.0001

Life Expectancy vs. Urbanisation Groups (1, 2, 3)

Distribution of lifeexpectancy

Life Expectancy vs. Urbanisation Groups (1, 2, 3)

Level of urban	N	lifeexpectancy	
		Mean	Std Dev
1	67	62.1542537	9.09098806
2	41	70.6090732	7.79592764
3	26	76.3351538	5.54211091

The ANOVA Procedure

Alcohol Consumption Category=2

Class Level Information		
Class	Levels	Values
urban	3	1 2 3

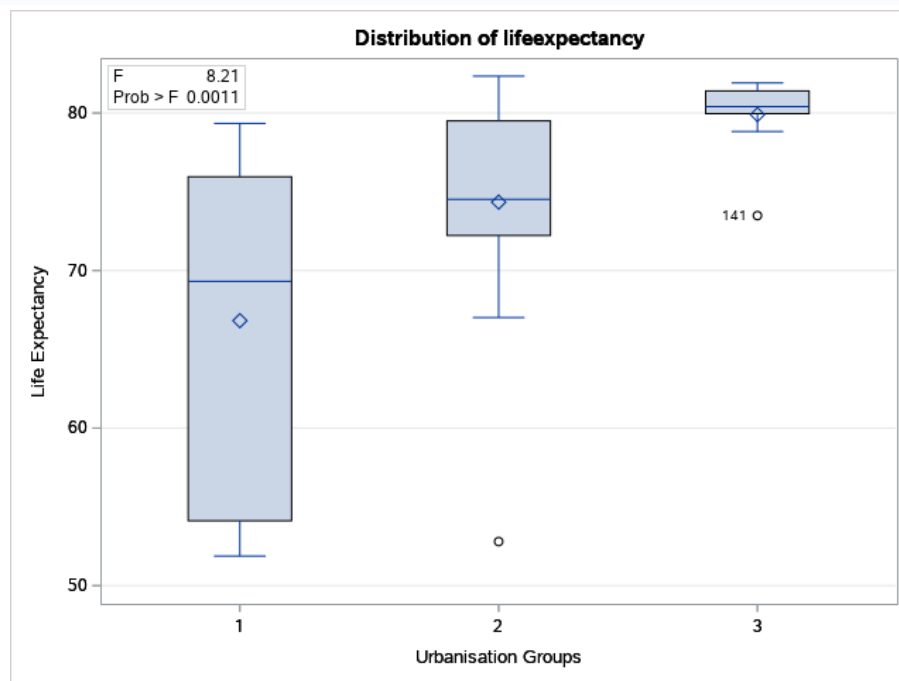| Number of Observations Read	42
Number of Observations Used	42

The ANOVA Procedure

Dependent Variable: lifeexpectancy Life Expectancy

Alcohol Consumption Category=2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	704.708155	352.354078	8.21	0.0011
Error	39	1674.701737	42.941070		
Corrected Total	41	2379.409892			

R-Square	Coeff Var	Root MSE	lifeexpectancy Mean
0.296169	8.806404	6.552944	74.41112

Source	DF	Anova SS	Mean Square	F Value	Pr > F
urban	2	704.7081553	352.3540776	8.21	0.0011

Distribution of lifeexpectancy

Distribution of lifeexpectancy

Level of urban	N	lifeexpectancy	
		Mean	Std Dev
1	7	66.8164286	10.8875772
2	25	74.3436800	6.1588230
3	10	79.8960000	2.4294330
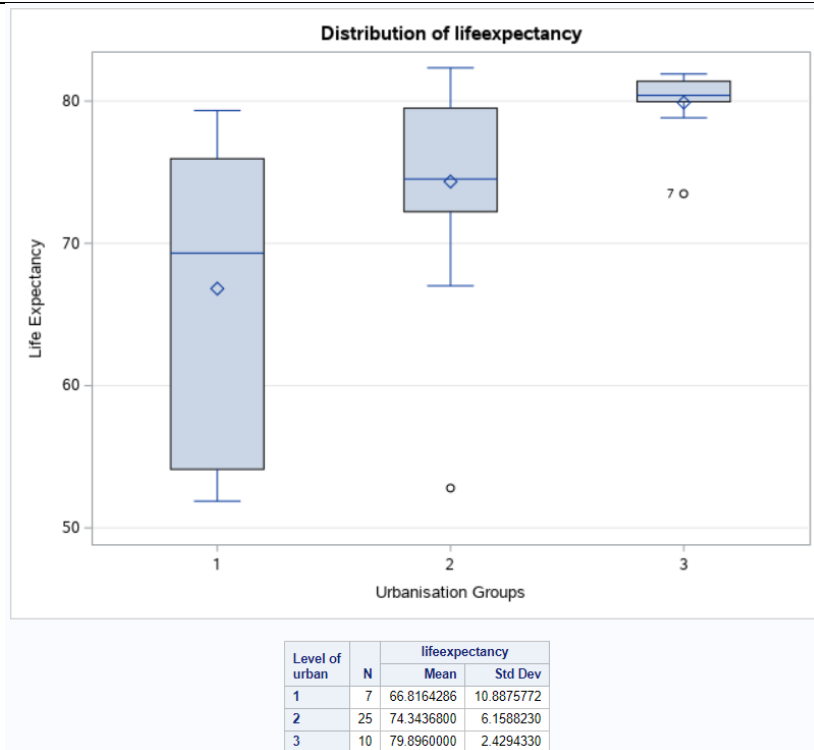
The FREQ Procedure

Alcohol Consumption Category=1

Frequency Percent Row Pct Col Pct	Table of le by urban			
	urban(Urbanisation Groups)			
le(Life Expectancy Groups)	1	2	3	Total
70Yrs or more	16 11.94 23.88 23.88	27 20.15 40.30 65.85	24 17.91 35.82 92.31	67 50.00
Less than 70 Yrs	51 38.06 76.12 76.12	14 10.45 20.90 34.15	2 1.49 2.99 7.69	67 50.00
Total	67 50.00	41 30.60	26 19.40	134 100.00

Statistics for Table of le by urban

Statistic	DF	Value	Prob
Chi-Square	2	41.0209	<.0001
Likelihood Ratio Chi-Square	2	45.3577	<.0001
Mantel-Haenszel Chi-Square	1	40.0813	<.0001
Phi Coefficient		0.5533	
Contingency Coefficient		0.4841	
Cramer's V		0.5533	

Sample Size = 134

The FREQ Procedure

Alcohol Consumption Category=2

Table of le by urban

Frequency Percent Row Pct Col Pct					
		urban(Urbanisation Groups)			
le(Life Expectancy Groups)		1	2	3	Total
70Yrs or more		3 7.14 8.82 42.86	21 50.00 61.76 84.00	10 23.81 29.41 100.00	34 80.95
Less than 70 Yrs		4 9.52 50.00 57.14	4 9.52 50.00 16.00	0 0.00 0.00 0.00	8 19.05
Total		7 16.67	25 59.52	10 23.81	42 100.00

Statistics for Table of le by urban

Statistic	DF	Value	Prob
Chi-Square	2	9.0918	0.0106
Likelihood Ratio Chi-Square	2	9.3565	0.0093
Mantel-Haenszel Chi-Square	1	7.8819	0.0050
Phi Coefficient		0.4653	
Contingency Coefficient		0.4218	
Cramer's V		0.4653	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 42

Pearson Correlation

The CORR Procedure

Alcohol Consumption Category=1

2 Variables:	lifeexpectancy urbanrate

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
lifeexpectancy	134	67.49269	9.88433	9044	47.79400	83.39400	Life Expectancy
urbanrate	134	51.75284	23.64462	6935	10.40000	100.00000	Urbanisation Rate

Pearson Correlation Coefficients, N = 134
Prob > |r| under H0: Rho=0

	lifeexpectancy	urbanrate
lifeexpectancy Life Expectancy	1.00000	0.59340 <.0001
urbanrate Urbanisation Rate	0.59340 <.0001	1.00000

Pearson Correlation

The CORR Procedure

Alcohol Consumption Category=2

2 Variables:	lifeexpectancy urbanrate

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
lifeexpectancy	42	74.41112	7.61803	3125	51.87900	82.33800	Life Expectancy
urbanrate	42	64.29762	17.47757	2701	12.98000	97.36000	Urbanisation Rate

Pearson Correlation Coefficients, N = 42
Prob > |r| under H0: Rho=0

	lifeexpectancy	urbanrate
lifeexpectancy Life Expectancy	1.00000	0.53689 0.0002
urbanrate Urbanisation Rate	0.53689 0.0002	1.00000

This analysis is intended to check whether the moderator "Alcohol Consumption" affects the relationship between Urbanization Rate (explanatory variable) and Life Expectancy (response variable). It appeared that the moderator does not influence the relationship between the two variables. The Life Expectancy was still higher with the increase of Urbanization Rate in both moderator sub-groups. Moreover, in each sub-group the relationship between Urbanization Rate and Life Expectancy remained statistically significant. More details concerning the results of Anova procedure is captured below:

<u>ANOVA on Urbanization Rate Vs Life Expectancy – in subgroup moderator "Countries with LOW alcohol consumption"</u>

F-statistic: 32.85

Prob (F-statistic): <0.0001

Since p is less than 0.05, we can say that there is a significant relationship between Urbanization Rate and Life Expectancy in countries with LOW alcohol consumption.

<u>ANOVA on Urbanization Rate Vs Life Expectancy – in subgroup moderator "Countries with HIGH alcohol consumption"</u>

F-statistic: 8.21

Prob (F-statistic): 0.0011

Again, as p is less than 0.05, there is a significant relationship between Urbanization Rate and Life Expectancy in countries with HIGH alcohol consumption.

Taking the above analysis into consideration, we can assume that the Alcohol Consumption variable does not moderate the relationship between Urbanization Rate and Life Expectancy. The said relationship remains positive and statistically significant in both sub-groups of the moderator.