Generating a Correlation Coefficient

(Data Analysis Tools Week 3 Assignment)

Expected Activities

- Generate a correlation coefficient.
- Submit syntax used to generate a correlation coefficient (copied and pasted from your program) along with corresponding output and a few sentences of interpretation.

Note 1: Two 3+ level categorical variables can be used to generate a correlation coefficient if the categories are ordered and the average (i.e. mean) can be interpreted. The scatter plot on the other hand will not be useful. In general, the scatterplot is not useful for discrete variables (i.e. those that take on a limited number of values).

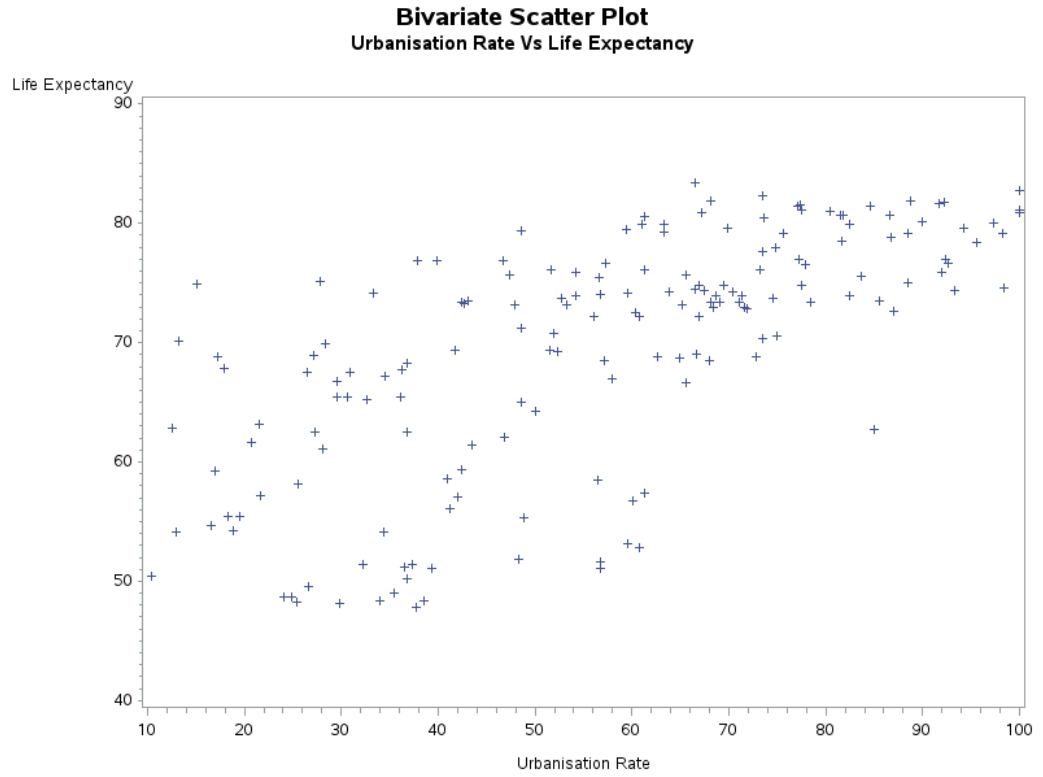
Note 2: When we square r, it tells us what proportion of the variability in one variable is described by variation in the second variable (a.k.a. RSquared or Coefficient of Determination).

SAS Program

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

DATA new;
      SET mydata.gapminder;
      KEEP country lifeexpectancy urbanrate femaleemployrate;
      LABEL lifeexpectancy="Life Expectancy";
      LABEL urbanrate="Urbanisation Rate";
      LABEL femaleemployrate="Female Employment Rate";

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

PROC SORT;
      BY country;

PROC GPLOT;
      PLOT lifeexpectancy*urbanrate;
      Title 'Bivariate Scatter Plot';
      Title2 'Urbanisation Rate Vs Life Expectancy';
```

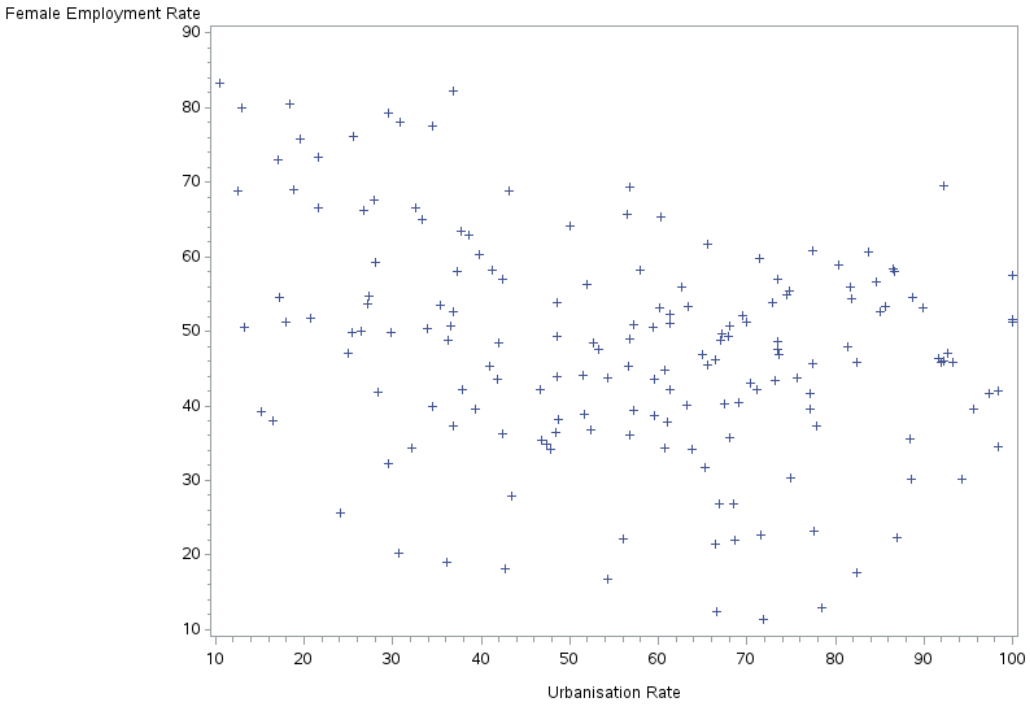
```
PROC GPLOT;
      PLOT femaleemployrate*urbanrate;
      Title 'Bivariate Scatter Plot';
      Title2 'Urbanisation Rate Vs Female Employment Rate';

PROC GPLOT;
      PLOT femaleemployrate*lifeexpectancy;
      Title 'Bivariate Scatter Plot';
      Title2 'Life Expectancy Vs Female Employment Rate';

      /* Pearson Correlation */
PROC CORR;
      VAR lifeexpectancy urbanrate femaleemployrate;
      Title 'Pearson Correlation';

RUN;
```

Output



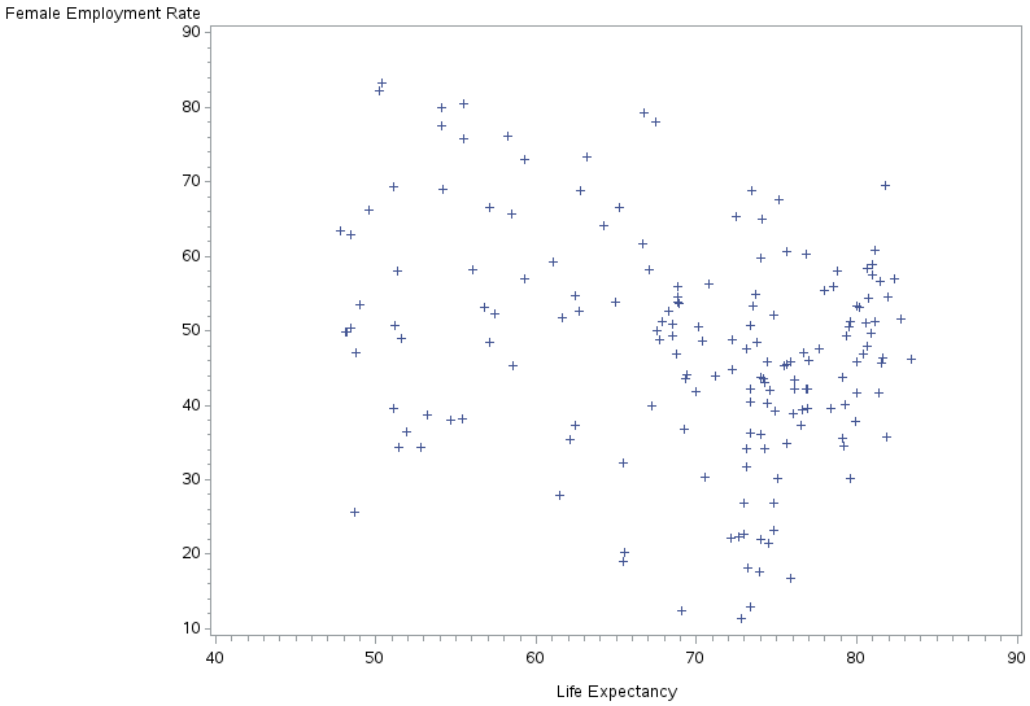
Bivariate Scatter Plot

Urbanisation Rate Vs Female Employment Rate



Bivariate Scatter Plot

Life Expectancy Vs Female Employment Rate



Pearson Correlation

The CORR Procedure

3 Variables:	lifeexpectancy urbanrate femaleemployrate

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
lifeexpectancy	173	69.37894	9.98926	12003	47.79400	83.39400	Life Expectancy
urbanrate	173	56.63653	23.26193	9798	10.40000	100.00000	Urbanisation Rate
femaleemployrate	173	47.75260	14.77495	8261	11.30000	83.30000	Female Employment Rate

Pearson Correlation Coefficients, N = 173 Prob > |r| under H0: Rho=0			
	lifeexpectancy	urbanrate	femaleemployrate
lifeexpectancy Life Expectancy	1.00000	0.67222 <.0001	-0.26498 0.0004
urbanrate Urbanisation Rate	0.67222 <.0001	1.00000	-0.30298 <.0001
femaleemployrate Female Employment Rate	-0.26498 0.0004	-0.30298 <.0001	1.00000

Urbanisation Rate Vs Life Expectancy

There is a significant relationship between Urbanization Rate and Life Expectancy. The p value is less than 0.0001. The Correlation Coefficient, R is 0.67222 which means that there´s a positive association between these variables – higher the Urbanization Rate, higher the Life Expectancy. R^2 is equals to 0.45188, indicating that if we know the Urbanization Rate (explanatory variable) we can predict 45% of the variability we will see in the Life Expectancy (response variable).

Urbanisation Rate Vs Female Employment Rate

Surprisingly, there is a slight negative correlation observed among the variables Urbanisation Rate and Female Employment Rate. The p value is less than 0.0001. The Correlation Coefficient, R is -0.30298 which means that there´s a negative association between these variables though it is minimal – higher the Urbanization Rate, lower the Female Employment

Rate. R^2 is equals to 0.09180, indicating that if we know the Urbanization Rate (explanatory variable) we can predict 9% of the variability we will see in the Female Employment Rate (response variable).

Life Expectancy Vs Female Employment Rate

There is a minimal negative correlation observed among the variables Life Expectancy and Female Employment Rate as well. The p value is 0.0004 and the Correlation Coefficient, R is observed as -0.26498 which indicates a very minimal negative association among these variables. R^2 equals 0.07021 which means that if we know the Female Employment Rate (explanatory variable) we can predict 7% of the variability we will see in the Life Expectancy (response variable) and vice versa.