8.9 Exercises

Mean Distribution of Data

Distribution of worked hours



Distribution of worked hours for Males



Distribution of hours worked for Females



Distribution of Education



Distribution of Education for Males



Distribution of Education for Females



Structure of the Data for Analysis to see if there are any missing values or special values

--------------------------------------------------------------------------------------------------------

hrs1 NUMBER OF HOURS WORKED LAST WEEK

--------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: HRS1, but 88 nonmissing values are not labeled

range: [1,89] units: 1

unique values: 88 missing .: 3,776/10,179

examples: 40

45

70

.

--------------------------------------------------------------------------------------------------------

educ HIGHEST YEAR OF SCHOOL COMPLETED

--------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: EDUC, but 21 nonmissing values are not labeled

range: [0,20] units: 1

unique values: 21 missing .: 32/10,179

examples: 12

12

14

16

--------------------------------------------------------------------------------------------------------

sex RESPONDENTS SEX

--------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: SEX

range: [1,2] units: 1

unique values: 2 missing .: 0/10,179

tabulation: Freq. Numeric Label

4,516 1 MALE

5,663 2 FEMALE

.

Conclusion of the Data Analysis:There are missing values in both the variables

Correlation test for both gendas

Males correlation test

| hrs1 educ

-------------+------------------

hrs1 | 1.0000

educ | 0.0681 1.0000

Females correlation test

| hrs1 educ

-------------+------------------

hrs1 | 1.0000

educ | 0.1220 1.0000

Regression Test

Source | SS df MS Number of obs = 6,387

-------------+---------------------------------- F(1, 6385) = 44.86

Model | 374.7339 1 374.7339 Prob > F = 0.0000

Residual | 53335.8949 6,385 8.35331165 R-squared = 0.0070

-------------+---------------------------------- Adj R-squared = 0.0068

Total | 53710.6288 6,386 8.41068412 Root MSE = 2.8902

------------------------------------------------------------------------------

educ | Coef. Std. Err. t P>|t| Beta

-------------+----------------------------------------------------------------

hrs1 | .0169341 .0025283 6.70 0.000 .0835279

\_cons | 13.06539 .1123928 116.25 0.000 .

------------------------------------------------------------------------------

Regression Test for Males

Source | SS df MS Number of obs = 3,195

-------------+---------------------------------- F(1, 3193) = 14.86

Model | 138.959577 1 138.959577 Prob > F = 0.0001

Residual | 29863.1706 3,193 9.35269985 R-squared = 0.0046

-------------+---------------------------------- Adj R-squared = 0.0043

Total | 30002.1302 3,194 9.39327808 Root MSE = 3.0582

------------------------------------------------------------------------------

educ | Coef. Std. Err. t P>|t| Beta

-------------+----------------------------------------------------------------

hrs1 | .0145222 .0037675 3.85 0.000 .0680563

\_cons | 13.0301 .1785737 72.97 0.000 .

------------------------------------------------------------------------------

.

Regression test for females

Source | SS df MS Number of obs = 3,192

-------------+---------------------------------- F(1, 3190) = 48.17

Model | 351.844687 1 351.844687 Prob > F = 0.0000

Residual | 23302.4598 3,190 7.30484634 R-squared = 0.0149

-------------+---------------------------------- Adj R-squared = 0.0146

Total | 23654.3045 3,191 7.41281871 Root MSE = 2.7027

------------------------------------------------------------------------------

educ | Coef. Std. Err. t P>|t| Beta

-------------+----------------------------------------------------------------

hrs1 | .0244788 .0035271 6.94 0.000 .1219608

\_cons | 12.91548 .1456582 88.67 0.000 .

Summary Analytics : Females have a higher beta value and correlation-coefficient than males therefore the assumptions that there is no need to educate them a they will just stay at home is wrong.

.

Structure of the Data

--------------------------------------------------------------------------------------------------------

sphrs1 NUMBER OF HRS SPOUSE WORKED LAST WEEK

--------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: SPHRS1, but 78 nonmissing values are not labeled

range: [1,89] units: 1

unique values: 78 missing .: 6,961/10,179

examples: 40

.

.

.

. codebook hrs1

--------------------------------------------------------------------------------------------------------

hrs1 NUMBER OF HOURS WORKED LAST WEEK

--------------------------------------------------------------------------------------------------------

type: numeric (byte)

label: HRS1, but 88 nonmissing values are not labeled

range: [1,89] units: 1

unique values: 88 missing .: 3,776/10,179

examples: 40

45

70

.

Data Structure Analysis :There are missing values between the two variables

Scatter-grams of the two variables



Scatter-gram for the two variable for Males



Scatter-gram of the two variable for Females



Regression Analysis

Source | SS df MS Number of obs = 2,393

-------------+---------------------------------- F(1, 2391) = 0.75

Model | 130.596748 1 130.596748 Prob > F = 0.3872

Residual | 417393.396 2,391 174.568547 R-squared = 0.0003

-------------+---------------------------------- Adj R-squared = -0.0001

Total | 417523.992 2,392 174.550164 Root MSE = 13.212

------------------------------------------------------------------------------

sphrs1 | Coef. Std. Err. t P>|t| Beta

-------------+----------------------------------------------------------------

hrs1 | .0164893 .0190642 0.86 0.387 .0176858

\_cons | 40.7701 .8522983 47.84 0.000 .

------------------------------------------------------------------------------

Regression Analysis for Males

Regression .

Source | SS df MS Number of obs = 1,136

-------------+---------------------------------- F(1, 1134) = 10.86

Model | 1752.10969 1 1752.10969 Prob > F = 0.0010

Residual | 182950.073 1,134 161.331634 R-squared = 0.0095

-------------+---------------------------------- Adj R-squared = 0.0086

Total | 184702.183 1,135 162.733201 Root MSE = 12.702

------------------------------------------------------------------------------

sphrs1 | Coef. Std. Err. t P>|t| Beta

-------------+----------------------------------------------------------------

hrs1 | .0991445 .0300848 3.30 0.001 .0973968

\_cons | 31.98229 1.472278 21.72 0.000 .

------------------------------------------------------------------------------

Analysis for Females

Source | SS df MS Number of obs = 1,257

-------------+---------------------------------- F(1, 1255) = 48.16

Model | 6765.1294 1 6765.1294 Prob > F = 0.0000

Residual | 176296.727 1,255 140.47548 R-squared = 0.0370

-------------+---------------------------------- Adj R-squared = 0.0362

Total | 183061.857 1,256 145.749886 Root MSE = 11.852

------------------------------------------------------------------------------

sphrs1 | Coef. Std. Err. t P>|t| Beta

-------------+----------------------------------------------------------------

hrs1 | .1644874 .0237025 6.94 0.000 .1922379

\_cons | 39.55875 .9600615 41.20 0.000 .

------------------------------------------------------------------------------

Correlation Analysis

| sphrs1 hrs1

-------------+------------------

sphrs1 | 1.0000

hrs1 | 0.0177 1.0000

Correlation Analysis for Males

| sphrs1 hrs1

-------------+------------------

sphrs1 | 1.0000

hrs1 | 0.0974 1.0000

Correlation Analysis for Females

| sphrs1 hrs1

-------------+------------------

sphrs1 | 1.0000

hrs1 | 0.1922 1.0000

Summary Analytics: From the Beta-value that is 0973968=Males and .1922379=Females it is true to say that married men work more than women.



Data Structure

(9,929 observations deleted)

SCATTERGRAM RELATING FATHER’S EDUCATION AND SON’S EDUCATION



SCATTERGRAM WITH JITTER FACTOR OF 3



Summary Analytics:When using the jitter(3) rekationship the graphs generated shows more groups or classes of overlap of datapoints than without the jitter 3. Therefore the jitter(3) option helps to bring out more inter-data-points relationships through the dataset.

1. .

Structure of Data

-----------------------------------------------------------------------------------------------

happy GENERAL HAPPINESS

-----------------------------------------------------------------------------------------------

type: numeric (byte)

label: HAPPY

range: [1,3] units: 1

unique values: 3 missing .: 70/250

tabulation: Freq. Numeric Label

47 1 VERY HAPPY

111 2 PRETTY HAPPY

22 3 NOT TOO HAPPY

70 .

-----------------------------------------------------------------------------------------------

hapmar HAPPINESS OF MARRIAGE

-----------------------------------------------------------------------------------------------

type: numeric (byte)

label: HAPMAR

range: [1,3] units: 1

unique values: 3 missing .: 168/250

tabulation: Freq. Numeric Label

56 1 VERY HAPPY

24 2 PRETTY HAPPY

2 3 NOT TOO HAPPY

168 .

-----------------------------------------------------------------------------------------------

health CONDITION OF HEALTH

-----------------------------------------------------------------------------------------------

type: numeric (byte)

label: HEALTH

range: [1,4] units: 1

unique values: 4 missing .: 61/250

tabulation: Freq. Numeric Label

40 1 EXCELLENT

108 2 GOOD

34 3 FAIR

7 4 POOR

61 .

.

Correlations between happy, hapmar, and health by using correlate

(obs=53)

Variable | Mean Std. Dev. Min Max

-------------+----------------------------------------------------

happy | 1.528302 .5407891 1 3

hapmar | 1.264151 .4450991 1 2

health | 2 .7337994 1 4

| happy hapmar health

-------------+---------------------------

happy | 1.0000

hapmar | 0.5276 1.0000

health | 0.0485 0.0000 1.0000

Correlations between happy, hapmar, and health by using pwcorr

| happy hapmar health

-------------+---------------------------

happy | 1.0000

hapmar | 0.5821 1.0000

health | 0.1881 0.0000 1.0000

Estimate the correlations by using pwcorr, and get the significance level and the number of observations for each case

| happy hapmar health

-------------+---------------------------

happy | 1.0000

|

| 180

|

hapmar | 0.5821 1.0000

| 0.0000

| 82 82

|

health | 0.1881 0.0000 1.0000

| 0.0349 1.0000

| 126 53 189

|

.

Repeat the pwcorr command so that all the Ns are the same (that is, there is casewise/listwise deletion).

| happy hapmar health

-------------+---------------------------

happy | 1.0000

|

| 53

|

hapmar | 0.5276 1.0000

| 0.0000

| 53 53

|

health | 0.0485 0.0000 1.0000

| 0.7304 1.0000

| 53 53 53

|

Summary Analytics: The results for pwcorr command are slightly significant because they account for the missing values therefore taking into consideration more observations. From the relationship analysis of the correlation and regression analysis happier marriages lead to more report of happiness that good health does.

Run the codebook command on these variables

-----------------------------------------------------------------------------------------------

happy7 (unlabeled)

-----------------------------------------------------------------------------------------------

type: numeric (byte)

label: happy7

range: [1,7] units: 1

unique values: 7 missing .: 1,605/2,765

tabulation: Freq. Numeric Label

141 1 completely happy

510 2 very happy

391 3 fairly happy

69 4 neither happy nor unhappy

32 5 fairly unhappy

16 6 very unhappy

1 7 completely unhappy

1,605 .

-----------------------------------------------------------------------------------------------

satfam7 family satisfaction in general

-----------------------------------------------------------------------------------------------

type: numeric (byte)

label: satfam7

range: [1,7] units: 1

unique values: 7 missing .: 1,617/2,765

tabulation: Freq. Numeric Label

265 1 completely satisfied

467 2 very satisfied

286 3 fairly satisfied

70 4 neither satisfied nor

dissatisfied

31 5 fairly dissatisfied

20 6 very dissatisfied

9 7 completely dissatisfied

1,617 .

.

Regression Analysis

Source | SS df MS Number of obs = 1,143

-------------+---------------------------------- F(1, 1141) = 840.15

Model | 640.744418 1 640.744418 Prob > F = 0.0000

Residual | 870.184716 1,141 .762650934 R-squared = 0.4241

-------------+---------------------------------- Adj R-squared = 0.4236

Total | 1510.92913 1,142 1.32305528 Root MSE = .8733

------------------------------------------------------------------------------

satfamnew | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

happynew | .7628363 .0263179 28.99 0.000 .7111993 .8144732

\_cons | .4367221 .0700948 6.23 0.000 .299193 .5742512

Regression Equation

Y = a + b(X)

Satfamnew=.43672+happynew(.762836)

The family has a default satisfaction unit of .43672 but with every unit increase change of happynew there is a proportional increase of .762836.

7.

SCATTERGRAM DIAGRAM OF AGE AGAINST DEPRESSION



BINSCATTER DIAGRAM OF AGE AGAINST DEPRESSION



Summary Analytics: e binscatter places a linear regression line on the figure allowing for a quick eye-glimpsing of the direction of the Relationship between Age and Depression.From the graph it is correct to state that depression decreases with increase in Age.

8.

Structure of the Data

----------------------------------------------------------------------------------------------

age Age (years)

----------------------------------------------------------------------------------------------

type: numeric (byte)

range: [22,60] units: 1

unique values: 39 missing .: 0/1,000

mean: 36.542

std. dev: 9.48009

percentiles: 10% 25% 50% 75% 90%

26 28 32 45 50

----------------------------------------------------------------------------------------------

TotalScore Sum of Questions 1-20

----------------------------------------------------------------------------------------------

type: numeric (byte)

range: [23,96] units: 1

unique values: 71 missing .: 0/1,000

mean: 57.411

std. dev: 11.9691

percentiles: 10% 25% 50% 75% 90%

41 49 58 65 73

.

BINSCATTER DIAGRAM IF AGE==50 AND ABOVE



Summary Analytics: The relationship between Age and Depression doesn’t linearly fit well with each other. Averagely many points are away from the regression line therefore a poor correlation score thus deprecating the linearity of the assumed relationship.

DO FILE

1. . histogram hrs1

. histogram hrs1 if sex==1

. histogram hrs1 if sex==2

. histogram educ

. histogram educ if sex==1

. histogram educ if sex==2

. codebook hrs1

. codebook educ .

. pwcorr hrs1 educ if sex==1

. pwcorr hrs1 educ if sex==2

. regress educ hrs1, beta

. regress educ hrs1 if sex==1, beta

. regress educ hrs1 if sex==2, beta

1. . histogram sphrs1

. codebook sphrs1

. histogram sphrs1

. histogram sphrs1 if sex==1

. histogram sphrs1 if sex==2

. codebook hrs1

. regress sphrs1 hrs1, beta

. regress sphrs1 hrs1 if sex==1, beta

. regress sphrs1 hrs1 if sex==2, beta

.pwcorr sphrs1 hrs1

. pwcorr sphrs1 hrs1 if sex==1

. pwcorr sphrs1 hrs1 if sex==2

1. . set seed 111

. sample 250 ,count

twoway (scatter educ paeduc if sex==1, msize(5)) if sex==1, ytitle(Son's Education) xtitle(Fa

> thers's Education) title(Scattergram relating father’s education to his son’s education) note

> (N=250 observation)

. twoway (scatter educ paeduc if sex==1, msize(5) jitter(3) jitterseed(111)) if sex==1, ytitle(

> Son's Education) xtitle(Fathers's Education) title(Scattergram relating father’s education to

> his son’s education) note(N=250 observation)

1. . codebook happy hapmar health

. correlate happy hapmar health, means

. pwcorr happy hapmar health

. pwcorr happy hapmar health, obs sig

. pwcorr happy hapmar health, obs sig listwise

. codebook happy7 satfam7

. rename satfam7 satfamnew

. label define satfam7 0 "nap" 1 "completely disatisfied" 2 "very satisfied" 3 "fairly satisfied" 4 "nei

> ther satisfied nor dissatisfied" 5 "fairly dissatisfied" 6 "very dissatisfied" 7 "completely satisfied

> " 8 "cannot choose" 9 "no answer", replace

. label define happy7 0 "nap" 1 "completely unhappy" 2 "very happy" 3 "fairly happy" 4 "neither happy no

> r unhappy" 5 "fairly unhappy" 6 "very unhappy" 7 "very happy" 8 "cannot choose" 9 "no answer", replace

. rename happy7 happynew

. regress satfamnew happynew

7.

. twoway (scatter TotalScore age), ytitle(Depression) xtitle(Age) title(Scattergram relating D

> epression and Age) note(Relation betweeen age and depression)

. binscatter age TotalScore

8.

. codebook age TotalScore

. binscatter age TotalScore if age>50

warning: nquantiles(20) was specified, but only 19 were generated. see help file under nquanti

> les() for explanation.