Optional Assignment Week 9

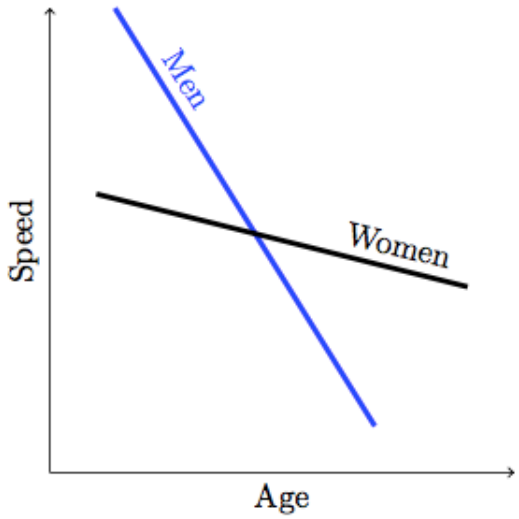
PIED 2711 | Analysing Data | 2018

November 26, 2018

Question 1

The figure below shows a hypothesized (i.e. predicted) relationship between how fast a person runs, the person's age, and sex. Consider the model formula:

$$\widehat{\text{Speed}}_i = \alpha + \beta_1 \text{Age}_i + \beta_2 \text{Female}_i + \beta_3 \text{Age}_i \times \text{Female}_i$$



1. According to the prediction, what do we expect the sign of the coefficient on Age_i to be?

   (a) Negative

   (b) Zero

   (c) Positive

   (d) No way to tell from the information given

2. According to the prediction, what do we expect the sign of the coefficient on Female_i to be?

   (a) Negative

   (b) Zero

   (c) Positive

   (d) No way to tell from the information given

3. According to the prediction, what do we expect the sign of the coefficient on the interaction term ($\text{Age}_i \times \text{Female}_i$) to be?

   (a) Negative

   (b) Zero

   (c) Positive

   (d) No way to tell from the information given

Question 2

A researcher has data from the 67 Florida countries on y = crime rate (number per residents), x1 = median income (thousands of dollars), and x2 = percentage in urban environment.

1. The first table shows the results when she regressed y on x1. Write down the model equation from the output and interpret the slope coefficient.

2. The second table shows the results when she regressed y on x1 and x2. Write down the model equation from the output and interpret the slope coefficients.

3. Take a look at the correlation matrix. Use these correlations to explain why the x1 effect seems so different in the first model and second model.

4. Write down the model equations relating crime rate to income at urbanization levels of 1) 0, 2) 50, and 3) 100. Interpret.

```
. reg Crime Income, noheader
```

Crime	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Income	2.611494	.6728733	3.88	0.000	1.267673	3.955315
_cons	-11.6059	16.78631	-0.69	0.492	-45.13048	21.91867

```
. reg Crime Income Urban, noheader
```

Crime	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Income	-.7905853	.8049117	-0.98	0.330	-2.398581	.8174107
Urban	.6418376	.1109602	5.78	0.000	.4201692	.863506
_cons	39.97226	16.35362	2.44	0.017	7.302154	72.64238

```
. corr Crime Income Urban
(obs=67)
```

	Crime	Income	Urban
Crime	1.0000		
Income	0.4338	1.0000	
Urban	0.6774	0.7307	1.0000

Question 3

Table 11.12 shows Stata output from fitting the multiple regression model to recent statewide data, excluding D.C., on y = violent crime rate (per 100,000 people), x1 = poverty rate (percentage with income below the poverty line), and x2 = percentage living in urban areas.

TABLE 11.14

```
. regress violent poverty urban

   Source |       SS    df       MS    Number of obs =      50
----------+------------------------------   F(2, 47)      =  31.249
    Model | 2448368.07   2  1224184.   Prob > F      =  0.0001
 Residual | 1841257.15  47  39175.68   R-squared     =  0.5708
----------+------------------------------   Adj R-squared =  0.5525
    Total | 4289625.22  49  87543.37   Root MSE      = 197.928

    violent |    Coef.     Std. Err.      t      P>|t|
    poverty |   32.622      6.677     4.885    0.0001
      urban |    9.112      1.321     6.900    0.0001
      _cons | -498.683    140.988    -3.537    0.0009

. corr violent poverty urban
                    violent        poverty        urban
       violent      1.0000
       poverty       .3688         1.0000
       urban         .5940         -.1556         1.0000
```

1. Write down the model equation (with the estimated coefficients).
2. Massachusetts had y = 805, x1 = 10.7, and x2 = 96.1. Find its predicted violent crime rate. Find the residual, and interpret.

3. Interpret the fit by showing the model equation relating y and x1 for states with 1) x2= 0, 2) x2=100. Interpret.

4. Conduct a hypothesis testing when the null hypothesis is that there is no relationship between violent crime and poverty.

5. When we add x3 = percentage of single-parent family to the model, we get the results in the table below. Why do you think the effect of poverty rate is much lower after x3 is added to the model?

TABLE 11.15

Variable	Coefficient	Std. Error
Intercept	−1197.538	
Poverty	18.283	6.136
Urban	7.712	1.109
Single parent	89.401	17.836
R^2	0.722	