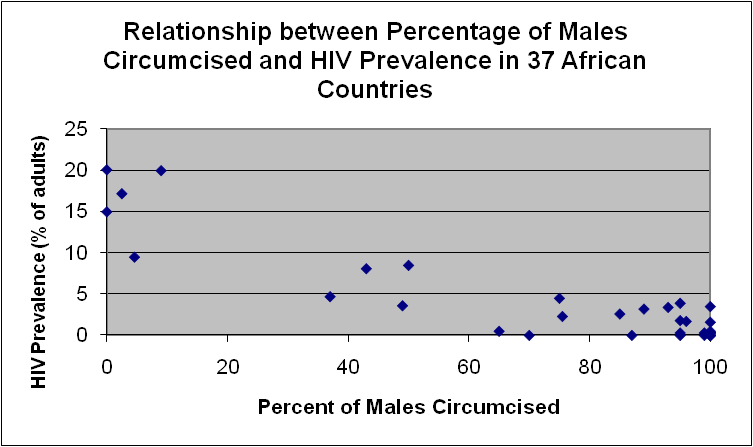
Homework 6 

Due Date:  **Thursday** November 4 by NOON

Many studies have investigated the association between male circumcision and HIV prevalence. One of the first studies (Bongaarts et al, 1989) examined the relationship between percentage of males who were circumcised in 37 African countries and the HIV seroprevalence in those countries based on estimates from the capital city. The authors report the two factors were strongly correlated (*p*<0.001).

For questions 1-8, use the following scatterplot and SAS output correspond to the data from Bongaarts et al. We reproduce some of the calculations in the original article in order to study the relationship between “Percent of Males Circumcised” and “HIV Seroprevalence”.



**Descriptive Statistics Parameter Estimate Table**

| *Variable* | *N* | *Mean* | *Std Dev* |
| --- | --- | --- | --- |
| HIV Prevalence Males Circumcised (%) | 37 37 | 3.7216 75.9081 | 5.6929 34.0605 |

| *Parameter* | *Estimate* | *Standard Error* | *t Value* | *Pr > |t|* |
| --- | --- | --- | --- | --- |
| *Intercept* | 15.143 | 1.0214 | 14.82 | <.0001 |
| *MalesCircumcised* | -0.151 | 0.0123 | -12.23 | <.0001 |

**Model Summary ANOVA**

| *R-Square* | *Coeff Var* | *Root MSE* | *prev Mean* |
| --- | --- | --- | --- |
| 0.8103 | 67.5682 | 2.5146 | 3.7216 |

| *Source* | *DF* | *Sum of Squares* | *Mean Square* | *F Value* | *Pr > F* |
| --- | --- | --- | --- | --- | --- |
| *Model* | 1 | 945.424337 | 945.424337 | 149.51 | <.0001 |
| *Error* | 35 | 221.318365 | 6.323382 |  |  |
| *Total* | 36 | 1166.742703 |  |  |  |

1. What is the equation of the least squares regression line?

2. What is the value of the correlation between “Males Circumcised” and “HIV Prevalence”?

3. Let ****** be the population correlation between “Males Circumcised” and “HIV Prevalence”. What is the value of the *t* statistic for testing the hypotheses H0: ****** = 0 versus Ha: ****** ≠ 0?

4. Is the slope significantly different from zero? Include the value of the test statistic and the corresponding *p*-value in your answer, along with a conclusion in language the investigator can understand.

5. What is the value of the estimate for ******, the standard deviation of the random deviations ***i***?

6. Give a 95% confidence interval for the population slope parameter ******.

7. Find the predicted “HIV Seroprevalence (%)” when the “Percent of Circumcised Males” is 70%.

8. Find the “Percent of Circumcised Males” predicted by the regression line when the “HIV Seroprevalence (%)” is 10%.

The residuals from this regression analysis are plotted versus the explanatory variable. The graph is provided here.

|  |  |
| --- | --- |
| 7. | Which of the following interpretations of this residual plot is correct? |
| A) | The fit is not good, because there is no linear relationship shown in this plot. |
| B) | The fit is good, because the residuals form a pattern around the *x*-axis. |
| C) | The fit is good, because the residuals indicate a departure from the linear regression line. |
| D) | The fit is good, because no clear pattern is evident. |