BIOSTATISTICS 600

Global Activity One

Male Circumcision and HIV Infection in African populations

INTRODUCTION

Many studies have investigated the association between male circumcision and HIV prevalence. One of the first studies (Bongaarts 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 ($p<0.001$). In the following exercises, students will explore the relationship between *Percent of Males Circumcised* and *HIV Seroprevalence* by reproducing many of the calculations in the original journal article and expanding on those findings.

SOURCE

Bongaarts J, Reining P, Way P, Conant F. 1989. The relationship between male circumcision and HIV infection in African populations. AIDS 3:373-7.

QUESTIONS

Use the data provided (published in Bongaart (1989) and given in GA_One_Bongaarts.xls) to complete the following questions:

1. Provide descriptive statistics for *x= Percent Males Circumcised* and for *y= Percent HIV Seroprevalence* in one neat, well-labeled table. Explain your reasoning for your choice of descriptive statistics.

2. a. Produce a histogram for *x=Percent Males Circumcised*. Do the data appear to be normally distributed?

b. Produce a histogram for *y=Percent HIV Seroprevalence*. Do the data appear to be normally distributed?

c. Which do you prefer to describe the factors, the descriptive statistics (1.) or the histograms (2a.,b.)? Why?

3. Produce a neat scatter plot for the given data with *x=Percent Males Circumcised* and *y = Percent HIV Seroprevalence*. Include the equation of the least squares regression line.

4. a. Check the assumptions for the linear regression analysis. Produce a residual plot and comment on the normality assumption.

b. Do *Percent Males Circumcised* and *Percent HIV Seroprevalence* need to be normally distributed? Explain.

5. Are there any countries which, in your opinion, are outliers? [a) with respect to x, b) with respect to y, c) with respect to the overall data distribution or d) an influential point.

6. Find the predicted *Percent HIV Seroprevalence* when the *Percent of Circumcised Males* is 75%.

7. Find the *Percent of Circumcised Males* predicted by the regression line when the *Percent HIV* is 10%.

8. Calculate r and r^2. Explain the interpretation of r^2. Conduct a statistical test for $\rho = 0$ and interpret the p-value.

9. Interpret the estimate for the slope. Conduct a hypothesis test for population slope equal to zero and interpret the p-value carefully.

10. Note that x= *Percent Males Circumcised* was estimated for each country while y=*Percent HIV Prevalence* was estimated based on the data from each capital city. The paper states, "the resulting errors produce an underestimate of the true level of correlation between the proportion of males who are circumcised and the level of HIV infection, because deviations from the regression line in a number of countries are due to discrepancies between urban and national circumcision practices." Explain.

11. The design of this study is an ecologic study with the unit of observation, "country". Discuss some limitations of an ecologic study design. For more information about ecologic studies see http://www.bmj.com/epidem/epid.6.html.
[Later studies have addressed this association using stronger study designs. We'll explore more about the topic in Global Activity Two.]

BIOSTATISTICS TOPICS:
SIMPLE LINEAR REGRESSION, CORRELATION, DESCRIPTIVE STATISTICS.