WWS 509 Generalized Linear Models: Precept 1 Section 2.1 through 2.5.6

Kristin E. Bietsch

Office of Population Research, Princeton University

September 2012

1 Introducing the data

This precept uses data from the United Nations about urbanization and fertility. Today we will focus on 50 Africa countries, the percentage of a population in each country classified as "urban," the county's total fertility rate (ask the demographer sitting closest to you for a definition), and the Human Development Index score in 2005. Please do not cite this data or the regressions.

2 Null and Saturated Models

- A null model postulates no systematic differences between units
  - In this week's example, the null model would be that all countries have the same percentage of population living in urban areas
- A saturated model has as many parameters in the linear predictor as it has observation.
  - In this week's example, imagine dummy variables for Ghana, Nigeria, South Africa, etc.

Neither of these models are very informative. Let's look in the middle!

3 Simple Linear Regression

3.1 Fertility and Urbanization

Presented below is the Stata output from a regression of TFR on urbanization:

```
. reg  propurbOS tfr

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  1,    48) =   21.61
       Model |  5161.91771     1  5161.91771            Prob > F      =  0.0000
    Residual |  11464.1449    48  238.836352            R-squared     =  0.3105
-------------+------------------------------           Adj R-squared =  0.2961
       Total |  16626.0626    49  339.3074              Root MSE      =  15.454

------------------------------------------------------------------------------
    propurbOS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
          tfr |  -7.188698   1.546304    -4.65   0.000    -10.29775   -4.079645
        _cons |   76.86343    8.00464     9.60   0.000     60.76902    92.95784
------------------------------------------------------------------------------
```

Figure 1: Regression of TFR on Proportion Urban

1. Interpret the coefficient for the constant and TFR
2. Construct the Likelihood ratio where this model is nested in the null
3. Test the significance for the coefficent for TFR
4. How would you calculate the R^2 by hand?
5. Calculate Pearson's r

3.2 HDI and Urbanization

Presented below is the Stata output from a regression of HDI on urbanization:

```
. reg  propurbOS hdi

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  1,    48) =   25.69
       Model |  5796.67508     1  5796.67508            Prob > F      =  0.0000
    Residual |  10829.3875    48  225.61224            R-squared     =  0.3486
-------------+------------------------------           Adj R-squared =  0.3351
       Total |  16626.0626    49  339.3074              Root MSE      =  15.02

------------------------------------------------------------------------------
    propurbOS |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
          hdi |   82.64527   16.30459     5.07   0.000     49.86269    115.4278
        _cons |   6.954846   7.056541     0.99   0.329    -7.233282    21.14297
------------------------------------------------------------------------------
```

Figure 2: Regression of HDI on Proportion Urban

1. Interpret the coefficient for the constant and HDI
2. Construct the Likelihood ratio where this model is nested in the null
3. Test the significance for the coefficent for HDI
4. How would you calculate the R^2 by hand?
5. Calculate Pearson's r

4 Multiple Linear Regression

In this section we look at the additive model which includes TFR and HDI. In an additive model, the effect of each perdictor on the response is assumed to be the same for all values of the other predictors.
Presented below is the Stata output from a regression of HDI and TFR on urbanization:

```
. reg  propurb05 tfr hdi

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  2,    47) =   13.63
       Model |  6103.2542      2  3051.6271             Prob > F      =  0.0000
    Residual |  10522.8084     47  223.88954             R-squared     =  0.3671
-------------+------------------------------           Adj R-squared =  0.3402
       Total |  16626.0626     49  339.3074              Root MSE      =  14.963

------------------------------------------------------------------------------
   propurb05 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         tfr |  -2.975143   2.542456    -1.17   0.248    -8.089904    2.139619
         hdi |   56.55783   27.58273     2.05   0.046     1.068542    112.0471
       _cons |   32.53766   22.96455     1.42   0.163    -13.66105    78.73637
------------------------------------------------------------------------------
```

Figure 3: Regression of HDI and TFR on Proportion Urban

1. Interpret the coefficient for the constant, HDI, and TFR
2. Test the significance for HDI and TFR
   - Has significance changed from the previous models?

4.1 Gross and Net Effects

- Gross effect: the change in the response that can be associated with a given predictor in a simple linear regression
  - What is the gross effect of TFR on proportion urban? (Fill in the box below)
  - What is the gross effect of HDI on proportion urban?
- Net effect: the change in the response that can be associated with a given predictor for fixed values of other predictors
  - What is the net effect of TFR on proportion urban?
  - What is the net effect of HDI on proportion urban?

Predictor	Gross	Net
TFR		
HDI		

4.2 ANOVA for Multiple Regression

Fill in the following table for the Analysis of Variance for Multiple Regression of Proportion Urban by TFR and HDI:

Source of variation	Sum of squares	Degrees of freedom	Mean squared	F-ratio
Regression				
Residual				
Total				

4.2.1 Hierarchical Table 1

Fill in the following table for the Hierarchical Analysis of Variance for Multiple Regression of Proportion Urban by TFR and HDI:

Source of variation	Sum of squares	Degrees of freedom	Mean squared	F-ratio
TFR				
HDI|TFR				
Residual				
Total				

4.2.2 Hierarchical Table 2

Fill in the following table for the Hierarchical Analysis of Variance for Multiple Regression of Proportion Urban by TFR and HDI:

Source of variation	Sum of squares	Degrees of freedom	Mean squared	F-ratio
HDI				
TFR‖HDI				
Residual				
Total				

4.3 Partial and Multiple Correlation

- Multiple Correlation Coefficient: square root of the proportion of variance explained
- Partial Correlation Coefficient: square root of the proportion of variation explained by the second variable out of the amount left unexplained by the first
  - What are the parital correlations for TFR and Urbanization?