

Questions about the last lab

September 10, 2008

What is the null hypothesis in a univariate normality test?

- E.g., PROC UNIVARIATE on the dependent variable
- Kolmogorov-Smirnov, Wald, Anderson-Darling, etc..
 - H0: data is normally distributed
 - E.g., p-value < 0.05 → reject normality hypothesis

What is the null hypothesis for the regression?

- E.g., PROC REG
- F-test is testing if the relationship between the dependent and independent variables is significant
 - Are *any* of the coefficients different from zero?
- Note: this test does *not* take into account whether assumptions have been met (in particular, this test should not be interpreted if there is heteroscedasticity or non-normality)

What are the implications of multicollinearity?

- When you have high VIF or multicollinearity, you are not violating any assumptions of regression
- Collinearity is a problem when isolating the effects of individual predictor variables

What are the null hypotheses for the predictor variables?

- E.g., PROC REG
- the t-test for each predictor variable tests if each is a significant predictor
 - H0: the independent variable is not a significant predictor of the dependent variable

What is the best way to plot the data?

- In PROC REG:

```
PROC REG DATA=data;
  MODEL height = age age_sqrd;
  PLOT height*age;
  PLOT height*age predicted.*age /OVERLAY;
  PLOT residual.*predicted.;
  OUTPUT out=out_reg r=resid p=pred;
RUN;
```

What is the best way to plot the data?

- 2

- In PROC GPLOT:
 - Use OUTPUT file created in PROC REG

```
PROC GPLOT DATA=out_reg;  
  PLOT height*age;  
  PLOT height*age pred*age /OVERLAY;  
  PLOT resid*pred;  
RUN;
```