

Fox Module 19: Heteroscedasticity

(The attached PDF file has better formatting.)

Homework assignment: residual plots and heteroscedasticity

This homework assignment simulates heteroscedastic data and forms residual plots. Use Excel or other statistical software.

You must use Excel or other statistical software for the student project. Learn the statistical tools in your software. Simulations are the quickest way to learn the statistical concepts.

- A. Choose a sample for the explanatory variables. The exhibits below use 1 to 100.
- B. Simulate response variables whose observed values are heteroscedastic. The exhibits below use $Y_j = X_j + X_j \times \Phi(0,1)$: the variance of the error term is proportional to the value of the X_j .
- C. Form the regression equation and derive the fitted values at each point.
- D. Plot the residuals against the observed values (left pane below).
- E. Plot the residuals against the fitted values (right pane below).
- F. Explain why the two plots are so different.

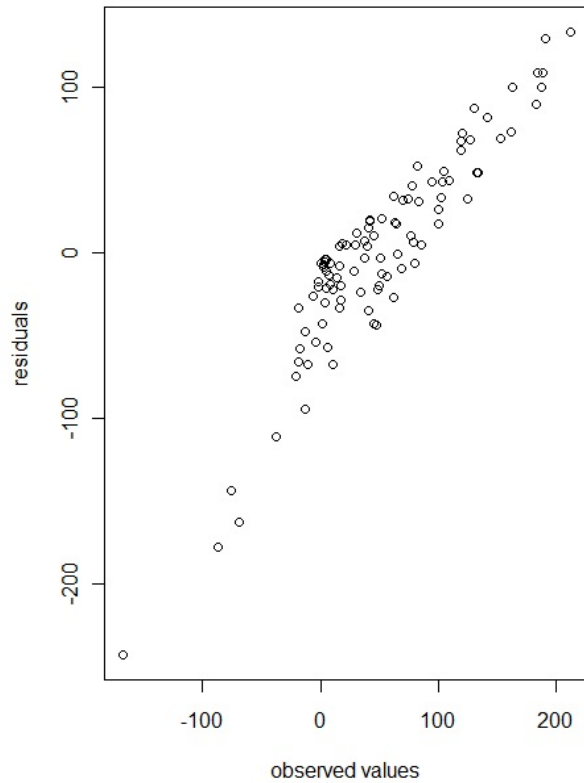
We show sample exhibits below. Your exhibits will look different, since you simulate points.

If your variance is too small, you won't see the effect that the textbook discusses.

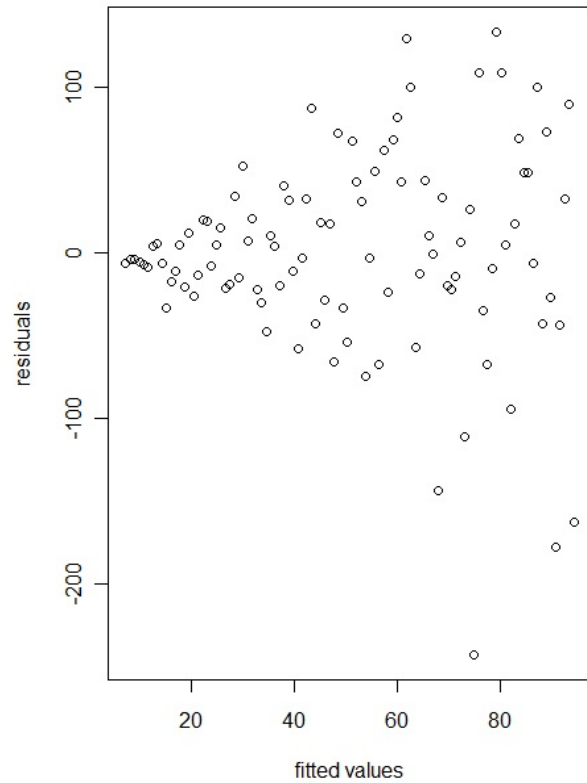
We should use $Y_j = X_j + \hat{y}_j \times \Phi(0,1)$: the variance of the error term is proportional to the fitted value of Y_j . This is harder to simulate, so we use the simpler formula above.

Form the exhibits in Excel, SAS, R, or any other software. Sample exhibits are below.

residuals vs observed values



residuals vs fitted values



```
xv <- 1:100
yv <- xv + xv * rnorm(100,0,1)
lm.hsce <- lm(yv ~ xv)
rds <- residuals(lm.hsce)
fits <- fitted(lm.hsce)
par(mfcol=c(1,2))
plot(yv, rds, xlab="observed values", ylab = "residuals", main="residuals vs observed values")
plot(fits, rds, xlab="fitted values", ylab = "residuals", main="residuals vs fitted values")
```