STAT 420: Methods of Applied Statistics

Model Diagnostics — Heteroscedasticity

Ruoqing Zhu, Ph.D. <rqzhu@illinois.edu>

Course website: https://sites.google.com/site/teazrq/teaching/STAT420

University of Illinois at Urbana-Champaign

March 14, 2017

Model Diagnostics

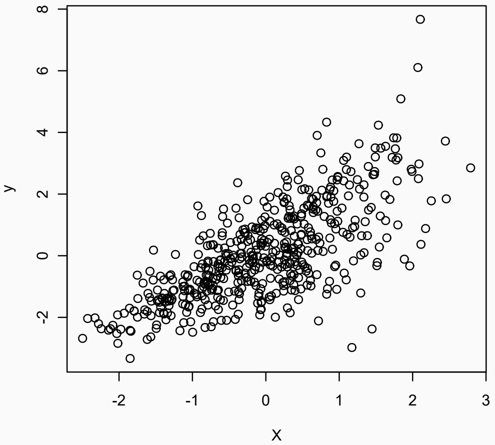
- Previously we learned how to check the normality assumption using graphical assessment and formal tests
- Even when the errors are normally distributed, they may not have the same variance
- A classical example is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\text{where} \quad \boldsymbol{\epsilon} = \mathbf{u}\ \exp(\mathbf{X}\boldsymbol{\beta}).$$

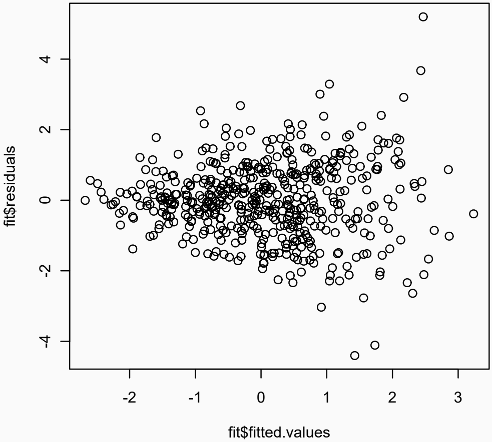
  Hence, the variance of the observation also depends on the true mean.
- Example: test the speed of a car using a radar detector. Detectors may give larger error when the speed is high.

Non-constant variances



Predictor vs. Outcome

Non-constant variances



Predicted value vs. Residual

Heteroscedasticity

- What are the potential problems of non-constant variances?
- We can still find unbiased and consistent estimates for the regression coefficients, but the estimates no longer have minimum variance, so we need to account for the differences in error variance.
- The confidence interval for a future outcome can also be too wide or too narrow, because our model doesn't account for the different variances of ϵ at each x_{new} value.

Heteroscedasticity

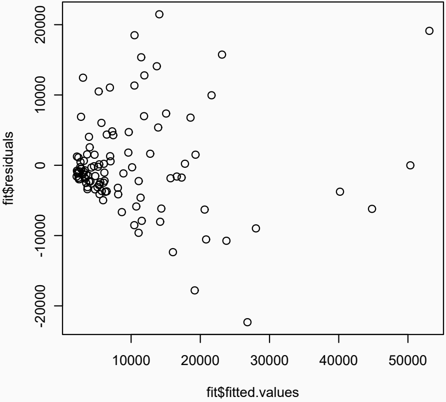
- If heteroscedasticity is present the variance of the residuals is no longer $\sigma^2\mathbf{I}$
- Instead, the variance for $\boldsymbol{\epsilon}$ is

$$\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

- Our previous estimation $\widehat{\sigma}^2$ is not estimating the correct component.
- We will discuss how to check the constant variance assumption, and also introduce a way to deal with non-constant variance situation.

Example: AT&T data

- In the AT&T data, there appears to be nonconstant variance with larger variance for larger predicted values
- How to test this?



Breusch-Pagan Test

- Suppose that we fit a linear model and obtain the errors e_i.
- We consider an auxiliary model that examines the relationship between e_i^2 and the predictors:

$$e_i^2 = \gamma_0 + \sum_{j=1}^{p} \gamma_j x_{ij} + u_i$$

- Intuition: If the errors have constant variance, then since they are already centered at 0, $E(e_i^2) = \sigma^2$. So it should not depend on any covariate.
- Hence, the above model should have $\gamma_j = 0$ for all $j = 1, \dots, p$, which leads to small coefficient of determination R^2.

Breusch-Pagan Test

- Let R^2_* be the percentage of variance that X explains e_i^2, the Breusch-Pagan statistic is defined as

$$\chi^2_{\text{BP}} = nR^2_*$$

- If $\chi^2_{\text{BP}} > \chi^2_{\alpha,p}$, we reject the null and conclude there is evidence of nonconstant variance.
- An extension of this test is the Whites test, which uses a more general function of predictors.

Example: gala data

```
# We fit the gala data with two predictors
> fit = lm(Species ~ Elevation + Adjacent, data = gala)
> library(lmtest)
>
> bptest(fit)

 studentized Breusch-Pagan test

data:  fit
BP = 7.6832, df = 2, p-value = 0.02146

> bp.fit = lm(fit$residuals^2 ~ gala$Elevation + gala$Adjacent)
> BP = summary(bp.fit)$r.squared*nrow(gala)
> BP
[1] 7.683191
> 1-pchisq(BP, 2)
[1] 0.02145934
```

Breusch-Pagan test

- Example: Calculate the BP test on the AT&T data.
- Do we reject the constant variance test?

A : Yes   B : No   C : Maybe?

White Test

- The Whites test is another popular one
- It follows exactly the same procedure as the BP test, except that all the squared and cross-produce terms of the original covariates are included.
- The reference distributed is still χ^2, but with degrees of freedom equal to the number of predictors.
- Warnings: When the white test rejects the null, there is possible heterocedasticity. But we do not have any information regarding what causes the heterocedasticity. It could be due to the model misspecification.

Example: gala data

```
# We calculate the white test
> white.fit = lm(fit$residuals^2 ~ gala$Elevation +
                                   gala$Adjacent +
                                   I(gala$Elevation^2) +
                                   I(gala$Adjacent^2) +
                                   gala$Elevation*gala$Adjacent)

> white = summary(white.fit)$r.squared*nrow(gala)

# the degrees of freedom is 5 in this case

> 1-pchisq(white, 5)
[1] 0.03022517
>
```

White test

- Example: Calculate the White test on the gala data using three predictors: Area, Elevation and Adjacent.
- Do we reject the constant variance test?

A : Yes B : No C : Maybe?

Weighted Least Squares

- Sometimes we still have to fit the linear regression model even when the variance are non-constant.
- Weighted least squares is an approach to deal with the problem and obtain a better estimate of the $\boldsymbol{\beta}$ parameters.
- Lets first consider the case when we know the covariance matrix of ϵ, then we discuss how to estimate it.

Weighted Least Squares

- Suppose the variance for $\boldsymbol{\epsilon}$ is

$$\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \doteq \mathbf{W}^{-1}$$

- By our "famous" linear transformation property, if we multiply $\mathbf{W}^{1/2}$ on both sides of our regression equation, we have

$$\mathbf{W}^{1/2}\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{1/2}\boldsymbol{\epsilon}$$

Weighted Least Squares

- The variance of the transformed residuals is

$$\begin{aligned}\text{Var}(\mathbf{W}^{1/2}\boldsymbol{\epsilon}) &= \mathbf{W}^{1/2}\text{Var}(\boldsymbol{\epsilon})\mathbf{W}^{1/2} \\ &= \mathbf{W}^{1/2}\mathbf{W}^{-1}\mathbf{W}^{1/2} \\ &= \mathbf{I}\end{aligned}$$

- Constant variance now!
- Then we can perform the linear regression using $\mathbf{W}^{1/2}\mathbf{Y}$ as the outcome, and $\mathbf{W}^{1/2}\mathbf{X}$ as the design matrix
- The problem is that we do not know $\mathbf{W}$ in practice. Is there a way to estimate it?

Estimating w_i

- The matrix $\mathbf{W}$ is essentially a diagonal matrix with w_i's on the diagonal. We can estimate them through the Breusch-Pagan test idea: model them using the predictors:
- Perform our BP test model, and then compute the predicted squared residuals:

$$\widehat{e}_i^2 = \widehat{\gamma}_0 + \sum_{j=1}^{p} \widehat{\gamma}_j x_{ij}$$

- Then our weights are defined as

$$\widehat{w}_i = \frac{1}{\widehat{e}_i^2}$$

- You can iteratively reweight the data and then fit the original regression model, until convergence is achieved. Our $\mathbb{R}$ code provide the fitting for one iteration.

Robust variance estimator

- Still, finding the exact form of heteroscedasticity is very difficult.
- An alternative strategy is to use a robust variance estimator:

$$\mathsf{Var}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$$

- And we can replace $\mathbf{W}^{-1}$ by an estimated value:

$$\mathbf{D} = \mathsf{Diag}(e_1^2, e_2^2, \ldots, e_n^2)$$

Hence the estimated variance-covariance matrix for $\widehat{\boldsymbol{\beta}}$ is

$$(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{D}\mathbf{X}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$$

Robust variance estimator

```
# The robust variance estimator can be calculated
> e = fit$residuals
> X = cbind(1, gala$Elevation, gala$Adjacent)
> sw = solve(t(X) %*% X) %*% t(X) %*% diag(e^2) %*% X %*% solve(
    t(X) %*% X)
> round(sw, 4)
        [,1]    [,2]    [,3]
[1,] 92.8838 -0.2705  0.0605
[2,] -0.2705  0.0022 -0.0006
[3,]  0.0605 -0.0006  0.0002
# sometimes a correction factor n / (n-p-1) is used to reduce
    the bias:
> round(sw*nrow(gala)/(nrow(gala) - 3), 4)
         [,1]    [,2]    [,3]
[1,] 103.2042 -0.3005  0.0672
[2,]  -0.3005  0.0024 -0.0007
[3,]   0.0672 -0.0007  0.0002
```

Robust variance estimator

```
# The car package provides functions to calculate them
> library(car)
> vcbeta = hccm(fit, type = "hc0")
> round(vcbeta, 4)
            (Intercept) Elevation Adjacent
(Intercept)     92.8838   -0.2705   0.0605
Elevation       -0.2705    0.0022  -0.0006
Adjacent         0.0605   -0.0006   0.0002

# The bias corrected version
> vcbeta = hccm(fit, type = "hc1")
> round(vcbeta, 4)
            (Intercept) Elevation Adjacent
(Intercept)    103.2042   -0.3005   0.0672
Elevation       -0.3005    0.0024  -0.0007
Adjacent         0.0672   -0.0007   0.0002
```