Homework 5

Jing Leng
*October 30, 2014*

1

a

```r
library(faraway)
lm <- lm(Lab~Field, data = pipeline)
lms <- summary(lm)
lms
```

```
##
## Call:
## lm(formula = Lab ~ Field, data = pipeline)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -21.98  -4.07  -1.43   2.50  24.33
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.9675     1.5748   -1.25     0.21
## Field         1.2230     0.0411   29.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.86 on 105 degrees of freedom
## Multiple R-squared:  0.894,  Adjusted R-squared:  0.893
## F-statistic:  887 on 1 and 105 DF,  p-value: <2e-16
```

```r
lminf <- influence(lm)
stud <- residuals(lm)/(lms$sig*sqrt(1-lminf$hat))
plot(fitted(lm), stud)
```
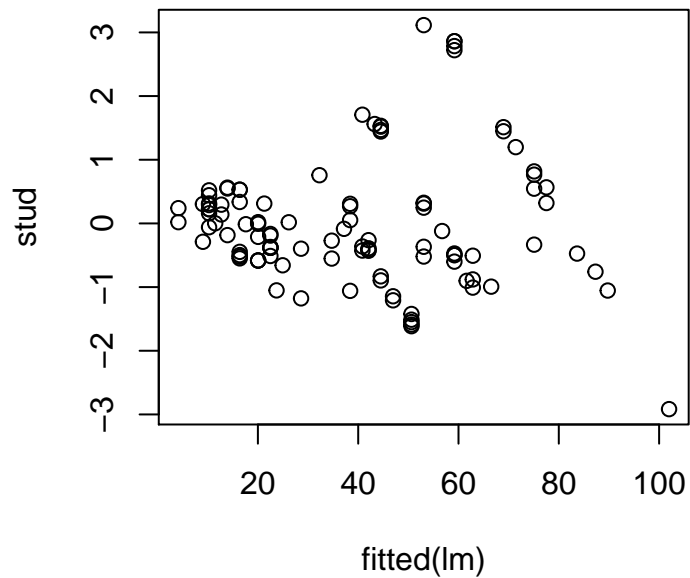
The constant variance assumption does not hold.

b

```
i <- order(pipeline$Field)
npipe <- pipeline[i,]
ff <- gl(12,9)[-108]
meanfield <- unlist(lapply(split(npipe$Field,ff),mean))
varlab <- unlist(lapply(split(npipe$Lab,ff),var))
x <- log(meanfield)
lm2 <- lm(log(varlab) ~ x)
summary(lm2)
```

```
##
## Call:
## lm(formula = log(varlab) ~ x)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -2.204 -0.673  0.166  0.721  1.189
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.354      1.571   -0.23    0.826
## x              1.124      0.462    2.44    0.035 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.02 on 10 degrees of freedom
## Multiple R-squared:  0.372,  Adjusted R-squared:  0.31
## F-statistic: 5.93 on 1 and 10 DF,  p-value: 0.0351
```

2

```
wt <- exp(predict(lm2, data.frame(x = log(pipeline$Field))))
lm3 <- lm(Lab~Field, pipeline, weights = wt)
summary(lm3)
```

```
##
## Call:
## lm(formula = Lab ~ Field, data = pipeline, weights = wt)
##
## Weighted Residuals:
##    Min    1Q Median    3Q    Max
## -206.4  -25.1  -10.5   10.6  173.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0779     2.4979   -0.03     0.98
## Field         1.1806     0.0519   22.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57.7 on 105 degrees of freedom
## Multiple R-squared:  0.832,  Adjusted R-squared:  0.83
## F-statistic:  518 on 1 and 105 DF,  p-value: <2e-16
```

$\hat{a_1} = \beta_1 = 1.124$
$\hat{a_0} = e^{\frac{\beta_0}{a_1}} = 0.73$

```
plot(pipeline$Field, pipeline$Lab)
abline(lm)
abline(lm3, lty = 2)
```
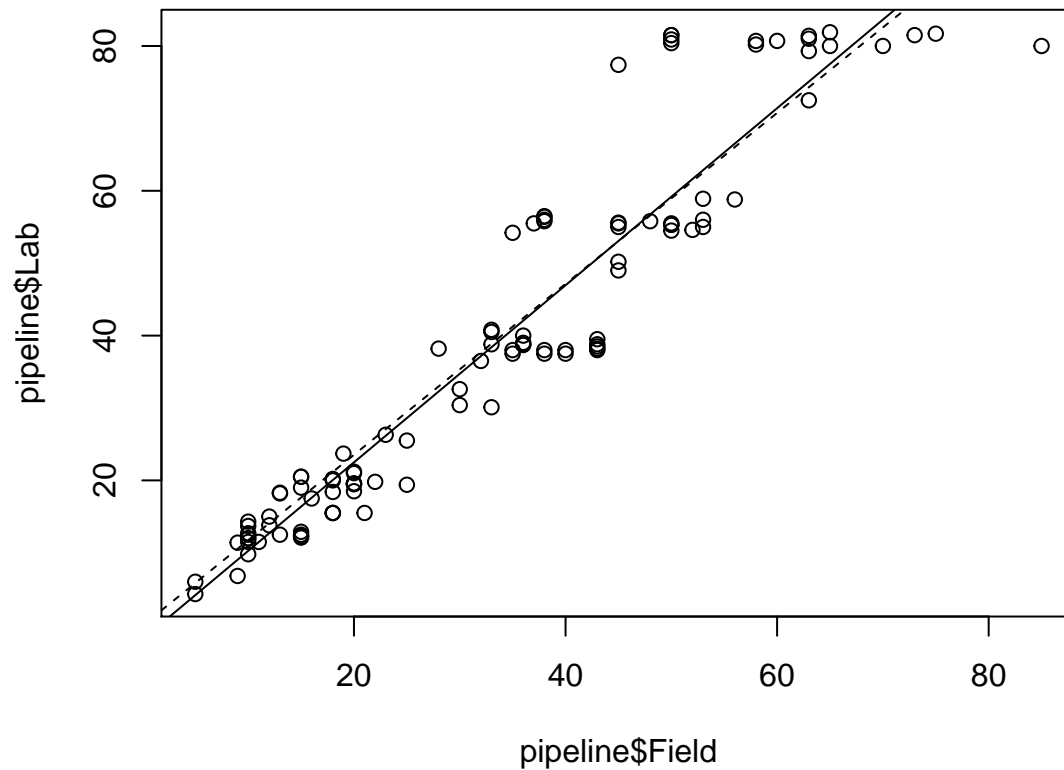
d

```
sum(lm3$residuals)
```

```
## [1] -49.84
```

```
sum(lm3$residuals*wt)
```

```
## [1] -6.189e-13
```

e

```
lm4 <- lm(Lab ~ 1, data = pipeline, weights = wt)
anova(lm4, lm3)
```

```
## Analysis of Variance Table
##
## Model 1: Lab ~ 1
## Model 2: Lab ~ Field
##   Res.Df     RSS Df Sum of Sq   F Pr(>F)
## 1    106 2072914
## 2    105  349156  1   1723758 518 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sst <- sum((lm4$residuals)^2*wt)
sst
```

```
## [1] 2072914
```

```
sse <- 349156
ssr <- sst - sse
ssr
```

```
## [1] 1723758
```

$F = 518.4$ p-value < 0.01 We reject the null hypothesis. sst = 2072914, ssr = 1723758, sse = 349156.

2

a

```
setwd("~/Documents/GoBlue/F14/F14R")
icu <- read.csv("icu.csv")
icu$race <- factor(icu$race)
glm <- glm(sta ~ ., family=binomial, data = icu)
summary(glm)
```

```
##
## Call:
## glm(formula = sta ~ ., family = binomial, data = icu)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.296  -0.682  -0.490  -0.283   2.682
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.1075     0.8601   -4.78  1.8e-06 ***
## age           0.0284     0.0115    2.47    0.014 *
## can           0.2607     0.6181    0.42    0.673
## cpr           1.5394     0.6206    2.48    0.013 *
## inf           0.8800     0.3958    2.22    0.026 *
## race2         0.9163     0.4532    2.02    0.043 *
## race3         0.4227     0.4912    0.86    0.389
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 176.29  on 193  degrees of freedom
## AIC: 190.3
##
## Number of Fisher Scoring iterations: 5
```

No, because of lack of distribution theory.

b

```
glm2 <- glm(sta~age + can + race, family = binomial, data = icu)
anova(glm2, glm)
```

```
## Analysis of Deviance Table
##
## Model 1: sta ~ age + can + race
## Model 2: sta ~ age + can + cpr + inf + race
##   Resid. Df Resid. Dev Df Deviance
## 1       195        189
## 2       193        176  2     12.8
```

```
 1-pchisq(12.84,2)
```

```
## [1] 0.001629
```

```
library(aod)
```

```
##
## Attaching package: 'aod'
##
## The following objects are masked from 'package:faraway':
##
##     rats, salmonella
```

```
wald.test(b = coef(glm), Sigma = vcov(glm), Terms = c(4,5))
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 12.4, df = 2, P(> X2) = 0.0021
```

LRT: deviance = 12.84, p-value = 0.0016. WALD: X2 = 12.4, p-value = 0.0021.
We reject the null hypothesis.

c

```
x <- model.matrix(glm)[,-1]
newdat <- data.frame(sta=glm$y,x)
glm3 <- glm(sta ~ age + can + cpr + inf + I(race2+race3), family = binomial, data = newdat)
anova(glm3, glm)
```

```
## Analysis of Deviance Table
##
## Model 1: sta ~ age + can + cpr + inf + I(race2 + race3)
## Model 2: sta ~ age + can + cpr + inf + race
##   Resid. Df Resid. Dev Df Deviance
## 1       194        177
## 2       193        176  1    0.962
```

```
1-pchisq(0.962,1)
```

```
## [1] 0.3267
```

```
lc <- cbind(0,0,0,0,0,1,-1)
wald.test(b = coef(glm), Sigma = vcov(glm), L=lc)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 0.95, df = 1, P(> X2) = 0.33
```

We fail to reject the null hypothesis.

d

```
confint(glm3)
```

```
## Waiting for profiling to be done...
```

```
##                      2.5 %    97.5 %
## (Intercept)       -5.981799 -2.58487
## age                0.008163  0.05357
## can               -1.124055  1.36190
## cpr                0.291248  2.76471
## inf                0.078733  1.61747
## I(race2 + race3)  -0.070339  1.48898
```

(-0.07, 1.49)