Regression Analysis
Chapter 4: Interpretation

Shaobo Jin

Department of Mathematics

Interpreting Parameter Estimates

Suppose that the estimated model is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

Note that

$$\begin{aligned}\hat{y}\left(x_1 = a\right) &= \hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 x_2, \\ \hat{y}\left(x_1 = a + 1\right) &= \hat{\beta}_0 + \hat{\beta}_1 \left(a + 1\right) + \hat{\beta}_2 x_2.\end{aligned}$$

We usually interpret $\hat{\beta}_j$, $j \neq 0$, as one unit increase in x_j results in an expected change of $\hat{\beta}_j$ in y, with all the other terms in the model held fixed.

An Example

```
summary(lm(Fuel ~ Tax + Dlic + Income + log2(Miles), data=fuel2001))

##
## Call:
## lm(formula = Fuel ~ Tax + Dlic + Income + log2(Miles), data = fuel2001)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -163.145  -33.039    5.895   31.989  183.499
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 154.1928   194.9062   0.791 0.432938
## Tax          -4.2280     2.0301  -2.083 0.042873 *
## Dlic          0.4719     0.1285   3.672 0.000626 ***
## Income       -6.1353     2.1936  -2.797 0.007508 **
## log2(Miles)  18.5453     6.4722   2.865 0.006259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Interpreting Parameter Estimates

Is it always possible to have "with all the other terms in the model held fixed"?

Yes/No?

Consider the examples

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2,$$
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1^2.$$

Can we interpret β_3 as an one-unit increase in x_1 corresponds to a change of β_1 in y, holding the other terms fixed?

Berkeley Guidance Study

```
data("BGSgirls", package = "alr4") # Load data
head(BGSgirls)

##     WT2  HT2  WT9   HT9  LG9 ST9 WT18  HT18 LG18 ST18 BMI18 Soma
## 67 13.6 87.7 32.5 133.4 28.4  74 56.9 158.9 34.6  143  22.5  5.0
## 68 11.3 90.0 27.8 134.8 26.9  65 49.9 166.0 33.8  117  18.1  4.0
## 69 17.0 89.6 44.4 141.5 31.9 104 55.3 162.2 35.1  143  21.0  5.5
## 70 13.2 90.3 40.5 137.1 31.8  79 65.9 167.8 39.3  148  23.4  5.5
## 71 13.3 89.4 29.9 136.1 27.7  83 62.3 170.9 36.3  152  21.3  4.5
## 72 11.3 85.5 22.8 130.6 23.4  60 47.4 164.9 31.8  126  17.4  3.0
```

Data from the Berkeley guidance study of children born in 1928-29 in Berkeley, CA.

1. BMI18: Body Mass Index at age 18
2. WT2: Age 2 weight (kg)
3. WT9: Age 9 weight (kg)
4. WT18: Age 18 weight (kg)

Berkeley Guidance Study

Even models with only linear terms can cause issues. Heavier girls at age 2 tend to have a lower BMI18.

```
summary(lm(BMI18 ~ WT2 + WT9 + WT18, data = BGSgirls))

##
## Call:
## lm(formula = BMI18 ~ WT2 + WT9 + WT18, data = BGSgirls)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1037 -0.7432 -0.1240  0.8320  4.3485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.30978    1.65517   5.020 4.16e-06 ***
## WT2         -0.38663    0.15145  -2.553    0.013 *
## WT9          0.03141    0.04937   0.636    0.527
## WT18         0.28745    0.02603  11.044  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 66 degrees of freedom
## Multiple R-squared:  0.7772,Adjusted R-squared:  0.767
## F-statistic: 76.73 on 3 and 66 DF,  p-value: < 2.2e-16
```

Berkeley Guidance Study: New Model

Now we define new variables and fit a new model using the new variables as regressors.

```
## Define new variables
BGSgirls$DW9 <- BGSgirls$WT9 - BGSgirls$WT2
BGSgirls$DW18 <- BGSgirls$WT18 - BGSgirls$WT9
summary(lm(BMI18 ~ WT2 + DW9 + DW18, data = BGSgirls))
```

Berkeley Guidance Study: New Model

These two models have the same R^2, etc, but different estimated coefficients and WT2 becomes not significant,

```
##
## Call:
## lm(formula = BMI18 ~ WT2 + DW9 + DW18, data = BGSgirls)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1037 -0.7432 -0.1240  0.8320  4.3485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.30978    1.65517   5.020 4.16e-06 ***
## WT2         -0.06778    0.12751  -0.532    0.597
## DW9          0.31886    0.03855   8.271 8.68e-12 ***
## DW18         0.28745    0.02603  11.044  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 66 degrees of freedom
## Multiple R-squared:  0.7772,Adjusted R-squared:  0.767
## F-statistic: 76.73 on 3 and 66 DF,  p-value: < 2.2e-16
```

Linear Combinations of Variables

Suppose that we have regressed y on $\boldsymbol{x}$ and obtained the OLS estimator

$$\hat{\boldsymbol{\beta}}_X = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}.$$

Say that we regress y on $\boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{C}\boldsymbol{x}$ for a known matrix $\boldsymbol{C}$ that is invertible. The OLS estimator is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_Z &= \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{y} \\ &= \left(\boldsymbol{C}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^T\boldsymbol{X}^T\boldsymbol{y} \\ &= \boldsymbol{C}^{-1}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}.\end{aligned}$$

Hence, the estimated coefficients are equivalent, but as long as predictors are correlated, interpretation of the effect of a predictor depends not only on the other predictors in a model but also upon which linear transformation of those variables is used.

Transformation

Besides the linear transformation, logarithms are commonly used both for the response and for regressors.

- If a linear model does not work well, we may take logarithm of the response.

If we take the log of the response variable, the model becomes

$$\mathrm{E}\left[\log\left(Y\right) \mid \boldsymbol{X} = \boldsymbol{x}\right] = \boldsymbol{x}^T \boldsymbol{\beta}.$$

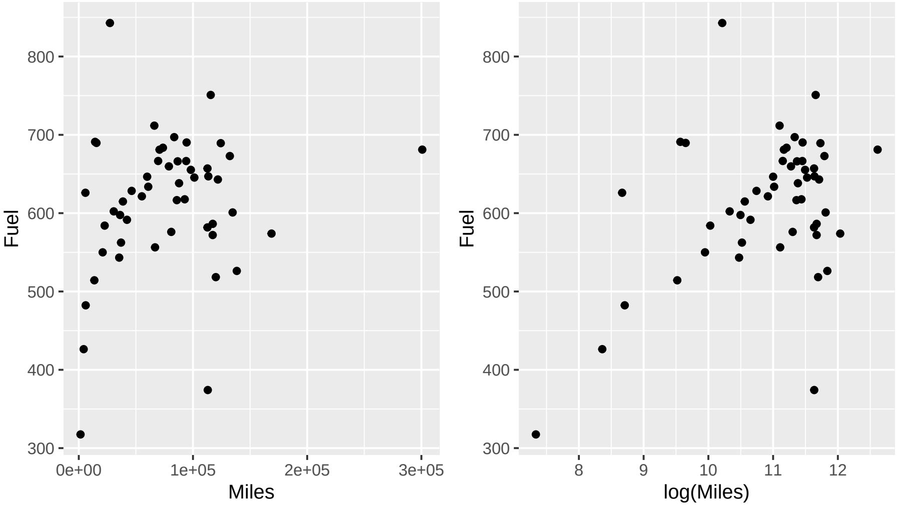
However,

$$\mathrm{E}\left[\log\left(Y\right) \mid \boldsymbol{X} = \boldsymbol{x}\right] \neq \log\left[\mathrm{E}\left(Y \mid \boldsymbol{X} = \boldsymbol{x}\right)\right],$$

even though the book suggests that

$$\mathrm{E}\left[\log\left(Y\right) \mid \boldsymbol{X} = \boldsymbol{x}\right] \approx \log\left[\mathrm{E}\left(Y \mid \boldsymbol{X} = \boldsymbol{x}\right)\right].$$

Illustration: Fuel Consumption



Example: Water Usage in Minnesota

We have a data set on yearly water consumption in Minnesota from 1988-2011. The variables are

1. muniUse: total municipal water consumption, statewide, in billions of gallons
2. year: year number
3. muniPrecip: average May to September precipiciation (inches)
4. muniPop: urban population

```
data("MinnWater", package = "alr4")
```

Water Usage in Minnesota: Model 1

```
summary(lm(log(muniUse) ~ year, data = MinnWater))

##
## Call:
## lm(formula = log(muniUse) ~ year, data = MinnWater)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.117381 -0.033371  0.004126  0.044729  0.089021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.048043   3.745726  -5.352 2.25e-05 ***
## year          0.012432   0.001873   6.636 1.13e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06353 on 22 degrees of freedom
## Multiple R-squared:  0.6669,Adjusted R-squared:  0.6517
## F-statistic: 44.04 on 1 and 22 DF,  p-value: 1.132e-06
```

Water Usage in Minnesota: Model 2

Adding muniPrecip does not change the estimated slope for year much.

```
summary(lm(log(muniUse) ~ year + muniPrecip, data = MinnWater))

##
## Call:
## lm(formula = log(muniUse) ~ year + muniPrecip, data = MinnWater)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10852 -0.03747  0.01067  0.03028  0.07094
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.158353   2.726451  -7.394 2.85e-07 ***
## year          0.012586   0.001364   9.228 7.77e-09 ***
## muniPrecip   -0.009932   0.002192  -4.531 0.000183 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04624 on 21 degrees of freedom
## Multiple R-squared:  0.8315,Adjusted R-squared:  0.8155
## F-statistic: 51.83 on 2 and 21 DF,  p-value: 7.557e-09
```

Water Usage in Minnesota: Model 3

However, adding also log(muniPop) greatly change the results.

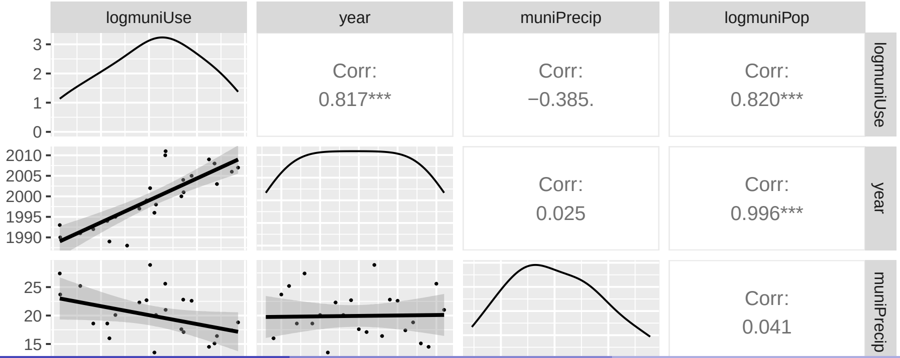
```
summary(lm(log(muniUse) ~ year + muniPrecip + log(muniPop), data = MinnWater))

##
## Call:
## lm(formula = log(muniUse) ~ year + muniPrecip + log(muniPop),
##     data = MinnWater)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.087690 -0.032781  0.000155  0.034694  0.080204
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.278394  11.508965  -0.111    0.913
## year         -0.011132   0.014141  -0.787    0.440
## muniPrecip   -0.010559   0.002135  -4.946 7.78e-05 ***
## log(muniPop)  1.917355   1.138236   1.684    0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04434 on 20 degrees of freedom
```

Water Usage in Minnesota

The correlation between year and log(muniPop) is large, and relation is very linear.

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```



Multicollinearity

Multicollinearity exists when some regressors are highly correlated. For example,

1. the correlation between two regressors is too large.
2. there exists a vector of constants $\boldsymbol{c}$ such that $\boldsymbol{X}\boldsymbol{c} \approx \mathbf{0}$.

When it exits,

1. the estimated regression coefficients may change dramatically,
2. we may not event get the estimates.

Non-Existence of OLS Estimator

For simplicity, suppose that

1. the response has a zero sample mean, i.e., $\bar{y} = 0$,
2. all regressors are demeaned, i.e., $\bar{x}_j = 0$ for all j,
3. the intercept is not included in the model.

Then, $\boldsymbol{X}^T\boldsymbol{X}/n$ is the sample covariance matrix of the regressors and $\boldsymbol{X}^T\boldsymbol{y}/n$ is the sample covariance between $\boldsymbol{x}$ and Y. If multicollinearity exits, say a column in $\boldsymbol{X}$ is a linear combination of other columns, then $\boldsymbol{X}^T\boldsymbol{X}$ is not of full rank and $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ does not exist.

Dropping Regressors

Consider the model

$$\mathrm{E}\left(Y \mid X_1 = x_1, X_2 = x_2\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Now we instead consider the model

$$\mathrm{E}\left(Y \mid X_1 = x_1\right) = \beta_0 + \beta_1 x_1.$$

We can show that

$$\mathrm{E}\left(Y \mid X_1 = x_1\right) = \beta_0 + \beta_1 x_1 + \beta_2 \mathrm{E}\left(X_2 \mid X_1 = x_1\right).$$

If $\mathrm{E}\left(X_2 \mid X_1 = x_1\right) = \gamma_0 + \gamma_1 x_1$, then

$$\mathrm{E}\left(Y \mid X_1 = x_1\right) = \left(\beta_0 + \beta_2 \gamma_0\right) + \left(\beta_1 + \beta_2 \gamma_1\right) x_1.$$

We are estimating $\beta_0 + \beta_2 \gamma_0$ and $\beta_1 + \beta_2 \gamma_1$.

An Example

```
data(usair , package = "gamlss.data")
LR0 <- lm(y ~ x1 + x2, data = usair)
LR1 <- lm(y ~ x1, data = usair)
LR2 <- lm(x2 ~ x1, data = usair)
coef(LR1)

## (Intercept)          x1
##  108.571058   -1.408133

c(coef(LR0)[1] + coef(LR0)[3] * coef(LR2)[1],
  coef(LR0)[2] + coef(LR0)[3] * coef(LR2)[2])

## (Intercept)          x1
##  108.571058   -1.408133
```

Role of Independence

$$\mathrm{E}\left(Y \mid X_1 = x_1\right) = \beta_0 + \beta_1 x_1 + \beta_2 \mathrm{E}\left(X_2 \mid X_1 = x_1\right).$$

1. If $\beta_2 = 0$, we are still estimating β_0 and β_1.
2. If X_1 and X_2 are independent, then $\mathrm{E}\left(X_2 \mid x_1\right) = \mathrm{E}\left(X_2\right)$ that does not depend on X_1. Hence,

$$\mathrm{E}\left(Y \mid X_1 = x_1\right) = \left[\beta_0 + \beta_2 \mathrm{E}\left(X_2\right)\right] + \beta_1 x_1.$$

   We are still estimating β_1, but not β_0.

Omitted Variable

Suppose that the truth is

$$Y = \boldsymbol{x}_1^T\boldsymbol{\beta}_1 + \boldsymbol{x}_2^T\boldsymbol{\beta}_2 + e,$$

where $\mathrm{E}\left(\boldsymbol{x}_1\right) = \mathbf{0}$ and $\mathrm{E}\left(\boldsymbol{x}_2\right) = \mathbf{0}$.

However, we assume $\mathrm{E}\left(Y \mid \boldsymbol{x}_1\right) = \boldsymbol{x}_1^T\boldsymbol{\beta}_1$. The OLS estimator satisfies

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= \left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\boldsymbol{y} \\ &= \left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\left(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}\right) \\ &= \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{\beta}_2 + \left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\boldsymbol{e}. \end{aligned}$$

Omitted Variable Bias

If $\mathrm{E}\left(\boldsymbol{e} \mid \boldsymbol{X}_1\right) = \mathbf{0}$, then

$$\mathrm{E}\left(\hat{\boldsymbol{\beta}}_1 \mid \boldsymbol{X}_1\right) = \boldsymbol{\beta}_1 + \left(\frac{1}{n}\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\left[\frac{1}{n}\boldsymbol{X}_1^T\mathrm{E}\left(\boldsymbol{X}_2 \mid \boldsymbol{X}_1\right)\right]\boldsymbol{\beta}_2.$$

Hence, the OLS estimator of $\boldsymbol{\beta}_1$ is biased if

1. $\boldsymbol{\beta}_2 \neq \mathbf{0}$,
2. or $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are correlated.