

STA 4320 CHAP 3.2.2

Prof. He Jiang

Sep 2024

Sec 3.2.2

Credit dataset

```
require(ISLR2)
```

```
## Loading required package: ISLR2
```

```
head(Credit)
```

```
##      Income Limit Rating Cards Age Education Own Student Married Region Balance
## 1  14.891  3606    283    2  34         11 No      No      Yes  South    333
## 2 106.025  6645    483    3  82         15 Yes     Yes     Yes   West    903
## 3 104.593  7075    514    4  71         11 No      No      No    West    580
## 4 148.924  9504    681    3  36         11 Yes     No      No    West    964
## 5  55.882  4897    357    2  68         16 No      No      Yes  South    331
## 6  80.180  8047    569    4  77         10 No      No      No    South   1151
```

Regression with only “Own”

```
reg = lm(Balance ~ Own, data = Credit)
```

```
summary(reg)
```

```
##
## Call:
## lm(formula = Balance ~ Own, data = Credit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -529.54 -455.35  -60.17   334.71 1489.20
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    509.80      33.13   15.389  <2e-16 ***
## OwnYes         19.73      46.05    0.429   0.669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 460.2 on 398 degrees of freedom
## Multiple R-squared:  0.0004611, Adjusted R-squared:  -0.00205
## F-statistic: 0.1836 on 1 and 398 DF, p-value: 0.6685
```

Use ifelse to turn categories into numerical values

Credit\$Own is a factor(category) vector. We can use ifelse to convert it to numerical values.

```
x_0_1 = ifelse(Credit$Own == "Yes", 1, 0)
x_m1_1 = ifelse(Credit$Own == "Yes", 1, -1)
```

Regression with -1 and 1

```
reg_m1_1 = lm(Credit$Balance ~ x_m1_1)
summary(reg_m1_1)
```

```
##
## Call:
## lm(formula = Credit$Balance ~ x_m1_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -529.54 -455.35  -60.17   334.71 1489.20
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   519.670     23.026   22.569  <2e-16 ***
## x_m1_1         9.867     23.026    0.429    0.669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 460.2 on 398 degrees of freedom
## Multiple R-squared:  0.0004611, Adjusted R-squared: -0.00205
## F-statistic: 0.1836 on 1 and 398 DF, p-value: 0.6685
```

Regression with only “Region”

Note that R automatically created the indicator variables for the 3 level factor Region.

```
reg = lm(Balance ~ Region, data = Credit)
summary(reg)
```

```
##
## Call:
## lm(formula = Balance ~ Region, data = Credit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -531.00 -457.08  -63.25   339.25 1480.50
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    531.00     46.32   11.464  <2e-16 ***
## RegionSouth    -12.50     56.68   -0.221    0.826
## RegionWest     -18.69     65.02   -0.287    0.774
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 460.9 on 397 degrees of freedom
## Multiple R-squared:  0.0002188, Adjusted R-squared: -0.004818
## F-statistic: 0.04344 on 2 and 397 DF, p-value: 0.9575
```

Model with both Income(numerical) and Student(categorical)

```

reg = lm(Balance ~ Income + Student, data = Credit)
summary(reg)

##
## Call:
## lm(formula = Balance ~ Income + Student, data = Credit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -762.37 -331.38  -45.04   323.60   818.28
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  211.1430    32.4572   6.505 2.34e-10 ***
## Income         5.9843     0.5566  10.751 < 2e-16 ***
## StudentYes   382.6705    65.3108   5.859 9.78e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 391.8 on 397 degrees of freedom
## Multiple R-squared:  0.2775, Adjusted R-squared:  0.2738
## F-statistic: 76.22 on 2 and 397 DF,  p-value: < 2.2e-16

```