

Lab Week 03: Linear Regression - Part 2

Pouria

1/26/2022

Lab Prompt

In this lab, we will study regression with 2 predictors, one continuous, and one qualitative, using the Credit data from the package ISLR

1. Build a linear model of `balance` as a function of `student` status. Compare with a t-test. What do you conclude?
2. Now build a model with `balance` as a function of `income`. What do you conclude?
3. Next we will build a model with both regressors, `student` status and `income`.
 - 3.a) Plot the data, using different markers for `student` status
 - 3.b) Develop a model corresponding to Figure 3.7 left. Plot the model with the data, as in a). What do you conclude?
 - 3.c) Now develop and plot the full model corresponding to Figure 3.7. Compare the result with those of an ANCOVA analysis. What do you conclude?

Lab 1

Balance ~ Student

##

Call:

lm(formula = Balance ~ Student)

##

Residuals:

##	Min	1Q	Median	3Q	Max
##	-876.82	-458.82	-40.87	341.88	1518.63

##

Coefficients:

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	480.37	23.43	20.50	< 2e-16 ***
##	StudentYes	396.46	74.10	5.35	1.49e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

##

Residual standard error: 444.6 on 398 degrees of freedom

Multiple R-squared: 0.06700 Adjusted R-squared: 0.0

Lab 1

Compare with t-test

```
##
##  Welch Two Sample t-test
##
## data:  Balance[Student == "Yes"] and Balance[Student ==
## t = 4.9028, df = 46.241, p-value = 1.205e-05
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  233.7088 559.2023
## sample estimates:
## mean of x mean of y
##  876.8250  480.3694

##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  Balance and Student
##
```

Lab 2

Balance ~ Income

##

Call:

lm(formula = Balance ~ Income)

##

Residuals:

##	Min	1Q	Median	3Q	Max
##	-803.64	-348.99	-54.42	331.75	1100.25

##

Coefficients:

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	246.5148	33.1993	7.425	6.9e-13 ***
##	Income	6.0484	0.5794	10.440	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

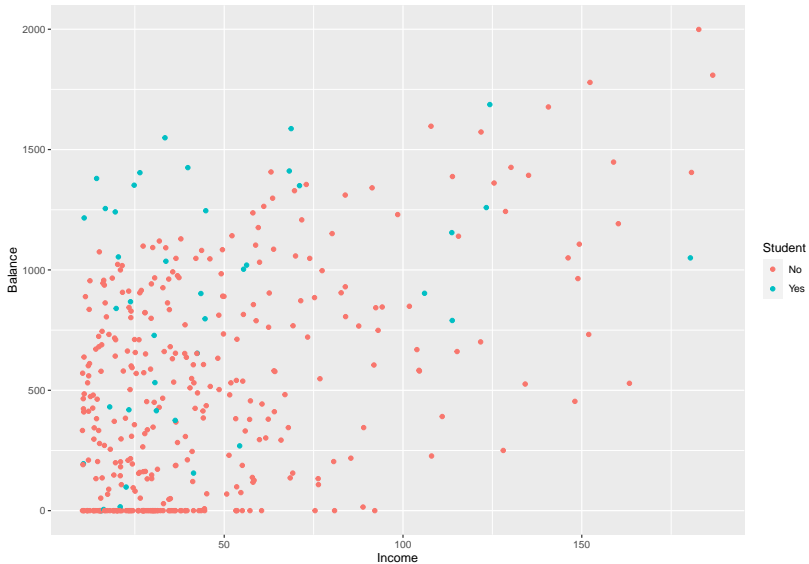
##

Residual standard error: 407.9 on 398 degrees of freedom

Multiple R-squared: 0.215 Adjusted R-squared: 0.212

Lab 3(a)

Plot the data points



Lab 3(b)

Balance ~ Income + Student

##

Call:

lm(formula = Balance ~ Student_dummy + Income)

##

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	***
##	Student_dummy1	382.6705	65.3108	5.859	9.78e-09	***
##	Income	5.9843	0.5566	10.751	< 2e-16	***

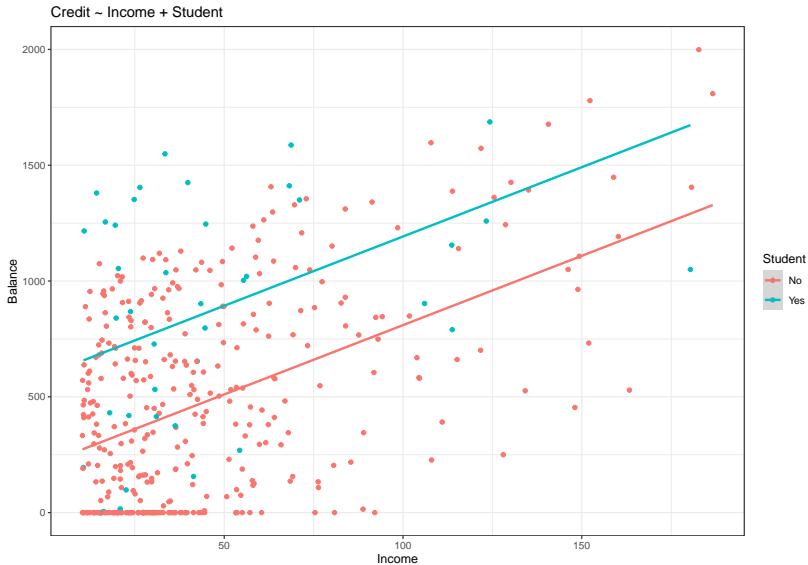
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

##

Residual standard error: 301.8 on 307 degrees of freedom

Lab 3(b)

Plot $\text{lm}(\text{Balance} \sim \text{Income} + \text{Student})$



Lab 3(c)

Balance ~ Income*Student

##

Call:

lm(formula = Balance ~ Income * Student_dummy)

##

Residuals:

##	Min	1Q	Median	3Q	Max
##	-773.39	-325.70	-41.13	321.65	814.04

##

Coefficients:

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	200.6232	33.6984	5.953	5.79e-
## Income	6.2182	0.5921	10.502	< 2e-
## Student_dummy1	476.6758	104.3512	4.568	6.59e-
## Income:Student_dummy1	-1.9992	1.7313	-1.155	0.2

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

##

Lab 3(c)

Balance ~ Student + Income:Student

##

Call:

lm(formula = Balance ~ Student_dummy + Income:Student_dummy,

##

Residuals:

##	Min	1Q	Median	3Q	Max
----	-----	----	--------	----	-----

##	-773.39	-325.70	-41.13	321.65	814.04
----	---------	---------	--------	--------	--------

##

Coefficients:

##

##		Estimate	Std. Error	t value	Pr(> t)
----	--	----------	------------	---------	----------

##	(Intercept)	200.6232	33.6984	5.953	5.79e-
----	-------------	----------	---------	-------	--------

##	Student_dummy1	476.6758	104.3512	4.568	6.59e-
----	----------------	----------	----------	-------	--------

##	Student_dummy0:Income	6.2182	0.5921	10.502	< 2e-
----	-----------------------	--------	--------	--------	-------

##	Student_dummy1:Income	4.2190	1.6269	2.593	0.009
----	-----------------------	--------	--------	-------	-------

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

##

Lab 3(c)

Compare with ANCOVA

```
## Analysis of Variance Table
##
## Model 1: Balance ~ Income * Student
## Model 2: Balance ~ Student + Income
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1      396 60734545
## 2      397 60939054 -1    -204509 1.3334 0.2489
```

Lab 3(c)

Plot the full model

