Session 11 – Linear Models

Assignment – 2

1. Use the given link below and locate the bank marketing dataset. Data Set Link

Perform the below operations:

1. Is there any association between Job and default?

Ho : There is NO association between Job and default

chisq.test(missing$job, missing$default)

Pearson's Chi-squared test

data: missing$job and missing$default

X-squared = 8.5797, df = 10, p-value = 0.5724

# Since P Value is greater than 0.05 ,

# there is no association between Job and default at 95% confidence level

# Since NA values are very less, are omitted

1. Is there any significant difference in duration of last call between people having housing loan or not?

library(dplyr)

by(missing[,-5],missing$duration,colMeans)

tapply(missing$duration,missing$housing,mean)

no yes

259.2905 267.5424

The mean duration of last call between people having housing loan or not is not significant

1. Is there any association between consumer price index and consumer?

HO : consumer price index across consumer is same

x <- aov(cons.price.idx~age, data = full)

> summary(x)

Df Sum Sq Mean Sq F value Pr(>F)

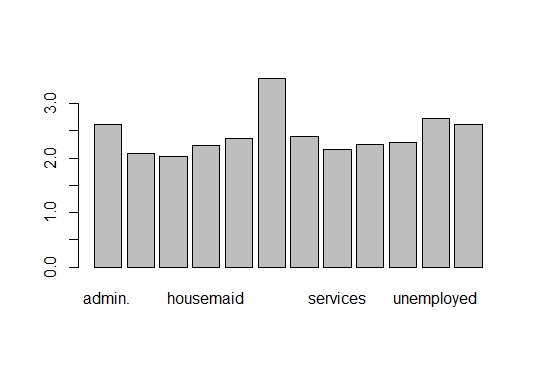
age 1 0 0.0101 0.03 0.862

Residuals 41186 13800 0.3351

Since p value is greater than 0.05, price index across the consumer is

same at 95% confidence level

1. Is the employment variation rate consistent across job types?
2. A<-tapply(full$emp.var.rate,full$job,var)
3. > A
4. admin. blue-collar entrepreneur housemaid management retired
5. 2.626108 2.084261 2.038192 2.230102 2.353027 3.457269
6. self-employed services student technician unemployed unknown
7. 2.402874 2.163692 2.248265 2.291772 2.727908 2.616184
8. > barplot(A)



The employment variation rate is not the same across all job types. Retired

Individuals employment variation rate is higher than any other individual

types

e. Is the employment variation rate same across education?

A<-tapply(full$emp.var.rate,full$education,mean)

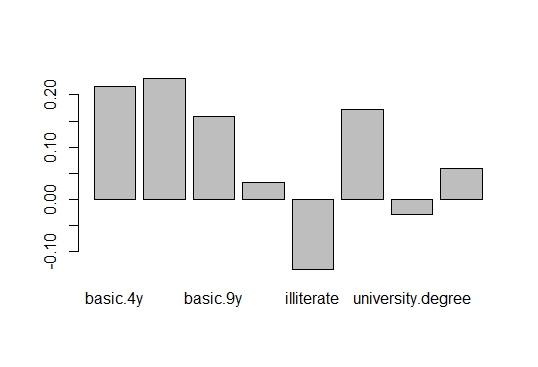
> A

basic.4y basic.6y basic.9y high.school

0.21592433 0.23115183 0.15923904 0.03293747

illiterate professional.course university.degree unknown

-0.13333333 0.17301163 -0.02809007 0.05909879



The employment variation rate is not same across education

f. Which group is more confident?

A<-tapply(full$cons.conf.idx,full$job,mean)

> A

admin. blue-collar entrepreneur housemaid management retired

-40.24543 -41.37582 -41.28365 -39.49528 -40.48947 -38.57308

self-employed services student technician unemployed unknown

-40.48811 -41.29005 -40.18754 -39.92757 -40.00759 -38.79788

The Blue collar is the most confident as per the data.