Assignment 2 Decs 435

Sebastian Salomon

2022-09-27

Assignment 1

* Learning Checks

Answer the following questions in your own words. None of your answers should be longer than one paragraph.

1. (easy) What is the standard error of an estimate? Why is it useful?

This represents the average distance that the observed values fall from the regression line. Useful since it tells us how good or bad our current predicted line is to the actual line.

2. (medium) Explain why it is wrong to claim that "if an estimate is statistically significant at the 5%-level, then the probability that the true value equals 0 is less than 5%."

When an estimate is statistically significant at 5% level then this means that there is a 5% risk of concluding that a difference exists when there is no actual difference. This is the probability of rejecting the null, not the probability of accepting the alternative. So statistical at 5% does not mean that the probability that the true value equals 0 is 5%.

3. (hard) What is the difference between a one-sided and a two-sided hypothesis test? Give an example of a business problem in which you should use a one-sided rather than a two-sided test.

One tailed tests allow for the possibility of an effect in one direction, where two-tailed tests account for the possibility of an effect in two directions (positive and negative). So an example of a business problem where we would use a one-sided test could be when determining the probability of increasing our current margins by a certain percentage. Since you are only taking into account the probability of increasing, you ignore completely the left hand side of the tail and it becomes one-sided. On the other hand if we want to take the probability the the margins are going to be within a certain bound, that's when we use a two-sided test since we have to take into account the negative and the positive side of the tails.

* Benchmarking Chikuho Bank

4. Choose an appropriate benchmarking group for Chikuho based only on subsidiary status (see the variable subsidiary).

```r
library(tidyverse)
library(ggplot2)
library(haven)
library(knitr)
library(kableExtra)
```

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 3 > 1' in coercion to 'logical(1)'
```

```r
library(sjPlot)

hospital_data <- read_dta('C:/Users/ssalo/Documents/2022FA_DECS_435-0_SEC1/banksjapan_benchmarking.dta')





df_benchm <- hospital_data %>% filter(., subsidiary == 0) %>%         #< exclude subsidiary banks
   mutate(av_margin = mean(margin),
         simple_rank = dense_rank(desc(margin))) %>%     #< rank observations by margin, FROM HIGH TO L
   arrange(simple_rank)                            #< sort by naïve rank

head(df_benchm %>% select(., name, margin, simple_rank))
```

```
## # A tibble: 6 x 3
##   name                                                          margin simpl~1
##   <chr>                                                          <dbl>   <int>
## 1 SHINKIN CENTRAL BANK                                            54.1       1
## 2 SHIZUOKA BANK                                                   53.2       2
## 3 AOZORA BANK LTD                                                 48.3       3
## 4 MITSUBISHI UFJ TRUST AND BANKING CORPORATION - MITSUBISHI UFJ ~  46.1       4
## 5 CHIBA BANK LTD.                                                 43.8       5
## 6 KIYO BANK                                                       42.8       6
## # ... with abbreviated variable name 1: simple_rank
```

The benchmarking group that I chose starts by excluding subsidiary banks, and then including a simple_rank chosen by the margin, and the mean margin is done to compare the current margin with the margin for specific banks.

How many banks are in this group?

```r
print(paste0("Number of banks in benchmarking group: ", nrow(df_benchm)))
```

```
## [1] "Number of banks in benchmarking group: 83"
```

```r
head(df_benchm) #< "best" banks
```

```
## # A tibble: 6 x 29
```

```
##   os_id_num~1 name  subsi~2 city  inmet~3 metro~4 metro~5 island year_~6 total~7
##   <chr>       <chr>   <dbl> <chr>   <dbl> <chr>     <dbl> <chr>    <dbl>   <dbl>
## 1 JP48842     SHIN~       0 TOKYO       1 Kanto     35.7  Honshu    1950   30.1
## 2 JP47937     SHIZ~       0 SHIZ~       1 Shizuo~    1.40 Honshu    1943   10.2
## 3 JP36702     AOZO~       0 TOKYO       1 Kanto     35.7  Honshu    1957    4.99
## 4 JP34194     MITS~       0 TOKYO       1 Kanto     35.7  Honshu    1927   29.3
## 5 JP47966     CHIB~       0 CHIB~       1 Kanto     35.7  Honshu    1943   11.3
## 6 JP17202     KIYO~       0 WAKA~       1 Keihan~   18.8  Honshu    1895    3.91
## # ... with 19 more variables: fixedassets <dbl>, executives <dbl>,
## #   customerloans <dbl>, problemloans <dbl>, probloanpct <dbl>,
## #   depositschecking <dbl>, depositssavings <dbl>, netinterest <dbl>,
## #   netcommisions <dbl>, otherincome <dbl>, totalincome <dbl>, taxes <dbl>,
## #   employees <dbl>, roae <dbl>, roa <dbl>, margin <dbl>, totalexpenses <dbl>,
## #   av_margin <dbl>, simple_rank <int>, and abbreviated variable names
## #   1: os_id_number, 2: subsidiary, 3: inmetroarea, 4: metroareaname, ...
```

What is Chikuho's position in this group when the group is naively ordered by profit margin?

```
df_benchm %>% select(name, av_margin, margin, simple_rank) %>%
  filter(name=="CHIKUHO BANK")                    #< find Chikuho Bank
```

```
## # A tibble: 1 x 4
##   name         av_margin margin simple_rank
##   <chr>            <dbl>  <dbl>       <int>
## 1 CHIKUHO BANK      23.9   10.0          76
```

Chikuho bank is ranked #76.

5. Use regression analysis to develop a meaningful ranking. Be sure to follow all of the necessary steps in model building. Be sure to have a cogent business justification for your variable choices. You should draw on information in the case. You can also feel free to use concepts that you have learned in other courses at Kellogg, or through your past experiences that might help you identify potentially relevant factors.

The variables that I chose for the regression are metropopulation since they mentioned that it could be a factor, also year_of_incorporation, since the article pointed to the fact that the year were they started might have an impact, and also the size of the bank. The article mentioned that medium sized banks might have an advantage over big and small banks.

```
df_benchm_improved <- df_benchm %>%
   mutate(size_tercile = factor(ntile(totalassets, 3))) #< create factor variable for tercile to contro
bm_reg <- lm(margin ~ metropop + size_tercile + year_of_incorporation, data = df_benchm_improved)
tab_model(bm_reg)
```

Profit Margin

Predictors

Estimates

CI

p

(Intercept)

-4.60

-152.05 – 142.85

0.951

population ofmetroarea/prefecture

0.13

-0.02 – 0.28

0.079

size_tercile:size_tercile2

7.39

2.42 – 12.36

0.004

size_tercile:size_tercile3

14.82

9.64 – 20.00

<0.001

Date of incorporation(Year)

0.01

-0.07 – 0.09

0.792

Observations

83

R2 / R2 adjusted

0.375 / 0.343

```
#summary(bm_reg)


df_benchm_improved <- df_benchm_improved %>%
 mutate(predicted_value = bm_reg$fitted.values,
       residual = margin - predicted_value) %>% #< calculate residuals
mutate(better_rank = dense_rank(desc(residual))) %>% #< need to rank HIGH to LOW!
      arrange(better_rank)                       #< sort by better rank
```

Where is Chikuho Bank in your ranking?

```
df_benchm_improved %>% select(name, margin, predicted_value, residual, better_rank) %>%
  filter(name=="CHIKUHO BANK")                     #< find Chikuho Bank
```

```
## # A tibble: 1 x 5
##   name          margin predicted_value residual better_rank
##   <chr>          <dbl>           <dbl>    <dbl>       <int>
## 1 CHIKUHO BANK    10.0            16.0    -5.93          66
```

List the top 5 banks according to your ranking.

```
head(df_benchm_improved %>% select(name, margin, predicted_value, residual, better_rank)) #< best banks
```

```
## # A tibble: 6 x 5
##   name                    margin predicted_value residual better_rank
##   <chr>                    <dbl>           <dbl>    <dbl>       <int>
## 1 SHIZUOKA BANK             53.2            30.1     23.1           1
## 2 SHINKIN CENTRAL BANK      54.1            34.8     19.3           2
## 3 KIYO BANK                 42.8            24.5     18.3           3
## 4 BANK OF OKINAWA           32.5            15.4     17.0           4
## 5 SURUGA BANK, LTD. (THE)   41.8            26.8     15.0           5
## 6 AOZORA BANK LTD           48.3            34.9     13.5           6
```

List the bottom 5 banks according to your ranking.

```
tail(df_benchm_improved %>% select(name, margin, predicted_value, residual, better_rank)) #< best banks
```

```
## # A tibble: 6 x 5
##   name                             margin predicted_value residual better_rank
##   <chr>                             <dbl>           <dbl>    <dbl>       <int>
## 1 SHIGA BANK, LTD (THE)             20.7             32.4    -11.6          78
## 2 TOYAMA BANK, LTD, (THE)            2.93            15.3    -12.4          79
## 3 SENSHU IKEDA BANK LTD             17.3             32.1    -14.8          80
## 4 NANTO BANK LTD. (THE)             14.8             32.4    -17.6          81
## 5 TOKYO TOMIN BANK, LTD. (THE)       8.08            27.4    -19.3          82
## 6 KANSAI URBAN BANKING CORPORATION   2.96            25.0    -22.0          83
```

6. Prepare a Deficiency Table based on your regression to help explain Chikuho Bank's relatively poor margins. What are the main conclusions you draw from the table you construct?

```
metropop <-df_benchm_improved %>% filter(name=="CHIKUHO BANK") %>% select(metropop)
 metropop_avg <-mean(df_benchm_improved$metropop)
 metropop_delta <- metropop-metropop_avg
 metropop_beta <- bm_reg$coefficients[2]
 size_tercile <-df_benchm_improved %>% filter(name=="CHIKUHO BANK") %>% select(size_tercile)
 size_tercile_avg <- mean(as.numeric(df_benchm_improved$size_tercile))
 size_tercile_delta <- as.numeric(size_tercile) - as.numeric(size_tercile_avg)
 size_tercile_beta <- 0# bm_reg$coefficients[3]
year_of_incorporation<- df_benchm_improved %>% filter(name=="CHIKUHO BANK") %>% select(year_of_incorpora
 year_of_incorporation_avg <-mean(df_benchm_improved$year_of_incorporation)
 year_of_incorporation_delta <- year_of_incorporation - year_of_incorporation_avg
```

	Chikuho_Bank	Sample_Average	Delta	B	(B * Delta)
metropop	5.6	12.08916	-6.489157	0.1340758	-0.8700386
size_tercile	1	1.987952	-0.9879518	0	0
year_of_incorporation	1952	1927.554	24.44578	0.01015207	0.2481753
Relative_to_Avg					-0.6218634

```
 year_of_incorporation_beta <-  bm_reg$coefficients[5]

deficiency_table <- matrix(c(metropop, metropop_avg,metropop_delta ,metropop_beta, metropop_delta*(metro
      size_tercile, size_tercile_avg, size_tercile_delta, size_tercile_beta,size_tercile_delta*(size_ter
      year_of_incorporation, year_of_incorporation_avg, year_of_incorporation_delta, year_of_incorporati
      '','','','',metropop_delta*(metropop_beta) + size_tercile_delta*(size_tercile_beta) + year_of_inco
rownames(deficiency_table) <- c('metropop','size_tercile','year_of_incorporation', 'Relative_to_Avg' )
colnames(deficiency_table) <- c('Chikuho_Bank','Sample_Average','Delta', 'B', '(B * Delta)')



as.data.frame(deficiency_table) %>%
  kbl() %>%
  kable_classic_2(full_width = F)
```

The deficiency table indicates that the metropop variable had the biggest effect for the margins on Chikuho_Bank. Compared to the average bank, Chikuho_Bank was predicted to have -.62 fewer margin percentage. We're finding some reason on why Chikuho_Bank is doing worse than other banks, but not enough to determine the magnitude of why it's doing as bad as it's actually doing.

7. Using the same benchmarking group as you did for Part I, augment your previous regression so that it can help you identify changes that Chikuho might make in order to improve its profit margins. What variables did you add and why?

```
df_benchm_final <- df_benchm_improved %>% mutate(execratio = executives/ employees)
bm_reg2 <- lm(margin ~ execratio + probloanpct + metropop + size_tercile, data = df_benchm_final)
#summary(bm_reg2)

tab_model(bm_reg2)
```

Profit Margin

Predictors

Estimates

CI

p

(Intercept)

28.02

20.86 – 35.19

<0.001

Number of corporateexecutives

-1.41

-3.15 – 0.32

0.108

problemloans/customerloans

-316.70

-477.57 – -155.83

<0.001

population ofmetroarea/prefecture

0.14

0.00 – 0.27

0.043

size_tercile:size_tercile2

5.10

0.53 – 9.67

0.029

size_tercile:size_tercile3

10.45

5.45 – 15.45

<0.001

Observations

83

R2 / R2 adjusted

0.489 / 0.456

For this question I used the same variables as before, but added execratio and probloanpct. These variables were important to see if the ratio between executives and employees had an impact on Banks margins. And also to check the ratio of problem loans to regular loans and decide if that ratio is a determining factor. I also took out the year_of_incorporation because it seemed to affect the mlr in a negative way. When I took it out the F-statistic rose, and the p-value of the model decreased. So this indicates that it was affecting the model negatively.

8. Are the variables you added statistically significant?

After taking out the year_of_incorporation, the rest of the variables with the exception of execration were statistically significant at a 95% confidence interval. When looking at execratio we can use other factors apart from the p-value to determine if it might be useful in the model. When excluding the variable execration I noticed that the p-value of the whole model very slightly decreased. Not enough to be a significant problem, but most of the indications point to the fact that execratio might not be helping our model.

What are the associated 95%-confidence intervals?

```
confint(bm_reg2)
```

```
##                       2.5 %       97.5 %
## (Intercept)    2.085632e+01   35.1866023
## execratio     -3.148546e+00    0.3188490
## probloanpct   -4.775743e+02 -155.8308487
## metropop       4.319838e-03    0.2686036
## size_tercile2  5.294575e-01    9.6685215
## size_tercile3  5.447950e+00   15.4514233
```

9. Prepare a Deficiency Table for Chikuho Bank based on your analysis above.

```
metropop <-df_benchm_final %>% filter(name=="CHIKUHO BANK") %>% select(metropop)
 metropop_avg <-mean(df_benchm_final$metropop)
 metropop_delta <- metropop-metropop_avg
 metropop_beta <- bm_reg2$coefficients[4]
 metropop_mult <- metropop_delta*metropop_beta
 size_tercile <-df_benchm_final %>% filter(name=="CHIKUHO BANK") %>% select(size_tercile)
 size_tercile_avg <- mean(as.numeric(df_benchm_final$size_tercile))
 size_tercile_delta <- as.numeric(size_tercile) - as.numeric(size_tercile_avg)
 size_tercile_beta <- 0 #bm_reg2$coefficients[5]
 size_tercile_mult <-size_tercile_delta*size_tercile_beta
#year_of_incorporation<- df_benchm_final %>% filter(name=="CHIKUHO BANK") %>% select(year_of_incorporat
# year_of_incorporation_avg <-mean(df_benchm_final$year_of_incorporation)
# year_of_incorporation_delta <- year_of_incorporation - year_of_incorporation_avg
# year_of_incorporation_beta <-  bm_reg2$coefficients[6]
# year_of_incorporation_mult <- year_of_incorporation_delta*year_of_incorporation_beta
 execratio <-df_benchm_final %>% filter(name=="CHIKUHO BANK") %>% select(execratio)
 execratio_avg <-mean(df_benchm_final$execratio)
 execratio_delta <- execratio-execratio_avg
 execratio_beta <- bm_reg2$coefficients[2]
 execratio_mult <- execratio_delta*execratio_beta
 probloanpct <-df_benchm_final %>% filter(name=="CHIKUHO BANK") %>% select(probloanpct)
 probloanpct_avg <-mean(df_benchm_final$probloanpct)
 probloanpct_delta <- probloanpct-probloanpct_avg
 probloanpct_beta <- bm_reg2$coefficients[3]
 probloanpct_mult <- probloanpct_delta*probloanpct_beta


defficiency_table <- matrix(c(metropop, metropop_avg,metropop_delta ,metropop_beta, metropop_mult,
      size_tercile, size_tercile_avg, size_tercile_delta, size_tercile_beta,size_tercile_mult,
      #year_of_incorporation, year_of_incorporation_avg, year_of_incorporation_delta, year_of_incorpora
      execratio, execratio_avg,execratio_delta ,execratio_beta, execratio_mult,
      probloanpct, probloanpct_avg,probloanpct_delta ,probloanpct_beta, probloanpct_mult,
      '','','','',metropop_mult + size_tercile_mult  +execratio_mult + probloanpct_mult  ), ncol=5, byr
rownames(defficiency_table) <- c('metropop','size_tercile','execratio', 'probloanpct', 'Relative_to_Avg
colnames(defficiency_table) <- c('Chikuho_Bank','Sample_Average','Delta', 'B', '(B * Delta)')
```

	Chikuho_Bank	Sample_Average	Delta	B	(B * Delta)
metropop	5.6	12.08916	-6.489157	0.1364617	-0.8855216
size_tercile	1	1.987952	-0.9879518	0	0
execratio	0.01960784	0.1800325	-0.1604247	-1.414849	0.2269767
probloanpct	0.04547169	0.03357674	0.01189496	-316.7026	-3.767163
Relative_to_Avg					-4.425708

```
as.data.frame(defficiency_table) %>%
  kbl() %>%
  kable_classic_2(full_width = F)
```

The deficiency table indicates that the probloanpct variable had the biggest effect for the margins on Chikuho_Bank. Compared to the average bank, Chikuho_Bank was predicted to have -4.42 fewer margin percentage. With probloanpct we found better reasons on why Chikuho_Bank is doing worse than other banks. This table is consistent with the regression that said that probloanpct had the biggest impact in the model.

10. What managerial recommendations would you make to Chikuho Bank after examining the new regression output and Deficiency Table?

After examining the latest regression results and deficiency table I would recommend the CEO to focus on reducing the percentage of bad loans. The current percentage for Chikuho Banks is much higher than the rest, and from our analysis this seems to be the main variable affecting margins.