National University of Lesotho

B.A. Examinations

EC4401 – Data Sciences for Economists

January 2024 **100 Marks** **3 Hours**

INSTRUCTIONS:

1. Answer ALL questions
2. All questions have 25 marks each.
3. Refer to Appendix 1 for the description of the variables used.

Question One

Assume that you have been employed into the economic team of the prime minister and your team has been tasked with the responsibility to ascertain the determinants of credit access. Knowing that you did data science using the R software, you are expected to head this particular task.

a. List 6 possible controls that could lead to credit access for firms in Lesotho. **[3 Marks]**
b. Explain why you have selected each of the controls in a. above. **[5 Marks]**
c. In an ideal situation (outside exams) explain what should normally guide the choice of control variables and explain why. **[2 Marks]**
d. Considering that a colleague who is assisting you with this task, estimates his own model, based on the data available to him, interprete the results of model fitness, logit, odds and marginal fixed effect estimates as presented in Appendix 2. **[8 Marks]**
e. Explain whether the results are as expected in the real life scenario of the Lesotho Economy or not. **[5 Marks]**
f. Will you approve this result as valid for presentation to the prime minister? Why or why not? **[2 Marks]**

Question Two

a. In each case, explain the what the R codes in Appendix 3 seeks to address. **[20 Marks]**
b. What are the similarities and differencies between a logit and probit model?**[5 Marks]**

Question Three

a. Imagine that, similar to the never-ending debate of the chicken and egg on which comes first, you have been asked to analyse whether it is the profit that influence growth in sales or vice versa. To do this, you employ the granger causality results stated in Appendix 4 below, Interpret the empirical results and clearly stating whether profit trends influence sales trends and vice versa. **[10 Marks]**
b. Clearly explain how to develop an object, argument, function and a data frame (with the specification of the codes) and explain what each is meant for. **[15 Marks]**

Question Four

Appendix 5 below, contains 3 graphs: a bar chart, a violin plot and a pair wise plot. Provide detailed interpretation of all 3 plots. **[25 Marks]**

APPENDICES

Appendix 1: Variable Descriptions

No	Variable name	Description
1	totalasset	Total assets of the firm
2	capital	Capital of the firms
3	agebus	Longevity or age of firm in years
4	employoperatives	Number of workers
5	education	Educational level where 2 = primary, 3 = secondary and 4 = tertiary
6	totalcredit	Total credit approved for firms
7	avsales	Average annual sales of firms
8	avprofit	Average annual profit of firms

Appendix 2: Logit Regression

```
glm(formula = genderfh ~ capital + agebus + totalcredit + factor(education),
    family = binomial(link = "logit"), data = Firm_Surv2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9094  -1.3156   0.8139   0.8968   1.4112

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.505e+00  8.844e-01   1.476    0.140
capital            -3.300e-08  3.101e-08  -1.258    0.208
agebus             -1.192e-02  2.861e-02  -0.417    0.677
totalcredit        -4.635e-07  2.988e-07  -1.551    0.121
factor(education)3  6.333e-02  7.918e-01   0.080    0.936
factor(education)4  5.051e-01  8.563e-01   0.590    0.555

(Dispersion parameter for binomial family taken to be 1)

Number of Fisher Scoring iterations: 4
```

```
> export_summs(model7)
```

	Model 1
(Intercept)	1.51
	(0.88)
capital	-0.00
	(0.00)
agebus	-0.01
	(0.03)
totalcredit	-0.00
	(0.00)
factor(education)3	0.06
	(0.79)
factor(education)4	0.51
	(0.86)
N	118
AIC	158.18
BIC	174.80
Pseudo R2	0.87

*** p < 0.001; ** p < 0.01; * p < 0.05.

Odds Ratio Estimates

```
> exp(model7$coefficients)
       (Intercept)            capital             agebus        totalcredit
         3.6878330          1.0000000          0.9881497          0.9999995
factor(education)3 factor(education)4
         1.0653825          1.6571209
```

Marginal Effect Estimates

```
> model8 <- mean(dlogis(predict(model7, type = "link")))
> model8* coef(model7)
       (Intercept)            capital             agebus        totalcredit
      2.803380e-01      -8.378708e-09      -2.560788e-03      -9.956127e-08
factor(education)3 factor(education)4
      1.360488e-02       1.084976e-01
```

Model Fitness

```
> chis=model7$null.deviance-model7$deviance
> dfdiff=model7$df.null-model7$df.residual
> #For pvalue of chi square pchisq(chis,dfdiff,lower.tail=F)
[1] 0.083242
```

Appendix 3

i. sample(x=1:6, size=2, replace=TRUE)

ii. rm(object)

iii. str(variable)

iv.
```
ggplot(Firm_surv) +
  geom_point(aes(x = agebus,
                 y = avsales, color = factor(education)),
             na.rm = TRUE) +
  geom_smooth(aes(x = agebus,
                  y = avsales),
              method = "loess",
              formula = y ~ x,
              na.rm = TRUE) +
  labs(title = "Relationship betwen Average sales and age of firm by education",
       x = "Age of business and education",
       y = "Average Sales",
       color = "Education of Firm Head")
```

v.
```
ggplot(Firm_surv) +
   geom_point(aes(x = agebus,
                  y = avsales),
              alpha = 0.1,
              na.rm = TRUE,
              )
```

vi. mutate (…)

vii. group_by(...)

viii. spread(object_1, object_2)

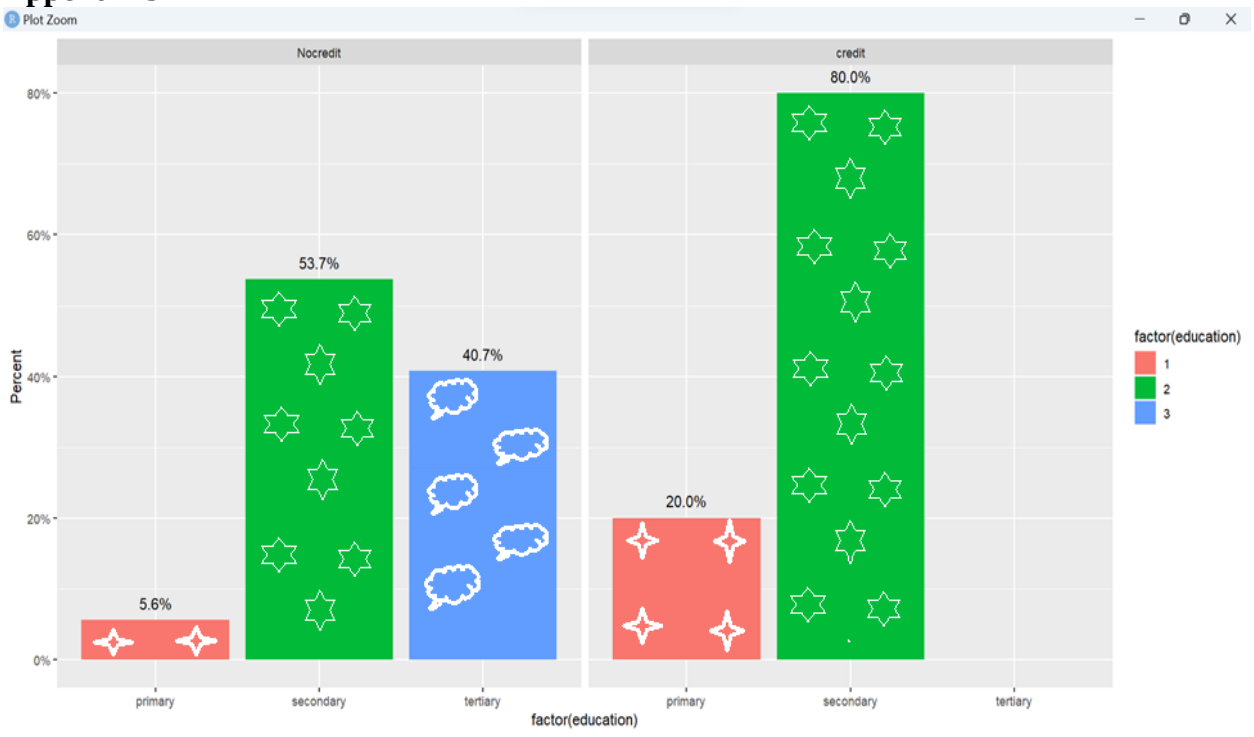
ix. waldtest(model5, model4)

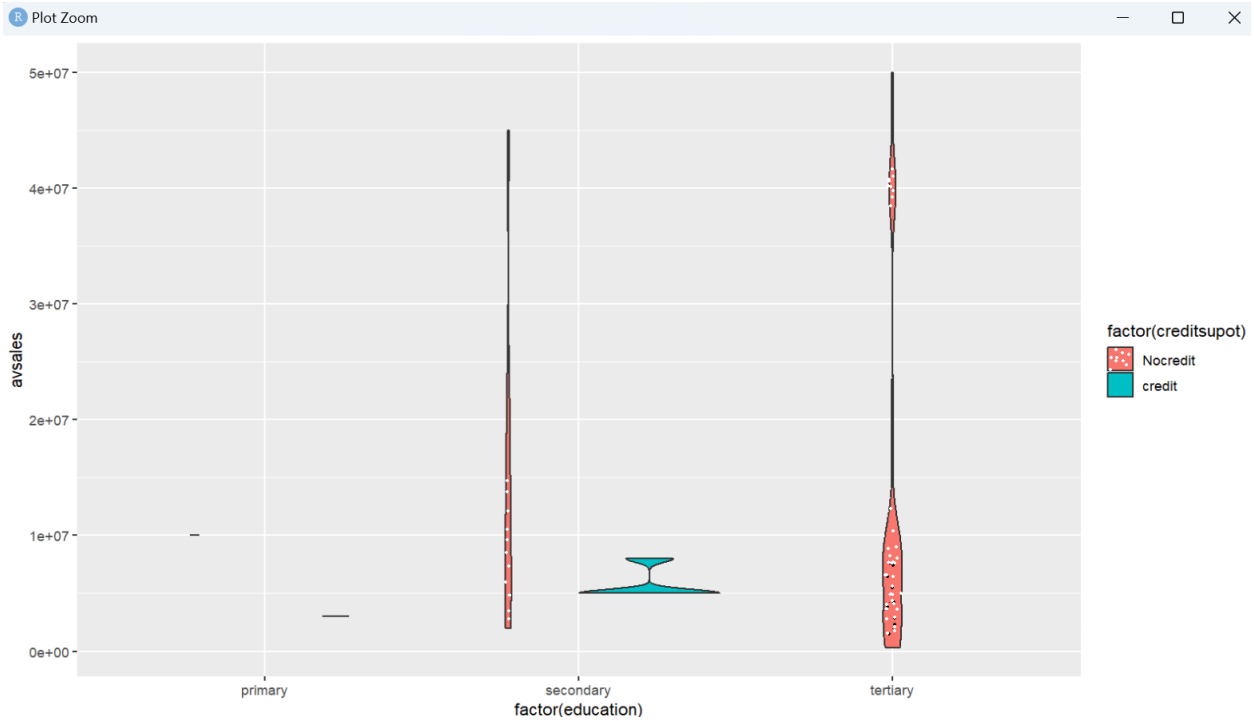
x. page %>% html_nodes("…") %>% html_text()

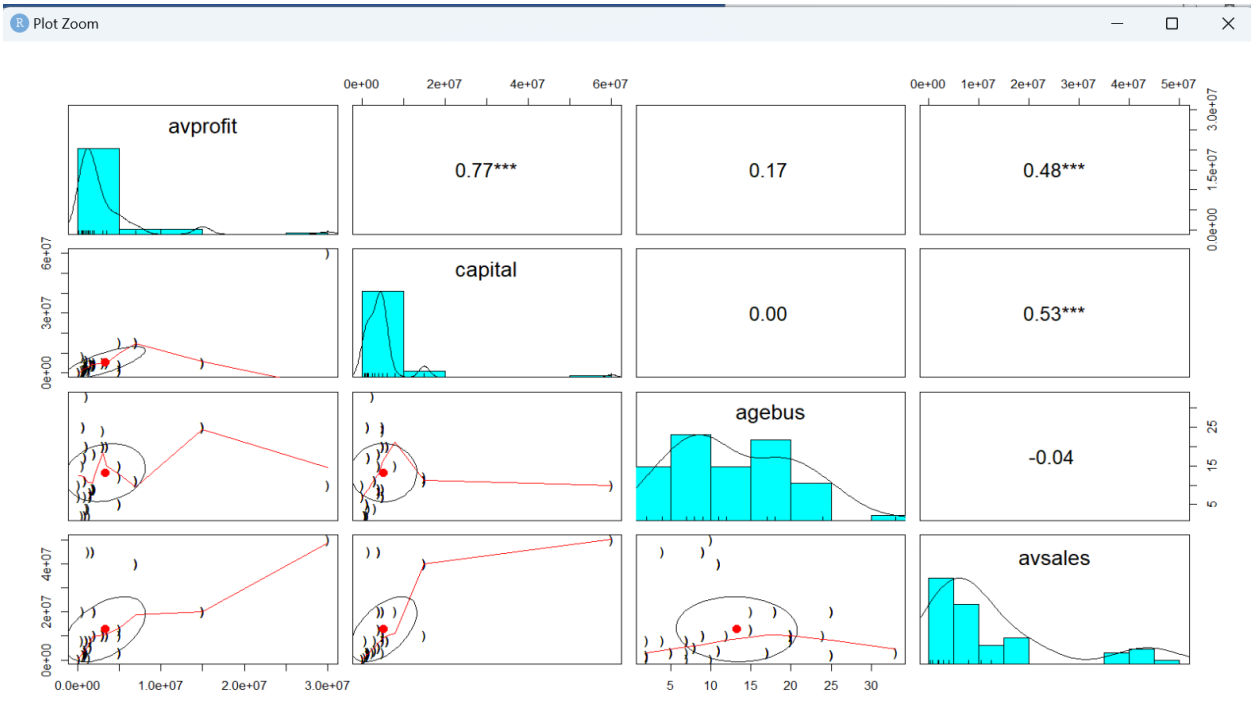
Appendix 4

```
>  grangertest(avsales ~ avprofit, order = 3, data = Firm_surv)
Granger causality test

Model 1: avsales ~ Lags(avsales, 1:3) + Lags(avprofit, 1:3)
Model 2: avsales ~ Lags(avsales, 1:3)
  Res.Df Df      F  Pr(>F)
1    108
2    111 -3 2.1509 0.09806 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> grangertest(avprofit ~ avsales, order = 3, data = Firm_surv)
Granger causality test

Model 1: avprofit ~ Lags(avprofit, 1:3) + Lags(avsales, 1:3)
Model 2: avprofit ~ Lags(avprofit, 1:3)
  Res.Df Df      F Pr(>F)
1    108
2    111 -3 0.5222 0.6679


>
```

Appendix 5




Plot Zoom
5e+07
4e+07
3e+07
2e+07
1e+07
0e+00
avsales
primary
secondary
tertiary
factor(education)
factor(creditsupot)
Nocredit
credit


Plot Zoom
avprofit
0.77***
0.17
0.48***
capital
0.00
0.53***
agebus
-0.04
avsales