logistic: Name _____ CLDa 8.25/12 → 4.125/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output (no R formulas) — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question (imprecise)

all variables defined

0.75/1

3. EDA:

univariate numerical (+ categorical) — bivariate numerical (cor) (+ crosstabs)

univariate graphical — bivariate graphical

1.25/2

4. Model fitting:

give mathematical definition of model (not R formulas)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected (be specific)

define all terms (LRT, AIC, Wald, Deviances)

1.25/2

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references (all tests mathematically)

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

1.25/2

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) (+ explanations)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

5.25/8

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions

1. recap analysis 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

logistic: Name _____ Adje 7.5/12 → 3.75/6

1. Formatting:

all margins 2.5cm — informative title

0.75/1 — 12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1 — 2. Introduction/Background: use paragraphing — too many digits

brief background and statement of scientific question — imprecise

all variables defined

1.75/2 — 3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical — square QQ plots — square pairs plots — Combine Figure 2a+b

1.25/2 — 4. Model fitting:

give mathematical definition of model — Before fitting description

state how model fitted (ie, maximum likelihood) — not 'tries to'

CLEARLY describe how model selected

define all terms AIC + all performance metrics

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

not done?

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

0.5/2 — carry out assessment (numerical / graphics): + explanations

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers) — not cook

vif (to identify multicollinearity) — Define

5/8

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions

1. recap analysis 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

(*) Don't need to re-do

logistic: Name _____ To Do 10/12 → 5/6 → 6/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

1/1

2. Introduction/Background:

0.75/1 brief background and statement of scientific question — imprecise

all variables defined

3. EDA: Don't need Mann-Whitney

1.5/2 univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical + pairs plots

4. Model fitting: excellent

2/2 give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms

5. Model assessment:

1.5/2 CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) (+exp/canotions)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

0 Define

6.75/8

1/1

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

1/1

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

not correct

1. recap analysis 2. state and interpret main findings

(implicit EDA)

0.75/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

3.25/4

logistic: Name _____ Il Fa 6.25/12 → 3.125/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

1/1

2. Introduction/Background:

0.5/1

brief background and statement of scientific question

all variables defined

3. EDA:

1.5/2

univariate numerical	bivariate numerical (cor)
univariate graphical	bivariate graphical

combine figures 2+3

4. Model fitting: Table 2 entries not clear + incomplete

0.75/2

give mathematical definition of model (in terms of logit)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected + what is 'backward and forward elimination'?

define all terms AIC/BIC + all performance metrics

5. Model assessment:

0.75/2

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) + explanations

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define

4.5/8

0.5/1 one model + put numeric values in equation

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.5/1 7. Plots:

label size (not too small) informative captions

placement a lot of blank space explanations

0.25/1 8. Conclusions

1. recap analysis 2. state and interpret main findings be specific

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references): rather superficial report spell check

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

1.75/4

logistic: Name _____ St Fe 7.75/12 → 3.875/6

1. Formatting:

1/1

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question (imprecise)

all variables defined

3. EDA:

1.5/2

univariate numerical — bivariate numerical (cor) (not just graphic)

univariate graphical — bivariate graphical

4. Model fitting:

1.5/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC (+ interpretation)

5. Model assessment: not Done?

0/2

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

4.75/8

4/4

6. Write out final *estimated* model **mathematically**

hat on *response* variable ($\hat{\ }$) max **2 sig digits** (after decimal) on coefs

7. Plots:

4/4

label size (not too small) informative captions

placement explanations

8. Conclusions

0.5/1

1. recap analysis (implicit EDA) 2. state and interpret main findings — be specific

0.5/1

9. Overall presentation (clarity of explanations, appropriate citations / references) : somewhat incomplete

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

3/4

logistic: Name _____ Am Ga 9/12 → 45/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output (Don't need R details) — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

too many digits — imprecise

0.75/1

brief background and statement of scientific question

all variables defined

3. EDA:

2/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting:

1.25/2

give mathematical definition of model (in terms of logit)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — Be specific — GOF P? (Don't need R details)

define all terms — Deviances, AIC

5. Model assessment:

1.25/2

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define

6/8

(ok) logit p

1/1 6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

1/1 label size (not too small) informative captions

placement explanations

0.25/1 8. Conclusions

be specific

1. recap analysis 2. state and interpret main findings

not entirely correct

0.75/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: Remove blank plot

3/4

logistic: Name ReGa-Av 9/12 → 4.5/6

1. Formatting:

0.75/1

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

Don't need R details

max **10** pages — no blurry plots (**NOT png**)

2. Introduction/Background:

0.75/1

brief background and statement of scientific question

all variables defined

3. EDA: (ok)

2/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting:

1.25/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC, Deviances

Incomplete results: Z, p

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

1.25/2

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) (only numeric)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define SQUARE ROC plot

6/8

1/1

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions Vague variable effects?

0.25

1. recap analysis + EDA 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

1/1

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

3/4

logistic: Name LaGi 6/12 → 3/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1 Throughout

2. Introduction/Background: — too many digits

0.75/1 imprecise

brief background and statement of scientific question

all variables defined

3. EDA: incomplete

0.25/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical + pairs

4. Model fitting: please display formulas

use more paragraphing + somewhat incomplete

give mathematical definition of model not ∈

1/2 + hard to follow

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms Firth penalized

5. Model assessment: not 'verified'

CLEARLY state model assumptions: + give PRIMARY references

+ explain

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

1/2 carry out assessment (numerical / graphics): + explanations

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

3.75/8

(ok) $\log\frac{p}{1-p}$

1/1

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.25/1

8. Conclusions

be specific not correct

1. recap analysis 2. state and interpret main findings

0.25/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

use your own words

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

please completely re-write copied material

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

2.25/4

logistic: Name _____ Is Go 8.5/12 → 4.25/6

1. Formatting: (remove excess blank space)

Y/1

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background: imprecise

brief background and statement of scientific question

all variables defined

3. EDA:

0.5/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical + pairs plots

4. Model fitting:

1.25/2

give mathematical definition of model – easier to read if defined in terms of logit

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected not exactly forward selection

define all terms AIC

5. Model assessment:

1.25/2

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) (add lines) + explanation

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

5.75/8

(ok)

1/1 6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75/1 7. Plots:

label size (not too small) informative captions

placement explanations

0.5/1 8. Conclusions

1. recap analysis (implicit EDA) 2. state and interpret main findings — not correct

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

2.75/4

logistic: Name _____ LiGr 8.5/12 → 4.25/6

1. Formatting:

0.75/1

all margins 2.5cm + informative title

12 pt size name on all pages

no raw R code or output all pages numbered

max **10** pages no blurry plots (**NOT png**)

No cover page + Contents

0.75/1

2. Introduction/Background:

brief background and statement of scientific question imprecise

all variables defined

3. EDA:

1.5/2

univariate numerical bivariate numerical (cor) (align cols)

univariate graphical bivariate graphical pairs plots

4. Model fitting:

1.25/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected - very unclear + p-values?

define all terms AIC

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

1.5/2

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define

First explain before interpret

5.75/8

(write in terms of logit)

1/1 6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75/1 7. Plots:

label size (not too small) informative captions

placement explanations

0.5/1 8. Conclusions

1. recap analysis 2. state and interpret main findings

be clear not correct

(implicit EDA)

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: Q1: put excessive plots in an appendix

Q2: You can use either 'I' or 'we'

Q3: You can cite an authoritative textbook, for example McCullagh+Nelder or Dobson

2.75/4

logistic: Name Or Gr-Pi 8.25/12 → 4.125/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**) — too many digits

0.75/1

2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

0.75/1

3. EDA:

univariate numerical — bivariate numerical (cor) — not just graphic

univariate graphical — bivariate graphical

1.75/2

4. Model fitting:

give mathematical definition of model — no ^ — Before fitting

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — Don't eliminate variables based on EDA

define all terms — Deviances (LRT 'actual contribution' =?? looks like incorrect interp.

1/2

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references — not 'verify'

1.25/2

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define

5.5/8

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions

1. recap analysis 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

logistic: Name La-Ro Ha 8/12 → 4/6

1. Formatting:

(Don't need AB-Ind Report)

all margins 2.5cm — informative title

12 pt size — name on all pages

0.75/1 **no raw R code or output** — all pages numbered

Don't need R details

max **10** pages — no blurry plots (**NOT png**)

2. Introduction/Background:

0.75/1 brief background and statement of scientific question — imprecise

all variables defined

1.5/2 3. EDA:

univariate numerical — bivariate numerical (cor) — out of place

univariate graphical — bivariate graphical — pairs plots for numeric vars

4. Model fitting:

1.25/2 give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — Don't choose variables based on EDA

define all terms — AIC, Deviances

5. Model assessment: not 'ensure'

1.25/2 CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics): + explanation

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define square QQ (but why do you expect them to be normal?)

5.5/8

(ok) (Don't need prob) $\text{logit}(\frac{p}{1-p})$

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions Vague be specific + not bullet points

1. recap analysis 2. state and interpret main findings

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

2.5/4

logistic: Name _____ Hulte 8.75/12 → 4.375/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size (Throughout) — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

3. EDA: Don't need hyp tests

univariate numerical — bivariate numerical (cor) — not just graphic

univariate graphical — bivariate graphical — pairs plots for numeric vars

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — Backward

define all terms — some incorrect interpretations

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references — all + explain

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

Then assess

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

+ explanation

Define

+ Define all performance measures

6/8

(ok) $\log\left(\frac{p}{1-p}\right)$

1/1

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

0.75/1

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions vague

1. recap analysis 2. state and interpret main findings (be specific)

9. Overall presentation (clarity of explanations, appropriate citations / references) :

0.5/1

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

2.75/4

logistic: Name _____ Mythi 7.25/12 → 3.625/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

1/1

2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

0.75/1

3. EDA:

univariate numerical — bivariate numerical (cor) (don't need this plot)

univariate graphical — bivariate graphical square

no boxplots

1.75/2

Don't need Tables 3+4+5+6

4. Model fitting:

incorrect interpretations

1/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — Very complicated 'procedure'

define all terms — LRT, AIC

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

1/2

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

+explanations

5.5/8

0.5/1

6. Write out final *estimated* model **mathematically** interaction term + why no main effects?

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.5/1

7. Plots:

label size (not too small) informative captions (+ number all figures

placement annex takes too many pages explanations

0.25/1

8. Conclusions

1. recap analysis 2. state and interpret main findings — be specific — not completely correct

0.5/1

9. Overall presentation (clarity of explanations, appropriate citations / references):

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other:

1.75/4

logistic: Name _____ Ca Im 9/12 → 4.5/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

1/1

2. Introduction/Background:

0.75/1

brief background and statement of scientific question — imprecise

all variables defined

3. EDA:

1.25/2

univariate numerical — All summary stats in a table — bivariate numerical (cor)

univariate graphical — bivariate graphical — pairs plots

4. Model fitting:

1.5/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — AIC — p-values (not <0.001)

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

1.5/2

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity) — Define

6/8

1/1

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

be specific

1. recap analysis 2. state and interpret main findings

(implicit EDA)

0.75/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

_____

_____

_____

_____

3/4

logistic: Name _____ CeKa 8.25/12 → 4.125/6

1. Formatting:

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

1/2

2. Introduction/Background: run on sentences - clarify imprecise

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA:

univariate numerical	bivariate numerical (cor) all
univariate graphical	bivariate graphical pairs

1.25/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected) incomplete - be specific

define all terms Wald, LRT, AIC mathematically

1.25/2

5. Model assessment:

assess not 'verify'

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

+ clear explanations

Define

5.5/8

1/1 (ok) logit π

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

7. Plots:

0.75/1 label size (not too small) informative captions

placement explanations

0.5/1 8. Conclusions

be specific not completely correct

1. recap analysis 2. state and interpret main findings

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

(ok) Technometrics

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

2.75/4

logistic: Name Mika 8.5/12 → 4.25/6

1. Formatting:

0.75/1

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

2. Introduction/Background: — too many digits

0.75/1

brief background and statement of scientific question — imprecise

all variables defined

3. EDA: all vars 1.5/2

univariate numerical — bivariate numerical (cor) Don't need Figure 2

univariate graphical — bivariate graphical — square

4. Model fitting: What is the purpose of Table 2? 1.5/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — Don't eliminate variables using EDA

define all terms — interpretations not entirely correct

5. Model assessment:

1.5/2

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define

6/8

1/1

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.25/1

8. Conclusions

be specific

1. recap analysis 2. state and interpret main findings not correct

0.5/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: -strange pagination

2.5/4

(*) Don't need to re-do

logistic: Name _____ Yakr 10.5/12 → ~~5.25/6~~ → 6/6

1. Formatting:

4/4

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief background and statement of scientific question imprecise

all variables defined

3. EDA:

2/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting:

2/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — Don't need all steps (Table 1)

OR

5. Model assessment:

1.75/2

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs
3. linear relation between logit and linear predictor
4. no multicollinearity 5. no outliers (6. large sample size)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) ✓

DEFINE -> Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

7.5/8

4/4 (OK) log P

6. Write out final *estimated* model **mathematically**

hat on *response* variable max **2 sig digits** (after decimal) on coefs

0.75/1 7. Plots:

label size (not too small) informative captions

placement explanations

0.25/1 8. Conclusions

be specific not entirely correct

1. recap analysis + EDA 2. state and interpret main findings

(1/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

B – cite PRIMARY refs (not course notes, not wikipedia, *etc.*)

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

E – Intro: **1.** Give context; **2.** Clearly state scientific question; **3.** Describe data

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

H – clearly EXPLAIN / INTERPRET PLOTS (don't just state conclusions)

I – plot size / aspect ratio (make 'pretty')

Other: _____

3/4