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

1. Formatting:

all margins 2.5cm

12 pt size

no raw R code or output

max **10** pages

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.75/1 Don't need R fn details

2. Introduction/Background:

0.75/1 brief background and statement of scientific question

all variables defined

3. EDA:

0.75/2 univariate numerical

univariate graphical

bivariate numerical (cor)

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

define all terms

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)

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

vif (to identify multicollinearity)

5.75/8

- too many digits
imprecise
methods, not R fns
AIC, LRT
+explanation
Define
- Define all performance metrics

4/4 (ok) $\log\left(\frac{p}{1-p}\right)$

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 use paragraphing be specific

1. recap analysis 2. state and interpret main findings (not done?)

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 _____ RoLa 7.75/12 → 3.875/6

1. Formatting:

all margins 2.5cm — informative title (-(too many digits))

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: imprecise

brief background and statement of scientific question

all variables defined

1.5/2

3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical + pairs plots

1.25/2

4. Model fitting:

give mathematical definition of model no ^

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms LRT 2 's in Table 2

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) (+ explanations)

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

vif (to identify multicollinearity)

Define

5.5/8

0.75 (ok) $\log it\left(\frac{p}{1-p}\right)$ + $\hat{\beta}$'s

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

0.25 be specific 'rigorous'?

1. recap analysis 2. state and interpret main findings clearly

+EDA

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

0.5

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: _____

2.25/4

(*) Don't need to re-do

logistic: Name Yola 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: imprecise

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA:

univariate numerical — bivariate numerical (cor) why Pearson, you have mainly categoric

univariate graphical — (bivariate graphical)

1.75/2

4. Model fitting: very good

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

2/2

CLEARLY describe how model selected

define all terms

5. Model assessment: (Inference from log reg relies ...) (not 'validate suitability')

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)

1.5/2

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 all performance metrics

7/8

1/1 (ok) log(p/1-p)

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 vague variables+effects?

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

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 _____ FrlaR 9.25/12 → 4.625/6

1. Formatting:

all margins 2.5cm — informative title

1/1 12 pt size — name on all pages

no raw R code or output — all pages numbered

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

2. Introduction/Background:

0.75/1 brief background and statement of scientific question imprecise

all variables defined

3. EDA: Don't need Missing / Complete Columns

1.5/2 univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical all pairs

4. Model fitting: interpretation of log-odds / OR predicted change

1.25/2 give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC

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 + all performance metrics

6/8

(OK) logit?

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

1/1

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

be specific, 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 Jiki 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:

brief background and statement of scientific question — somewhat imprecise

all variables defined

0.75/1

3. EDA:

univariate numerical — bivariate numerical (cor) (as a matrix)

univariate graphical — bivariate graphical

1.75/2 square

4. Model fitting:

give mathematical definition of model — not +ϵ + define logit

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — explain your stepwise procedure

define all terms

1.5/2

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)

1.5/2

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

- SQUARE ROC curve
- Define performance metrics

6/8

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 be specific estimated risk/likelihood

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 Ra Li 8.5/12 → 4.75/6

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

Don't need R fns, don't use R formulas

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

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: - All results in a table, hard to follow

1.25/2 give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected stepwise = ?

define all terms AIC

1.75/2 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)

carry out assessment (numerical / graphics): clearly explain

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

└ + Define

6/8

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

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 (labels)

placement explanations

0.5/1 8. Conclusions (be specific)

1. recap analysis 2. state and interpret main findings

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : (+EDA) many run-on sentences

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

(*) Don't need to re-do

logistic: Name An Ma 10.75/12 → 5.375 → 6/6

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**)

2. Introduction/Background: 0.75/1

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: very clear

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:

CLEARLY state model assumptions: + give PRIMARY references

2/2

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)

7.75/8

0.75

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

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

1 / 1

7. Plots:

label size (not too small) informative captions

placement explanations

1 / 1

8. Conclusions

0.25

1. recap analysis + EDA 2. state and interpret main findings — be specific — be clear

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

logistic: Name _____ Al Mq de Fr 6.75/12 → 3.375/6

1. Formatting:

0.75/1

all margins 2.5cm — informative title (- too many digits)

12 pt size — name on all pages

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

max 10 pages — put big plots/tables in appendix — no blurry plots (**NOT png**)

2. Introduction/Background: why assumptions above intro?

0.5/1

brief background and statement of scientific question (imprecise)

all variables defined

3. EDA:

1.75/2 all summary stats

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting: - incorrect interpretations

1/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — $\hat{\beta}$ is for numerical values

define all terms — AIC, deviances

5. Model assessment:

1/2

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

4/8

1/1

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

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

1/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.25/1

8. Conclusions

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:

2.75/4

logistic: Name _____ An Mi 7.75/12 → 3.875/6

1. Formatting:

all margins 2.5cm

12 pt size

no raw R code or output

max **10** pages

informative title (+ Don't need contents)

name on all pages

all pages numbered

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: exploratory (not 'initial')

univariate numerical

univariate graphical

bivariate numerical (cor) (not Pearson for categorical)

bivariate graphical

1.5/2

4. Model fitting:

give mathematical definition of model (use paragraphing your explanation is very hard to follow + unclear)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected (explain stepwise)

define all terms (AIC, variable importance, LRT)

1/2

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)

(First, then separate p)

1.25/2

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption) (don't need for categorical predictors)

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

vif (to identify multicollinearity)

Define

5.25/8

0.75/1 (not clear)

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 specific incorrect

0.5/1 (+EDA)

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 _____ ArMo 8.75/12 → 4.375/6

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

Don't need R details

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

too many digits

0.75/1 2. Introduction/Background:

imprecise

brief background and statement of scientific question

all variables defined

3. EDA: Don't need Figures 5/6 plus they look rotated

2/2 (ok)

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

no Figure 7

4. Model fitting:

1.25/2 give mathematical definition of model First, before fitting

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms

add ORs to LRT results table

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):

+ explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

6/8

(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

0.5/1

8. Conclusions

be specific — vague what does 'stability' mean?

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

0.5/1

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

2.75/4

logistic: Name MaMo 7.75/12 → 3.875/6

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**)

0.5/1

2. Introduction/Background: - too many digits

very unclear + imprecise

brief background and statement of scientific question

all variables defined

1.5/2

3. EDA: Plots too small

univariate numerical — bivariate numerical (cor)

univariate graphical — ~~bivariate graphical~~ pairs plots

1.25/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — why these models only?

define all terms — AIC

put results in a table

5. Model assessment:

not 'ensure validity'

CLEARLY state model assumptions: + give PRIMARY references

1.25/2

(+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)

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

5.25/8

(ok) log(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: Blank space

label size (not too small) informative captions

placement explanations

0.25/1

8. Conclusions vague be specific not 'confirms'

1. recap analysis 2. state and interpret main findings

0.5/1

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

_____

_____

_____

_____

2.5/4

logistic: Name _____ SaMo 9.75/12 → 4.875/6

1. Formatting:

all margins 2.5cm — informative title

1/1 — 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: imprecise

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:

2/2 — give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — criterion AIC (not a)

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)

7.25/8

1/1 (y not defined)

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 vague

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

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 (please re-write copied sections)

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 _____ FrHu 9.5/12 → 4.75/6

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 — tumor size + age also ordinal

3. EDA:

2/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting: some unclear + incorrect interpretations of results

1.25/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC / LRT

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) (ok if don't do for categorical)

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

vif (to identify multicollinearity)

L + Define

6.5/8

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: _____

_____

_____

_____

_____

3/4

logistic: Name _____ JaMu 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**)

1/1

2. Introduction/Background: use paragraphing

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA: all summary stats [plot layout]

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical pairs plots

1.25/2

4. Model fitting:

① give mathematical definition of model

② state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected stepwise not explained

define all terms AIC / p-values not = 0.050

1.25/2

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)

1.25/2

carry out assessment (numerical / graphics):

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

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

vif (to identify multicollinearity) not 'confirming'

Define

5.5/8

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

8. Conclusions

0.25/1

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

be specific not 'validated'

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 _____ ArNa 8.75/12 → 4.375/6

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**)

2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

3. EDA: — not just graphic

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical — pairs

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms

Results incomplete

5. Model assessment: 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)

6/8

0.75/1

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

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 — not 'validated'

1. recap analysis — 2. state and interpret main findings — not 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:

2.75/4

logistic: Name _____ ViNa 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**)

0.5/1

2. Introduction/Background:

brief background and statement of scientific question

all variables defined

0.25/1

3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical (+ pairs plots)

1.75/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms

1.25/2

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)

1.25/2

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

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 _____ EI No 8/12 → 4/6

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)

univariate graphical — bivariate graphical

1/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — AIC, BIC + numerical results?

1.5/2

5. Model assessment: not 'ensure validity'

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) not 'prove'

+ explanations

+ Define all performance metrics

5.25/6

1/1

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

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

1/1

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) : + superficial in parts

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 _____ Sa No 8.25/12 → 4.125/6

1. Formatting:

no cover page / EPFL logo

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:

0.75/1

brief background and statement of scientific question — imprecise

all variables defined

3. EDA:

1.25/2

univariate numerical + counts — bivariate numerical (cor)

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 — why just 2?

define all terms — AIC - imprecise / pseudo R^2 / LRT

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) (+ explanation)

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

vif (to identify multicollinearity)

Define

5.5/8

1/1 6. Write out final *estimated* model **mathematically** (in terms of logit but OK)

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 not 'confirm')

1. recap analysis 2. state and interpret main findings

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

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 _____ SyNo 8.25/12 → 4.125/8

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:

0.75/1

brief background and statement of scientific question — imprecise

all variables defined

3. EDA:

1.75/2

univariate numerical — bivariate numerical (cor)

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 — explain stepwise

define all terms — AIC

Results table

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)

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

vif (to identify multicollinearity)

+ explanations

Define

5.75/8

1/1 (ẑ not defined)

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 Vague be specific

1. recap analysis 2. state and interpret main findings unclear

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 Please re-write copied sections

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 _____ IP Pa 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

bivariate numerical (cor)

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 — procedure? don't eliminate variables based on EDA

define all terms

give p-values, not < 0.001; incorrect odds interp

5. Model assessment:

use paragraphing

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

5.75/8

(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

8. Conclusions

0.75/1

1. recap analysis 2. state and interpret main findings

0.75/1

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

3.25/4

logistic: Name _____ ThPa 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:

0.75/1

brief background and statement of scientific question — imprecise

all variables defined

3. EDA:

1.5/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical + pairs plots

4. Model fitting: (Variable, not 'Facteur')

1.5/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — mathematical

define all terms — LRT: complete description of test

5. Model assessment: not 'confirm'

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) + explanations

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

vif (to identify multicollinearity)

Define

5.75/8

(ok) logit(p)

1/1

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

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

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

not completely correct

1. recap analysis 2. state and interpret main findings

(implicit EDA) (not 'confirmed')

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 _____ LePi 8.25/12 → 4.125/6

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**)

2. Introduction/Background:

0.75/1

unclear + imprecise

brief background and statement of scientific question

all variables defined

3. EDA:

1.5/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical pairs

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 need R details

define all terms much is vague + use your own words

GVIF

5. Model assessment:

1/2

CLEARLY state model assumptions: + give PRIMARY references

not 'ensuring' - what if pts related?

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)

+ explanations (superficial)

Define

5.5/8

1/1 (ok) logit p why change notation??

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 specific

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

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

clarity

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 _____ Da Pr 7.5/12 → 3.75/6

1. Formatting:

all margins 2.5cm

12 pt size

no raw R code or output

max **10** pages

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background:

brief background and statement of scientific question

all variables defined

3. EDA:

univariate numerical

univariate graphical

bivariate numerical (cor)

bivariate graphical

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms

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)

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

4.5/8

1/1

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

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

7. Plots: (ok)

1/1

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

1. recap analysis 2. state and interpret main findings

be specific

not 'confirm'

(implicit EDA)

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

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

more here

new paragraph

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 _____ MaPr 8.75/12 → 4.375/6

1. Formatting:

0.5/1

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**)

– too many digits

2. Introduction/Background:

1/1

brief background and statement of scientific question

all variables defined — age / tumor size / degree all ordinal cat vars

3. EDA:

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 / LRT / Deviance

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

1/2

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) (ok if don't do for categorical predictors)

DEFINE → Cook's distance / standardized residuals (outliers)

vif (to identify multicollinearity)

Define

5.75/8

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 _____ Ra-St Pu 6.25/12 → 3.125/6

1. Formatting:

all margins 2.5cm — informative title — too many digits

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: remove refs from intro

brief background and statement of scientific question

all variables defined — unclear + imprecise

0.75/1

3. EDA: incomplete

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

0.5/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood) — Don't need numerical details

CLEARLY describe how model selected

define all terms

1/2

5. Model assessment: + Define Training/Test and all performance metrics

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)

0.75/2

carry out assessment (numerical / graphics):

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

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

vif (to identify multicollinearity)

Define

3.75/8

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

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

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

0.75/1 7. Plots: remove excessive blank space

label size (not too small) informative captions

placement explanations

0.25/1 8. Conclusions - vague - be specific - not 'simple'

1. recap analysis 2. state and interpret main findings - what does 'robust' mean?

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: _____

_____

_____

_____

_____

2.5/4