logistic: Name _____ Ia Ra 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**)

1/1

re-arrange plots and/or move to an appendix

2. Introduction/Background:

0.5/1

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 plots

4. Model fitting:

1.5/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood) seems strange to end section with flag like stepwise

CLEARLY describe how model selected

define all terms

5. Model assessment: results table? not 'verify'

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): + explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

6/8

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

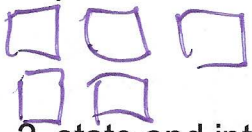
6. Write out final *estimated* model **mathematically**

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

0.75/1 7. Plots:

label size (not too small) informative captions

placement – Figure 1: explanations



0.5/1 8. Conclusions not correct

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

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

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: [You could say more about how model relates, but not strictly necessary]

2.75/4

logistic: Name _____ Je Ra 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**)

0.75/1

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) — not Pearson + numeric

univariate graphical — bivariate graphical — not just graphic

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 — explain stepwise

define all terms — AIC / test set (+ selection)

5. Model assessment: not 'ensure validity' not 'verify'

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

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:

3/4

(*) Don't need to re-do

logistic: Name _____ FeRo 10.5/12 → 5.25/6 → 6/6

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:

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:

1.75/2 — give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — Performance metrics

5. Model assessment:

incomplete

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):

don't need to do for categorical vars

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

7/8

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: + some small plots

label size (not too small) informative captions

placement explanations

0.75/1

8. Conclusions be specific

1. recap analysis 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:

3.5/4

logistic: Name AnSa 7.75/12 → 3.875/6

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: — write out words

0.5/2

univariate numerical	bivariate numerical (cor)
univariate graphical	bivariate graphical

incomplete - put all plots + tables in an appendix

4. Model fitting:

1.5/2

give mathematical definition of model — (f(x) = linear predictor / not 'tries to') logit: $\frac{p}{1-p}$

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — specify LRTs mathematically

define all terms — AIC, LRT, μ, β

— Results incomplete

5. Model assessment:

1.5/2

CLEARLY state model assumptions: + give PRIMARY references — Here, not above

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) — not cook

vif (to identify multicollinearity)

(square QQ)

5.25/8

0.75/1

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

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

1/1

7. Plots: (some plots too big)

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

+ 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:

2.5/4

logistic: Name _____ JaSa 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**)

0.75/1

(too many digits)

2. Introduction/Background:

0.75/1

brief background and statement of scientific question — imprecise

all variables defined

3. EDA: Don't need Figure 2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

1.5/2

4. Model fitting:

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:

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) (numerical only)

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

vif (to identify multicollinearity)

Define + Define all performance measures

6.25/8

(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 7. Plots:

label size (not too small) informative captions

placement explanations

0.75/1 8. Conclusions

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

(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 _____ FeSa-Sc 8/12 → 4/6

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**)

- too many digits

0.75/1

2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

2/2

3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical square pairs plots

square QQ

1/2

4. Model fitting:

Table 3 after model + fitting description

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — describe entire procedure

define all terms — AIC, LRT + all 'pseudo' measures

incorrect interpretation(s)

5. Model assessment:

1/2

CLEARLY state model assumptions: + give PRIMARY references

First + explain, THEN assess

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

0.5/1

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

equation

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

0.5/1

8. Conclusions

be specific

not correct

1. recap analysis

2. state and interpret main findings

spell check

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

logistic: Name _____ He 5c 8.75/12 → 4.375/6

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:

0.5/1

brief background and statement of scientific question — unclear

all variables defined

3. EDA:

2/2

univariate numerical — 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 — This procedure does not seem particularly reproducible

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): + explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

6.25/8

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 _____ LeSe 5.75/12 → 2.875/6

1. Formatting: Don't need contents

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**)

0.25/1

2. Introduction/Background: — too many digits imprecise

0.75/1

brief background and statement of scientific question

all variables defined

3. EDA: All vars

0.75/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical – all pairs

4. Model fitting: no R output + clearly interpret

0.75/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC, LRT, deviances

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

0.75/2

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): clear + explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

3.25/8

1/1 (ok)

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

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

+ R output

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 _____ Ja So 11/12 → 5.5/6 → 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

Don't need R details

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

2/2

3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical (square)

2/2

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:

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) (vide more paragraphing)

Define gvif

7.5/8

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

0.5/1

not correct

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:

3.5/4

logistic: Name _____ La So 7.75/12 → 3.875/6

1. Formatting:

0.25/1

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: results not in logical order

1.5/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical — pairs plots

4. Model fitting:

1/2

give mathematical definition of model no ^

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — not 'significantly'

define all terms — AIC, + incorrect interpretation

5. Model assessment: Lrt, of coef, Deviances

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):

+ explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

4.75/8

(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

1/1 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 'confirmed' 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: - Table of missing values unnecessary

3/4

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

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:

brief background and statement of scientific question — imprecise

all variables defined

3. EDA: not 'influence' / 'impact' - incorrect interpretation

don't use the word 'significantly' outside the context of a statistical test

1.75/2

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 — stepwise

define all terms — Deviances

5. Model assessment: not 'verify'

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

6.25/8

0.75/1

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

Define y

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

7. Plots:

Y/1

label size (not too small) informative captions

placement explanations

0/1

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) :

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: Write more in words/paragraphs, not bullet points

2.25/4

logistic: Name _____ ThSt 9.5/12 → 4.75/6

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:

brief background and statement of scientific question explicitly

all variables defined

1.25/2 3. EDA:

univariate numerical — bivariate numerical (cor) not just graphic

univariate graphical — bivariate graphical all pairs

4. Model fitting: I don't understand Table 2; why noted for model ? ? Is it because all < 0.05 ?

use paragraphing

1.5/2 give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected 'arbitrary'?

define all terms

Table 3: incomplete + $\hat{\beta}$ / $e^{\hat{\beta}}$

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

7. Plots:

1/1

label size (not too small) informative captions

placement explanations

0.75/1

8. Conclusions

1. recap analysis 2. state and interpret main findings

be specific

for

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:

3.5/4

logistic: Name _____ SaTa 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:

0.5/1 brief background and statement of scientific question not correct + imprecise

all variables defined

3. EDA:

1.5/2 univariate numerical bivariate numerical (cor)

univariate graphical bivariate graphical all pairs

4. Model fitting:

1.25/2 give mathematical definition of model not correct/define in terms of logit

state how model fitted (ie, maximum likelihood) (imprecise description)

CLEARLY describe how model selected

define all terms LRTs + which models/hyps are tested in results table

should not have 0.00

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references

1/2 1. binary outcome 2. independent obs not vars
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

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:

2.75/4

logistic: Name _____ ArTe 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

2. Introduction/Background: too many digits imprecise

0.75/1

brief background and statement of scientific question

all variables defined

3. EDA: mostly not done

0.5/2

univariate numerical — bivariate numerical (cor)

univariate graphical) or — bivariate graphical pairs

Don't need univariate

4. Model fitting:

0.75/2

give mathematical definition of model — First, BEFORE fitting

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected explain don't just state procedure

define all terms

5. Model assessment: évaluer pas 'vérifier'

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): clearly

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

+explanations

3.75/8

0.75/1 (p̂)

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

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

7. Plots: 0.75

label size (not too small) informative captions

placement explanations

8. Conclusions 0.25/1

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

9. Overall presentation (clarity of explanations, appropriate citations / references) : 0.5/1 clearly somewhat incomplete + superficial

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

logistic: Name _____ LoTs 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**)

0.25/1

2. Introduction/Background: — too many digits

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

all variables defined

3. EDA: all

1.5/2 univariate numerical — bivariate numerical (cor) — not just graphics

univariate graphical — bivariate graphical + all pairs

4. Model fitting:

0.75/1 give mathematical definition of model (define in terms of logit)

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

CLEARLY describe how model selected — stepwise

define all terms — AIC

5. Model assessment:

0.75/1 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): — incomplete explanation

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity) — Define

Rstuds vs leverage plot too big

4/8

(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: Figure 1 too large

label size (not too small) informative captions no R formulas

placement explanations

0/1 8. Conclusions not done

1. recap analysis 2. state and interpret main findings

0.5/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : incomplete + no R output

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

logistic: Name _____ Ya Tu 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**)

0.75/1 Throughout; Don't need R details

2. Introduction/Background:

brief background and statement of scientific question — unclear and imprecise

all variables defined

0.5/1

3. EDA:

univariate numerical — bivariate numerical (cor) (too small)

univariate graphical — bivariate graphical

1.75/2

4. Model fitting:

give mathematical definition of model — no ~

state how model fitted (ie, maximum likelihood) — what measure for categorical?

CLEARLY describe how model selected — explain, explicitly

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)

carry out assessment (numerical / graphics): + explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

1.5/2

6/8

0.75/1 (ok) logit(p) (which are baseline groups)

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

0.75/1 8. Conclusions

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

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

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 _____ CQVe 8.5/12 → 4.25/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages p. 1

no raw R code or output — all pages numbered

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

0.75/1 Don't need R details — too many digits

2. Introduction/Background: imprecise

0.75/1 brief background and statement of scientific question

all variables defined

3. EDA: Don't need hyp tests Don't need figure 3B

1.75/2 univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical square QQ

4. Model fitting:

1.25/2 give mathematical definition of model (not log-linear relationship with log odds)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected why these models

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): + explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

5.75/8

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

be specific

1. recap analysis 2. state and interpret main findings

not entirely 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 _____ Zo Yo 6.25/12 ⇒ 3.125/6

1. Formatting: use your own words

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**)

0/1

2. Introduction/Background:

brief background and statement of scientific question imprecise

all variables defined

0.75/1

3. EDA: explanation - use your own words

univariate numerical — bivariate numerical (cor) square cor mat + too small to read

univariate graphical — bivariate graphical

1/2

4. Model fitting:

give mathematical definition of model First, before estimation

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC, Deviances, + other methods (mathematically)

1.25/2

5. Model assessment: not 'ensure', what do you mean by 'accurate'?

use paragraphing

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):

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

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

vif (to identify multicollinearity)

Define + Define all performance metrics

4/8

(ok) logit(p)

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

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

0.5/1 7. Plots: many plots too small to be legible

label size (not too small) informative captions Figure not Fig.

placement explanations

0.75/1 8. Conclusions

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

0/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : Use your own words throughout

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: (*) please re-write all copied/plagiarized sections using your own words

2.25/4

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

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:

brief background and statement of scientific question — imprecise

all variables defined

0.75/1

3. EDA: Don't need hyp tests / very incomplete

univariate numerical — bivariate numerical (cor) all vars

univariate graphical all vars — bivariate graphical pairs

0.5/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — be specific

define all terms AIC, Deviance

1.25/2

5. Model assessment: 'valid + reliable' means what?

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): all numeric

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define + Define all performance metrics

4.75/8

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

0.75/1 7. Plots:

label size (not too small) informative captions

placement explanations

0.5/1 8. Conclusions vague be specific + precise

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:

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 _____ Zhya 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: 'robust'? + clarify imprecise

brief background and statement of scientific question

all variables defined Don't need hyp tests

0.75/1

3. EDA: Figure 2. too big; Figure 3 stratified boxplots

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

1.75/2

4. Model fitting:

give mathematical definition of model first, before fitting

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected not done?

define all terms

1/2

5. Model assessment: not 'ensure'

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):

1/2

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

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

vif (to identify multicollinearity)

Define

5.5/8

0.75/1

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

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

logistic: Name _____ Anze 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**)

0.75/1

2. Introduction/Background: — too many digits, imprecise

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA:

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

univariate graphical — bivariate graphical

1.5/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — incomplete

define all terms

1.75/2

5. Model assessment: — not 'ensure'

CLEARLY state model assumptions: + give PRIMARY references

1. binary outcome 2. independent obs — incorrect assessment
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)

+ explanations

Define

6/8

0.75/1

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

0.75/1

8. Conclusions vague - 'valuable insights'? be specific

1. recap analysis 2. state and interpret main findings

(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: _____

2.75/4

logistic: Name _____ Hazh 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: imprecise

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical pairs

1.5/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — explain backward (completely) Before fitting

define all terms

Don't need Table 2 AIC

1.25/2

5. Model assessment:

CLEARLY state model assumptions: + give PRIMARY references not 'ensure'

First, then explain + assess

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): + explanations

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

5.75/8

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

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

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

0.5/1 7. Plots: Blank space p. 2

label size (not too small) informative captions

placement explanations

0.5/ 8. Conclusions vague

be specific

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

0.5/ 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 _____ Huzh 6.75/12 → 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**)

0.75/1

– too many digits

2. Introduction/Background:

brief background and statement of scientific question

all variables defined

0.5/1

3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical (histograms not boxplots) — bivariate graphical

1.5/2

4. Model fitting:

give mathematical definition of model – odds also

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms

1/2

AIC / LRT – what are you testing specifically?

5. Model assessment: – assessed not verified

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): + explain plots, don't just conclude

scatterplots of logit vs. predictors (linearity assumption)

DEFINE → Cook's distance / standardized residuals (outliers) – not cook's

vif (to identify multicollinearity)

(*) 'variance importance' – do you mean variable importance? + not defined – define

4.5/8

0.75/1

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

0.25/1

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) :

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

logistic: Name _____ Ru Zh 8.5/12 → 4.25/6

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 — all pages numbered

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

2. Introduction/Background: use paragraphing 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: square QQ

1/2

give mathematical definition of model — Define + interpret ORs

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — not done?

define all terms — AIC, McFadden, Wald, variable importance

Put results in a Table for clarity

5. Model assessment: not 'ensure'

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

+ Define all performance metrics

(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

0.5

7. Plots: + many plots are too small

label size (not too small)

informative captions

PCA of 'Data'?

placement

explanations

PC1 + PC2 (not PCA)

0.5/1

8. Conclusions

be specific

(ok 1. recap analysis)

2. state and interpret main findings

not correct

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:

2.5/4

logistic: Name _____ Tizh 7.75/12 → 3.875/6

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**)

2. Introduction/Background:

0.5/1

brief background and statement of scientific question — unclear + imprecise

all variables defined

3. EDA:

1/2

univariate numerical — 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 — incomplete explanation

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)

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)

↳ Define

5/8

(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

8. Conclusions

0.5/1

be specific 'robust'?

1. recap analysis 2. state and interpret main findings

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