logistic: Name _____ Jeal 8.25/12 → 7.125/6

1. Formatting:

all margins 2.5cm — informative title

12 pt size throughout (Figure 10) — 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

imprecise

2. Introduction/Background:

0.75/1

brief background and statement of scientific question

all variables defined

3. EDA:

what measure?

1.75/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting:

1.25/2

give mathematical definition of model no ^

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected 'gradually' ??

define all terms AIC

5. Model assessment: not 'verify' / not 'confirm'

1/2

CLEARLY state model assumptions: + give PRIMARY references

↳ First, 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)

5.5/8

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

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 not correct

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

logistic: Name _____ Na Al FR 5/12 → 2.5/6

1. Formatting:

Don't need contents

0.75/1

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

Don't need R details

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

2. Introduction/Background: use paragraphing

0.5/1

brief background and statement of scientific question

all variables defined

3. EDA:

1/2

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

4. Model fitting:

1/2

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC, McFadden, LRT, H-L test, Wald

Importance

5. Model assessment: + Define all performance metrics

CLEARLY state model assumptions: + give PRIMARY references

not Done?

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

carry out assessment (numerical / graphics):

0/2

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

+explanations

3.25/8

not done?

0/1 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

0.25/1 8. Conclusions

1. recap analysis 2. state and interpret main findings

be specific

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

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: very incomplete

1.75/4

logistic: Name MeAt 9.5/12 → 4.75/6

1. Formatting:

all margins 2.5cm — informative title (- no cover page)

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

0.75/1

3. EDA:

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

univariate graphical — bivariate graphical (square)

1.75/2

4. Model fitting:

give mathematical definition of model (no ^)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected (stepwise)

define all terms (BIC, LRT)

1.25/2

5. Model assessment: (not 'confirm')

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)

Define + Define all performance metrics

6/8

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

8. Conclusions

0.75/1

be specific

1. recap analysis 2. state and interpret main findings

(implicit EDA)

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

0.75/1

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 _____ Al Ay 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: imprecise

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA: Table 1 illegible + plots too small

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

1.5/2

4. Model fitting: seems incomplete

give mathematical definition of model no $\hat{}$; y not defined

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected not done?

define all terms performance metrics

0.75

5. Model assessment: not 'validate'

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)

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

0.5/1

7. Plots: + several very small plots

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

be specific 'robust'?

1. recap analysis 2. state and interpret main findings

0.5/1

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

order not logical, please follow order in evaluation criteria

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: - use 1-column format

2.25/4

logistic: Name Pa Au 8.5/12 → 4.25/6

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

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA:

all summary stats

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical all pairs

1.5/2

4. Model fitting:

give mathematical definition of model

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected stepwise / + start with more vars

define all terms AIC

1.25/2

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)

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

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

be specific

1. recap analysis 2. state and interpret main findings

not completely 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:

3/4

logistic: Name _____ MoAz 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**)

Y / Don't need R details

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:

1.25/2

give mathematical definition of model — no ^

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — 'relevant'

define all terms

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):

Clearly

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define

+ clear explanations

SQUARE QQ, but why assume normality?

6/8

1/1 (ok) (unit?)

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 Vague

0.5/1 1. recap analysis 2. state and interpret main findings (be specific)

(implicit EDA) Vague

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: _____

3/4

logistic: Name NiBa 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**)

0.75/1 2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

1.75/2 3. EDA:

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical — all pairs

1.25/2 4. Model fitting:

give mathematical definition of model — First, before results

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — LRT, McFadden

5. Model assessment:

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

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

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

+ explanations

Define

5.75/8

6.75/1

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

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

7. Plots:

1/1

label size (not too small) informative captions

placement explanations

0.5/1

8. Conclusions

1. recap analysis 2. state and interpret main findings

be specific

+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 _____ CaBe 7/12 → 3.5/6

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

Don't need R fns

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: incomplete

0.5/2 — univariate numerical — bivariate numerical (cor) (crosstabs)

univariate graphical — bivariate graphical

4. Model fitting: Don't need 'Data Type Conversion' details

1/2 (*) — give mathematical definition of model (in terms of logit)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms — AIC / metrics 'confirming' — Don't 'stat verify' hyp, not 'may be appropriate'?

5. Model assessment:

1.25/2 — CLEARLY state model assumptions: + give PRIMARY references — Just say what you do and why, don't be vague

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)

(*) explain CV — Define

4.5/8

(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 + small plots

label size (not too small) informative captions

placement explanations

0.25/1 8. Conclusions be specific not correct

1. recap analysis 2. state and interpret main findings

(+ 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 Ma Be-Oz 8/12 → 4/6

1. Formatting: - no abstract / 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**)

0.75/1

2. Introduction/Background: use paragraphing

brief background and statement of scientific question imprecise

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

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

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): not 'ensuring'

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity) + explanations

Define

5.5/8

(ok) $\text{logit}\ \frac{p}{1-p}$

1/1

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

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

7. Plots:

0.75/1

label size (not too small) informative captions

placement explanations

8. Conclusions – vague be specific

0.25/1

1. recap analysis 2. state and interpret main findings

incomplete

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: _____

_____

_____

_____

_____

2.5/4

logistic: Name _____ El Ba-C 7.75/12 → 3.875/6

1. Formatting:

Don't need contents

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:

0.5/1

brief background and statement of scientific question — imprecise

all variables defined — 'dimensionless'?

1.25/2

3. EDA: — ALL VARS

univariate numerical — ~~all summary stats~~ — 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 — converges more 'reliably'?

define all terms CV, ~~AIC~~

5. Model assessment:

1.25/2

CLEARLY state model assumptions: + give PRIMARY references

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

carry out assessment (numerical / graphics):

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

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

vif (to identify multicollinearity)

Define + Define all performance metrics

5/8

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

1/1

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 (ok) explanations

8. Conclusions

0.5/1

be specific

1. recap analysis 2. state and interpret main findings

+ EDA

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

0.5/1

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

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

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

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

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

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

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

Other:

2.75/4

logistic: Name _____ Pa Bi 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

name on all pages

all pages numbered

no blurry plots (**NOT png**)

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

2. Introduction/Background:

brief background and statement of scientific question — imprecise

all variables defined

0.75/1

3. EDA: write out words

univariate numerical — all summary stats

univariate graphical

bivariate numerical (cor) — what measure? you have mostly categorical vars

bivariate graphical

introduce then put after text

1.5/2

4. Model fitting:

give mathematical definition of model — no ^

state how model fitted (ie, maximum likelihood)

1/2

CLEARLY describe how model selected — explain stepwise

define all terms — AIC, LRT (χ^2 test) — Results incomplete: $\hat{\beta}$, $SE(\hat{\beta})$, z, p

5. Model assessment:

CLEARLY state model assumptions; + give PRIMARY references (First, + explain, then assess)

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) (only numeric)

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**

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

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 – structure + run-on sentences

1. recap analysis 2. state and interpret main findings

be specific

0.5/1

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

+somewhat incomplete

poor satisfactory good excellent

10. Other comments:

A – no / incomplete / insufficient references

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

C – interpretation (cannot conclude causation, only association)

D – use your OWN WORDS / no apparently unattributed quotations

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

F – univariate graphical: histograms not boxplots

G – (mathematical) model misspecified / unclear

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

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

Other:

2.25/4

logistic: Name _____ ImBo 9/12 → 4.5/8

1. Formatting:

all margins 2.5cm — informative title

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

no raw R code or output — all pages numbered

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

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 just graphic)

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 AIC

5. Model assessment:

1.25/6 — 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)

(+ explanations)

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 (Figure 3)

placement explanations

0.25/1

8. Conclusions

1. recap analysis 2. state and interpret main findings (be specific — 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: _____

3/4

logistic: Name _____ Me Bo 8/12 → 4/6

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

Don't need R details

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 just graphic

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 — methods, not R commands

define all terms AIC

5. Model assessment: - no R formulas

1.25/2

- not 'ensure reliability'

CLEARLY state model assumptions: + give PRIMARY references

- not 'verify'

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) — not 'confirming'

Define

5.5/8

(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

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

8. Conclusions

be specific 'clinically relevant'?

0.25/1

1. recap analysis 2. state and interpret main findings

not correct

(+ 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 Ma Bo-H 8.5/12 → 4.25/6

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

brief background and statement of scientific question

all variables defined

0.75/1

3. EDA: (ok)

univariate numerical — bivariate numerical (cor)

univariate graphical — bivariate graphical

2/2

4. Model fitting: why binarize?

give mathematical definition of model incomplete

state how model fitted (ie, maximum likelihood) incorrect interp

CLEARLY describe how model selected

define all terms AIC / deviances

1/2

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)

-> + Define metrics not defined

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

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: _____

2.75/4

logistic: Name _____ Ca Bu 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**)

(– too many digits imprecise)

2. Introduction/Background:

0.75/1

brief background and statement of scientific question

all variables defined

3. EDA:

1.5/2

univariate numerical — bivariate numerical (cor) (what measure?)

univariate graphical — bivariate graphical + pairs (numeric)

4. Model fitting:

1/2

give mathematical definition of model – First (explain all hyps mathematically)

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected F-stat?

define all terms Wald, McFadden!

Results table incomplete

5. Model assessment:

1.25/2

CLEARLY state model assumptions: + give PRIMARY references

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

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

Define performance metrics

5.25/8

0.75/1 put coefs in equation

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 'relevance'?

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 _____ DeCa 8.5/12 → 4.25/6

1. Formatting:

0.75/1

all margins 2.5cm — informative title

12 pt size throughout — 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: — too many digits 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 numerical

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

5. Model assessment:

1.25/2

First, before assessments

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

+ explanation

5.5/8

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

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

not correct

1. recap analysis 2. state and interpret main findings

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

poor satisfactory good excellent

10. Other comments:

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: put tables 4+5 together into 1 table

3/4

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

1. Formatting:

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

0.25/1

2. Introduction/Background: use paragraphing

0.75/2

brief background and statement of scientific question imprecise

all variables defined

3. EDA:

1.5/2

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

4. Model fitting:

1.25/2

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 LRT stepwise

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

5/8

1/1 (why 2 times?)

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

0/1 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) : no R output 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:

2.25/4

logistic: Name _____ LyCo 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

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: why keep both logHOMA + logIns? VIFs?

1.25/2

give mathematical definition of model — somewhat incomplete

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected — explain

define all terms accuracy/recall/precision/F1/AUC/VIF

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): what test do you use to compare models?

clearly explain

scatterplots of logit vs predictors (linearity assumption)

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

vif (to identify multicollinearity) — define

6.25/8

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

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: _____

_____

_____

_____

_____

2.5/4

logistic: Name _____ GaCo-Lo 8.5/12 → 4.25/6

1. Formatting:

0.5/2

all margins 2.5cm — informative title

12 pt size — name on all pages

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

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

too many digits

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:

1.25/2

give mathematical definition of model — no R formulas

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms → Wald, Deviance

(*)

5. Model assessment:

1.25/2

CLEARLY state model assumptions: + give PRIMARY references

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

carry out assessment (numerical / graphics):

scatterplots of logit vs. predictors (linearity assumption)

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

vif (to identify multicollinearity)

+ explanation

5.75/8 (*) only include interaction if you also include the associated main effects

6. Write out final *estimated* model **mathematically** (but wrong equation / already penalized) 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

8. Conclusions 0.25/0.5

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

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 _____ AxG 7.5/12 → 3.75/6

1. Formatting:

- Don't need abstract

all margins 2.5cm — informative title

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

no raw R code or output — all pages numbered

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

2. Introduction/Background:

0.75/1 brief background and statement of scientific question — imprecise (+ 'robust'?)

all variables defined

3. EDA:

1.5/2 univariate numerical — bivariate numerical (cor) — what measure?

univariate graphical — bivariate graphical

4. Model fitting: very incomplete + hard to follow

0.75/2 give mathematical definition of model BEFORE fitting

state how model fitted (ie, maximum likelihood)

CLEARLY describe how model selected

define all terms AIC, BIC + all performance metrics + Deviances

5. Model assessment: not 'ensure validity'

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
- SQUARE ROC plot

5/8

(ok) Write in terms of logit

1/1

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

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

0.75/1

7. Plots:

label size (not too small) informative captions

placement explanations

0.25/1

8. Conclusions

vague

be specific

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

0.5/1

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

not in logical order, please follow order in evaluation criteria

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