GWAS: Name 282278 5.25/12 → 2.625/6

1. Formatting:

0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~ ok

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.5/0.5

2. ~~Introduction~~/Background:

brief statement of scientific question

all variables defined

3. PCA:

B

0.75/1.5

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2

4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons

C D

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

0.75/2

5. Association / post-association analysis:

G H

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) E K

LD heatmap (optional – does NOT count); measure of LD F

3.25

0/1.5 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP (not done)

0.5/1 7. Plots: (O)

label size (not too small) captions

placement **NOT BLURRY** (L)

0.75/1.5 8. Conclusions

recap analysis state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory good excellent

0.5/1 10. Other comments:

See attached

(*) please email to me the code you used to generate each plot as well as the output from

sessionInfo()

2

Comments **Name:** 282278

00 – informative title

A – eda ok

B – PCA + explain

C – define and explain HWE

PC2

PC1

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation) ⇒ $-\log_{10} p$

J – identify significant markers (show in table)

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- use 12 pt size
- re-arrange layout so that there is less blank space
- Define ALL terms: MAF, IBD, LD, etc + measures + explain
- your explanations are very incomplete
- write out the model you are fitting and explain what test you are using for assoc - where do the p-values come from?
- clearly explain reasons for filtering steps

GWAS: Name 29602 4.75/12 → 2.375/6

1. Formatting:

0.25/0.5

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- ~~max 7 pages~~ ok
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

0.5/0.5

2. Introduction/Background:

- brief statement of scientific question
- all variables defined

0.5/1.5

3. PCA:

- explain relation between PCs and population stratification (B)
- plot pc2 (y-axis) vs pc1 (x-axis) + Square

0.75/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons (C, D)
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

0,75/2

5. Association / post-association analysis:

- Describe association analysis in words and mathematically (G, H)
- Manhattan plot (I)
- lambda analysis (including **SQUARE** QQ-normal plots) (E, K)
- LD heatmap (optional – does NOT count); measure of LD (F)

2,75

0/1.5 6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP (4)

0.5/1 7. Plots:

(0) label size (not too small)

placement

captions

NOT BLURRY (L)

0.75/1.5 8. Conclusions

recap analysis

(0K) state main findings

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor (satisfactory) good excellent

0.5/1 10. Other comments:

see attached

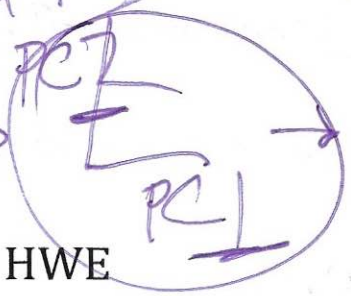
(*) please email to me the code you used to generate each plot as well as the output of sessionInfo()

a2

Comments **Name:** 29602

00 – informative title

A – eda – don't need file names

B – PCA + explain → 

→ square

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test not clear

H – write out final model *mathematically*

I – Manhattan plot (and explanation) – log(10) p-value

J – identify significant markers (ok)

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small — some

P – plot size (see text) — square λ, QQ plots — some too big

Q – plot layout (see text) — include in text

R – overall organization and explanation of procedure

S – other:

- You need to explain the reasons for each step
- There is a lot missing here - you need to define all terms and explain + measure (eg, MAF, IBD, LD, etc)
- clearly explain reasons for filtering steps

GWAS: Name 296954 6.75/12 → 3.375/6

1. Formatting:

0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max **7** pages~~ ok

informative title

name on all pages

all pages numbered

no blurry plots ~~(**NOT png**)~~

0.5/0.5 2. Introduction/Background:

brief statement of scientific question

all variables defined

0.5/1.5 3. PCA:

explain relation between PCs and population stratification (B)

plot pc2 (y-axis) vs pc1 (x-axis) + Square

4. Pre-processing / QC steps:

0.75/2

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality (C) (D)

Overall QC explanation

1.25/2 5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)

LD heatmap (optional – does NOT count); measure of LD (F)

3.5

1/1.5 2.5

6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

0.5/1 7. Plots: +Coefs

label size (not too small) captions

placement (O) **NOT BLURRY** (L)

0.75/1.5 8. Conclusions

recap analysis (state main findings) (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor (satisfactory) good excellent

0.5/1 10. Other comments:

- see attached

(*) please email me the code you used to generate each plot as well as the output from sessionInfo()

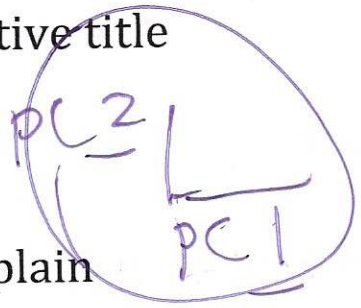
3.25

Comments **Name:** 296954

00 – informative title

A – eda

B – PCA + explain



+ Square

C – define and explain HWE (mathematically)

D – define and explain HWE **test**

E – define λ

F – define LD measure – it looks like you are confusing individuals with SNPs

G – explain association test

H – write out final model *mathematically* – what is sex. (numerically)?

I – Manhattan plot (and explanation)

J – identify significant markers → $-\log_{10} p$

K – square QQ plots ↳ show in table

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions (ok)

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- use 12 pt size
- Clearly define ALL terms (MAF, etc) and measures + explain
- explain assoc test - where do the p-values come from?
- clearly explain reasons for all filtering steps

GWAS: Name 300467 6.25/12 → 3.125/6

1. Formatting: 0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~

informative title

name on all pages

all ~~pages~~ numbered

no blurry plots (**NOT png**)

2. Introduction/Background: 0.5/0.5

brief statement of scientific question

all variables defined

3. PCA: 0.75/1.5

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis) B

4. Pre-processing / QC steps: 0.75/2

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality C D

Overall QC explanation

5. Association / post-association analysis: 1/2 G H

Describe association analysis in words and mathematically

Manthattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) E

LD heatmap (optional – does NOT count); measure of LD F

3.5

0.75/1.5

6. Write out final estimated model **mathematically** (for a given SNP) include coefs (H)

hat on response variable MUST RELATE TO SNP

0.5/1

7. Plots:

label size (not too small) (O) captions

placement **NOT BLURRY** (L)

1.25/1.5

8. Conclusions

recap analysis state main findings (J)

0.25/0.5

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

poor satisfactory good excellent

10. Other comments:

see attached

(*) please email to me the code you used to generate each plot as well as the output from sessionInfo()

2.75

Comments **Name:** 300467

00 – informative title

A – eda) OK

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* – incorrectly specified

I – Manhattan plot (and explanation) → -log10p

J – identify significant markers – put in table

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

PC1 PC2 (y vs. x)

O – plot labels too small

P – plot size (see text) Ok

Q – plot layout (see text) OK

R – overall organization and explanation of procedure

S – other:

- You need to define all terms (MAF, LD, etc), measures and explain all methods
- model incorrectly specified – does not include SNP. Read the R fn. help
- Should look something like:

$$\widehat{\text{transHDL}} = \hat{\beta}_0 + \hat{\beta}_1 \text{age} + \hat{\beta}_2 \underset{\text{(code)}}{\text{sex}} + \hat{\beta}_3 (\text{var}_i \text{ for the SNP}) + \hat{\beta}_4 PC1 + \cdots + \hat{\beta}_{13} PC10$$

- Clearly explain reasons for the filtering steps

GWAS: Name 301486 5.75/12 → 2.875/6

1. Formatting:

0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~ (ok)

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.5/0.5 2. Introduction/Background:

brief statement of scientific question

all variables defined

0.5/1.5 3. PCA:

explain relation between PCs and population stratification (B)

plot pc2 (y-axis) vs pc1 (x-axis) + Square

0.75/2 4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality (C) (D)

Overall QC explanation

0.75/2 5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)

LD heatmap (optional – does NOT count); measure of LD (F)

0.75/1.5

6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

0.5/1

7. Plots:

label size (not too small) captions

placement (O) **NOT BLURRY** L

0.75/1.5

8. Conclusions

recap analysis state main findings (J)

0.25/0.5

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

poor satisfactory good excellent

0.5/1

10. Other comments:

see attached

(*) please email to me the code you used to generate each plot as well as the output of

sessionInfo()

2.75

Comments **Name:** 301436

00 – informative title

A – eda

B – PCA + explain PC2 PC1 + square

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* incorrect explanation of model

I – Manhattan plot (and explanation) $-\log_{10} p$

J – identify significant markers put in table

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- ~~You need to define all terms (MAF, IBD, HWE, LD, λ, etc) + measures + explain~~
- ~~you need to explain the methods~~
- ~~your model is incorrectly specified;~~ Should be:

pheno =
↑
transHDL,
NOT LDL
NOT TG

(+ Define what the covariates are)

- no g → it's not a ~~parameter~~ of the model fitted by the R fn → see the help

→ only 1 snp per linear model, not all in same model

- clearly explain reasons for filtering steps

GWAS: Name 301876 4.25/12 → 2.125/6

1. Formatting:

0.5/0.5

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- ~~max 7 pages~~ ok
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

2. Introduction/Background:

0.5/0.5

- brief statement of scientific question
- all variables defined

0.5/1.5

3. PCA:

- explain relation between PCs and population stratification (B)
- plot pc2 (y-axis) vs pc1 (x-axis) + square

0.75/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons (C) (D)
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

0.75/2

5. Association / post-association analysis:

- Describe association analysis in words and mathematically (G) (H)
- Manhattan plot (I)
- lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)
- LD heatmap (optional – does NOT count); measure of LD (F)

0/1.5 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP (not done ??)

0.75/1 7. Plots:

label size (not too small) (O) captions

placement **NOT BLURRY** (L)

0/1.5 8. Conclusions

recap analysis state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory - incomplete good excellent

0.25/1 10. Other comments:

- see attached

(*) please email me the code you used to generate each plot as well as the output from sessionInfo()

1.25

Comments **Name:** 301876

00 – informative title

A – eda – Don't need file names

B – PCA + explain – dimension reductionality is not why we do PCA here

C – define and explain HWE

D – define and explain HWE **test**

+ square

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation) → $-\log_{10} P$

J – identify significant markers → show in table

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions – no conclusions? recap procedure + give results + interp.

N – **no raw R**

O – plot labels (too small) but ok

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- define 'call rate' more carefully
- Define all terms (IBD, HWE, etc) and measures + explain
- report incomplete, see comments other side
- For help with the model, see the R fn. you use
- needs References

GWAS: Name 301954 5.5/12 → 2.75/6

1. Formatting:

0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max **7** pages~~

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.5/0.5 2. Introduction/Background:

brief statement of scientific question

all variables defined

0.5/1.5 3. PCA: (B)

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis) + square

0.75/2 4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality (C) (D)

Overall QC explanation

1/2 5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)

LD heatmap (optional – does NOT count); measure of LD (F)

3.25

1/1.5 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

0.5/1 7. Plots: (O)

label size (not too small) captions

placement **NOT BLURRY** (L)

0.5/1 8. Conclusions

recap analysis state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory good excellent

0.5/1 10. Other comments:

- see attached

(*) please email to me the code you used to generate each plot as well as the output from

sessionInfo()

2.25

Comments **Name:** 301954

00 – informative title

A – eda ok – Don't need table 1

B – PCA + explain PC2 PC1 + square

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* – model should include 10 PCs

I – Manhattan plot (and explanation) → $-\log_{10} P$

J – identify significant markers – should come after modeling

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- use 12 pt size
- Don't need table 1
- more carefully define ALL terms + measures
- explain methods - where do the p-values come from? EXPLAIN
- clearly explain reasons for the filtering steps

GWAS: Name 303060 5/12 → 2.5/6

1. Formatting: 0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~ ok

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background: 0.5/0.5

brief statement of scientific question

all variables defined

3. PCA: 0.5/1.5

explain relation between PCs and population stratification (B)

plot pc2 (y-axis) vs pc1 (x-axis) + square

4. Pre-processing / QC steps: 0.75/2

SNP QC: criteria and reasons

sample QC: criteria and reasons (C)

Hardy-Weinberg equilibrium: what it means and how it relates to quality (D)

Overall QC explanation

5. Association / post-association analysis: 0.75/2

Describe association analysis in words and mathematically (G) (H)

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (k)

LD heatmap (optional – does NOT count); measure of LD (F)

0.5/1.5

6. Write out final estimated model **mathematically** (for a given SNP) H

hat on response variable MUST RELATE TO SNP

7. Plots:

label size (not too small) O captions

placement **NOT BLURRY** L

0.75/1.5

8. Conclusions

recap analysis state main findings J

0.25/0.5

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

poor satisfactory good excellent

0.5/1

10. Other comments:

see attached

* please email to me the code you used to generate each plot as well as the output from sessionInfo()

Comments **Name**: 303060

00 – informative title

A – eda (OK)

B – PCA + explain PC2 PC1 + square

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* NOT R formula

I – Manhattan plot (and explanation) + $-\log_{10} p$

J – identify significant markers – put in table

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- use 12 pt size
- Define ALL terms: MtF, IBD, LD, etc
- Don't need section 5 + measures + explain
- Write model mathematically NOT as R formula
- Clearly explain reasons for the filtering steps

GWAS: Name 311822 8.75/12 → 4.375/6

good job!

1. Formatting:

0.5/0.5

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- ~~max 7 pages~~ ok
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

0.5/0.5

2. Introduction/Background:

- brief statement of scientific question
- all variables defined

0.5/1.5

3. PCA: (B)

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) + square

1.5/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality (D)
- Overall QC explanation

1.25/2

5. Association / post-association analysis: (G) (H)

- Describe association analysis in words and mathematically
- Manhattan plot (I)
- lambda analysis (including **SQUARE** QQ-normal plots) (E)
- LD heatmap (optional – does NOT count); measure of LD (F)

4.25

1.25/1.5 6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP (4)

0.5/1 7. Plots:

label size (not too small) captions

placement **NOT BLURRY** (2)

1.5/1.5 8. Conclusions

recap analysis state main findings

0.5/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (2)

poor satisfactory good excellent

0.75/1 10. Other comments:

1 - see attached

(*) please email to me the code you used to generate each plot as well as the output from

sessionInfo()

4.5

Comments **Name**: 311822

00 – informative title

A – eda ok

B – PCA + explain + square

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* – need intercept,

I – Manhattan plot (and explanation) what is numerical value of 'sex'?

J – identify significant markers – log(10) p (label)

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions – split into paragraphs

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- use 12 pt size
- Define all terms - MAF, IBD, LD, etc
- model not complete + measures to explain
- clearly explain reasons for filtering steps
- needs references

GWAS: Name 316305 5.75/12 → 2.875/6

1. Formatting:

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- ~~max 7 pages~~
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

0.5/0.5

2. Introduction/Background:

- brief statement of scientific question
- all variables defined

0.5/0.5

3. PCA:

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) + square (B)

0.5/1.5

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons (C) (D)
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

0.75/2

5. Association / post-association analysis: (G) (H)

- Describe association analysis in words and mathematically
- Manhattan plot (I)
- lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)
- LD heatmap (optional – does NOT count); measure of LD (F)

0.75/2

└ incorrect model

6. Write out final estimated model **mathematically** (for a given SNP) (H) incorrect model

1/1.5 **hat** on response variable MUST RELATE TO SNP

7. Plots:

0.5/1 label size (not too small) (O) captions (L)

placement **NOT BLURRY**

0.5/1.5 8. Conclusions

recap analysis state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory good excellent

0.5/1 10. Other comments:

see attached

* please email to me the code you used to generate each plot as well as the output from

sessionInfo()

2.75

Comments **Name:** 316305

00 – informative title — what are the data set variables?

A – eda

B – PCA + explain — PC2 / PC1 + square (what are the red dots for?)

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* — don't need R-formula; Define 'SNP' + 'sex' numerically → model incorrect

I – Manhattan plot (and explanation)

J – identify significant markers — ↳ $-\log_{10}(p)$ — show in table

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions — not clear

N – **no raw R**

O – plot labels too small

P – plot size (see text) OK

Q – plot layout (see text) OK

R – overall organization and explanation of procedure

S – other:
- Define all terms (MAF, IBP, etc) + measures + explain
- Your conclusions should come at the end not in intro
- your model should only have 1 SNP at a time – Read the help for the R fn.

model:

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 age + \cdots$ (and since it's the estimated model no error term)

for each SNP separately, not combined

- Define (numerically) the variables SNP and SEX in the model
- Clearly explain reasons for the filtering steps

GWAS: Name 32657 | 8.25/12 → 4.125/6

good job!

1. Formatting: 0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~ Ok

~~informative title~~

name on all pages

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background: 0.5/0.5

brief statement of scientific question

all variables defined

3. PCA: 1.25/1.5

explain relation between PCs and population stratification (B)

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: 0.75/2

SNP QC: criteria and reasons

sample QC: criteria and reasons (C)

Hardy-Weinberg equilibrium: what it means and how it relates to quality (D)

Overall QC explanation

5. Association / post-association analysis: 1.5/2

Describe association analysis in words and mathematically (G) (H)

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E)

LD heatmap (optional – does NOT count); measure of LD (F)

4.5

1.25/2 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

0.5/1 7. Plots: (O)

label size (not too small) captions

placement **NOT BLURRY** (L)

0.75/1.5 8. Conclusions

recap analysis (expand) state main findings (Start new paragraph) (J)

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

poor satisfactory good excellent

0.75/1 10. Other comments:

- see attached

(*) please email me the code you used to generate each plot: as well as the output from

sessionInfo()

3.75

Comments **Name:** 326571

00 – informative title

A – eda – don't need file names

B – PCA + explain Good!! Plot PC2 against PC1 – what are the red dots?

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* – sex numerically = ?

use ^ onset. cofs,

I – Manhattan plot (and explanation) → $-\log_{10} p$

J – identify significant markers

K – square QQ plots ← show in table

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions ⇒ 1-2 paragraphs recap then 1-2 paragraphs conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- Define ALL terms (MAF, HWE, LD, etc) + measures + explain
- use 12 pt size
- clearly explain reasons for filtering steps
- needs references

GWAS: Name 32771 6.75/12 → 3.375/6

1. Formatting: 0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

max ~~7~~ pages ok

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.5/0.5 2. Introduction/Background:

brief statement of scientific question

all variables defined

0.5/1.5 3. PCA:

explain relation between PCs and population stratification (B)

plot pc2 (y-axis) vs pc1 (x-axis) + square

0.75/2 4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons (C) (D)

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

1.25/2 5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)

LD heatmap (optional – does NOT count); measure of LD (F)

3.5

1.25/1.5

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP (H)

0.75/1

7. Plots: (O)

label size (not too small) captions

placement **NOT BLURRY** (L)

0.5/1.5

8. Conclusions

recap analysis state main findings (J)

0.25/0.5

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

poor satisfactory good excellent

0.5/1

10. Other comments:

see attached

(*) please email to me the code you used to generate each plot as well as the output of sessionInfo()

3.25

Comments **Name:** 32777

00 – informative title

A – eda – don't need file names

B – PCA + explain → square

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically* (y = trans. HDL) + PC - {10} + sex = ?

I – Manhattan plot (and explanation)

J – identify significant markers (ok) - log(10) p-value

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) ok

Q – plot layout (see text) ok

R – overall organization and explanation of procedure

S – other:

- You need to define all terms: MAF, IBD, etc + measures + explain
- use pt size 12
- you don't need to state the functions you use, you should explain the methods

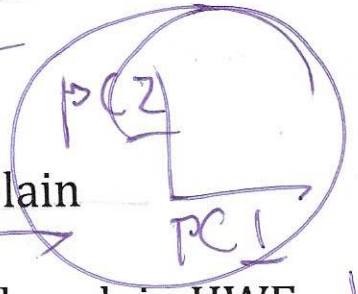
- not correct to say 'using linear reg. we get 10 pc's'
  - ⇒ you use 10 pc's in the reg
  - ⇒ explain Why
- clearly explain reasons for filtering steps
- needs references

Comments **Name:** 330603

00 – informative title

A – eda

B – PCA + explain



+ Square

C – define and explain HWE mathematically

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD mathematically

G – explain association test

H – write out final model *mathematically* – Define all terms
– what is 'sex' numerically

I – Manhattan plot (and explanation) – log(10) P

J – identify significant markers – show in table

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) OK

Q – plot layout (see text) OK

R – overall organization and explanation of procedure

S – other:

- Use 12 pt size
- Be more precise about heterozygosity explanation, yours is vague
- Define all terms (IBD, HWE, etc) and measures (mathematically) & explain
- needs references

GWAS: Name 330603 6.25/12 → 3.125/6

1. Formatting:

0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7~~ pages ok

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.5/0.5

2. Introduction/Background:

brief statement of scientific question

all variables defined

0.5/1.5

3. PCA: (B)

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis) + square

0.75/2

4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons (C) (D)

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

1.25/2

5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)

LD heatmap (optional – does NOT count); measure of LD (F)

3.5

1.25/1.5

6. Write out final estimated model **mathematically** (for a given SNP) (M)

hat on response variable MUST RELATE TO SNP

define all terms

0.5/1

7. Plots: (G)

label size (not too small) captions

placement **NOT BLURRY** (L)

0.75/1.5

8. Conclusions

recap analysis expand new paragraph

state main findings (J)

0.25/0.5

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

poor satisfactory good excellent

10. Other comments:

- see attached

(*) please email to me the code you used to generate each plot as well as the output from sessionInfo()

2.75

GWAS: Name 359952 5/12 → 2.5/6

1. Formatting:

0.5/0.5

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- ~~max 7 pages~~ ok
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

2. Introduction/Background:

0.5/0.5

- brief statement of scientific question
- all variables defined

0.75/1.5

3. PCA: (B)

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality (C) (D)
- Overall QC explanation

0.75/2

5. Association / post-association analysis:

- Describe association analysis in words and mathematically (G) (+1) incorrect model
- Manhattan plot
- lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)
- LD heatmap (optional – does NOT count); measure of LD (F)

3.25

0.25/1.5 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

7. Plots:

0.25/1 label size (not too small) (O) +(P) captions

placement **NOT BLURRY** (L)

0.5/1.5 8. Conclusions

recap analysis — say more state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory good excellent

0.5/ 10. Other comments:

1 - see attached

(*) please email to me the code you used to generate each plot as well as the output from

sessionInfo()

(**) I have some doubts about the originality of your report, there is some evidence of plagiarism (plagiat).

1.75 Please make an appointment to meet with me.

Comments **Name:** 359952

00 – informative title

A – eda – Don't need file names

B – PCA + explain * PC2 – PC1 * what is PCA + how related to ancestry + mathematically

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure – mathematically + LD

G – explain association test

H – write out final model *mathematically* – model incorrect – see help for R fn for model-fitting

I – Manhattan plot (and explanation) **

J – identify significant markers

K – square QQ plots → show in table → $-\log_{10} p$

L – fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions (ok)

N – **no raw R**

** Fig 2 below explanation of Manhattan plot

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text) ok

Fig 4 too short and appears to be copied

R – overall organization and explanation of procedure

S – other:

- Use 12 pt size
- Carefully define all terms (IBD, LD, etc) + measures (mathematically) + explain
- model incorrect - you need to correct for age, sex, 10 PCs, and include an intercept plus a coefficient for the SNP
- explain the test of association — where do the p-values come from?
- Q-Q plots not labeled
- You regress trait on SNP
- general linear model

GWAS: Name 360536 4.75/12 → 2.375/6

1. Formatting:

0.5/0.5

all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~ (ok)

informative title

name on all pages

all pages numbered

no blurry plots (~~**NOT png**~~)

2. Introduction/Background:

0.5/0.5

brief statement of scientific question

all variables defined

0.5/1.5

3. PCA:

explain relation between PCs and population stratification (B)

plot pc2 (y-axis) vs pc1 (x-axis) + square

4. Pre-processing / QC steps:

0.75/2

SNP QC: criteria and reasons

sample QC: criteria and reasons (C) (D)

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

0,75

5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically incorrect model

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (K)

LD heatmap (optional – does NOT count); measure of LD (F)

0.25/1.5 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

0.25/1 7. Plots: (O)

label size (not too small) captions

placement +(P) **NOT BLURRY** (L)

0.5/1.5 8. Conclusions

recap analysis (vague) state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory good excellent

0.5/1 10. Other comments:

- see attached

(*) please email to me the code you used to generate each plot as well as the output from sessionInfo()

(**) I have some doubt about the originality of your report, there is some evidence of plagiarism (plagiat). Please make an appointment to meet with me.

1.75

Comments **Name:** 360536

00 – informative title

A – eda — don't need file names

B – PCA + explain — PC2 ⊥ PC1 + square

C – define and explain HWE — mathematically

D – define and explain HWE **test**

E – define λ

F – define LD measure — mathematically

G – explain association test

H – write out final model *mathematically* — NOT R formula → write each variable explicitly NOT hidden in matrix

I – Manhattan plot (and explanation) — $-\log_{10} p$

J – identify significant markers — show in table

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions (ok)

N – **no raw R**

O – plot labels too small ok

P – plot size (see text) Manhattan too short

Q – plot layout (see text) – Manhattan should come after assoc analysis

R – overall organization and explanation of procedure → QQ plots should come after manhattan

S – other:

- Use 12pt size
- its LOW hetzyg that indicates possible inbreeding – high het may indicate possible low sample quality
- Define all terms (LBD, HWE etc) and measures (mathematically + explain)
- What do the lines on the Manhattan plot represent?
- Makes more sense to have Q-Q's:  unadj adj
- Formula bottom p. 6 not correct – it omits age, sex, 10 PCs; SNP is NOT binary – See the help for the R fn for the model fitting
- fitted model does not include error term

GWAS: Name 366674 3.5/12 → 1.875/6

1. Formatting:

0.5/0.5 all margins 2.5cm

12 pt size

no raw R code or output

~~max 7 pages~~

informative title

name on all pages

all pages numbered

no blurry plots (**NOT png**)

0.5/0.5 2. Introduction/Background:

brief statement of scientific question

all variables defined

0.75/1.5 3. PCA: (B)

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2 4. Pre-processing / QC steps:

SNP QC: criteria and reasons

sample QC: criteria and reasons (C) (D)

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

0.75/2 5. Association / post-association analysis: (G) (H)

Describe association analysis in words and mathematically

Manhattan plot (I)

lambda analysis (including **SQUARE** QQ-normal plots) (E) (J)

LD heatmap (optional – does NOT count); measure of LD (F)

3.25

0/1.5 6. Write out final estimated model **mathematically** (for a given SNP) (H)

hat on response variable MUST RELATE TO SNP

0.25/1 7. Plots:

label size (not too small) captions

placement (O) **NOT BLURRY** (L)

0/1.5 8. Conclusions

recap analysis state main findings (J)

0.25/0.5 9. Overall presentation (clarity of explanations, appropriate citations / references) : (R)

poor satisfactory incomplete good excellent

0.25/1 10. Other comments:

see attached

* please email to me the code you used to generate each plot as well as the output of

sessionInfo()

0.5

Comments **Name:** 366674

00 – informative title

A – eda

B – PCA + explain — PC2 — PC1

C – define and explain HWE (OK – but include assumptions)

D – define and explain HWE **test**

E – define λ

F – define LD measure + LD

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation) (ok — explain lines)

J – identify significant markers — put in table

K – square QQ plots

L – fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) – Manhattan plot too short

Q – plot layout (see text) (OK)

R – overall organization and explanation of procedure

S – other:

– please explain the methods

– incomplete

– write as a scientific report, NOT in question/answer form

– clearly explain reasons for filtering steps

– Define all terms (MAF, IBD, etc.) and measures and explain