The effects of an HIV partner on a patient's HIV severity

Fan Yang, Jia Liu, Josh Wiltsie, Monica Patrin, Rodrigo Lovaton, Seongkwon Lyu

Introduction

- Problem: A researcher wants to know if partner's HIV status has an effect on three blood measurements for patients with HIV?
- Data includes: patient's gender, partner status (HIV or not), and three blood measurements (CD4, CD8, and RNA).
- N = 278 patients with HIV (83 missing values for RNA, 2 missing for CD4 and CD8).

Responses

- **CD4** or white blood cell count - measures the severity of sickness. Higher is healthier.
- **CD8** cell count - measures the patient's ability to fight infections. Higher is healthier.
- **RNA** test - measures how much virus is in the patient's blood. Lower is healthier.

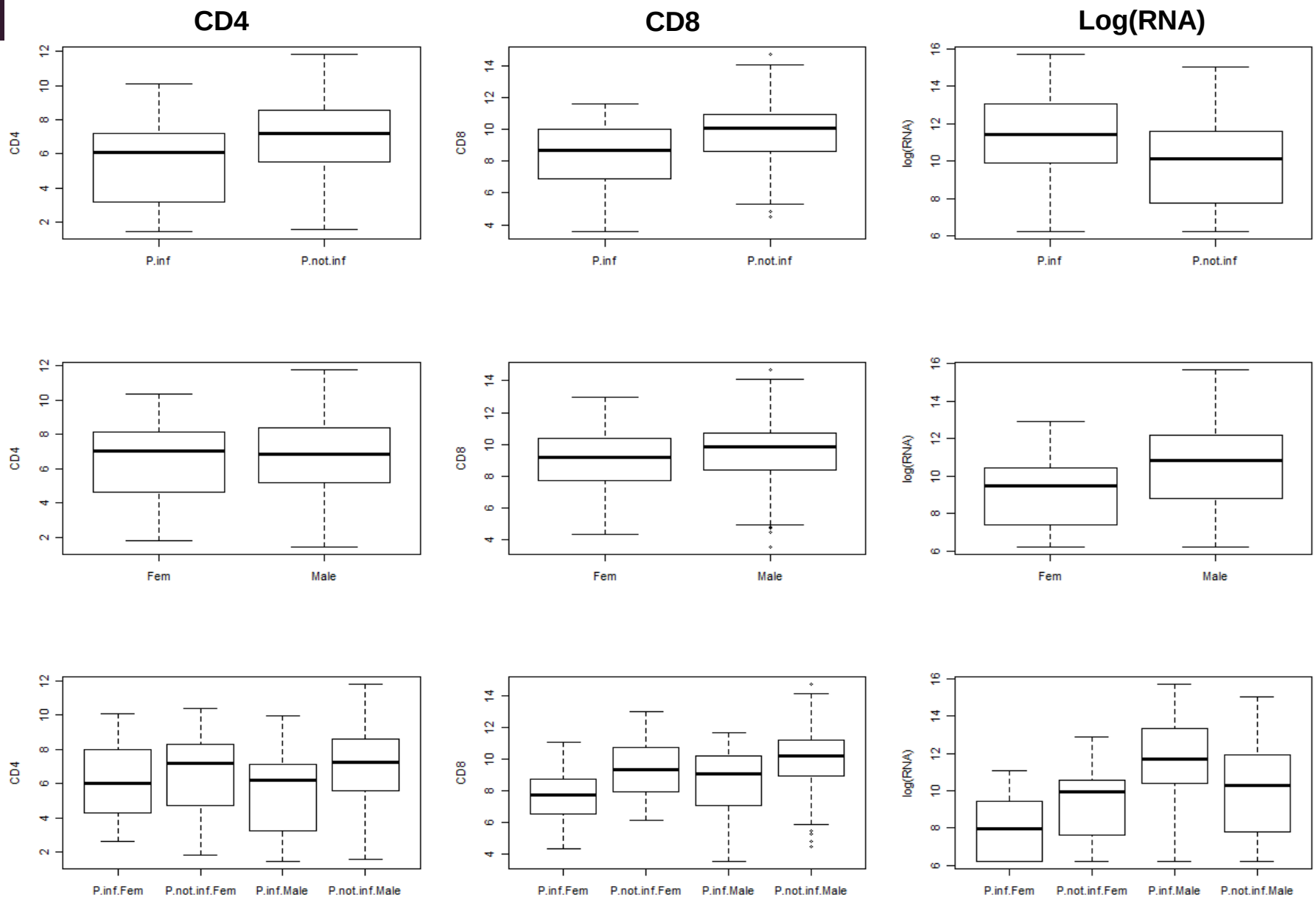
Data collection

- The data was collected over a one-month period at a single clinic.
- Participation was voluntary. A gift card was given to all participants.
- The sample is expected to be similar to the overall HIV infected population of interest for the clinic.

Background

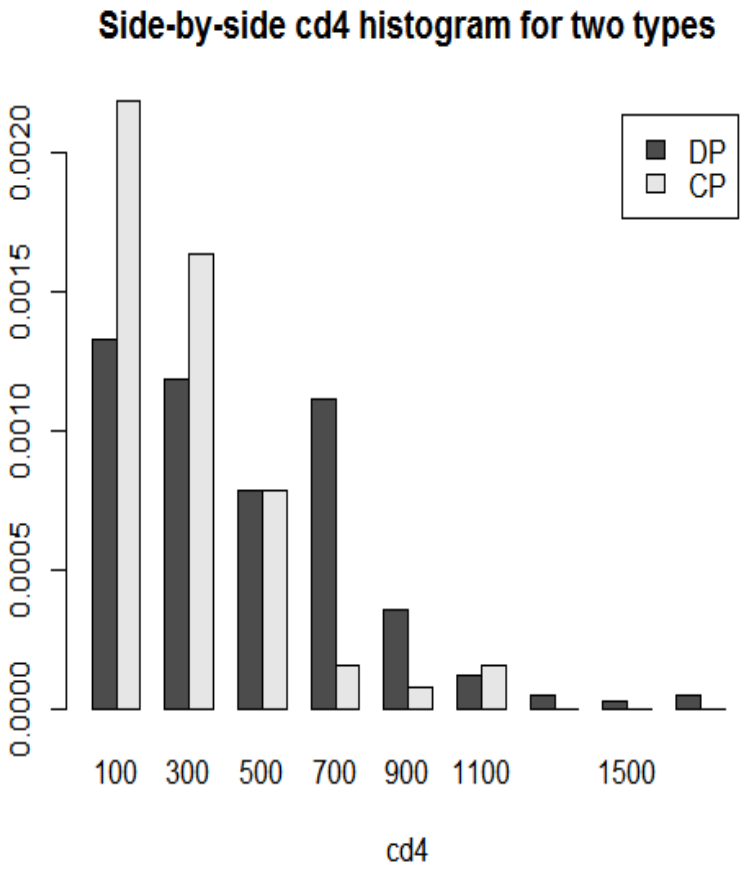
- Two other unmeasured factors could possibly affect the blood measurements:
  - Type of treatment being administered and compliance with treatment
  - Length of time since diagnosis
- Sampling bias is possible but the client thinks the sample represents the population well.
- Other demographics are not thought to affect responses (eg. ethnicity, income).

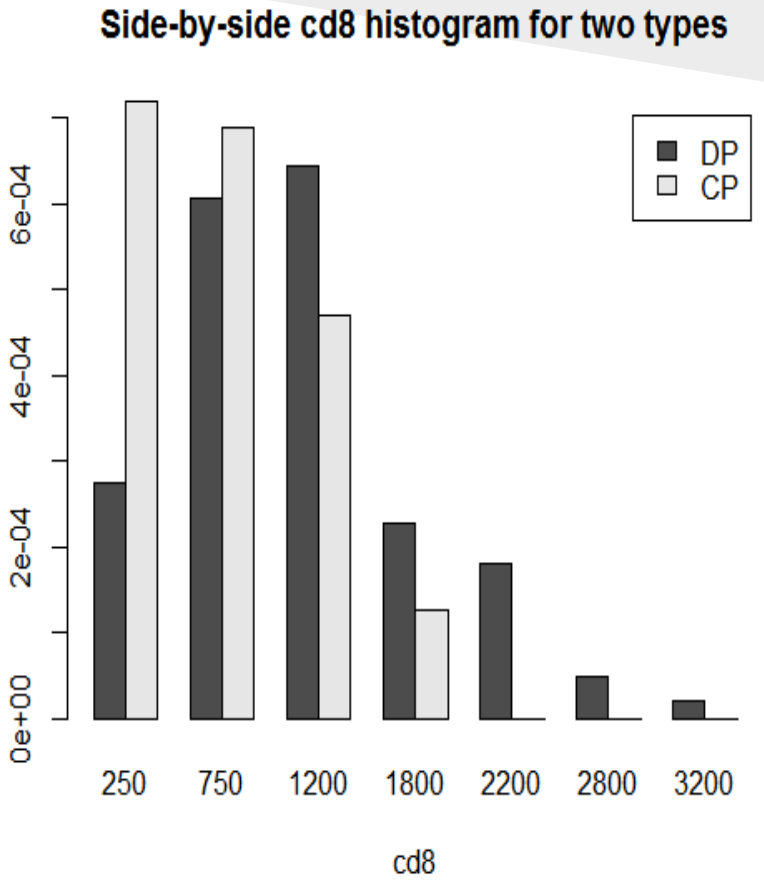
Box plots for CD4, CD8, and log(RNA) by partner's HIV status and gender



Part II: Data Exploration

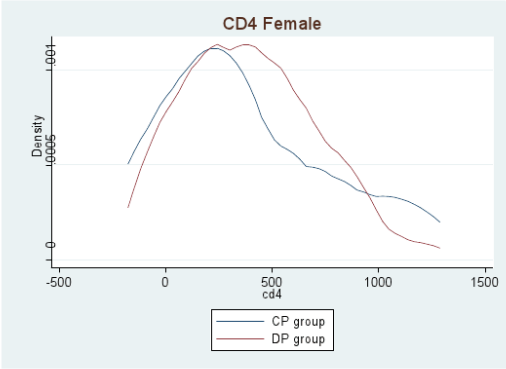
Side-by-side histogram

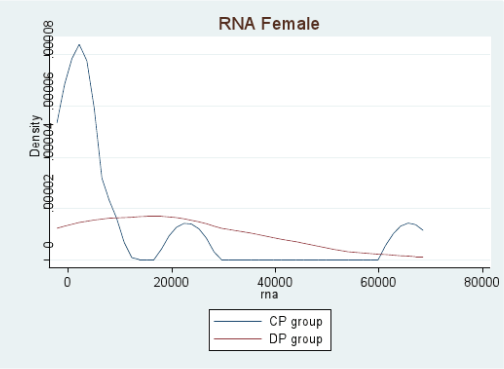


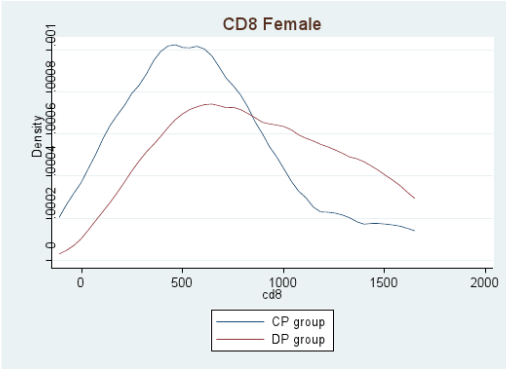


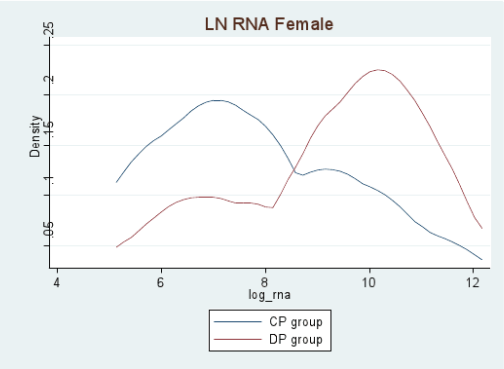
Data visualization

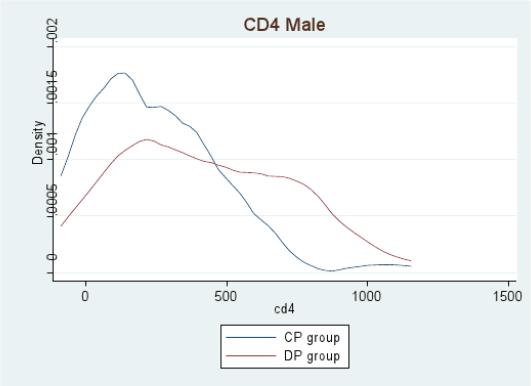
- Kernel density curves

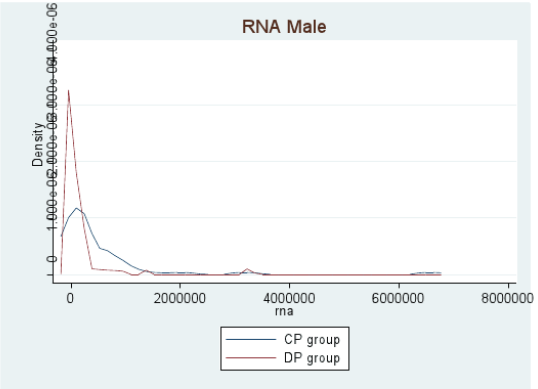


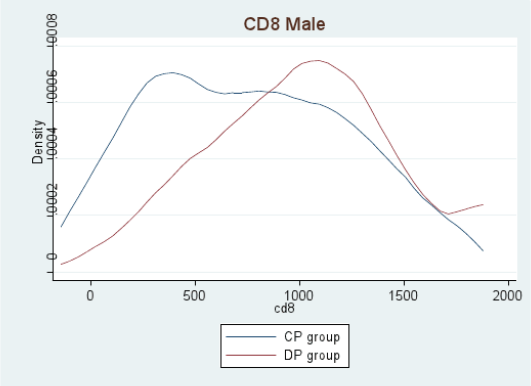


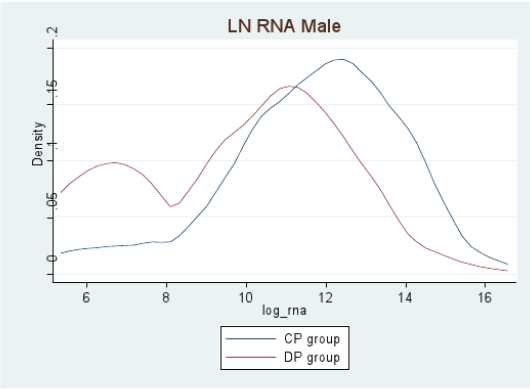






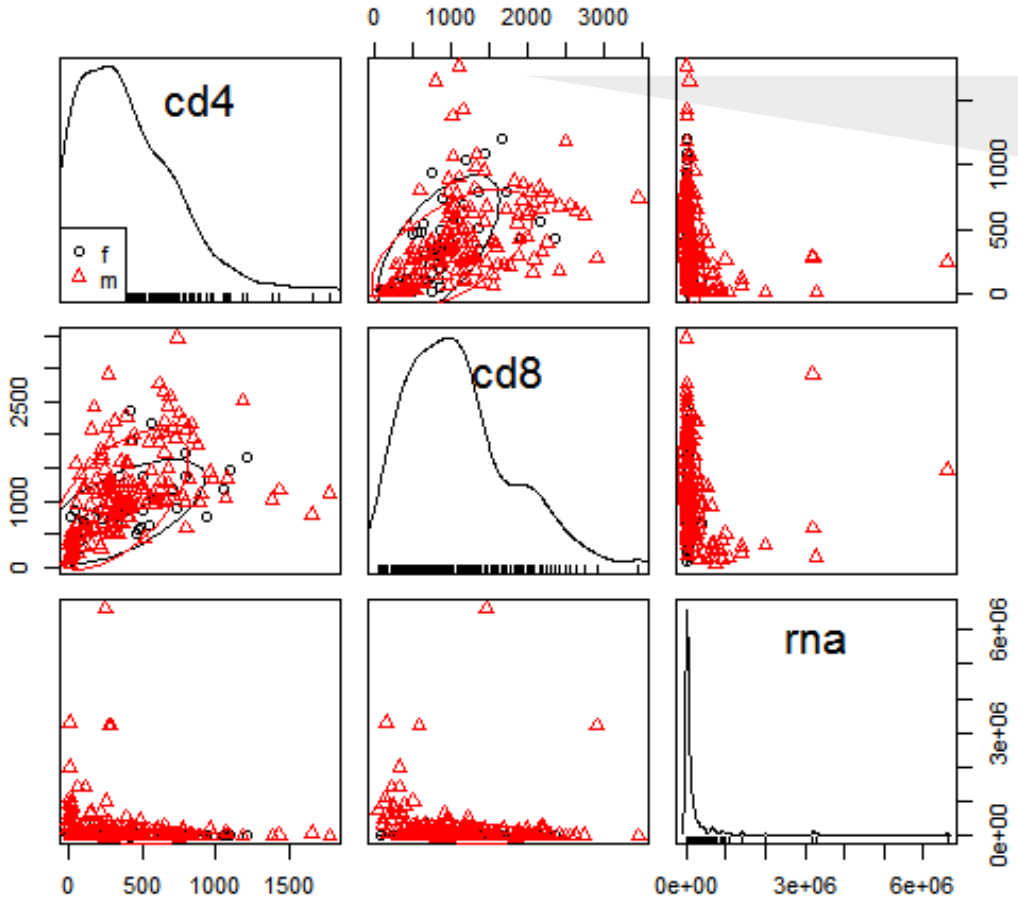






From the graphs we can see that sex and type seem to have effect on some responses.

Scatterplot matrix



RNA is very skewed,

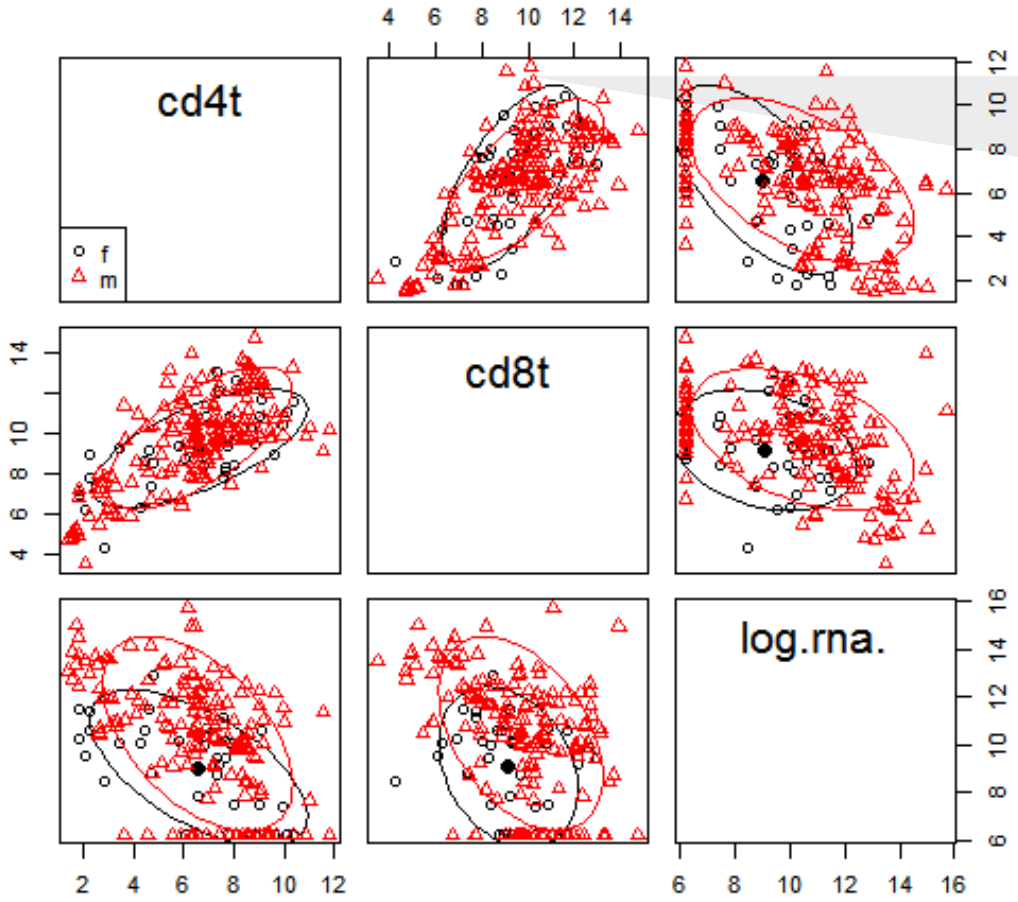
and the multivariate normality seems violated.

Transformation

- For later analysis, we want our responses to be multivariate normal. Using R we figured out to do the following

```
> cd4 <- cd4^0.33
> cd8 <- cd8^0.33
> rna <- log(rna)
```

Scatterplot Matrix



The multivariate normality assumption seem to hold now.

Part III: Data Analysis

MANOVA

```
> man1 <- manova(cbind(cd4.tr,cd8.tr,rna.tr)~type*sex,AIDS))
> summary(man1)
           Df   Pillai approx F num Df den Df    Pr(>F)
type        1 0.136803   9.8260      3    186 4.771e-06 ***
sex         1 0.077059   5.1765      3    186  0.001856 **
type:sex    1 0.053073   3.4749      3    186  0.017173 *
Residuals 188
```

- All variables are statistically significant.

Linear regression result

Outcome	$CD4^{0.33}$		$CD8^{0.33}$		log(RNA)	
	Estimate	Std. Er	Estimate	Std. Er	Estimate	Std. Er
Intercept	6.191 ***	[0.758]	7.718***	[0.639]	8.066 ***	[0.820]
partner without HIV	0.327	[0.821]	1.658*	[0.692]	1.225	[0.929]
Male	-0.784	[0.817]	0.795	[0.689]	3.489***	[0.881]
Interaction	1.185	[0.894]	-0.089	[0.754]	-2.929**	[1.009]
R2	0.063		0.12		0.14	
Observations	271		271		191	

More problems ...

- Missing values

```
> dim(AIDS)
[1] 278   6
> (MI.AIDS<-mi.info(AIDS))
  names include order number.mis all.mis                type collinear
1    id      No    NA          0      No positive-continuous        No
2   sex     Yes    NA          0      No              binary        No
3  type     Yes    NA          0      No              binary        No
4   cd4     Yes     1          3      No positive-continuous        No
5   cd8     Yes     2          3      No positive-continuous        No
6   rna     Yes     3         83      No positive-continuous        No
```

Missing data imputation

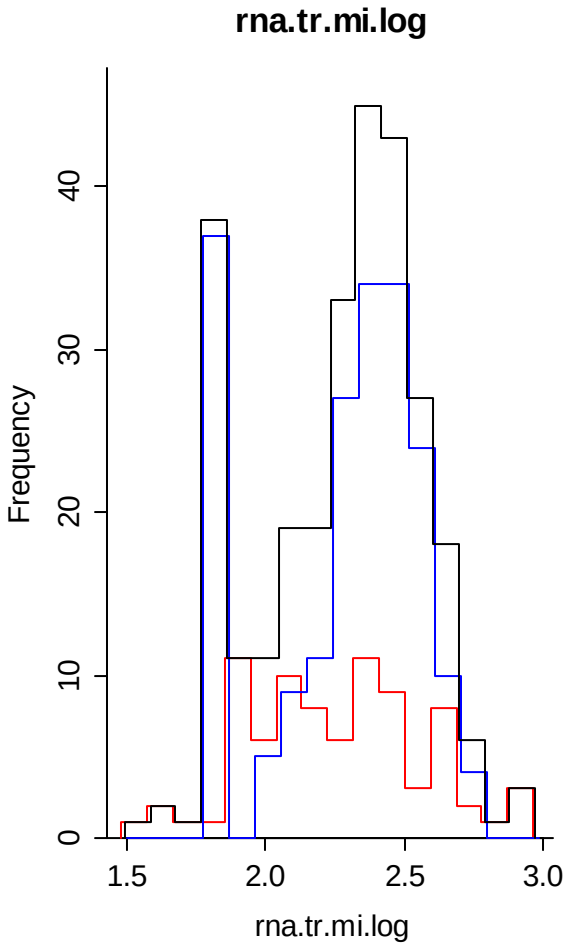
```
> missings <- which(is.na(AIDS$rna)=="TRUE")
> predictRNA <- predict(lm(rna.tr~cd4.tr+cd8.tr,data=AIDS,
      na.action=na.omit),AIDS[missings,c("cd4.tr","cd8.tr")])
>
> AIDS$rna.pred <- AIDS$rna.tr
> AIDS$rna.pred[missings] <- predictRNA
>
> summary(m6 <- lm(cbind(cd4.tr,cd8.tr,rna.pred)~type*sex,
AIDS))
```

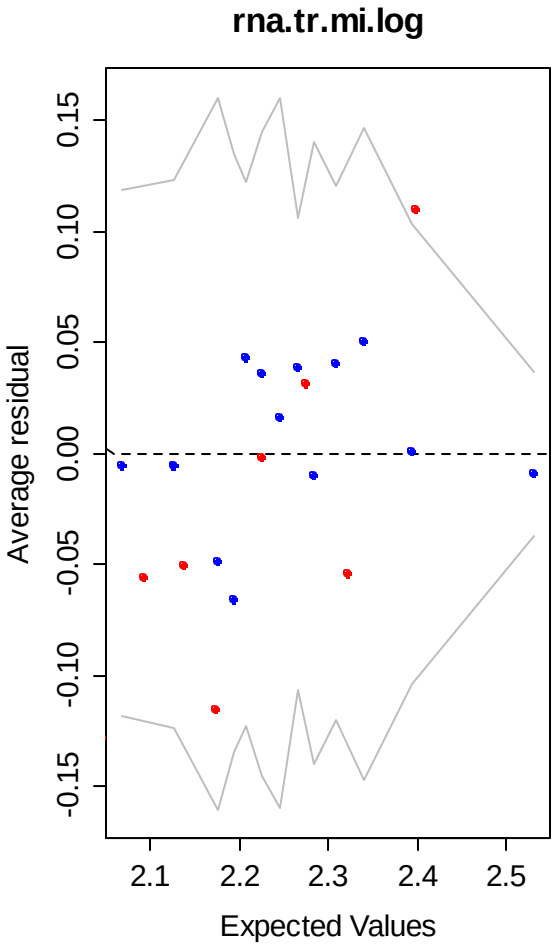
Response rna.pred:

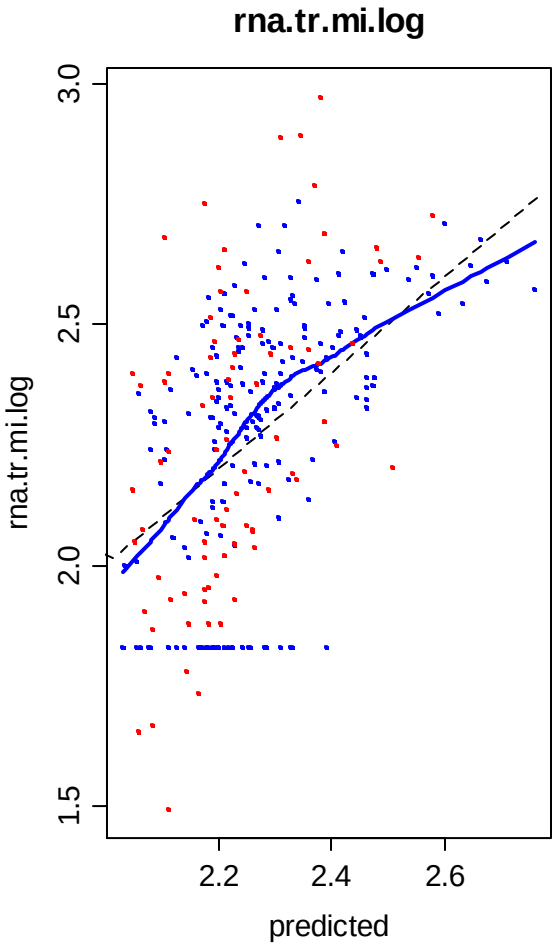
```
Call: lm(formula = rna.pred ~ type * sex, data = AIDS)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.5209     0.6833  12.471  < 2e-16 ***
typeDP         1.1009     0.7400   1.488 0.138004
sexm           2.9329     0.7370   3.979 8.88e-05 ***
typeDP:sexm   -2.7149     0.8065  -3.366 0.000873 ***

Residual standard error: 2.05 on 271 degrees of freedom
Multiple R-squared: 0.1135, Adjusted R-squared: 0.1037
F-statistic: 11.56 on 3 and 271 DF,  p-value: 3.718e-07
```

“MI” package

```
> AIDS.mi <-AIDS[,c("sex","type","cd4.tr","cd8.tr","rna.tr")]
> info = mi.info(AIDS.mi)

> AIDS.new = mi.preprocess(AIDS.mi)
> attr(AIDS.new,"mi.info")
> info.upd = update(info,"imp.formula",list("cd4.tr" =
"cd4.tr ~ sex * type + cd8.tr + rna.tr"))
> info.upd = update(info.upd,"imp.formula",list("cd8.tr" =
"cd8.tr ~ sex * type + cd4.tr + rna.tr"))
> info = update(info.upd,"imp.formula",list("rna.tr" =
"rna.tr ~ sex * type + cd4.tr + cd8.tr"))

> AIDS.new = mi.preprocess(AIDS.mi)
> IMP <- mi(AIDS.new, n.iter=50)
```


rna.tr.mi.log
Frequency
0
10
20
30
40
1.5
2.0
2.5
3.0
rna.tr.mi.log


rna.tr.mi.log
Average residual
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
2.1
2.2
2.3
2.4
2.5
Expected Values

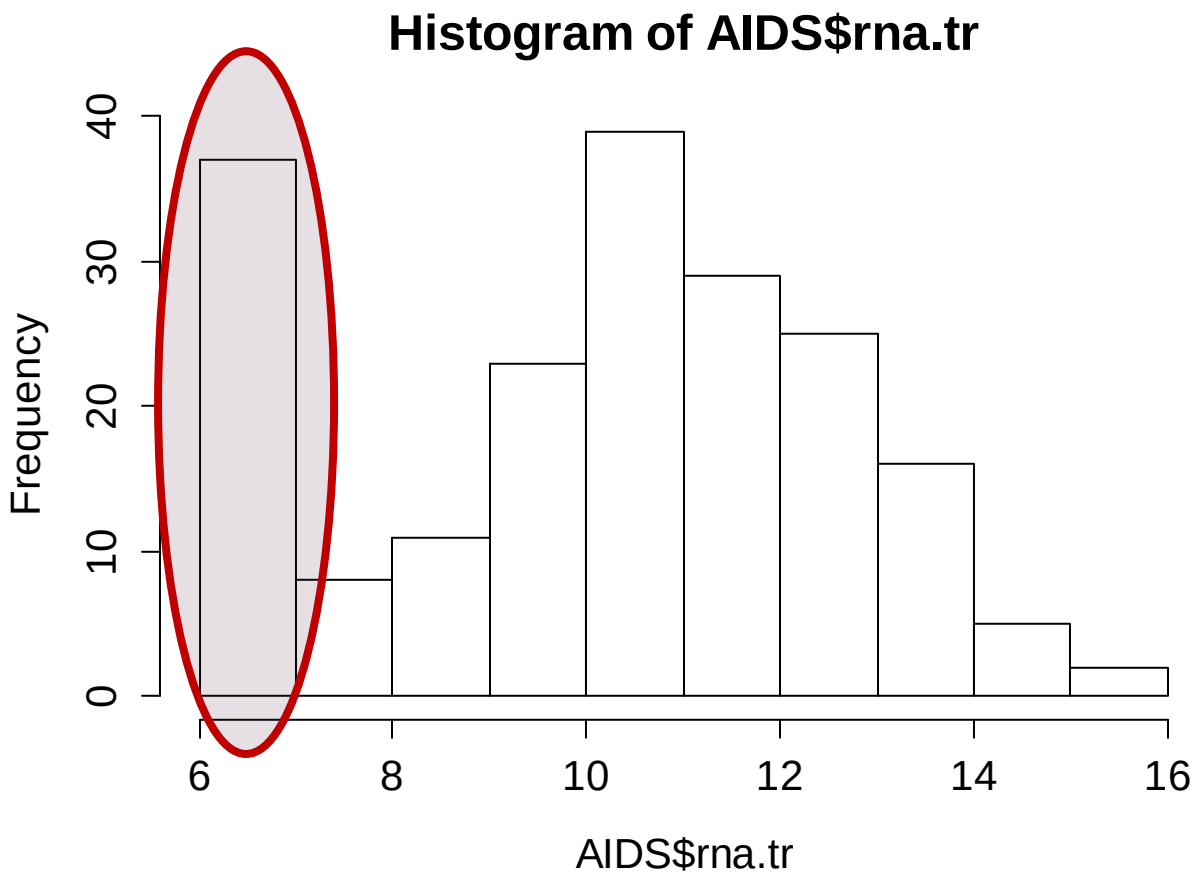

rna.tr.mi.log
rna.tr.mi.log
1.5
2.0
2.5
3.0
2.2
2.4
2.6
predicted

Linear regression result

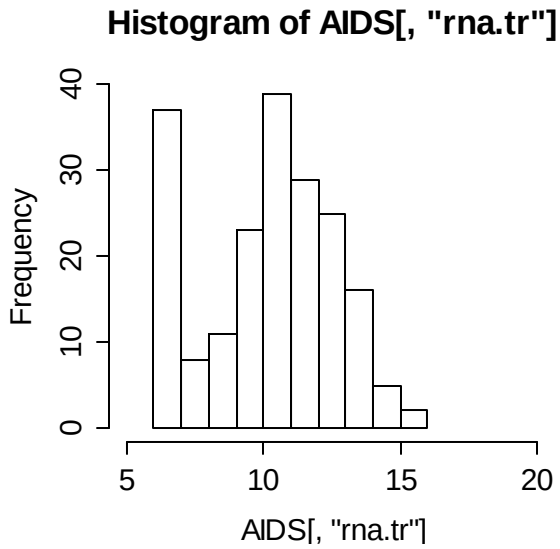
Outcome	Log(RNA)		Log(RNA.pred)		log(RNA.mi)	
	Estimate	Std. Er	Estimate	Std. Er	Estimate	Std. Er
Intercept	8.066 ***	[0.820]	8.521***	[0.683]	8.066 ***	[0.824]
partner without HIV	1.225	[0.929]	1.101*	[0.740]	1.225	[0.935]
Male	3.489***	[0.881]	2.933	[0.737]	3.417***	[0.889]
Interaction	-2.929**	[1.009]	-2.7147	[0.807]	-2.856**	[1.018]
R2	0.14		0.10		0.12	
Observations	191		271		274	

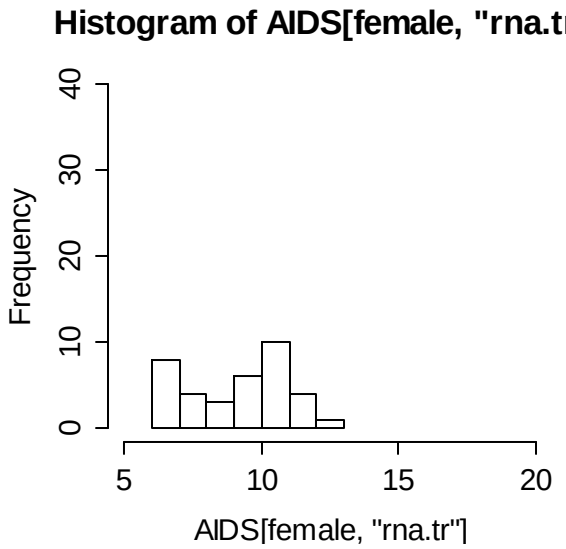
More problems ...

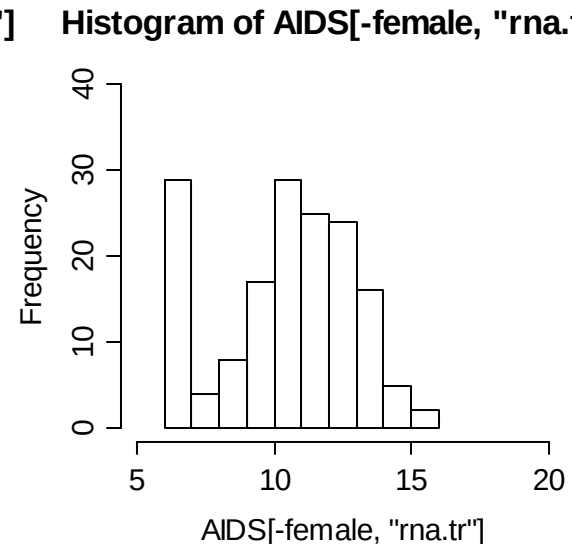
- Lowest value inflation
  -> against nomality assumption

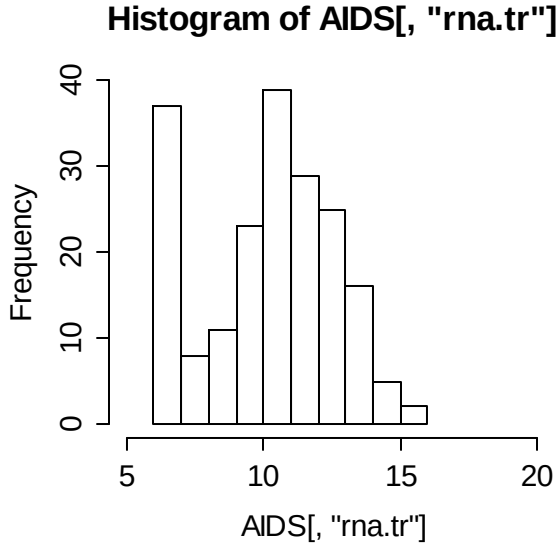


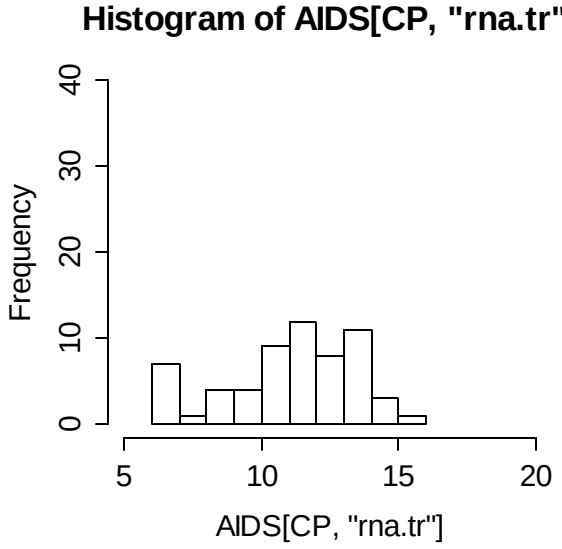
However,

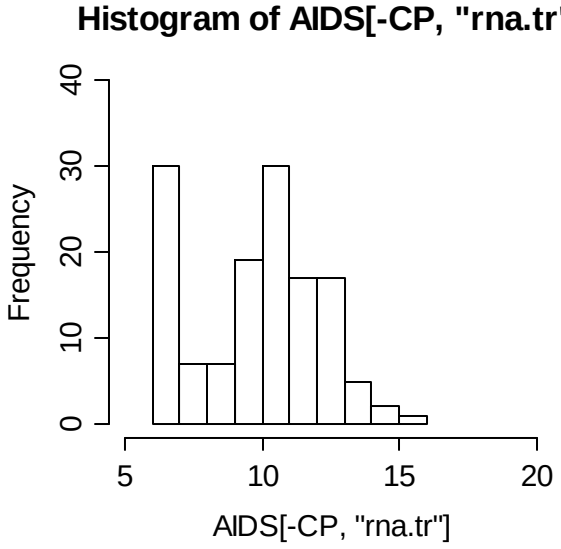

Histogram of AIDS[, "rna.tr"]
Frequency
0
10
20
30
40
5
10
15
20
AIDS[, "rna.tr"]


Histogram of AIDS[female, "rna.tr"]
Frequency
0
10
20
30
40
5
10
15
20
AIDS[female, "rna.tr"]


Histogram of AIDS[-female, "rna.tr"]
Frequency
0
10
20
30
40
5
10
15
20
AIDS[-female, "rna.tr"]


Histogram of AIDS[, "rna.tr"]
Frequency
0
10
20
30
40
5
10
15
20
AIDS[, "rna.tr"]


Histogram of AIDS[CP, "rna.tr"]
Frequency
0
10
20
30
40
5
10
15
20
AIDS[CP, "rna.tr"]

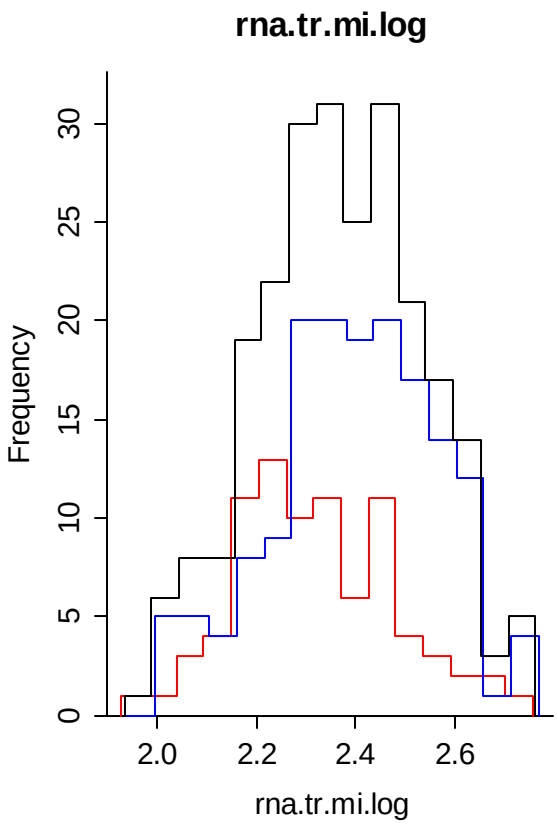

Histogram of AIDS[-CP, "rna.tr"]
Frequency
0
10
20
30
40
5
10
15
20
AIDS[-CP, "rna.tr"]

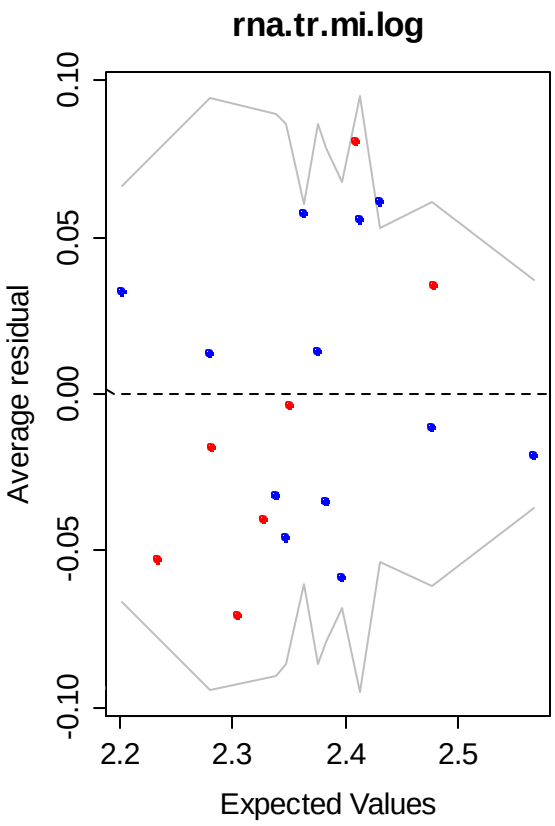
However,

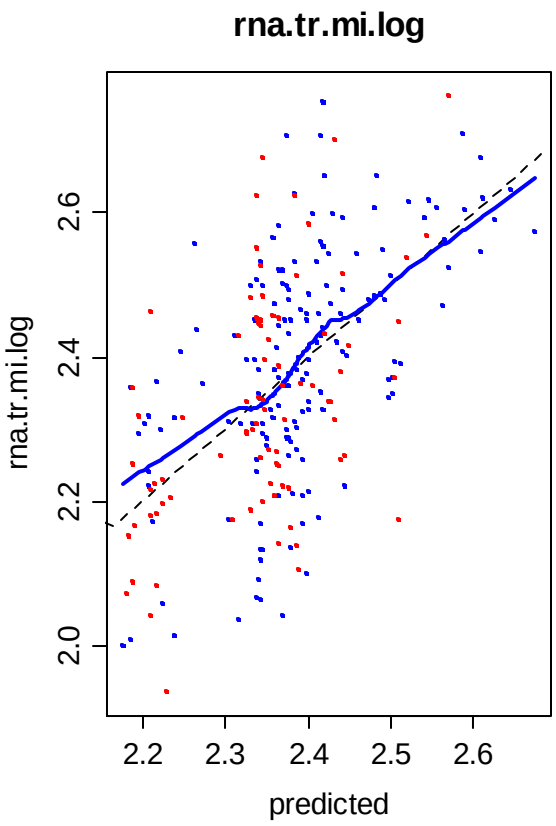
- The lowest values seem to be distributed randomly over each IV.
- The hypothesis, that their proportions are same, could not be rejected.
- It seems fair that we use data without those lowest points.

Comparison of results

Outcome	Log(RNA)		Log(RNA.pred)		log(RNA.mi)	
	Estimate	Std. Er	Estimate	Std. Er	Estimate	Std. Er
Intercept	**9.177*** **	[0.718]	**9.686*** **	[0.566]	**9.014*** **	[0.636]
partner without HIV	0.613	[0.799]	0.744	[0.601]	0.540	[0.675]
Male	**2.793*** **	[0.764]	**2.225*** **	[0.599]	**3.038*** **	[0.670]
Interaction	**-1.608 .**	[0.862]	**-1.783** **	[0.643]	**-1.697* **	[0.721]
R2	0.14		0.10		0.12	
Observations	191		271		274	


rna.tr.mi.log
Frequency
0
5
10
15
20
25
30
2.0
2.2
2.4
2.6
rna.tr.mi.log


rna.tr.mi.log
Average residual
0.10
0.05
0.00
-0.05
-0.10
2.2
2.3
2.4
2.5
Expected Values


rna.tr.mi.log
rna.tr.mi.log
2.0
2.2
2.4
2.6
2.2
2.3
2.4
2.5
2.6
predicted

Part IV: Conclusion

Conclusion

- Goal : to check whether an AIDS carrier partner influences a patient's progression of the HIV virus.
- IVs : partner’s gender and status (HIV or not)
- DVs : three different blood measurements (CD4, CD8, and RNA)
- We applied MANOVA, multiple regression and missing data imputation.

- A partner without HIV has a positive effect on CD8.
- The partner status is not statistically significant for CD4 or RNA.
- Males that have a partner without HIV have lower values for RNA (i.e. healthier response) with respect to males that have a partner with HIV.