Stat 412/512

RESIDUALS & PARTIAL RESIDUALS

Feb 4th 2015

Charlotte Wickham

stat512.cwick.co.nz

Case 11.01 Alcohol Metabolism

Not many alcoholics, and an extra sum of square F-test with reduced model:

μ{ *First pass metabolism* | *gast*, female, *alcoholic*} =

gast + female + gast x female

has a p-value of 0.93.

So, use this reduced model, i.e. no effect of alcoholism.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.06952    0.80195   0.087 0.931580
Gastric             1.56543    0.40739   3.843 0.000704 ***
SexFemale          -0.26679    0.99324  -0.269 0.790352
Gastric:SexFemale  -0.72849    0.53937  -1.351 0.188455
```

#1

#2

No evidence of different intercepts, if different slopes are in the model. #1

No evidence of different slopes, if different intercepts are in the model. #2

Here:

a multiplicative rather than additive difference makes more sense,

both having the same zero intercept makes sense,

μ\{ *First pass metabolism* | *gast*, female, *alcoholic*\} =

$\beta_1$gast + $\beta_2$gast x female

F-test to complicated

slope Men

	Estimate	Std. Error	t value	Pr(>|t|)	
Gastric	1.5989	0.1249	12.800	3.20e-13	***
Gastric:SexFemale	-0.8732	0.1740	-5.019	2.63e-05	***

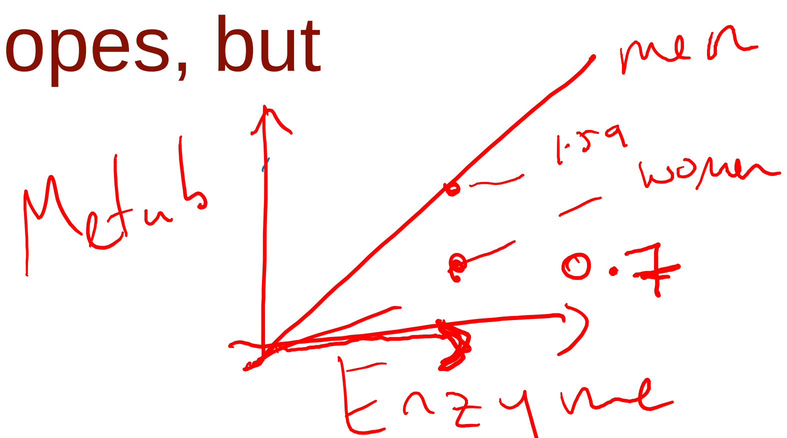
slope Women = 0.7

"Males had a higher first-pass metabolism than females even after accounting for differences in gastric AD activity (two-sided p-value = 0.0003 for a t-test for equality of male and female slopes when both intercepts are zero.)

For a given level of gastric AD activity the mean first-pass metabolism for men is estimated to be 2.20 times as large as the mean first-pass alcohol metabolism for women."

a different way to interpret a difference in slopes, but only if there are **no intercepts**.

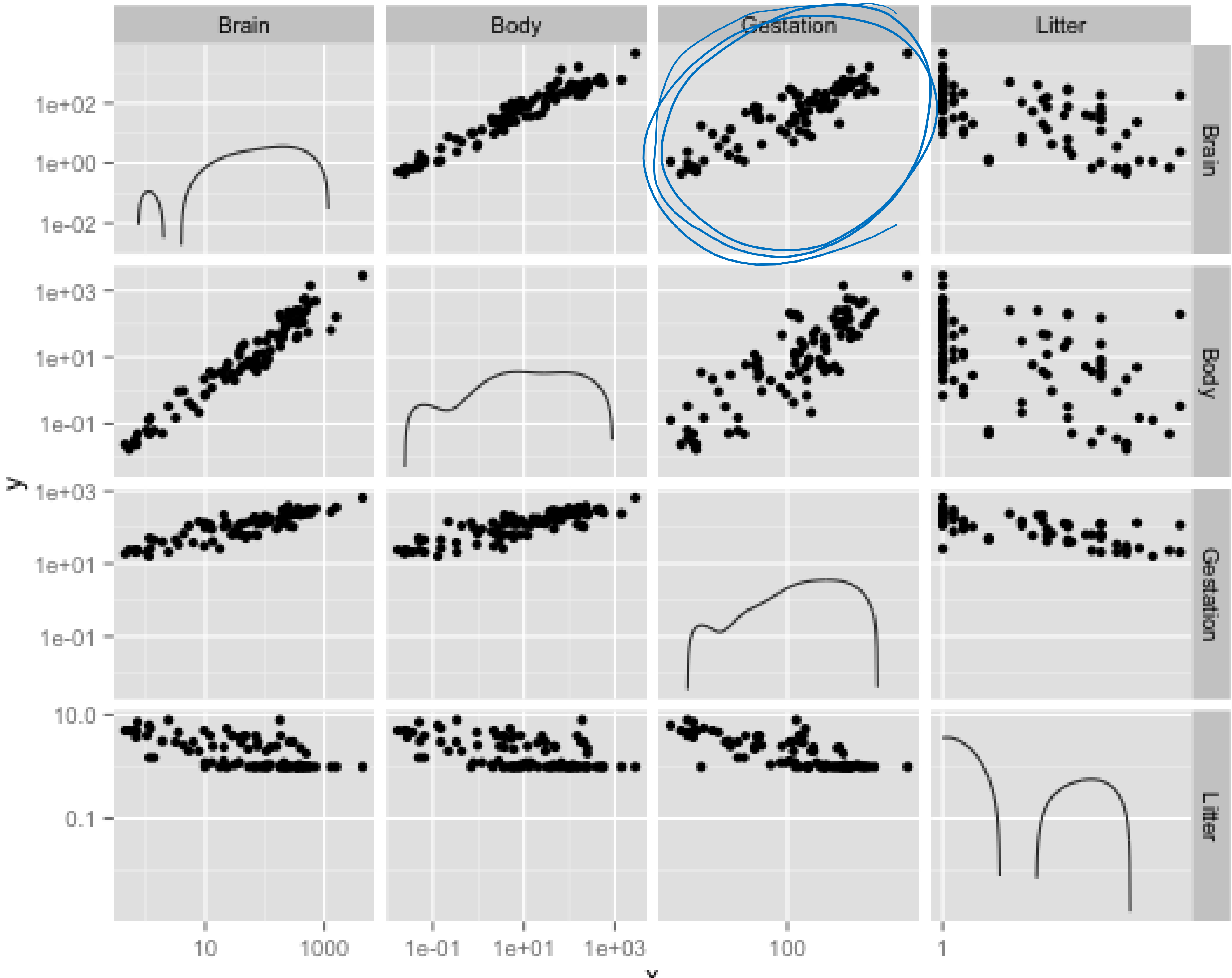
$$\frac{1.59}{0.7}$$



Strategy

1. Fit a tentative model — err on the side more complicated
2. Check residual plots for problems and outliers
3. Investigate influential points
4. Find a simple good fitting model
5. Answer questions of interest

Case Study 9.2 Mammalian Brain Size



How does gestation length relate to brain size after accounting for body weight?

Overview of regression

A model for the mean:

$$\mu\{Y \mid X_1, \ldots X_p\} = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 \ldots + \beta_p X_p$$

+ assumptions:

There is a Normally distributed subpopulation at each combination of explanatory variables values.

The means of the subpopulations fall on the line/surface defined above ($\mu\{Y \mid X_1, \ldots X_p\}$) Linearity

The subpopulation standard deviations are all equal to σ

The selection of an observation from one subpopulation is independent of the selection of any other observation.

The deviation of an observation from the mean, is independent of the deviation from the mean for any other observation.

equivalent

Residual plots to check

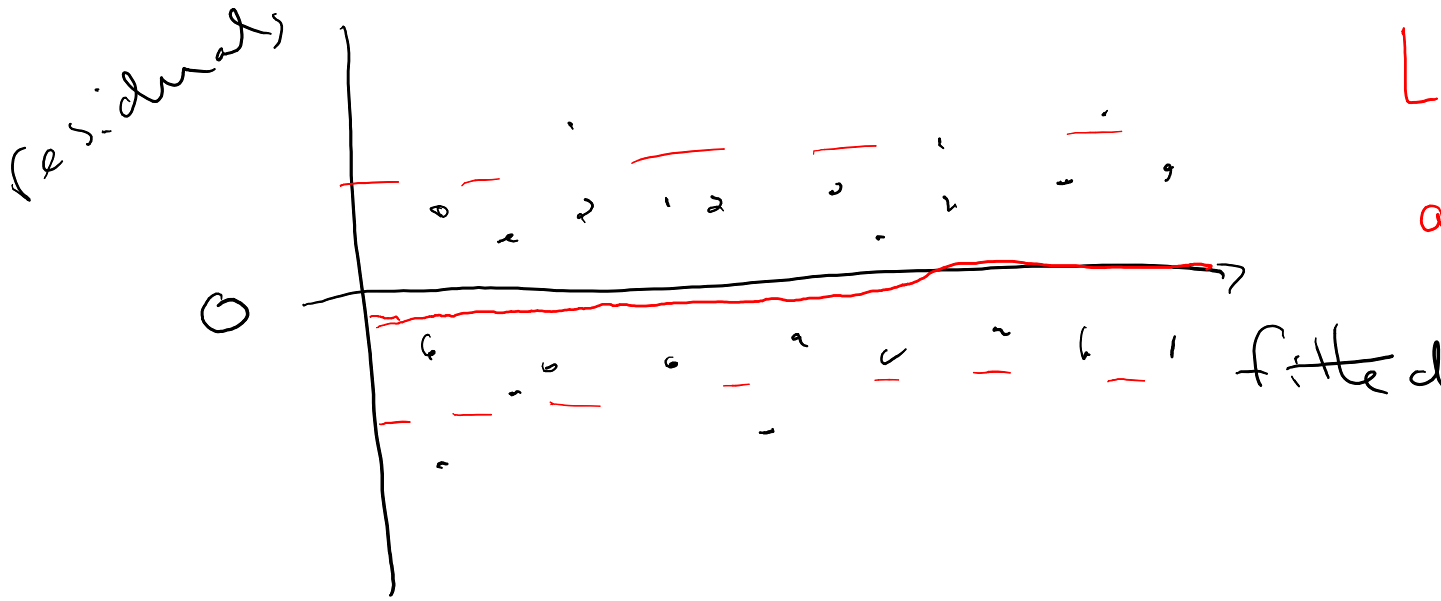
Residuals against fitted values

Residuals against explanatory variables

↳ all of them

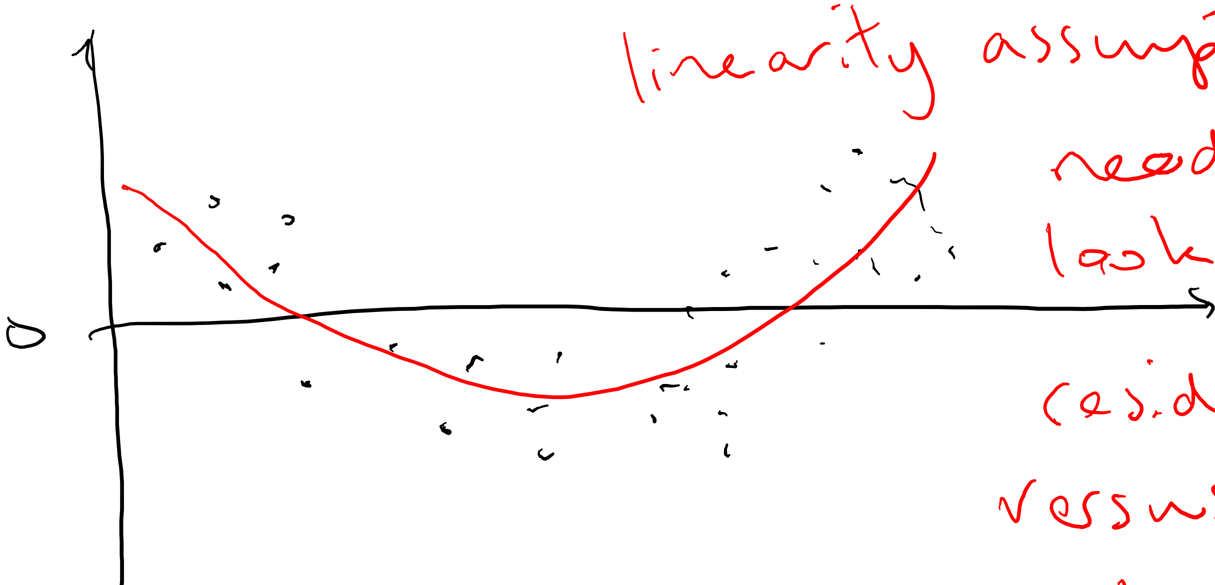
Normal probability plot of residuals

Residuals against fitted values



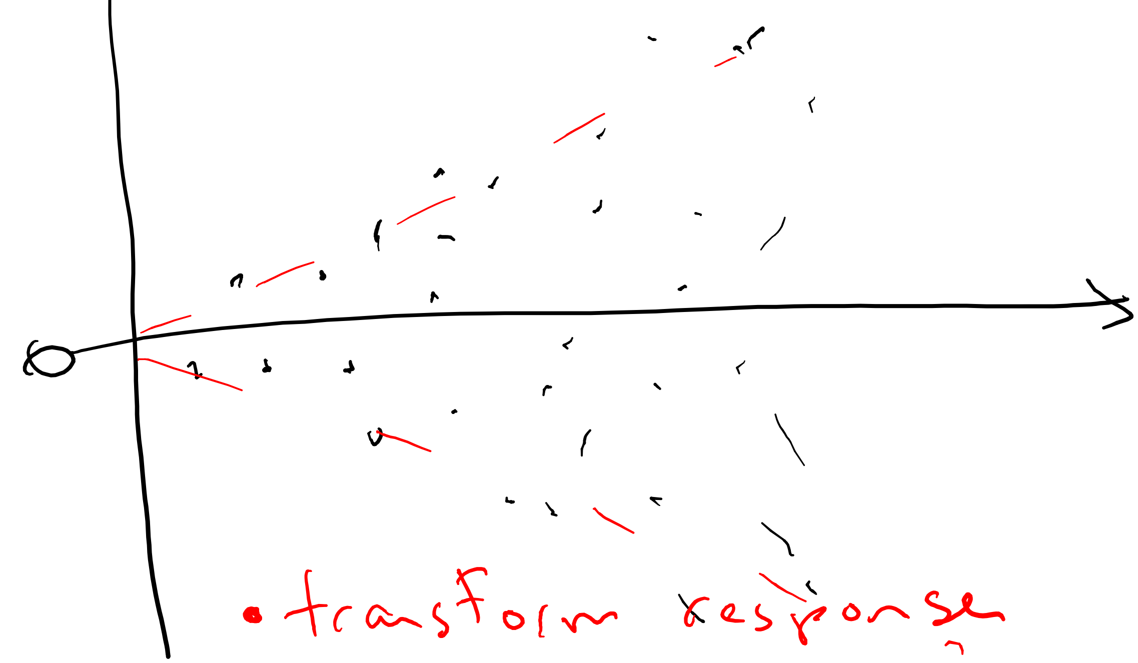
Looking for:
an even width band centered around zero

evidence against linearity assumption



need to look at residuals versus explanatories

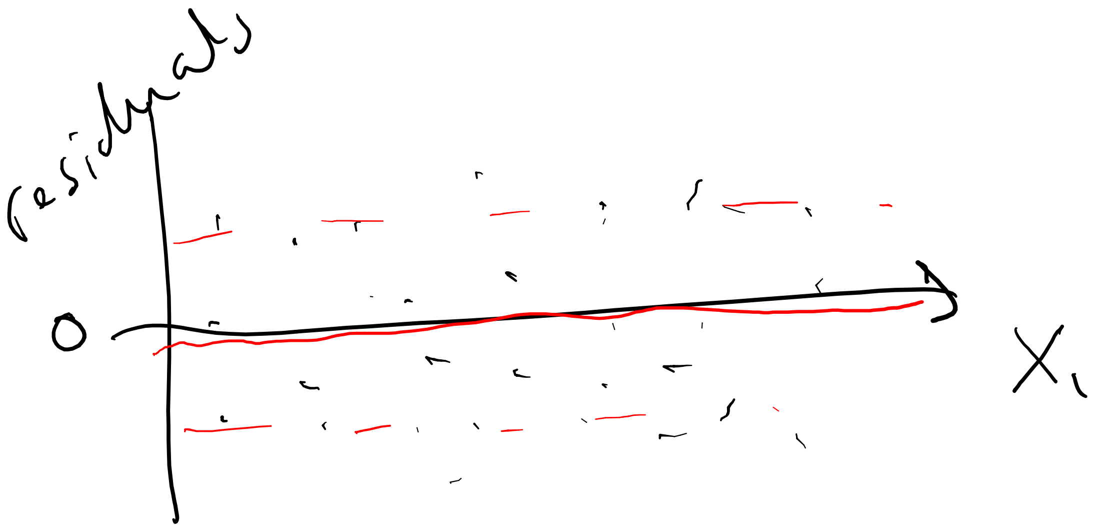
evidence against constant spread

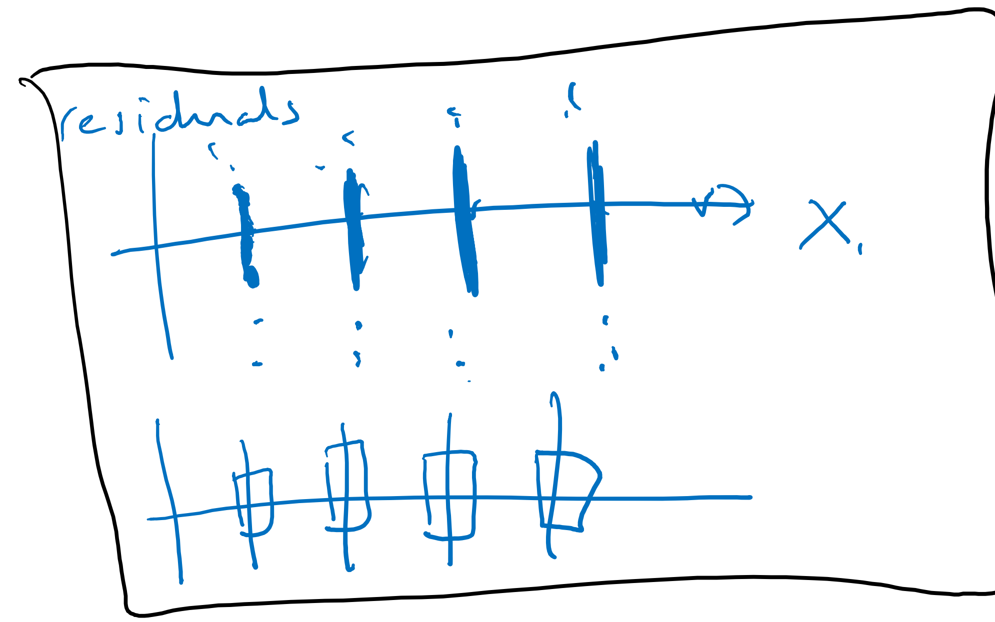


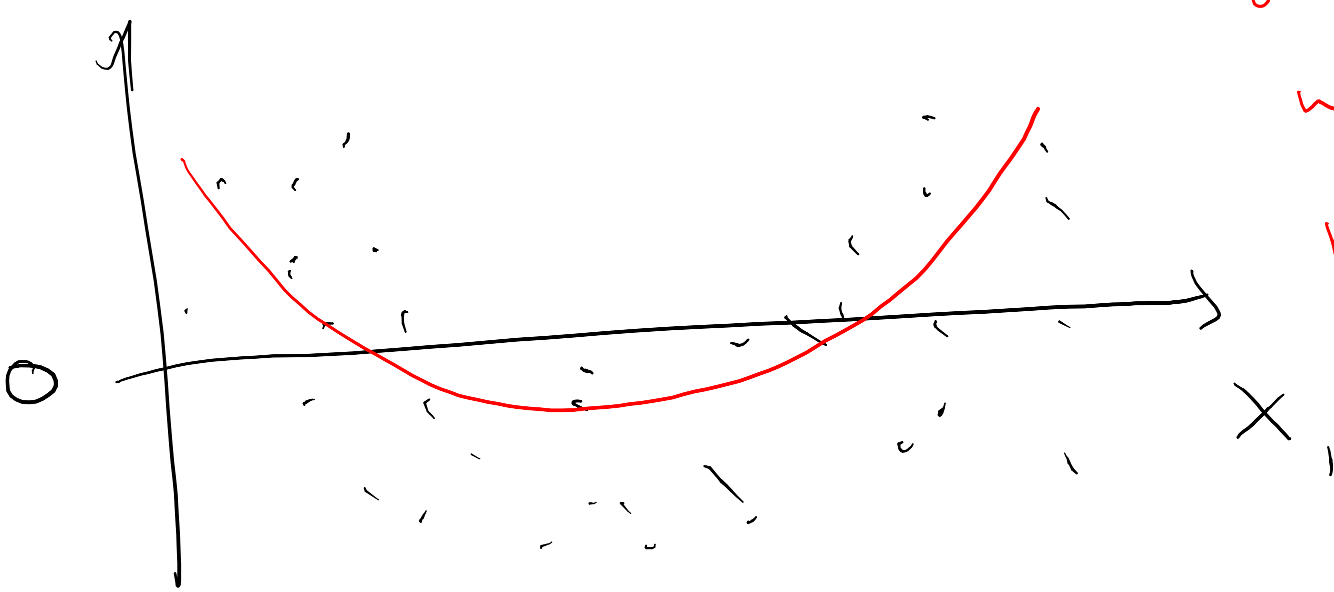
- transform response
- missing something in model

Residuals against explanatory values

equivalent to fitted versus residuals in SLR



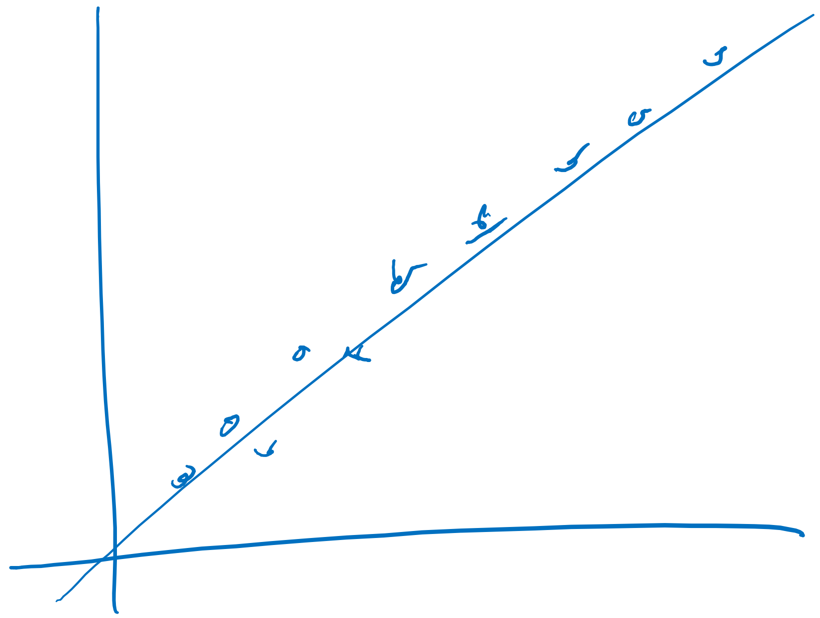




our problem with non-linearity is probably due to X_1 not entering the model in the right way.

- X_1^2
- complicated curves
- transform X_1

Normal probability plot



Less important for large sample sizes

- especially if everything else is OK

Partial residuals

Sometimes you want to "**look**" at the relationship between an explanatory and the response, after taking into **account** the other variables.

Partial residuals are always relative to an explanatory variable.

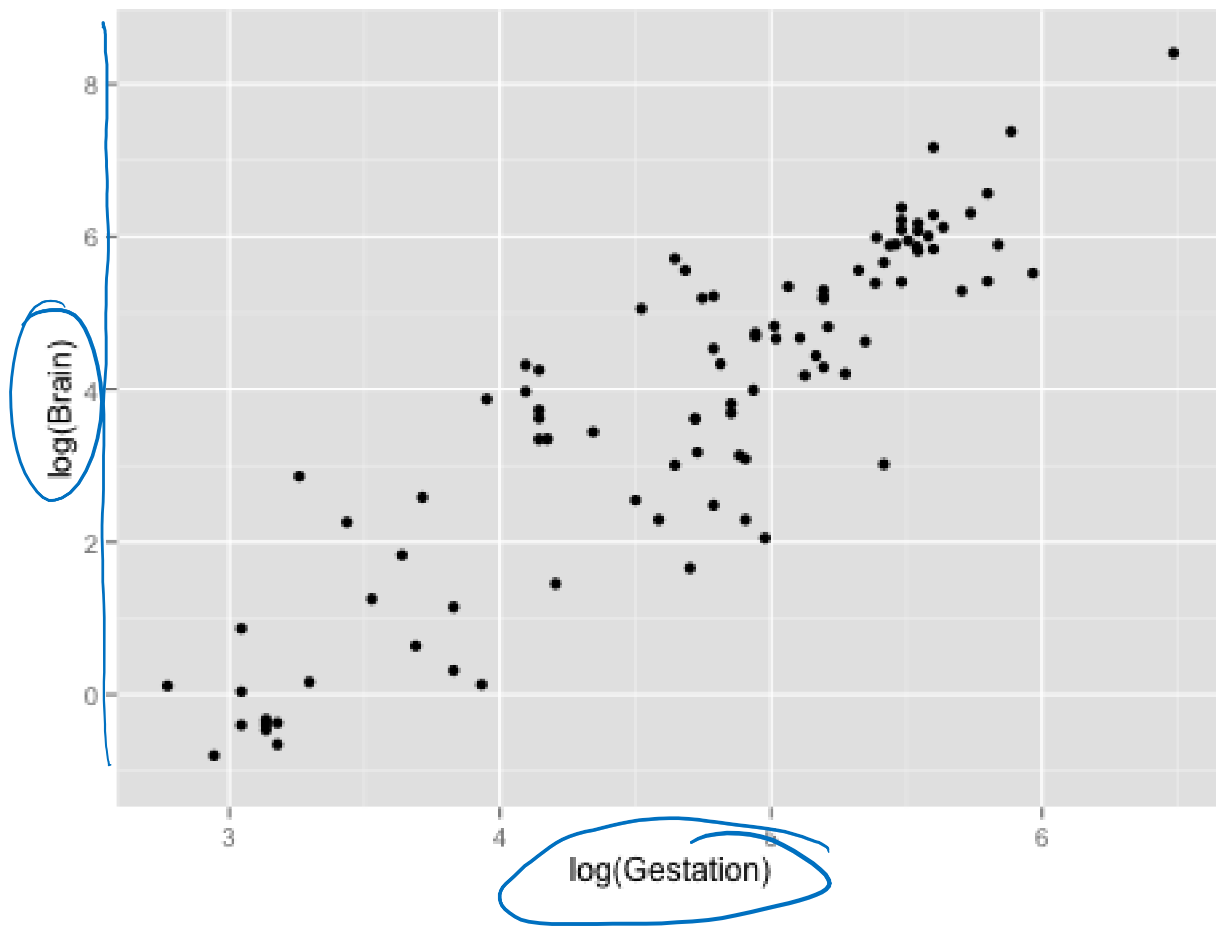
They represent the residual after subtracting off the contribution from **all the other** explanatory variables.

For example, in the mammalian case study:

The partial residuals with respect to gestation length, tell us about the relationship between log brain mass and gestation length after accounting for the effects of body mass (and anything else in the model).

They can be useful to verify the form of terms in a model (e.g. linear versus non-linear).

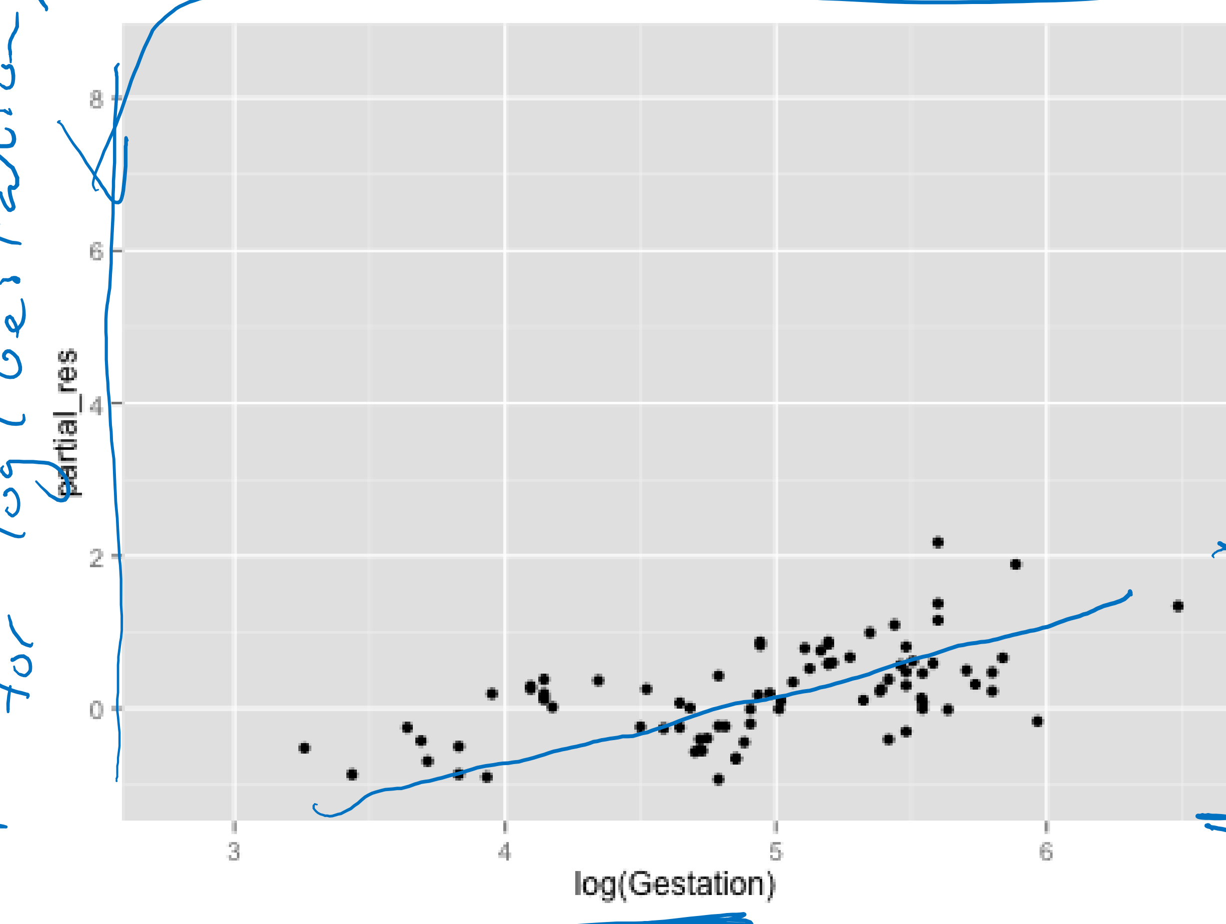
response



unadjusted log(brain weight) against Gestation length

response minus the contribution of other variables, log body mass

partial residual for log(Gestation)



log(brain weight) adjusted for body weight against Gestation length

looks like a straight line, = β_2LG

Defining partial residuals

We imagine that:

some function of gestation length

$\mu\{\log(\text{brain}) \mid \text{gestation, body, litter}\} =$

$\beta_0 + \beta_1 \log(\text{body}) + f(\log(\text{gestation}))$

implies

$f(\log(\text{gestation})) = \mu\{\log(\text{brain}) \mid \text{gestation, body, litter}\} -$

$(\beta_0 + \beta_1 \log(\text{body}))$

we approximate with:

$\text{partial res}_i = \text{response}_i - (\hat{\beta}_0 + \hat{\beta}_1 \log(\text{body}_i))$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ come from a model with gestation length in it too (usually linear).

Case1102: Blood Brain Barrier

The brain is protected from bacteria and toxins by the blood-brain barrier, but it also blocks cancer medicine.

A new method to disrupt the blood-brain barrier - BD solution.

Rats inoculated with human lung cancer to induce brain tumors, then given either the BD solution or placebo (saline solution).

After some time, sacrificed and the amounts of antibody in the brain are measured.

Response is the ratio of antibody concentration in the brain to the liver.

Do for homework.

Display 11.4 manipulated by the researcher p. 308

Response variable, design variables, and several covariates for 34 rats in the blood-brain barrier disruption experiment

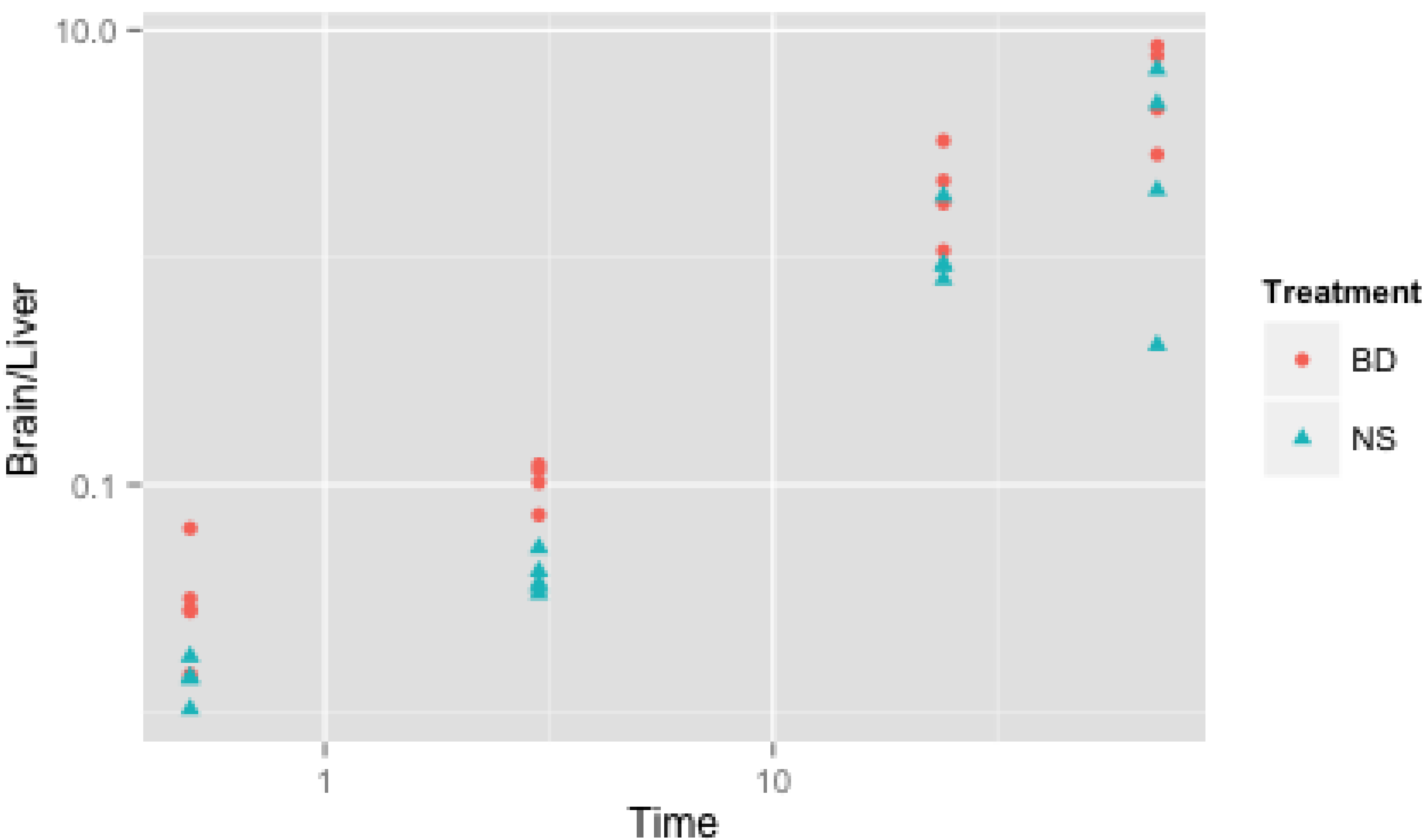
not controllable

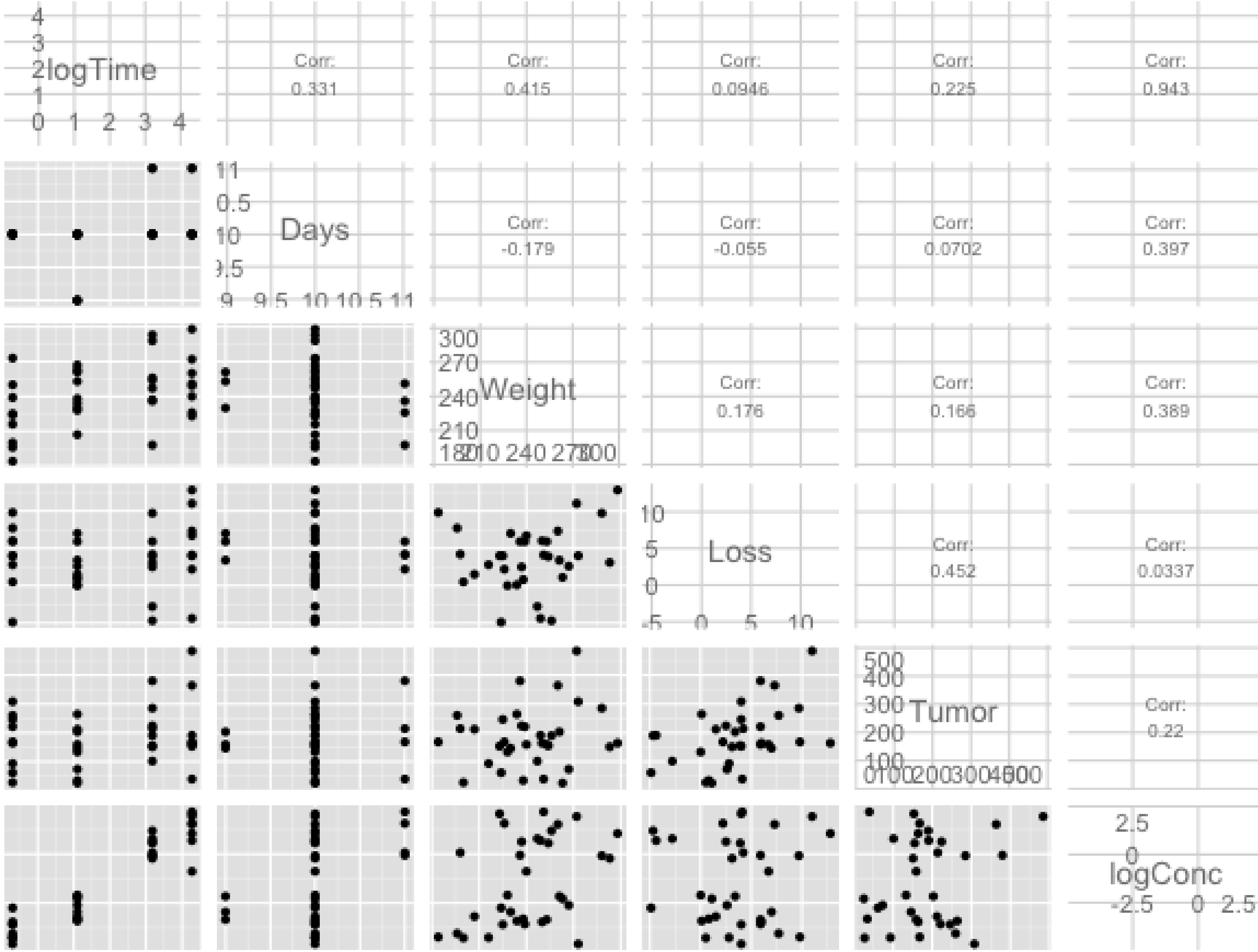
Case	*Response Variable*: Brain tumor Count (per gm) / Liver Count (per gm)	*Design Variables*: Sacrifice Time (hours)	*Design Variables*: Treatment	*Covariates*: Days Post Inoculation	*Covariates*: Sex	*Covariates*: Initial Weight (grams)	*Covariates*: Weight Loss (grams)	*Covariates*: Tumor Weight (10^{-4} grams)
1	41081 / 1456164	0.5	BD	10	F	239	5.9	221
2	44286 / 1602171	0.5	BD	10	F	225	4.0	246
3	102926 / 1601936	0.5	BD	10	F	224	-4.9	61
4	25927 / 1776411	0.5	BD	10	F	184	9.8	168
5	42643 / 1351184	0.5	BD	10	F	250	6.0	164
6	31342 / 1790863	0.5	NS	10	F	196	7.7	260
7	22815 / 1633386	0.5	NS	10	F	200	0.5	27
8	16629 / 1618757	0.5	NS	10	F	273	4.0	308
9	22315 / 1567602	0.5	NS	10	F	216	2.8	93
10	77961 / 1060057	3	BD	10	F	267	2.6	73
11	73178 / 715581	3	BD	10	F	263	1.1	25
12	76167 / 620145	3	BD	10	F	228	0.0	133
13	123730 / 1068423	3	BD	9	F	261	3.4	203
14	25569 / 721436	3	NS	9	F	253	5.9	159
15	33803 / 1019352	3	NS	10	F	234	0.1	264
16	24512 / 667785	3	NS	10	F	238	0.8	34
17	50545 / 961097	3	NS	9	F	230	7.0	146
18	50690 / 1220677	3	NS	10	F	207	1.5	212
19	84616 / 48815	24	BD	10	F	254	3.9	155
20	55153 / 16885	24	BD	10	M	256	-4.7	190
21	48829 / 22395	24	BD	10	M	247	-2.8	101
22	89454 / 83504	24	BD	11	F	198	4.2	214
23	37928 / 20323	24	NS	10	F	237	2.5	224
24	12816 / 15985	24	NS	10	M	293	3.1	151
25	23734 / 25895	24	NS	10	M	288	9.7	285
26	31097 / 33224	24	NS	11	F	236	5.9	380
27	35395 / 4142	72	BD	11	F	251	4.1	39
28	18270 / 2364	72	BD	10	F	223	4.0	153
29	5625 / 1979	72	BD	10	M	298	12.8	164
30	7497 / 1659	72	BD	10	M	260	7.3	364
31	6250 / 928	72	NS	10	M	272	11.0	484
32	11519 / 2423	72	NS	11	F	226	2.2	168
33	3184 / 1608	72	NS	10	M	249	-4.4	191
34	1334 / 3242	72	NS	10	F	240	6.7	159

Was the antibody concentration in the tumor increased by the use of the brain-blood barrier disruption solution? If so, by how much?

Do the answers depend on the length of time after the infusion?

What is the effect of treatment on antibody concentration after weight loss, total tumor weight and other covariates are accounted for?


10.0
0.1
Brain/Liver
1
10
Time
Treatment
BD
NS



logTime
Days
Weight
Loss
Tumor
logConc
Corr:
0.331
0.415
0.0946
0.225
0.943
-0.179
-0.055
0.0702
0.397
0.176
0.166
0.389
0.452
0.0337
0.22

Strategy

1. Fit a tentative model
2. Check residual plots for problems and outliers
3. Investigate influential points
4. Find a simple good fitting model
5. Answer questions of interest