

Mismatch data: example of Gamma response

1. Description of problem To study the effect of tissue match on survival time after heart transplant. Reference: Mosteller and Tukey, "Data Analysis and Regression", p. 571 ff.

2. Data set Observations on patients admitted to Stanford heart transplant program during period January 1968 - April 1974.

Variables:

surv_time survival time (days) following heart transplant

reject 1 if heart was rejected, 0 otherwise

mismatch measure of matching of tissue type between donor and recipient. high values indicate poor match.

age age of patient (years)

wait_time waiting time for donor heart (days)

cal_time number of days after 1.1.68, that the operation was performed.

sur	rej	mis	age	wai	cal
15	0	1.11	54.3	0	6
3	0	1.66	40.4	35	123
624	1	1.32	51	50	244
46	1	0.61	42.5	11	235
127	0	0.36	48	25	253
64	1	1.89	54.6	16	279
1350	1	0.87	54.1	36	300
...					
161	1	1.2	43.8	3	2087

3. Results of linear regression

Call:

```
lm(formula = surv_time ~ reject + mismatch + age +  
    wait_time + cal_time)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-446.73	-165.76	-118.31	50.04	931.45

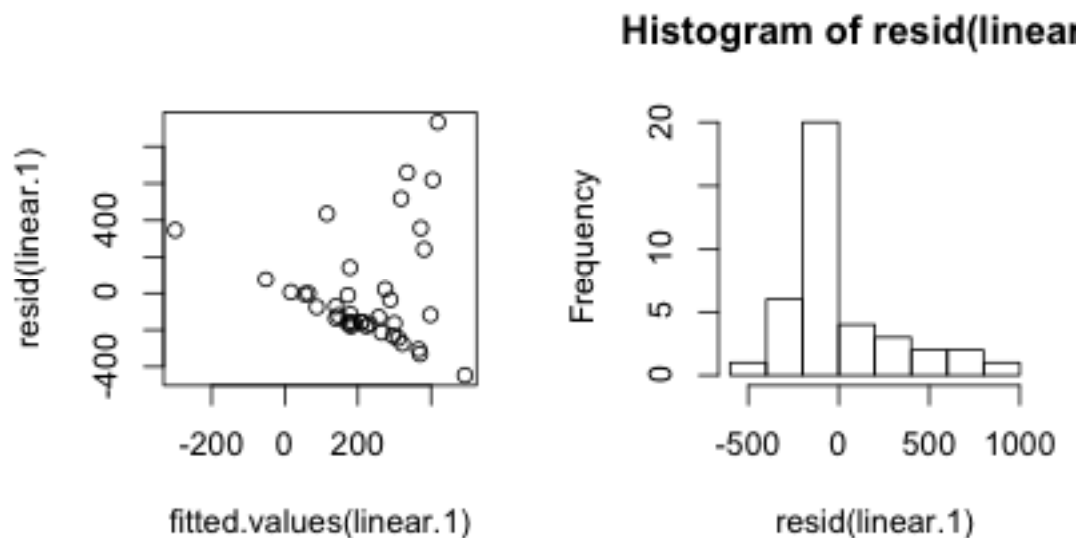
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	476.73391	425.35937	1.121	0.2705
reject	247.21704	126.33184	1.957	0.0589
mismatch	-120.88501	81.53030	-1.483	0.1476
age	-2.97777	8.64235	-0.345	0.7326
wait_time	0.01574	1.66464	0.009	0.9925
cal_time	-0.13233	0.08583	-1.542	0.1327

Residual standard error: 323.5 on 33 degrees of freedom

Multiple R-squared: 0.1942, Adjusted R-squared: 0.07209

F-statistic: 1.59 on 5 and 33 DF, p-value: 0.1901



4. Now try Gamma regression

a. `glm(surv_time ~
reject + mismatch + age + wait_time + cal_time,
family = Gamma(link="inverse"),
start=rep(1, 6), control=glm.control(trace = TRUE))`

Gives 50 warnings: step size truncated due to divergence (NaNs); halves the step size for the first 8 iterations. Finally converges

b. **Try obtaining a starting value from a linear regression of 1/surv_time on X**

`beta0=(1.2e-01 -1.3e-01 -1.3e-02 1.3e-03 8.7e-04 -4.7e-05)`

c. **Using beta0 as starting value gives error message: "cannot find valid starting values: please specify some"**

d. **Truncating negative elements in beta0 to 0 gives starting value which works. Now we only have 25 warnings, and step-halving for the first 4 iterations. Converges to the same estimate as in a. Since all elements of X*beta0 are positive, it's not clear what the problem is.**

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6823	-1.3166	-0.7889	0.1322	2.5009

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.209e-03	6.824e-03	0.910	0.369
reject	-8.670e-03	5.153e-03	-1.682	0.102
mismatch	3.443e-03	2.241e-03	1.536	0.134
age	1.030e-05	1.092e-04	0.094	0.925
wait_time	6.936e-06	4.831e-05	0.144	0.887
cal_time	2.206e-06	1.763e-06	1.251	0.220

(Dispersion parameter for Gamma family taken to be 1.90811)

Null deviance: 81.724 on 38 degrees of freedom
Residual deviance: **60.603** on **33** degrees of freedom
AIC: 491.56

NOTE: 60.603/33 = 1.836

```
pearson.resids=(surv_time - gamma.4$fitted)/gamma.4$fitted  
min, max(pearson.resids) = -0.9898183 4.9026819>  
phi.hat = sum(pearson.resids^2) / 33 = 1.908089
```