

Andrew

Adam

Factors Influencing IQ

	SA1	IQ	HC	BO	GENOME	SEX	TSA	VOL	BWT
1	6.08	96	54.7	1	1	2	1913.88	1005	57.607
2	5.73	89	54.2	2	1	2	1684.89	963	58.968
3	6.22	87	53.0	1	2	2	1902.36	1035	64.184
4	5.80	87	52.9	2	2	2	1860.24	1027	58.514
5	7.99	101	57.8	1	3	2	2264.25	1281	63.958
6	8.42	103	56.9	2	3	2	2216.40	1272	61.690
7	7.44	103	56.6	1	4	2	1866.99	1051	133.358
8	6.84	96	55.3	2	4	2	1850.64	1079	107.503
9	6.48	127	53.1	1	5	2	1743.04	1034	62.143
10	6.43	126	54.8	2	5	2	1709.30	1070	83.009
11	7.99	101	57.2	2	6	1	1689.60	1173	61.236
12	8.76	96	57.2	1	6	1	1806.31	1079	61.236
13	6.32	93	57.2	2	7	1	2136.37	1067	83.916
14	6.32	88	57.2	1	7	1	2018.92	1104	79.380
15	7.60	94	55.8	2	8	1	1966.81	1347	97.524
16	7.62	85	57.2	1	8	1	2154.67	1439	99.792
17	6.03	97	57.2	1	9	1	1767.56	1029	81.648
18	6.59	114	56.5	2	9	1	1827.92	1100	88.452
19	7.52	113	59.2	2	10	1	1773.83	1204	79.380
20	7.67	124	58.5	1	10	1	1971.63	1160	72.576

General Linear Model

```
> GLM = lm(IQ ~ SA1 + HC + factor(BO) + factor(GENOME) + factor(SEX) +  
TSA + VOL + BWT)
```

```
> GLM
```

Call:

```
lm(formula = IQ ~ SA1 + HC + factor(BO) + factor(GENOME) + factor(SEX)  
+  
    TSA + VOL + BWT)
```

Coefficients:

(Intercept)	SA1	HC	factor(BO) 2
factor(GENOME) 2	factor(GENOME) 3		
770.85535	-9.87665	-14.83835	-1.37761
-35.85767	38.39867		
factor(GENOME) 4	factor(GENOME) 5	factor(GENOME) 6	factor(GENOME) 7
factor(GENOME) 8	factor(GENOME) 9		
-36.01664	16.52103	65.05458	-0.96262
-29.66046	18.87517		
factor(GENOME) 10	factor(SEX) 2	TSA	VOL
BWT			
76.85812	NA	0.04747	0.03652
1.15367			

Knowing that the General Linear Model is too complex to present, a search for an effective reduced linear model is started. `drop1` function is used to determine factors that appear to have no or little influence on the model.

```
> drop1(GLM)
```

Single term deletions

Model:

```

IQ ~ SA1 + HC + factor(BO) + factor(GENOME) + factor(SEX) + TSA +
VOL + BWT

```

	Df	Sum of Sq	RSS	AIC
<none>			43.03	47.322
SA1	1	24.58	67.61	54.361
HC	1	157.42	200.44	76.096
factor(BO)	1	3.94	46.97	47.075
factor(GENOME)	8	2441.03	2484.06	112.438
factor(SEX)	0	0.00	43.03	47.322
TSA	1	86.09	129.11	67.299
VOL	1	14.63	57.66	51.176
BWT	1	130.84	173.87	73.252

Presented here is the Reduced Linear Model with the initial factors dropped.

```
> RLM1 = lm(IQ ~ SA1 + HC + factor(GENOME) + TSA + BWT)
```

```
> RLM1
```

Call:

```
lm(formula = IQ ~ SA1 + HC + factor(GENOME) + TSA + BWT)
```

Coefficients:

(Intercept)	SA1	HC	factor(GENOME) 2
factor(GENOME) 3	factor(GENOME) 4		
702.65280	-9.32191	-13.03018	-31.52729
41.22883	-30.88217		
factor(GENOME) 5	factor(GENOME) 6	factor(GENOME) 7	factor(GENOME) 8
factor(GENOME) 9	factor(GENOME) 10		

21.23124	64.37867	-1.38297	-16.64236
19.75976	76.55780		
TSA	BWT		
0.05146	1.06017		

This model is still complex, so a more involved search in determining un-important factors is started. The question then becomes what factor is un-important enough to get rid of while still maintaining a simple yet effective model. Trial and error is used to test different models and their associated p-values against each other.

```
> RRLM3=lm(IQ~HC+factor(GENOME)+TSA+BWT)
> RRLM5=lm(IQ~SA1+HC+factor(GENOME)+BWT)
> RRLM6=lm(IQ~SA1+HC+factor(GENOME)+TSA+BWT)
> anova(RRLM5,RRLM6)
Analysis of Variance Table

Model 1: IQ ~ SA1 + HC + factor(GENOME) + BWT
Model 2: IQ ~ SA1 + HC + factor(GENOME) + TSA + BWT
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1       7 163.748
2       6  58.186  1    105.56 10.885 0.01642 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> RRLM3=lm(IQ~HC+factor(GENOME)+TSA+BWT)
> anova(RRLM3,RRLM6)
Analysis of Variance Table

Model 1: IQ ~ HC + factor(GENOME) + TSA + BWT
Model 2: IQ ~ SA1 + HC + factor(GENOME) + TSA + BWT
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1       7  87.118
2       6  58.186  1    28.932  2.9835 0.1349
```

From the analysis above you can see that Reduced Linear Model 3 maintains the effectiveness of the model while simplifying it a little more from the original model. We further looked at eliminating BWT as it looks like it had the same effect on IQ as GENOME did (high BWT resulted in high IQ). However analysis of GENOME, IQ, and BWT showed that BWT could in fact not be eliminated. It is interesting to note that BWT and GENOME matched up very well; there seems to be some interaction between the two. This makes sense as obviously someone's genes influence their overall body weight. The analysis is shown below.

```
> anova(lm(IQ~BWT))
```

```
Analysis of Variance Table
```

```
Response: IQ
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
BWT	1	0	0.024	1e-04	0.991
Residuals	18	3316	184.221		

```
> anova(lm(IQ~factor(GENOME)))
```

```
Analysis of Variance Table
```

```
Response: IQ
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(GENOME)	9	2994	332.67	10.331	0.0005497 ***
Residuals	10	322	32.20		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(lm(BWT~factor(GENOME)))
```

```
Analysis of Variance Table
```

```
Response: BWT
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(GENOME)	9	7000.8	777.87	12.334	0.0002568 ***
Residuals	10	630.7	63.07		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The important concept to see here is that this model reduction is not saying certain factors do or don't play a role in IQ scores. Rather it is saying that a reduced model can be just as effective as the full model while simplifying the analysis for further research. Simpler is always better so the more factors we can take out of an equation *and still maintain the meaning of the model* the better and easier further analysis will be.