Study on the Effect of Height and Brain Size on Intelligence

Andrew Curtis

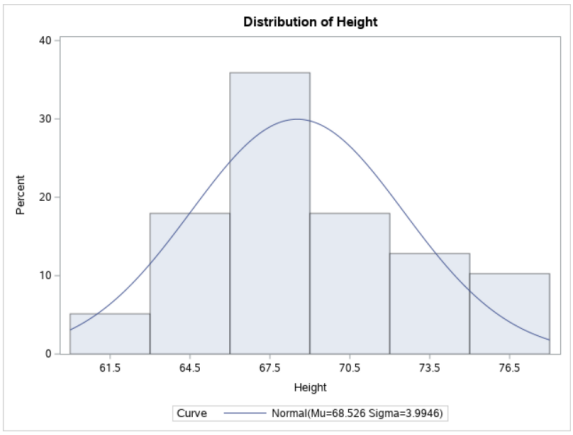
March 4, 2022

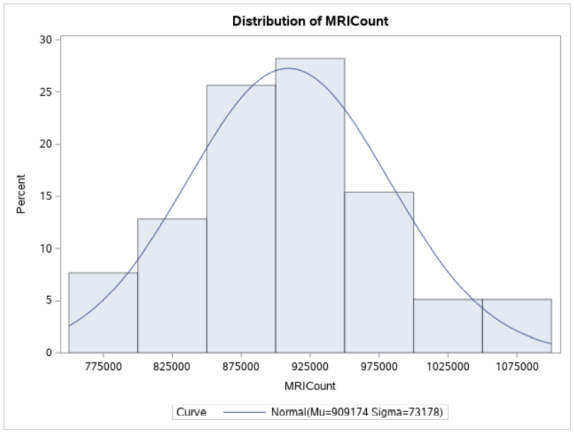
1 Introduction

The brain is a complicated organ and its connection to intelligence is still not fully understood. A natural hypothesis is that bigger brain size would positively affect intelligence. Willerman et al. (1991) used MRI machines to come up with a metric for measuring brain size. In conjunction with height as a measure of body size, a study of the effect of brain and body size on intelligence can be conducted. The results of which can perhaps shed some light on this complicated organ and its connection to intelligence.

2 Data

The data used in this experiment was broken down into three variables. Height is a measurement of body size and was recorded in inches. MRI Count is an integer count of the number of pixels the brain takes up in the MRI scans. Willerman et al. (1991) makes a case for MRI Count as a measure of brain size. Below are histogram plots of height and MRI Count measurements in the sample, fit with a normal curve for visualization,





Figures: a) left, b) right

From figures a) and b), the distributions of height and MRI Count are both fairly normal given the relatively small sample size.

Measurements of intelligence for individuals were deduced from FSIQ, or Full-Scale Intelligence Quotient test scores. These are visualized similarly below,

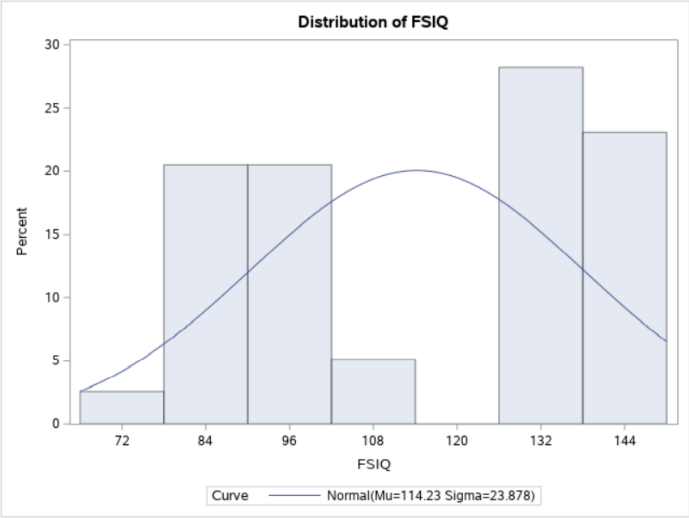


Figure c)

In figure c), the distribution of FSIQ scores has more variance than that of height or MRI Count. It does not appear to be well represented by a normal distribution, instead looking as though it might have two modes.

The relationship of the sample between the predictor height, representing body size, and FSIQ, representing intelligence is plotted below,

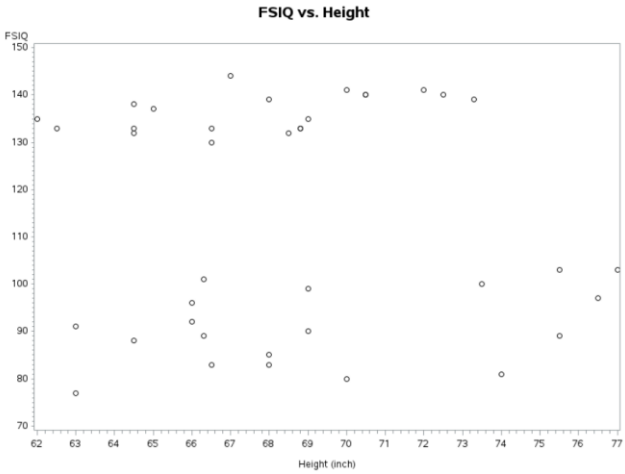


Figure d)

In figure d), height initially may not appear to have a strong correlation with intelligence. There are high and low FSIQ scores across heights with no clear trend.

The relationship of the sample between the predictor MRI Count, representing brain size, and FSIQ, representing intelligence is plotted below,

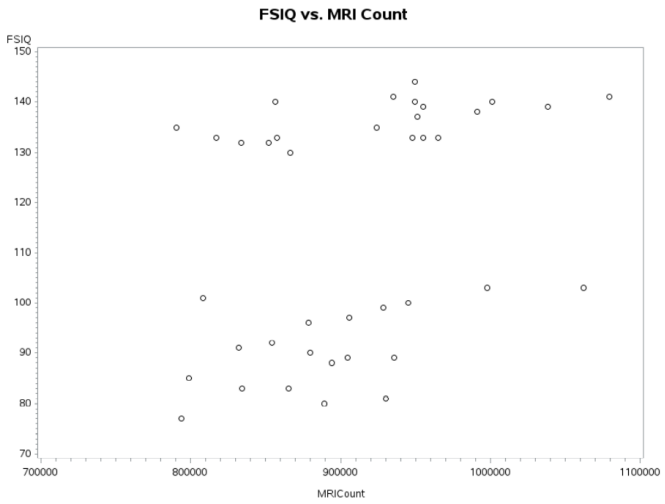


Figure e)

Similar to height, in figure e), it is difficult to immediately see a relationship between MRI Count and FSIQ score. Although there maybe be some positive correlation.

3 Model Assumption

The initial full model for our linear regression is as follows,

$$Y_{FSIQ} = \beta_o + \beta_{Height}X_1 + \beta_{MRICount}X_2 + \varepsilon$$

$$Y_{FSIQ} = 117.22 - 2.8241X_1 + 0.00020957X_2 + \varepsilon$$

A set of graphical interpretations of the error residuals are plotted below,

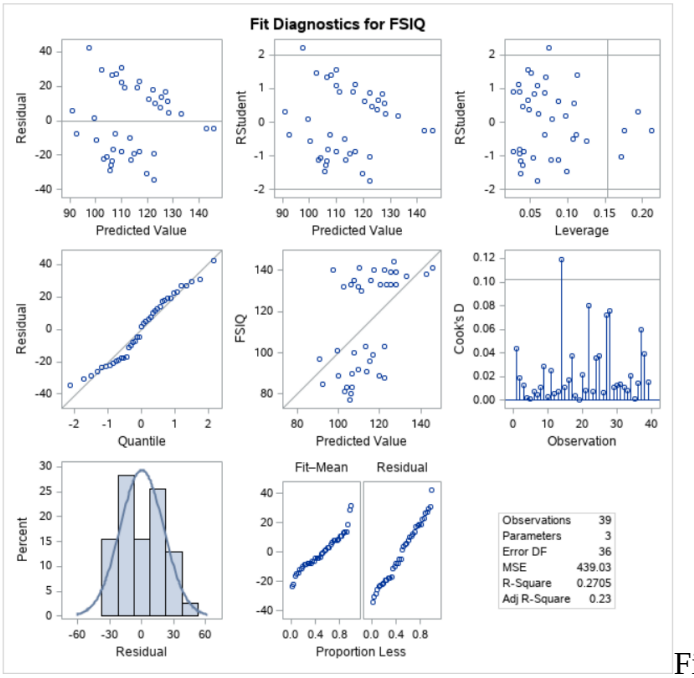


Figure f)

In the residual vs predicted value plot in figure f), residuals appear that they could be multi-modal. The fit of the residuals to the quantile line would again leads to possibly question normality. Finally, the histogram of the residuals is perhaps not exactly normal. The scatterplots of the residuals for the predictors are plotted below,

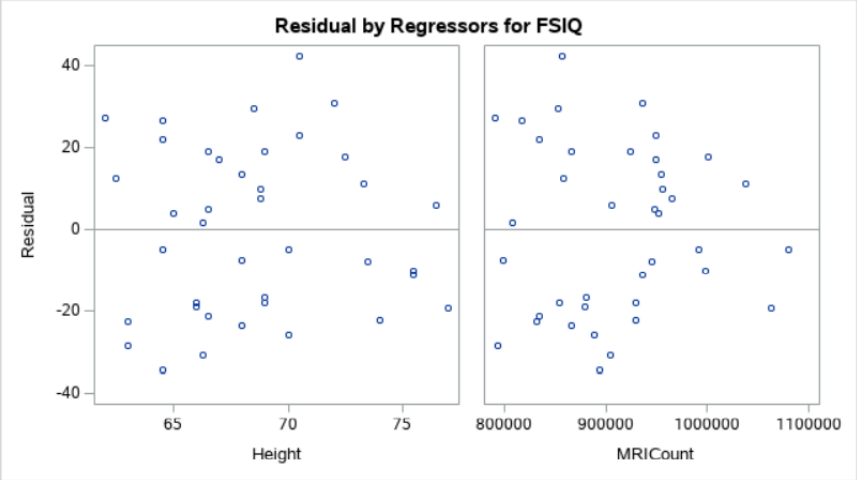


Figure g)

In figure g), the scatterplots of the residuals look like maybe there is no trend, indicating possibly constant variance and normality. In order to further test these assumptions, the Brown-Forsythe test for constant variance and the correlation test for normality were conducted and the results are shown below,

P-value for Brown-Forsythe test of constant variance in residual vs. predicted

Obs	t_BF	BF_pvalue
1	1.04626	0.30223

Output for correlation test of normality of residual (Check text Table B.6 for threshold)

The CORR Procedure

2 Variables:	resid expectNorm

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
resid	39	0	20.39430	0	-34.41614	42.38981	residual
expectNorm	39	0	0.97764	0	-2.14632	2.14632	

Pearson Correlation Coefficients, N = 39
Prob > |r| under H0: Rho=0

	resid	expectNorm
resid residual	1.00000	0.98374 <.0001
expectNorm	0.98374 <.0001	1.00000

Figure h)

The Brown-Forsythe test p-value is above the threshold, so fail to reject null hypothesis of constant variance. The value of the normality test of residuals is 0.98374 this is greater than the table value for a = 0.05, of 0.972. So, fail to reject null hypothesis of normality of residuals. From passing these tests, there appear to be no violations present (that we will deal with). Thus, this multiple linear regression model will act as the final model.

4 Model Inference and Validation

For our model regression, the model, height and MRI Count have p-values < 0.05. As a result, both parameters and the model are significant. Height as a predictor has a negative value, indicating that the taller a person is, the lower their corresponding FSIQ score. This is perhaps surprising given that larger body size would perhaps lead to larger brain size. The brain size predictor, MRI Count has a small positive value, but the MRI Count itself is a large number for pixels. This indicates that there is a positive correlation from brain size to FSIQ scores.

An investigation of multicollinearity was conducted, the variance inflation of both parameters was about ~ 1.57 which indicates fail to reject no multicollinearity. However, for the condition indexes, Height is just over 30 and MRI Count is about 46.4. This indicates that there may be some evidence that multicollinearity is occurring.

The 90% joint confidence interval of the regression coefficients are as follows:

Height (-4.98474, -0.66352) MRI Count (0.00009163, 0.00032751)

The R^2 value of the model is 0.2705. This indicates that with these two predictors, 27.05% of the total variance of the sample is being captured by the model.

Below is a 95% confidence interval for the mean response,

Simultaneous 95% interval estimation of mean response at three X-levels, using Bonferroni

Obs	Height	MRICount	Yhat	seYhat	Lower	Upper
1	64.5	816932	106.269	5.54439	95.024	117.513
2	72.5	1001121	122.276	5.52549	111.070	133.482
3	73.3	1038437	127.837	6.90105	113.841	141.833
4	68.8	965353	125.229	4.56810	115.965	134.494
5	65.0	951545	133.067	6.52492	119.834	146.301
6	69.0	928799	117.004	3.48155	109.943	124.065
7	64.5	991305	142.812	8.78150	125.002	160.622
8	66.0	854258	109.855	4.28430	101.166	118.544
9	66.3	904858	119.612	4.02816	111.442	127.781
10	68.8	955466	123.157	4.20040	114.639	131.676
11	64.5	833868	109.818	5.12121	99.432	120.204
12	70.0	1079549	145.773	9.65235	126.197	165.348
13	69.0	924059	116.011	3.42572	109.063	122.958
14	70.5	856472	97.610	5.73000	85.989	109.231
15	66.0	878897	115.018	3.98697	106.932	123.104
16	68.0	865363	106.534	4.04293	98.334	114.733
17	68.5	852244	102.372	4.70216	92.836	111.909
18	73.5	945088	107.709	5.51209	96.530	118.888
19	66.3	808020	99.317	5.89058	87.371	111.264
20	70.0	889083	105.857	4.15909	97.422	114.292
21	76.5	905940	91.032	9.24052	72.292	109.773
22	62.0	790619	107.814	7.03146	93.554	122.075
23	68.0	955003	125.320	4.52441	116.144	134.496
24	63.0	831772	113.615	5.85526	101.740	125.490
25	72.0	935494	109.935	4.52574	100.756	119.113
26	68.0	798612	92.545	6.96989	78.409	106.680
27	77.0	1062462	122.423	8.68212	104.815	140.031
28	63.0	793549	105.604	6.59467	92.230	118.979
29	66.5	866662	111.042	3.95044	103.030	119.054
30	62.5	857782	120.478	6.18946	107.925	133.031
31	67.0	949589	127.009	4.90105	117.069	136.949
32	75.5	997925	113.134	6.85218	99.237	127.031
33	69.0	879987	106.774	3.92759	98.809	114.740
34	66.5	834344	104.269	4.85288	94.427	114.111
35	66.5	948066	128.102	5.18688	117.583	138.622
36	70.5	949395	117.084	3.90281	109.169	124.999
37	64.5	893983	122.416	5.08636	112.101	132.732
38	74.0	930016	103.138	6.18324	90.598	115.679
39	75.5	935863	100.128	7.41579	85.088	115.167

Figure i)

Notice that in figure i), the mean response depends on the y-hat response variable, hence the full print out of observations. The lower column is the lower bounds at that y-hat value, similarly for the upper being the upper bound.

Three new people were added, and a 90% simultaneous prediction band is shown below in figure j),

Simultaneous 90% intervals of individualprediction at two X-profiles, using Scheffe and Bonferroni

Obs	Height	MRICount	Yhat	S_lower	S_upper	B_lower	B_upper
40	76	1000000	112.157	54.6829	169.630	63.1235	161.190
41	70	900000	108.144	52.8744	163.414	60.9913	155.298
42	65	950000	132.744	75.8707	189.616	84.2230	181.264

Figure j)

From figure j), we choose to use the more efficient Bonferroni prediction band. B_lower is the lower bound for that prediction band and B_upper is the upper bound. Notice these values should be wider than the 95% confidence interval for the mean response.

5 Conclusions

Intelligence of individuals is a complicated value to both measure and predict. A simple prediction is that larger brain size might have an impact on it. In this study, intelligence was represented by FSIQ scores and used as the response variable. A multiple linear regression model was fit with Height and MRI Pixel Count as predictors. Height is associated with larger body size and MRI Pixel Count is associated with brain size. After analysis, there is evidence for significance of this model and both predictors. Although it may be surprising that the regression coefficient for height is negative. On the other hand, as expected, brain size is positively correlated with FSIQ scores.

Overall, the model may have some issues with the FSIQ scores possibly being multi-modal and multicollinearity with the predictors. The multicollinearity of the predictors may be an expected due to a biological consequence. The R squared value of the model shows that it does capture about 27% of the variance of FSIQ scores. Future work could include finding a different measure of intelligence to regress on or adding other sensible predictors to the model.

References

Wechsler, D. (1981), “Historical review - The Psychometric Tradition: Developing the Wechsler Adult intelligence Scale”, Contemporary Education Psychology 6, 82-85.

Willerman, L., Shultz, R., Rutledge, J. N., and Bigler, E. (1991), “In Vivo Brain size and Intelligence”, Intelligence 15, pp. 223-228.

A Appendix

SAS Code Appendix

```
data brain;
    input FSIQ Height MRICount @@; cards;
133 64.5 816932
140 72.5 1001121
139 73.3 1038437
133 68.8 965353
137 65 951545
99 69 928799
138 64.5 991305
92 66 854258
89 66.3 904858
133 68.8 955466
132 64.5 833868
141 70 1079549
135 69 924059
140 70.5 856472
96 66 878897
83 68 865363
132 68.5 852244
100 73.5 945088
101 66.3 808020
80 70 889083
97 76.5 905940
135 62 790619
139 68 955003
91 63 831772
141 72 935494
85 68 798612
103 77 1062462
77 63 793549
130 66.5 866662
133 62.5 857782
144 67 949589
103 75.5 997925
90 69 879987
83 66.5 834344
133 66.5 948066
140 70.5 949395
88 64.5 893983
81 74 930016
89 75.5 935863
;
run;

proc Gplot data=brain;
    plot FSIQ*Height='circle';
    label FSIQ='FSIQ';
    label Height='Height (inch)';
    title 'FSIQ vs. Height';
run;

proc Gplot data=brain;
    plot FSIQ*MRICount='circle';
    label FSIQ='FSIQ';
    label Height='MRI Count (pixels)';
    title 'FSIQ vs. MRI Count';
run;

proc univariate data=brain;
   histogram / normal;
run;

proc reg data=brain;
  model FSIQ = Height MRICount;
  output out=out1 residual=resid predicted=pred;
  title1 'Full Model';
run;

%macro resid_num_diag(dataset,datavar,label='requested variable'
%resid_num_diag(dataset=out1, datavar=resid,
   label='residual', predvar=pred, predlabel='predicted');

proc transreg data=brain;
 model boxcox(FSIQ / lambda=-1 to 1 by 0.1)
       = identity(Height MRICount);
 title1 'Box-Cox Transformation: Regressing Level on Age';
run;

proc reg data=brain;
    model FSIQ = Height MRICount / vif collin;
  title1 'Test for multicollinearity';
run;


/* log transformation  */
data brain; set brain;
  log_FSIQ = log(FSIQ);
run;

proc reg data=brain;
  model log_FSIQ = Height MRICount;
  output out=out2 r=resid p=pred;
  title1 'Transformed model for log FSIQ data';
run;

/*Joint 90% intervals for Regession Coefficients*/
proc reg data=brain;
  model FSIQ = Height MRICount /
      clb alpha=.05;
  title1 'Simultaneous 90% intervals for two predictors
effects';
run;

/*95% Confindence Interval for the Mean Response*/
data temp; set brain;
proc reg data=temp noprint;
        model FSIQ = Height MRICount;
        output out=out1 p=Yhat stdp=seYhat;
data out1; set out1;
        alpha = 0.05;
        p = 3;
        n = 39;
        t = tinv((1-alpha/2),n-p);
        Upper = Yhat + t*seYhat;
        Lower = Yhat - t*seYhat;
proc print data=out1;
        var Height MRICount Yhat seYhat Lower Upper;
        title1
        'Simultaneous 95% interval estimation of mean response';
        title2
        'at three X-levels, using Bonferroni';
run;

/*90% Simultaneous Prediction Band*/
data dummy; input Height MRICount check; cards;
        76 1000000 1
        70 900000 1
        65 950000 1
        ;
data temp; set brain dummy;
run;

proc reg data=temp noprint;
  model FSIQ = Height MRICount;
  output out=out1 p=Yhat stdi=seYhatnew;
data out1; set out1;
  alpha = 0.10;
  p = 3;
  n = 39;
  g = 3;
  S = sqrt(g*finv(1-alpha,g,n-p));
  t = tinv(1-alpha/(2*g),n-p);
  S_upper = Yhat + S*seYhatnew;
  S_lower = Yhat - S*seYhatnew;
  B_upper = Yhat + t*seYhatnew;
  B_lower = Yhat - t*seYhatnew;
proc print data=out1;
  where check = 1;
  var Height MRICount Yhat S_lower S_upper
      B_lower B_upper;
  title1 'Simultaneous 90% intervals of individual
prediction';
  title2 'at two X-profiles, using Scheffe and Bonferroni';
run;
```