Runa Veigas
runaveigas@gmail.com

Stepwise Regression for IQ Data

About the Data

Some researchers (Willerman, et al, 1991) collected the IQ data on a sample of *n* = 38 college students to research if a person's brain size and body size predictive of his or her intelligence. The data comprised of the following:

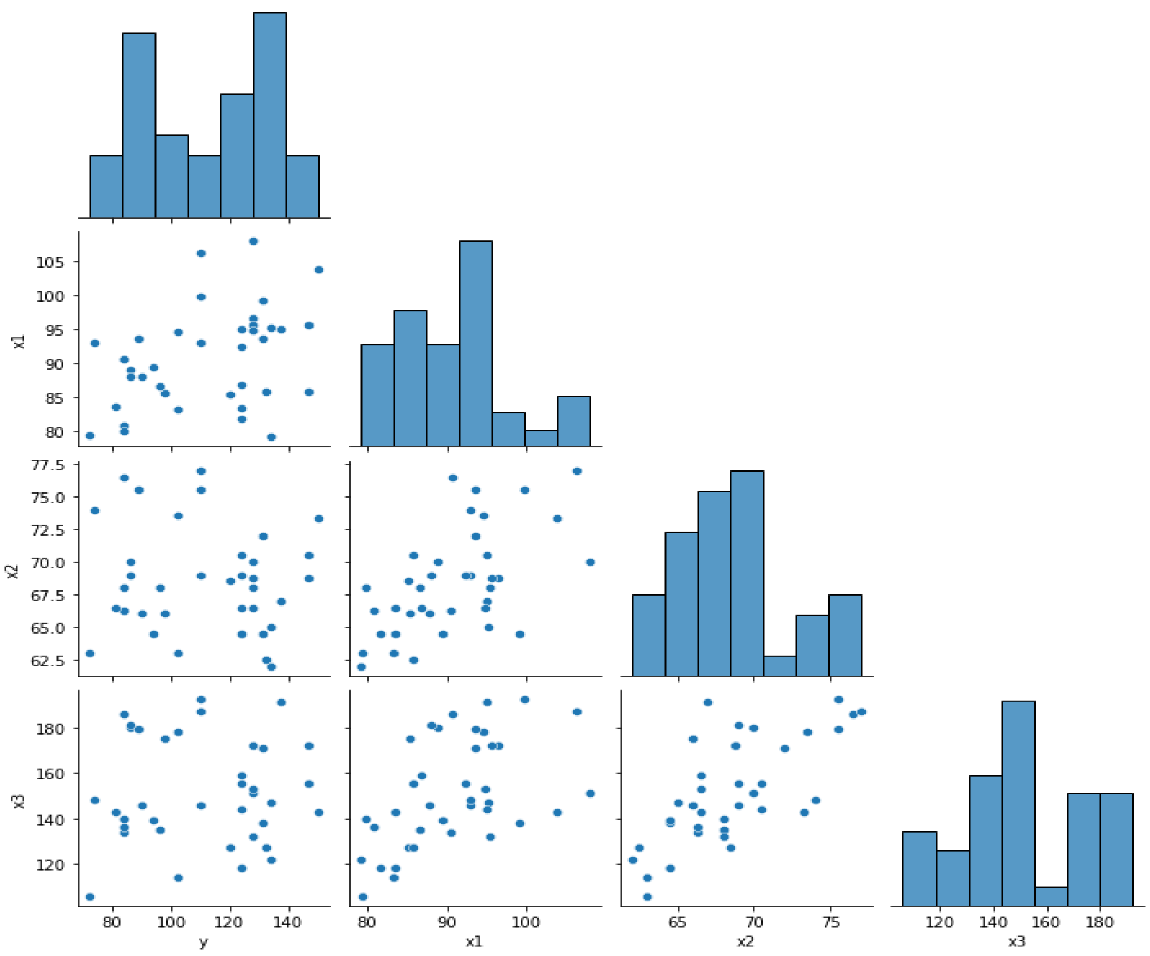
1. Response (**y**): Performance IQ scores (PIQ) from the revised Wechsler Adult Intelligence Scale. This variable served as the investigator's measure of the individual's intelligence.
2. Potential predictor (**x1**): Brain size based on the count obtained from MRI scans (given as count/10,000).
3. Potential predictor (**x2**): Height in inches.
4. Potential predictor (**x3**): Weight in pounds.

Descriptive Statistics

	Y	X1	X2	X3
count	38	38	38	38
mean	111.34	90.67	68.42	151.05
count	22.59	7.25	3.99	23.47
min	72	79	62	106
25%	89.25	85.48	66	135.25
50%	115	90.54	68	146.5
75%	128	94.95	70.37	172
max	150	107.95	77	192

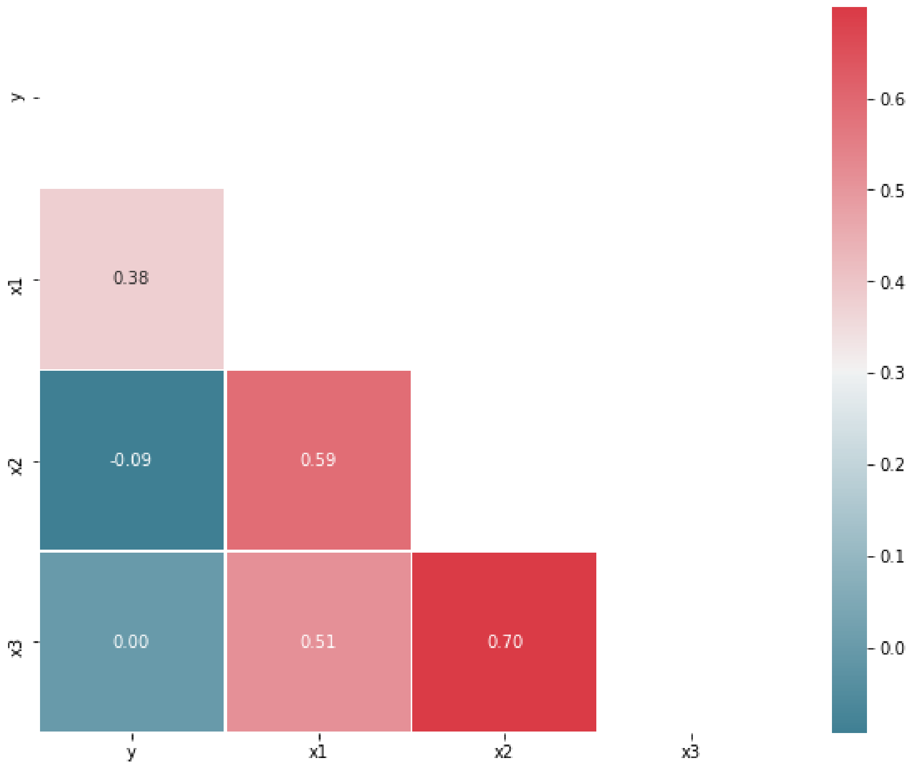
Correlation Analysis

1. Pairs Plot



Runa Veigas
runaveigas@gmail.com

2. Correlation Plot



Interpretation: There is a positive correlation between y and x_1 there is no strong relationship between the other variables and considering $\alpha_E = 0.15$ and $\alpha_R = 0.15$.

Stepwise Linear Regression

Regressing *y* on x_1, regressing *y* on x_2, regressing *y* on x_3, we obtain:

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.6519	43.712	0.106	0.916	-84.000	93.304
x1	1.1766	0.481	2.448	0.019	0.202	2.151

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	147.4067	64.350	2.291	0.028	16.899	277.914
x2	-0.5271	0.939	-0.561	0.578	-2.431	1.377

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	110.9769	24.514	4.527	0.000	61.259	160.694
x3	0.0024	0.160	0.015	0.988	-0.323	0.328

x_1 predictors is a candidate to be entered into the stepwise model because each t-test P-value is less than $\alpha_E = 0.15$.

As a result of the first step, **we enter x_1 into our stepwise model**. Now we fit each of the two-predictor models that include x_1 as a predictor that is, we regress *y* on x_1 and x_2, regress *y* on x_1 and x_3 obtaining:

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	111.2757	55.867	1.992	0.054	-2.141	224.692
x1	2.0606	0.547	3.770	0.001	0.951	3.170
x2	-2.7299	0.993	-2.749	0.009	-4.746	-0.714

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.7520	43.025	0.110	0.913	-82.593	92.097
x1	1.5925	0.551	2.889	0.007	0.473	2.712
x3	-0.2503	0.170	-1.469	0.151	-0.596	0.096

The predictor x_3 is not eligible for entry into the stepwise model because its t-test P-value (0.151) is greater than $\alpha_E = 0.15$. The predictors x_1 and x_2 are candidates because each t-test P-value is less than $\alpha_E = 0.15$. **As a result of the second step, we enter x_2 into our stepwise model.**

Runa Veigas
runaveigas@gmail.com

Now, since x_1 was the first predictor in the model, we must step back and see if entering x_2 into the stepwise model affected the significance of the x_4 predictor. It did not the t-test P-value for testing $\beta_1 = 0$ is less than 0.001, and thus smaller than $\alpha_R = 0.15$. **Therefore, we proceed to the third step with both x_1 and x_2 as predictors in our stepwise model.**

Now, we fit each of the three-predictor models that include x_1 and x_2 as predictors that is, we regress y on x_1, x_2, and x_3,

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	111.3536	62.971	1.768	0.086	-16.619	239.326
x1	2.0604	0.563	3.657	0.001	0.915	3.205
x2	-2.7319	1.229	-2.222	0.033	-5.230	-0.233
x3	0.0006	0.197	0.003	$\boxed{0.998}$	-0.400	0.401

The predictor x_3 is not eligible for entry into the stepwise model because its t-test P-value (0.998) is greater than $\alpha_E = 0.15$. Our final regression model, based on the stepwise procedure contains only the predictors x1 and x2:

OLS Regression Results

Dep. Variable:	y	R-squared:	0.295
Model:	OLS	Adj. R-squared:	0.255
Method:	Least Squares	F-statistic:	7.321
Date:	Tue, 15 Jun 2021	Prob (F-statistic):	0.00221
Time:	17:40:11	Log-Likelihood:	-165.25
No. Observations:	38	AIC:	336.5
Df Residuals:	35	BIC:	341.4
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	111.2757	55.867	1.992	0.054	-2.141	224.692
x1	2.0606	0.547	3.770	0.001	0.951	3.170
x2	-2.7299	0.993	-2.749	0.009	-4.746	-0.714

Omnibus:	1.377	Durbin-Watson:	1.827
Prob(Omnibus):	0.502	Jarque-Bera (JB):	1.087
Skew:	0.408	Prob(JB):	0.581
Kurtosis:	2.860	Cond. No.	2.01e+03

Conclusion

In order to investigate if a person's brain size and body size predictive of his or her intelligence a Stepwise Regression was carried. The scatter plot showed that there was a positive correlation between size of the brain and IQ score. This is was verified with Pearson's correlation coefficient as 0.38. Further Stepwise Regression was carried to investigate the factors that help best predict the IQ score. The final model obtained was,

$\hat{y}_i = \boldsymbol{\beta_o} + \boldsymbol{Brain_i \beta_1} + \boldsymbol{Height_i \beta_2}$ where i = 1,…,38 β_o is 111.28, coefficient of Brain is 2.06 and coefficient of Height is -2.73.

This means that when height of the brain is held constant for each 2.06 increase in the brain size the IQ score increases by 111.28 and for every 2.73 inch decrease in height of the brain the IQ score increases by 111.28 on average. The adjusted R^2 value of0.255 means there 25.5% variability in IQ score can be predicted by this model having Brain Size and Height.

************Thank You************