Runa Veigas
202410006
MDS 504
Linear Regression Models

Stepwise Regression for Blood Pressure Data

About the Data

The researchers were interested in determining if a relationship exists between blood pressure and age, weight, body surface area, duration, pulse rate and/or stress level for a group of 20 individuals

1. blood pressure (y = BP, in mm Hg)
2. age (x1 = Age, in years)
3. weight (x2 = Weight, in kg)
4. body surface area (x3 = BSA, in sq m)
5. duration of hypertension (x4 = Dur, in years)
6. basal pulse (x5 = Pulse, in beats per minute)
7. stress index (x6 = Stress)

Descriptive Statistics

	Blood Pressure	Age	Weight	Body Surface Area	Duration Hypertension	Basal Pulse	Stress
count	20.000000	20.000000	20.000000	20.000000	20.000000	20.000000	20.00000
mean	114.000000	48.600000	93.090000	1.998000	6.430000	69.600000	53.35000
count	5.428967	2.500526	4.294905	0.136482	2.145276	3.803046	37.08635
min	105.000000	45.000000	85.400000	1.750000	2.500000	62.000000	8.00000
25%	110.000000	47.000000	90.225000	1.897500	5.250000	67.750000	17.00000
50%	114.000000	48.500000	94.150000	1.980000	6.000000	70.000000	44.50000
75%	116.250000	49.250000	94.850000	2.075000	7.600000	72.000000	95.00000
max	125.000000	56.000000	101.300000	2.250000	10.200000	76.000000	99.00000

Correlation Analysis

1. **Pairs Plot**
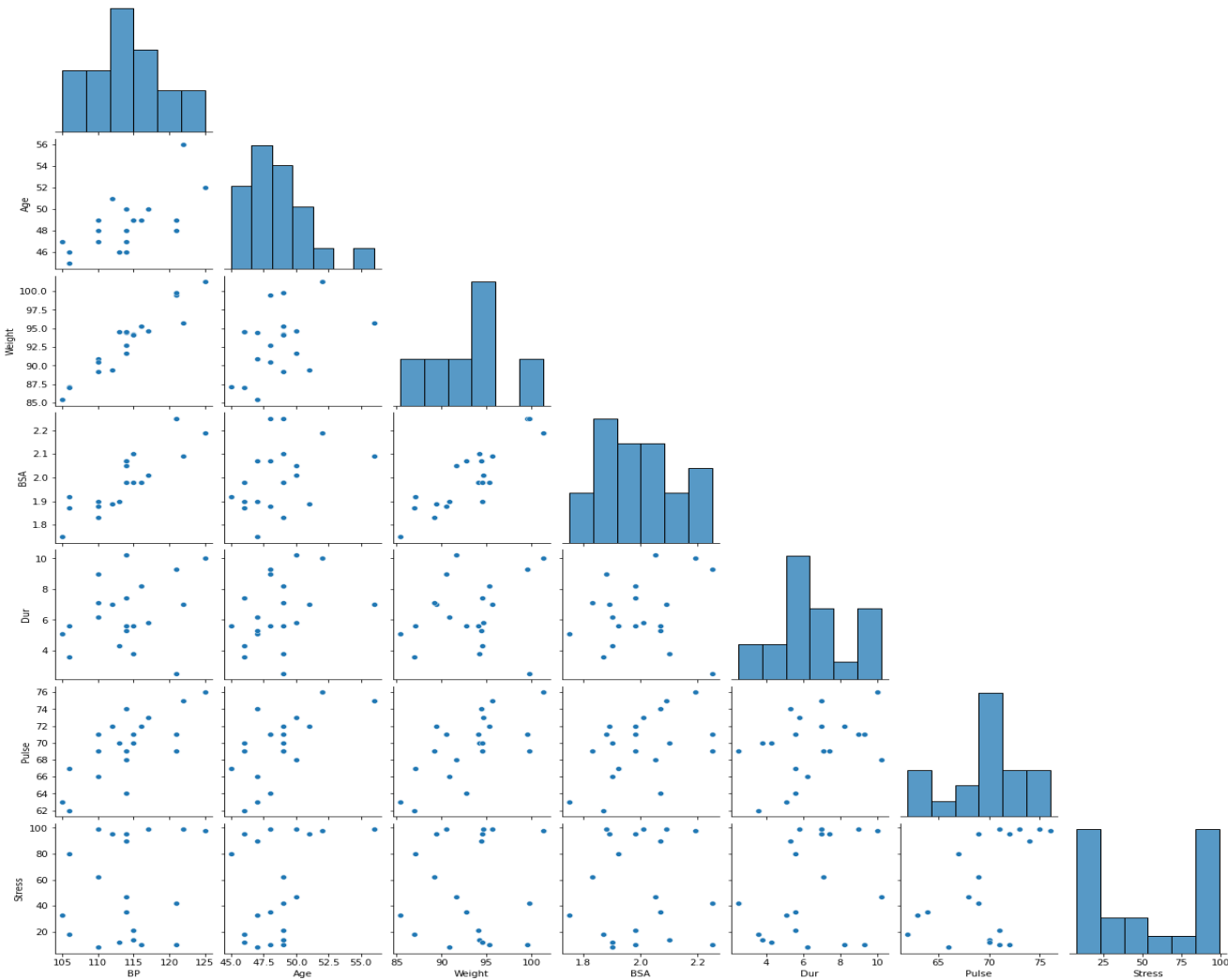
2. **Correlation Plot**
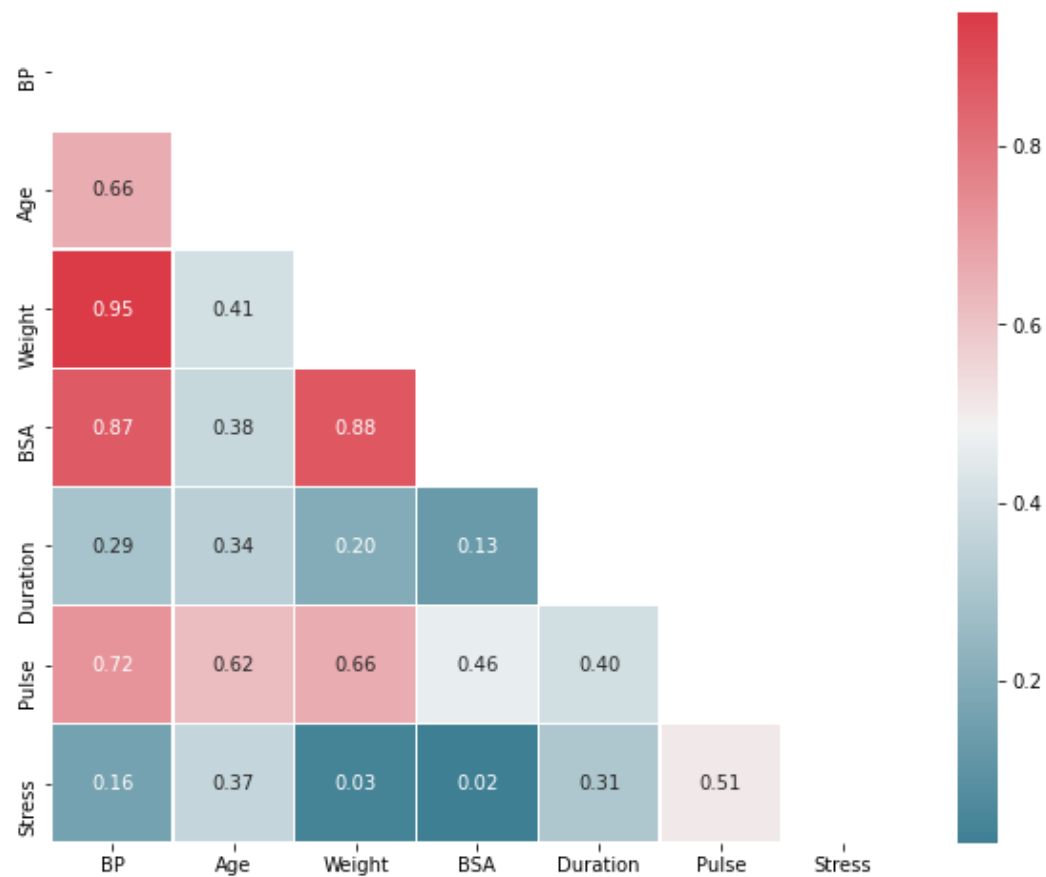
Interpretation: There is a strong positive correlation between blood pressure and age, weight, body surface aread and pulse and considering $\alpha_E = 0.15$ and $\alpha_R = 0.15$.

Stepwise Linear Regression

Regressing y on x_1, regressing y on x_2, regressing y on x_3, regressing y on x_4, regressing y on x_5, regressing y on x_6, we obtain:

```
==============================================================================
                 coef      std err           t        P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      44.4545       18.728       2.374       0.029       5.109      83.800
Age             1.4310        0.385       3.718       0.002       0.622       2.240
==============================================================================

==============================================================================
                 coef      std err           t        P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept       2.2053        8.663       0.255       0.802     -15.996      20.406
Weight          1.2009        0.093      12.917       0.000       1.006       1.396
==============================================================================

==============================================================================
                 coef      std err           t        P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     112.7200        2.193      51.389       0.000     108.112     117.328
Stress          0.0240        0.034       0.705       0.490      -0.048       0.096
==============================================================================

==============================================================================
                 coef      std err           t        P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     109.2350        3.856      28.327       0.000     101.133     117.337
Duration        0.7411        0.570       1.299       0.210      -0.457       1.939
==============================================================================

==============================================================================
                 coef      std err           t        P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      42.3231       16.240       2.606       0.018       8.203      76.443
Pulse           1.0298        0.233       4.420       0.000       0.540       1.519
==============================================================================

==============================================================================
                 coef      std err           t        P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      45.1833        9.392       4.811       0.000      25.452      64.915
BSA            34.4428        4.690       7.343       0.000      24.589      44.297
==============================================================================
```

Age, Weight, Body Surface Area and Pulse predictors are candidate to be entered into the stepwise model because each t-test P-value is less than α_E = 0.15. But weight, Body Surface Area and pulse have the least P value and weight has a higher t value which means it has the lower P Value compared to pulse. Thus **we enter Age to our stepwise model.**

Now we fit each of the two-predictor models that include Age as a predictor that is, we regress Weight, BP on Age, Pulse; BP on Age, Duration; BP on Age, BSA. BP on Age, Weight.

```
===============================================================================
                 coef     std err         t       P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept      -16.5794     3.007     -5.513      0.000     -22.925     -10.234
Age              0.7083     0.054     13.235      0.000       0.595       0.821
Weight           1.0330     0.031     33.154      0.000       0.967       1.099
===============================================================================
===============================================================================
                 coef     std err         t       P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept       16.0037     9.657      1.657      0.116      -4.370      36.378
Age              0.8398     0.199      4.225      0.001       0.420       1.259
BSA             28.6199     3.642      7.859      0.000      20.937      36.303
===============================================================================
===============================================================================
                 coef     std err         t       P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept       45.9573    19.572      2.348      0.031       4.664      87.250
Age              1.3749     0.420      3.275      0.004       0.489       2.261
Duration         0.1901     0.489      0.388      0.703      -0.842       1.223
===============================================================================
===============================================================================
                 coef     std err         t       P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept       27.1441    17.654      1.538      0.143     -10.103      64.391
Age              0.7483     0.427      1.752      0.098      -0.153       1.650
Pulse            0.7254     0.281      2.582      0.019       0.133       1.318
===============================================================================
===============================================================================
                 coef     std err         t       P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept       41.6248    20.081      2.073      0.054      -0.743      83.993
Age              1.5038     0.423      3.553      0.002       0.611       2.397
Stress          -0.0133     0.029     -0.468      0.646      -0.074       0.047
===============================================================================
```

Duration and Stress are not eligible into stepwise model because its t-test P value is greater than 0.15. The predictors weight, BSA and pulse are candidates because each t-test P-value is less than α_E = 0.15. But has the lowest p value as the t value is highest for weight. **As a result of the second step, we enter weight into our stepwise model**.

Now, since Age was the first predictor in the model, we must step back and see if entering weight into the stepwise model affected the significance of the Age predictor. It did not, the t-test P-value for testing β_1 = 0 is less than 0.001, and thus smaller than α_R = 0.15. **Therefore, we proceed to the third step with both Age and Weight as predictors in our stepwise model.**

Now, we fit each of the three-predictor models that include *Age* and *Weight* as predictors.

```
===============================================================================
                 coef     std err         t       P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
Intercept      -13.6672     2.647     -5.164      0.000     -19.278      -8.057
Age              0.7016     0.044     15.961      0.000       0.608       0.795
Weight           0.9058     0.049     18.490      0.000       0.802       1.010
BSA              4.6274     1.521      3.042      0.008       1.403       7.852
===============================================================================
```

```
==============================================================
                 coef      std err        t      P>|t|     [0.025      0.975]
--------------------------------------------------------------
Intercept     -16.0949      3.104     -5.185     0.000    -22.676      -9.514
Age             0.6953      0.057     12.280     0.000      0.575       0.815
Weight          1.0312      0.032     32.639     0.000      0.964       1.098
Duration        0.0482      0.062      0.784     0.445     -0.082       0.179
==============================================================
```

```
==============================================================
                 coef      std err        t      P>|t|     [0.025      0.975]
--------------------------------------------------------------
Intercept     -16.6900      2.938     -5.681     0.000    -22.917     -10.463
Age             0.7502      0.061     12.350     0.000      0.621       0.879
Weight          1.0614      0.037     28.722     0.000      0.983       1.140
Pulse          -0.0657      0.049     -1.353     0.195     -0.169       0.037
==============================================================
```

```
==============================================================
                 coef      std err        t      P>|t|     [0.025      0.975]
--------------------------------------------------------------
Intercept     -16.1963      3.090     -5.242     0.000    -22.747      -9.646
Age             0.6912      0.059     11.748     0.000      0.566       0.816
Weight          1.0362      0.032     32.518     0.000      0.969       1.104
Stress          0.0027      0.004      0.748     0.465     -0.005       0.010
==============================================================
```

The predictor Duration, Pulse and Stress are not eligible for entry into the stepwise model because its t-test P-value is greater than $\alpha_E = 0.15$. Our final regression model, based on the stepwise procedure contains only the predictors Age, Weight and BSA .

Now, since Age, Weight were the previous predictor in the model, we must step back and see if entering BSA into the stepwise model affected the significance of the Age and Weight 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 Age, Weight and BSA as predictors in our stepwise model.**

```
                          OLS Regression Results
==============================================================
Dep. Variable:                  BP   R-squared:                   0.995
Model:                         OLS   Adj. R-squared:              0.994
Method:              Least Squares   F-statistic:                 971.9
Date:             Tue, 15 Jun 2021   Prob (F-statistic):        2.62e-18
Time:                     19:22:11   Log-Likelihood:             -9.5930
No. Observations:               20   AIC:                         27.19
Df Residuals:                   16   BIC:                         31.17
Df Model:                        3
Covariance Type:         nonrobust
==============================================================
                 coef      std err        t      P>|t|     [0.025      0.975]
--------------------------------------------------------------
Intercept     -13.6672      2.647     -5.164     0.000    -19.278      -8.057
Age             0.7016      0.044     15.961     0.000      0.608       0.795
Weight          0.9058      0.049     18.490     0.000      0.802       1.010
BSA             4.6274      1.521      3.042     0.008      1.403       7.852
==============================================================
Omnibus:                     1.164   Durbin-Watson:               1.896
Prob(Omnibus):               0.559   Jarque-Bera (JB):            1.011
Skew:                       -0.363   Prob(JB):                    0.603
Kurtosis:                    2.172   Cond. No.                  2.93e+03
==============================================================
```

Conclusion

In order to determining if a relationship exists between blood pressure and age, weight, body surface area, duration, pulse rate and/or stress level for a group of 20 individuals Stepwise Linear Regression was carried. The scatter plot showed that there is a strong positive correlation between blood pressure and age, weight, body surface area and pulse. Further this is was verified with Pearson's correlation. Stepwise Regression was carried to investigate the factors that help best predict the Blood Pressure. The final model obtained was,

Runa Veigas
202410006
MDS 504
Linear Regression Models

$\hat{y}_i = \beta_o + Age_i\beta_1 + Weight_i\beta_2 + BSA_i\beta_2$ where i = 1,...,20 β_o is -13.67, coefficient of Age was 0.706, Weight was 0.9058 and Body Surface Area was 4.627.

This means that when Weight and BSA of the individual is held constant BP reduces by 13 mm Hg as Age increases by 0.71 years on average. When Age and Weight is held constant the BP reduces by 13mm Hg as Body Surface Area increases by 4.63 sq m. When Age and BSA is held constant BP reduces by 13 mm Hg as Weight increases by 0.91 kg. The adjusted R^2 value of 99.4 means that 99.4 % f variability in Blood Pressure is explained by the model.

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