Statistics for Analytics (BAN 100)

Assignment 4

by : Aaron Gonsalves

(161288196)

CODES

```
data fitness;
input age weight oxy runtime rstpulse runpulse maxpulse;
case = _n_;
label age='Age of man'
weight='Weight of man'
oxy='Oxygen consumption'
runtime='Time to run 1.5 miles'
rstpulse='Resting pulse rate'
runpulse='Pulse rate while running'
maxpulse='Maximum pulse rate'
case='Case no.';

datalines;
44 89.47 44.609 11.37 62 178 182
40 75.07 45.313 10.07 62 185 185
44 85.84 54.297 8.65 45 156 168
42 68.15 59.571 8.17 40 166 172
38 89.02 49.874 9.22 55 178 180
47 77.45 44.811 11.63 58 176 176
40 75.98 45.681 11.95 70 176 180
43 81.19 49.091 10.85 64 162 170
44 81.42 39.442 13.08 63 174 176
38 81.87 60.055 8.63 48 170 186
44 73.03 50.541 10.13 45 168 168
45 87.66 37.388 14.03 56 186 192
45 66.45 44.754 11.12 51 176 176
47 79.15 47.273 10.60 47 162 164
54 83.12 51.855 10.33 50 166 170
49 81.42 49.156 8.95 44 180 185
```

```
51 69.63 40.836 10.95 57 168 172
51 77.91 46.672 10.00 48 162 168
48 91.63 46.774 10.25 48 162 164
49 73.37 50.388 10.08 67 168 168
57 73.37 39.407 12.63 58 174 176
54 79.38 46.080 11.17 62 156 165
52 76.32 45.441 9.63 48 164 166
50 70.87 54.625 8.92 48 146 155
51 67.25 45.118 11.08 48 172 172
54 91.63 39.203 12.88 44 168 172
51 73.71 45.790 10.47 59 186 188
57 59.08 50.545 9.93 49 148 155
49 76.32 48.673 9.40 56 186 188
48 61.24 47.920 11.50 52 170 176
52 82.78 47.467 10.50 53 170 172
run;

title "Details of fitness dataset:";
proc contents data=fitness varnum;
run;

title "Listing Fitness data:";
proc print data=fitness;
run;

title "Performing Corr procedure:";
proc corr data=fitness plots=matrix(hist nvar=7);
var age weight oxy runtime rstpulse runpulse maxpulse;
run;
```

Details of fitness dataset:

The CONTENTS Procedure

Data Set Name	WORK.FITNESS	Observations	31
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	07/17/2021 15:24:06	Observation Length	64
Last Modified	07/17/2021 15:24:06	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	2043
Obs in First Data Page	31
Number of Data Set Repairs	0
Filename	/saswork/SAS_work4437000118F6_odaws04-usw2.oda.sas.com/SAS_workB600000118F6_odaws04-usw2.oda.sas.com/fitness.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	536977375
Access Permission	rw-r--r--
Owner Name	u58712040
File Size	256KB
File Size (bytes)	262144

Variables in Creation Order				
#	Variable	Type	Len	Label
1	age	Num	8	Age of man
2	weight	Num	8	Weight of man
3	oxy	Num	8	Oxygen consumption
4	runtime	Num	8	Time to run 1.5 miles
5	rstpulse	Num	8	Resting pulse rate
6	runpulse	Num	8	Pulse rate while running
7	maxpulse	Num	8	Maximum pulse rate
8	case	Num	8	Case no.

Listing Fitness data:

Obs	age	weight	oxy	runtime	rstpulse	runpulse	maxpulse	case
1	44	89.47	44.609	11.37	62	178	182	1
2	40	75.07	45.313	10.07	62	185	185	2
3	44	85.84	54.297	8.65	45	156	168	3
4	42	68.15	59.571	8.17	40	166	172	4
5	38	89.02	49.874	9.22	55	178	180	5
6	47	77.45	44.811	11.63	58	176	176	6
7	40	75.98	45.681	11.95	70	176	180	7
8	43	81.19	49.091	10.85	64	162	170	8
9	44	81.42	39.442	13.08	63	174	176	9
10	38	81.87	60.055	8.63	48	170	186	10
11	44	73.03	50.541	10.13	45	168	168	11
12	45	87.66	37.388	14.03	56	186	192	12
13	45	66.45	44.754	11.12	51	176	176	13
14	47	79.15	47.273	10.60	47	162	164	14
15	54	83.12	51.855	10.33	50	166	170	15
16	49	81.42	49.156	8.95	44	180	185	16
17	51	69.63	40.836	10.95	57	168	172	17
18	51	77.91	46.672	10.00	48	162	168	18
19	48	91.63	46.774	10.25	48	162	164	19
20	49	73.37	50.388	10.08	67	168	168	20
21	57	73.37	39.407	12.63	58	174	176	21
22	54	79.38	46.080	11.17	62	156	165	22
23	52	76.32	45.441	9.63	48	164	166	23
24	50	70.87	54.625	8.92	48	146	155	24
25	51	67.25	45.118	11.08	48	172	172	25
26	54	91.63	39.203	12.88	44	168	172	26
27	51	73.71	45.790	10.47	59	186	188	27
28	57	59.08	50.545	9.93	49	148	155	28
29	49	76.32	48.673	9.40	56	186	188	29
30	48	61.24	47.920	11.50	52	170	176	30
31	52	82.78	47.467	10.50	53	170	172	31

Performing Corr procedure:

The CORR Procedure

7 Variables:	age weight oxy runtime rstpulse runpulse maxpulse

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
age	31	47.67742	5.21144	1478	38.00000	57.00000	Age of man
weight	31	77.44452	8.32857	2401	59.08000	91.63000	Weight of man
oxy	31	47.37581	5.32723	1469	37.38800	60.05500	Oxygen consumption
runtime	31	10.58613	1.38741	328.17000	8.17000	14.03000	Time to run 1.5 miles
rstpulse	31	53.45161	7.61944	1657	40.00000	70.00000	Resting pulse rate
runpulse	31	169.64516	10.25199	5259	146.00000	186.00000	Pulse rate while running
maxpulse	31	173.77419	9.16410	5387	155.00000	192.00000	Maximum pulse rate

Pearson Correlation Coefficients, N = 31
Prob > |r| under H0: Rho=0

	age	weight	oxy	runtime	rstpulse	runpulse	maxpulse
age Age of man	1.00000	-0.23354 0.2061	-0.30459 0.0957	0.18875 0.3092	-0.16410 0.3777	-0.33787 0.0630	-0.43292 0.0150
weight Weight of man	-0.23354 0.2061	1.00000	-0.16275 0.3817	0.14351 0.4412	0.04397 0.8143	0.18152 0.3284	0.24938 0.1761
oxy Oxygen consumption	-0.30459 0.0957	-0.16275 0.3817	1.00000	-0.86219 <.0001	-0.39936 0.0260	-0.39797 0.0266	-0.23674 0.1997
runtime Time to run 1.5 miles	0.18875 0.3092	0.14351 0.4412	-0.86219 <.0001	1.00000	0.45038 0.0110	0.31365 0.0858	0.22610 0.2213
rstpulse Resting pulse rate	-0.16410 0.3777	0.04397 0.8143	-0.39936 0.0260	0.45038 0.0110	1.00000	0.35246 0.0518	0.30512 0.0951
runpulse Pulse rate while running	-0.33787 0.0630	0.18152 0.3284	-0.39797 0.0266	0.31365 0.0858	0.35246 0.0518	1.00000	0.92975 <.0001
maxpulse Maximum pulse rate	-0.43292 0.0150	0.24938 0.1761	-0.23674 0.1997	0.22610 0.2213	0.30512 0.0951	0.92975 <.0001	1.00000

Scatter Plot Matrix
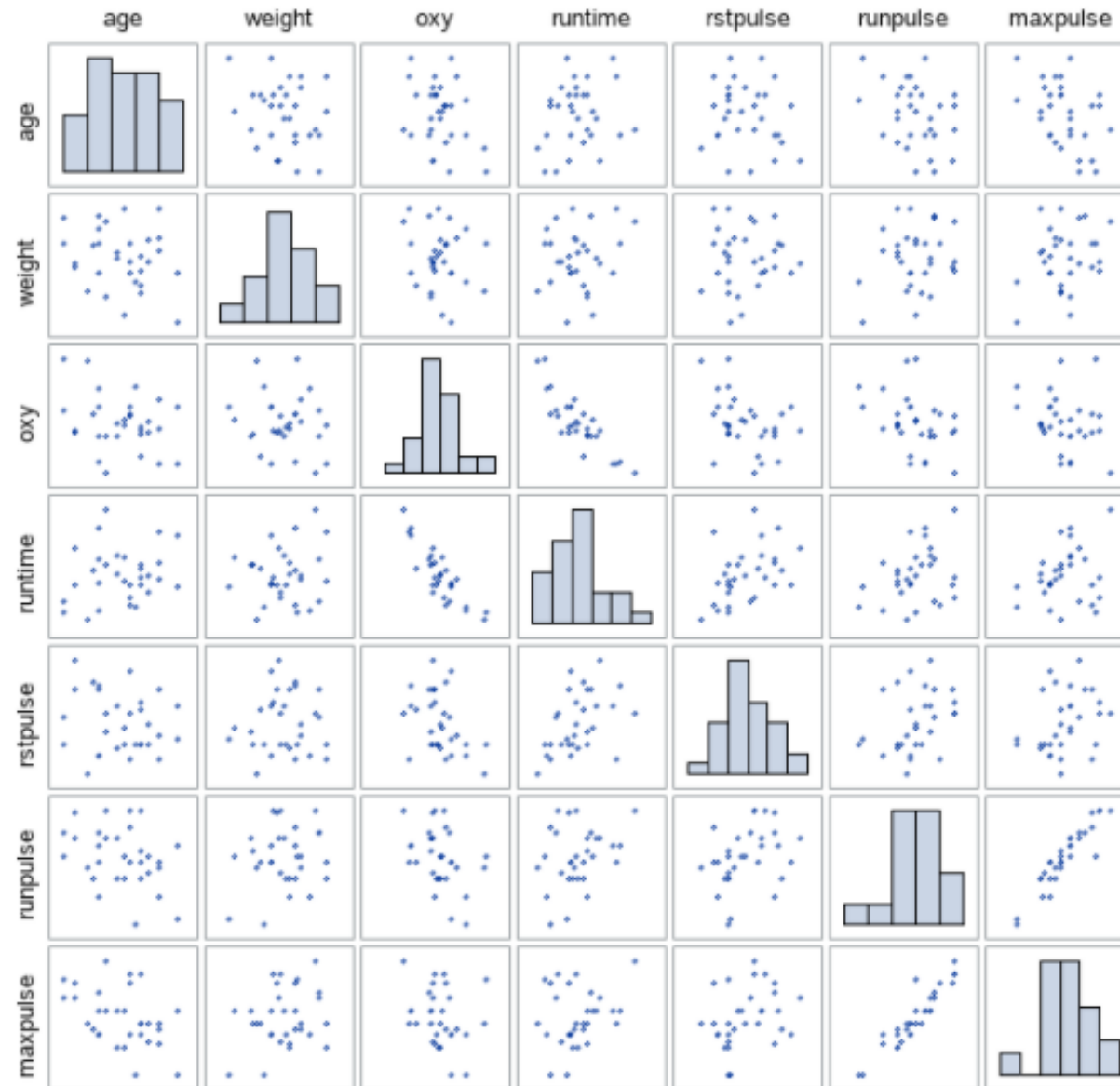
HYPOTHESIS 1

Determining whether there is relationship between age and maximum pulse rate.

Null Hypothesis(H_0):	Age and maximum pulse rate are independent variables.
Alternative Hypothesis(H_a):	Age and maximum pulse rate are dependent variables.

Interpretation

- There is a 5% significance level, P-value is 0.0150 and value of correlation co-efficient is -0.43292.

- As P-value<0.05, Null hypothesis (H_0) is rejected which considered age and maximum pulse rate as independent variables, and there is much evidence in support of Alternative Hypothesis(H_a).

- As correlation co-efficient is -0.43292 i.e. between -0.25 & -0.5 , there is a weak negative association.

Conclusion

- As Null hypothesis (H_0) is rejected from our interpretation, the conclusion is in support of Alternative Hypothesis(H_a).

- This deduces that age and maximum pulse rate are dependent on each other, and as Age increases maximum pulse rate decreases.

HYPOTHESIS 2

Determining whether there is relationship between Oxygen consumption and runtime.

Null Hypothesis(H_0):	**Oxygen consumption and runtime are independent variables.**
Alternative Hypothesis(H_a):	**Oxygen consumption and runtime are dependent variables.**

Interpretation

- There is a 5% significance level, P-value is less than 0.0001 and value of correlation co-efficient is -0.86219.

- As P-value<0.05, Null hypothesis (H_0) is rejected which considered oxygen consumption and runtime as independent variables, and there is much evidence in support of Alternative Hypothesis(H_a).

- As correlation co-efficient is -0.86219 i.e. between -0.75 & -1 , there is a strong negative association.

Conclusion

- As Null hypothesis (H_0) is rejected from our interpretation, the conclusion is in support of Alternative Hypothesis(H_a).

- This deduces that oxygen consumption and runtime are dependent on each other, and as runtime decreases oxygen consumption increases.

HYPOTHESIS 3

Determining whether there is relationship between Oxygen consumption and Resting pulse rate.

Null Hypothesis(H_0):	**Oxygen consumption and Resting pulse rate are independent variables.**
Alternative Hypothesis(H_a):	**Oxygen consumption and Resting pulse rate are dependent variables.**

Interpretation

- There is a 5% significance level, P-value is 0.0260 and value of correlation co-efficient is -0.39936.

- As P-value<0.05, Null hypothesis (H_0) is rejected which considered Oxygen consumption and Resting pulse rate as independent variables, and there is much evidence in support of Alternative Hypothesis(H_a).

- As correlation co-efficient is -0.39936 i.e. between -0.25 & -0.5 , there is a weak negative association.

Conclusion

- As Null hypothesis (H_0) is rejected from our interpretation, the conclusion is in support of Alternative Hypothesis(H_a).

- This deduces that Oxygen consumption and Resting pulse rate are dependent on each other, and as resting pulse decreases oxygen consumption increases.

HYPOTHESIS 4

Determining whether there is relationship between Oxygen consumption and Pulse rate while running.

Null Hypothesis(H_0):	Oxygen consumption and Pulse rate while running are independent variables.
Alternative Hypothesis(H_a):	Oxygen consumption and Pulse rate while running are dependent variables.

Interpretation

- There is a 5% significance level, P-value is 0.0266 and value of correlation co-efficient is -0.39797.

- As P-value<0.05, Null hypothesis (H_0) is rejected which considered Oxygen consumption and Pulse rate while running as independent variables, and there is much evidence in support of Alternative Hypothesis(H_a).

- As correlation co-efficient is -0.39797 i.e. between -0.25 & -0.5 , there is a weak negative association.

Conclusion

- As Null hypothesis (H_0) is rejected from our interpretation, the conclusion is in support of Alternative Hypothesis(H_a).

- This deduces that Oxygen consumption and Pulse rate while running are dependent on each other, and as Pulse rate while running decreases oxygen consumption increases.

HYPOTHESIS 5

Determining whether there is relationship between runtime and Resting Pulse Rate.

Null Hypothesis(H_0):	Runtime and Resting Pulse Rate are independent variables.
Alternative Hypothesis(H_a):	Runtime and Resting Pulse Rate are dependent variables.

Interpretation

- There is a 5% significance level, P-value is 0.0110 and value of correlation co-efficient is 0.45038.

- As P-value<0.05, Null hypothesis (H_0) is rejected which considered runtime and Resting Pulse Rate as independent variables, and there is much evidence in support of Alternative Hypothesis(H_a).

- As correlation co-efficient is 0.45038 i.e. between 0.25 & 0.5 , there is a weak positive association.

Conclusion

- As Null hypothesis (H_0) is rejected from our interpretation, the conclusion is in support of Alternative Hypothesis(H_a).

- This deduces that runtime and Resting Pulse Rate are dependent on each other, and as runtime increases, Resting Pulse Rate also increases.

HYPOTHESIS 6

Determining whether there is relationship between Pulse rate while running and maxpulse.

Null Hypothesis(H_0):	**Pulse rate while running(runpulse) and maxpulse are independent variables.**
Alternative Hypothesis(H_a):	**Pulse rate while running(runpulse) and maxpulse are dependent variables.**

Interpretation

- There is a 5% significance level, P-value is less than 0.0001 and value of correlation co-efficient is 0.92975.

- As P-value<0.05, Null hypothesis (H_0) is rejected which considered Pulse rate while running and maxpulse as independent variables, and there is much evidence in support of Alternative Hypothesis(H_a).

- As correlation co-efficient is 0.92975 i.e. between 0.75 & 1 , there is a strong positive association.

Conclusion

- As Null hypothesis (H_0) is rejected from our interpretation, the conclusion is in support of Alternative Hypothesis(H_a).

- This deduces that Pulse rate while running and maxpulse are dependent on each other, and as Pulse rate while running increases, maxpulse also increases.

Results from Correlation procedure and scatterplot

- Based on the interpretations and conclusions from above hypothesis' a clear relation between following variables is found.

1) Age and maxpulse

2) Oxygen consumption and runtime

3) Oxygen consumption and Resting Pulse Rate

4) Oxygen consumption and Pulse rate while running (runpulse)

5) Runtime and restpulse

6) Pulse rate while running (runpulse) and maxpulse

Limitations of Correlation Coefficient for Fitness Data

- They are based on linear association, sample size and outliers.

- r - co-efficient of correlation is a measure of linear association, and in fitness data all curves are linear, which shows that r value is good enough to predict the association.

- The sample size is 31, which is sufficient enough, so the value of correlation co-efficient is reliable and good enough to predict the association.

- Fitness data does not have outliers in it, so value of correlation co-efficient is good enough to predict the association.

THANK YOU