The SAS System

11:25 Sunday, April 02, 2023

The SURVEYSELECT Procedure

Luke Beebe
HW2 SAS

My code with its comments is posted below. The following pages are the outputs I exported. This time I highlighted the significant numbers the questions asked for. I used the same sample data I used for HW1. I did some extra coding in SAS for this project to get the quintiles to separate the values of player_height and player_weight into their respective positions.

```
ods rtf file="C:\Users\lb943\Box\HW2_LukeBeebe_lb943_nbaplayers" style=journal;
/* import data */
proc import out=all_seasons
datafile="C:\Users\lb943\Box\all_seasons.csv"
dbms=csv
replace;
getnames=YES;
run;
/* same simple random sample from HW1 */
proc surveyselect data=all_seasons
out=sample
method=srs
sampsize=30
seed=123;
run;
/* avg_height=200.61 cm avg_weight=100.37 kg (from population data) */
data sample1;
set sample;
height1=player_height-200.61;
weight1=player_weight-100.37;
run;
/* t-test on player's heights and weights */
proc means data=sample1 n mean std t prt;
var height1 weight1;
run; /* We do not reject the null on either */
/* finds quintiles */
proc univariate data=sample;
var player_height;
output out=quintile_height
pctlpts = 0 20 40 60 80 100
pctlpre = Q_;
run;
proc univariate data=sample;
var player_weight;
output out=quintile_weight
pctlpts = 0 20 40 60 80 100
pctlpre = Q_;
run;
proc print data=quintile_height;
run;
proc print data=quintile_weight;
run;
/* uses quintiles to seperate player's positions based on height and weight */
data sample2;
```

```
set sample;
if player_height>=209.55 then position=5;
if 203.2<=player_height<209.55 then position=4;
if 198.12<=player_height<203.2 then position=3;
if 193.04<=player_height<198.12 then position=2;
if player_height<193.04 then position=1;
run;
data sample3;
set sample;
if player_weight>=115.212 then position=5;
if 107.048<=player_weight<115.212 then position=4;
if 98.2027<=player_weight<107.048 then position=3;
if 92.3060<=player_weight<98.2027 then position=2;
if player_weight<92.3060 then position=1;
run;
proc print data=sample2;
var player_name position;
run;
proc print data=sample3;
var player_name position;
run;
/* frequency charts based on bins I made above */
proc freq data=sample2;
tables position;
run;
proc freq data=sample3;
tables position;
run;
/* interesting the quintiles didn't work out for player_height,
might've messed up somewhere */
proc corr data=sample;
var player_height;
with player_weight;
run;
/* strong, positive correlation between the variables at R^2=0.71655 */
ods rtf close;
```

The SAS System

11:25 Sunday, April 02, 2023

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	ALL_SEASONS
Random Number Seed	123
Sample Size	30
Selection Probability	0.002438
Sampling Weight	410.16667
Output Data Set	SAMPLE

The SAS System

The MEANS Procedure

Variable	*N*	*Mean*	*Std Dev*	*t Value*	*Pr > |t|*
height1	30	-0.1193333	8.4882438	-0.08	0.9392
weight1	30	2.8221800	12.5345611	1.23	0.2274

The SAS System

The UNIVARIATE Procedure
Variable: player_height

Moments			
N	30	*Sum Weights*	30
Mean	200.490667	*Sum Observations*	6014.72
Std Deviation	8.48824377	*Variance*	72.0502823
Skewness	-0.5496373	*Kurtosis*	-0.1254856
Uncorrected SS	1207984.68	*Corrected SS*	2089.45819
Coeff Variation	4.23373512	*Std Error Mean*	1.5497342

Basic Statistical Measures			
Location		*Variability*	
Mean	200.4907	*Std Deviation*	8.48824
Median	200.6600	*Variance*	72.05028
Mode	193.0400	*Range*	33.02000
		Interquartile Range	12.70000

Note: The mode displayed is the smallest of 2 modes with a count of 5.

Tests for Location: Mu0=0				
Test		*Statistic*	*p Value*	
Student's t	*t*	129.371	*Pr > |t|*	<.0001
Sign	*M*	15	*Pr >= |M|*	<.0001
Signed Rank	*S*	232.5	*Pr >= |S|*	<.0001

Quantiles (Definition 5)	
Level	*Quantile*
100% Max	213.36
99%	213.36
95%	210.82
90%	210.82
75% Q3	205.74
50% Median	200.66
25% Q1	193.04
10%	190.50
5%	182.88

The SAS System

The UNIVARIATE Procedure
Variable: player_height

Quantiles (Definition 5)	
Level	*Quantile*
1%	180.34
0% Min	180.34

Extreme Observations			
Lowest		*Highest*	
Value	*Obs*	*Value*	*Obs*
180.34	12	210.82	7
182.88	21	210.82	15
187.96	4	210.82	19
193.04	27	210.82	30
193.04	25	213.36	16

The SAS System

The UNIVARIATE Procedure
Variable: player_weight

Moments			
N	30	*Sum Weights*	30
Mean	103.19218	*Sum Observations*	3095.7654
Std Deviation	12.5345611	*Variance*	157.115223
Skewness	0.10939109	*Kurtosis*	-0.5890755
Uncorrected SS	324015.122	*Corrected SS*	4556.34145
Coeff Variation	12.146813	*Std Error Mean*	2.28848729

Basic Statistical Measures			
Location		*Variability*	
Mean	103.1922	*Std Deviation*	12.53456
Median	100.6974	*Variance*	157.11522
Mode	117.9339	*Range*	52.16308
		Interquartile Range	19.05086

Tests for Location: Mu0=0				
Test		*Statistic*	*p Value*	
Student's t	*t*	45.09187	*Pr > |t|*	<.0001
Sign	*M*	15	*Pr >= |M|*	<.0001
Signed Rank	*S*	232.5	*Pr >= |S|*	<.0001

Quantiles (Definition 5)	
Level	*Quantile*
100% Max	131.5417
99%	131.5417
95%	117.9339
90%	117.9339
75% Q3	114.3052
50% Median	100.6974
25% Q1	95.2543
10%	86.6361
5%	83.9145
1%	79.3786
0% Min	79.3786

The SAS System

The UNIVARIATE Procedure
Variable: player_weight

Extreme Observations

Lowest		*Highest*	
Value	*Obs*	*Value*	*Obs*
79.3786	21	117.934	3
83.9145	12	117.934	7
86.1825	4	117.934	15
87.0897	8	117.934	23
89.8112	24	131.542	10

The SAS System

Obs	*Q_0*	*Q_20*	*Q_40*	*Q_60*	*Q_80*	*Q_100*
1	180.34	193.04	198.12	203.2	209.55	213.36

The SAS System

Obs	*Q_0*	*Q_20*	*Q_40*	*Q_60*	*Q_80*	*Q_100*
1	79.3786	92.3060	98.2027	107.048	115.212	131.542

The SAS System

Obs	player_name	position
1	Jerry Stackhouse	3
2	Marty Conlon	5
3	Danny Fortson	3
4	Charlie Ward	1
5	Chris King	4
6	Tom Hammonds	4
7	Herb Williams	5
8	Eldridge Recasner	2
9	Keith Van Horn	4
10	Jahidi White	4
11	Jerry Stackhouse	3
12	Chucky Atkins	1
13	Brian Cardinal	4
14	James White	3
15	Tim Duncan	5
16	JaVale McGee	5
17	O.J. Mayo	2
18	Andre Iguodala	3
19	Kevin Garnett	5
20	Terrence Jones	4
21	Darren Collison	1
22	Kyle Korver	3
23	Lavoy Allen	4
24	Matthew Dellavedo	2
25	David Nwaba	2
26	Jarrell Brantley	2
27	Talen Horton-Tuck	2
28	Udonis Haslem	4
29	Paul Watson	3
30	Santi Aldama	5

The SAS System

Obs	player_name	position
1	Jerry Stackhouse	3
2	Marty Conlon	4
3	Danny Fortson	5
4	Charlie Ward	1
5	Chris King	2
6	Tom Hammonds	3
7	Herb Williams	5
8	Eldridge Recasner	1
9	Keith Van Horn	5
10	Jahidi White	5
11	Jerry Stackhouse	3
12	Chucky Atkins	1
13	Brian Cardinal	4
14	James White	1
15	Tim Duncan	5
16	JaVale McGee	4
17	O.J. Mayo	2
18	Andre Iguodala	2
19	Kevin Garnett	4
20	Terrence Jones	4
21	Darren Collison	1
22	Kyle Korver	2
23	Lavoy Allen	5
24	Matthew Dellavedo	1
25	David Nwaba	3
26	Jarrell Brantley	4
27	Talen Horton-Tuck	3
28	Udonis Haslem	3
29	Paul Watson	2
30	Santi Aldama	2

The SAS System

The FREQ Procedure

position	Frequency (height)	Percent	Cumulative Frequency	Cumulative Percent
1	3	10.00	3	10.00
2	6	20.00	9	30.00
3	7	23.33	16	53.33
4	8	26.67	24	80.00
5	6	20.00	30	100.00

The SAS System

The FREQ Procedure

position	Frequency (weight)	Percent	Cumulative Frequency	Cumulative Percent
1	6	20.00	6	20.00
2	6	20.00	12	40.00
3	6	20.00	18	60.00
4	6	20.00	24	80.00
5	6	20.00	30	100.00

The SAS System

The CORR Procedure

1 With Variables:	player_weight
1	*Variables:*	player_height

Simple Statistics

Variable	*N*	*Mean*	*Std Dev*	*Sum*	*Minimum*	*Maximum*
player_weight	30	103.19218	12.53456	3096	79.37860	131.54168
player_height	30	200.49067	8.48824	6015	180.34000	213.36000

Pearson Correlation Coefficients, N = 30
Prob > |r| under H0: Rho=0

	player_height
player_weight	0.71655 <.0001