The SAS System

10:12 Sunday, April 23, 2023

The SURVEYSELECT Procedure

Luke Beebe
HW4 SAS

My code with its comments is posted below. The following pages are the outputs I exported. I highlighted the answers to the questions asked for. I used the same sample data I used for HW1.

```
ods rtf file="C:\Users\lb943\Box\HW4_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_
seasonsout=sample
method=srs
sampsize=30
seed=123;
run;
/* uses quintiles to seperate player's positions beased on height and weight */
data sample2;
set sample;
if player_height>=209.55 then height_position=5;
if 203.2<=player_height<209.55 then height_position=4;
if 198.12<=player_height<203.2 then height_position=3;
if 193.04<=player_height<198.12 then height_position=2;
if player_height<193.04 then height_position=1;
if player_weight>=115.212 then weight_position=5;
if 107.048<=player_weight<115.212 then weight_position=4;
if 98.2027<=player_weight<107.048 then weight_position=3;
if 92.3060<=player_weight<98.2027 then weight_position=2;
if player_weight<92.3060 then weight_position=1;
run;
/* create cross tab, chisq pval=0.0419 statistically significant relationship at a=0.05 */
proc freq data=sample2;
tables height_position*weight_position / chisq;
run;
/* ANOVA table, pval<.0001 the means of 'size' are not equal at a=0.05 */
data sample3;
set sample;
if player_height>=205 then size='L';
if 195<=player_height<205 then size='M';
```

```
if player_height<195 then size='S';
run;
proc glm data=sample3;
class size;model player_height = size;
run;
/* paired t-test, pval<.0001 the means of player_height and player_weight are not equal */
proc ttest data=sample3 alpha=0.05;
paired player_height*player_weight;
run;
/* scatter plot, pval<.0001 r=0.7166 player_weight and player_height have a statistically
significant linear relationship */
proc corr data=sample3 plots=scatter(nvar=all);
var player_weight player_height;
run;
ods rtf close;
```

The SAS System

10:12 Sunday, April 23, 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 FREQ Procedure

Table of height_position by weight_position

height_position	weight_position					
Frequency Percent Row Pct Col Pct	1	2	3	4	5	Total
1	3 10.00 100.00 50.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	3 10.00
2	2 6.67 33.33 33.33	1 3.33 16.67 16.67	2 6.67 33.33 33.33	1 3.33 16.67 16.67	0 0.00 0.00 0.00	6 20.00
3	1 3.33 14.29 16.67	3 10.00 42.86 50.00	2 6.67 28.57 33.33	0 0.00 0.00 0.00	1 3.33 14.29 16.67	7 23.33
4	0 0.00 0.00 0.00	1 3.33 12.50 16.67	2 6.67 25.00 33.33	2 6.67 25.00 33.33	3 10.00 37.50 50.00	8 26.67
5	0 0.00 0.00 0.00	1 3.33 16.67 16.67	0 0.00 0.00 0.00	3 10.00 50.00 50.00	2 6.67 33.33 33.33	6 20.00
Total	6 20.00	6 20.00	6 20.00	6 20.00	6 20.00	30 100.00

Statistics for Table of height_position by weight_position

Statistic	DF	Value	Prob
Chi-Square	16	26.9643	0.0419
Likelihood Ratio Chi-Square	16	29.4608	0.0210
Mantel-Haenszel Chi-Square	1	12.3694	0.0004
Phi Coefficient		0.9481	
Contingency Coefficient		0.6880	
Cramer's V		0.4740	

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 30

The SAS System

The GLM Procedure

Class Level Information

Class	*Levels*	*Values*
size	3	L M S

Number of Observations Read	30
Number of Observations Used	30

The SAS System

The GLM Procedure

Dependent Variable: player_height

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	2	1752.435400	876.217700	70.20	<.0001
Error	27	337.022786	12.482325		
Corrected Total	29	2089.458187			

R-Square	*Coeff Var*	*Root MSE*	*player_height Mean*
0.838703	1.762193	3.533033	200.4907

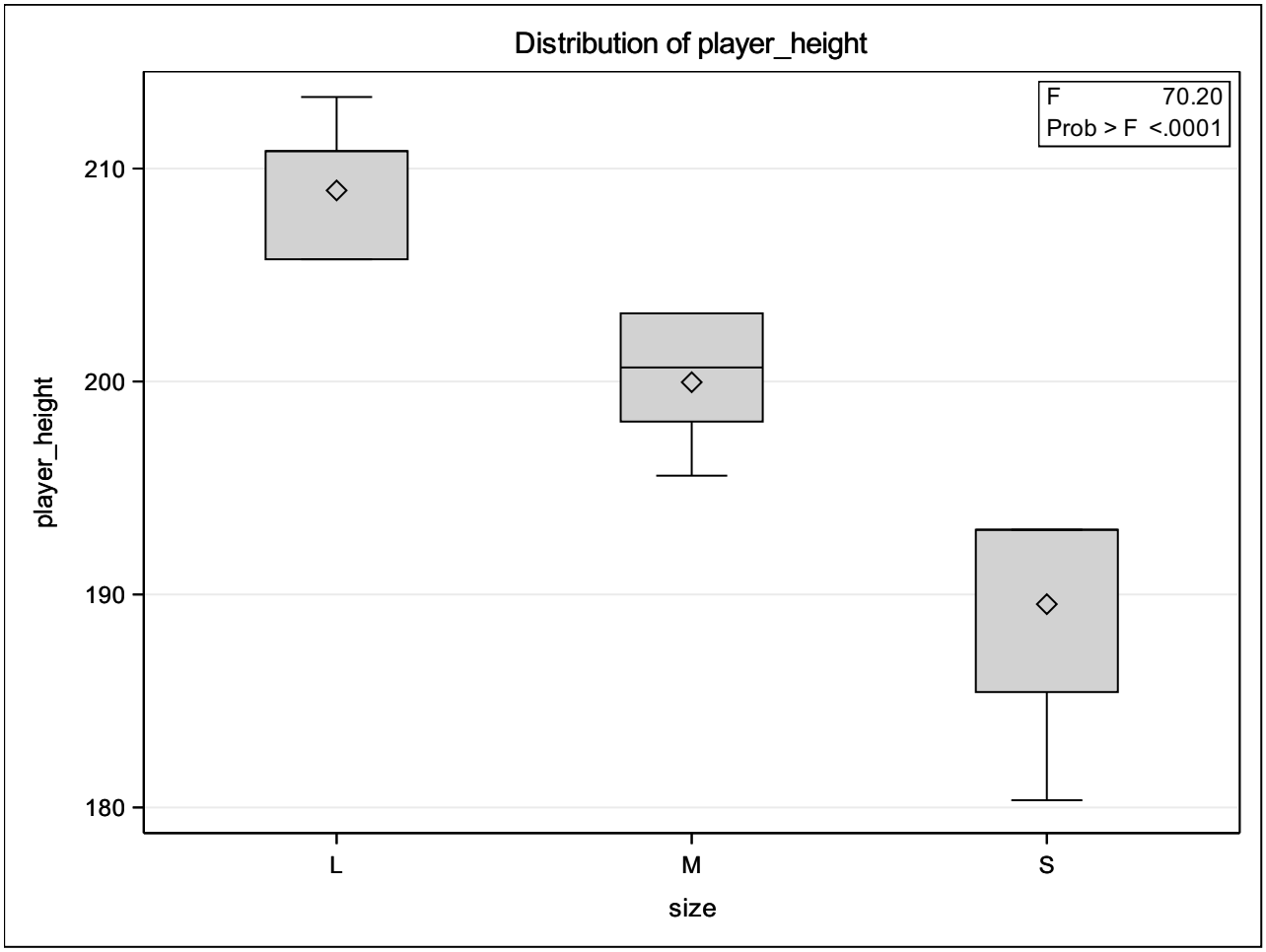
Source	*DF*	*Type I SS*	*Mean Square*	*F Value*	*Pr > F*
size	2	1752.435400	876.217700	70.20	<.0001

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
size	2	1752.435400	876.217700	70.20	<.0001

The SAS System

The GLM Procedure

Dependent Variable: player_height



The SAS System

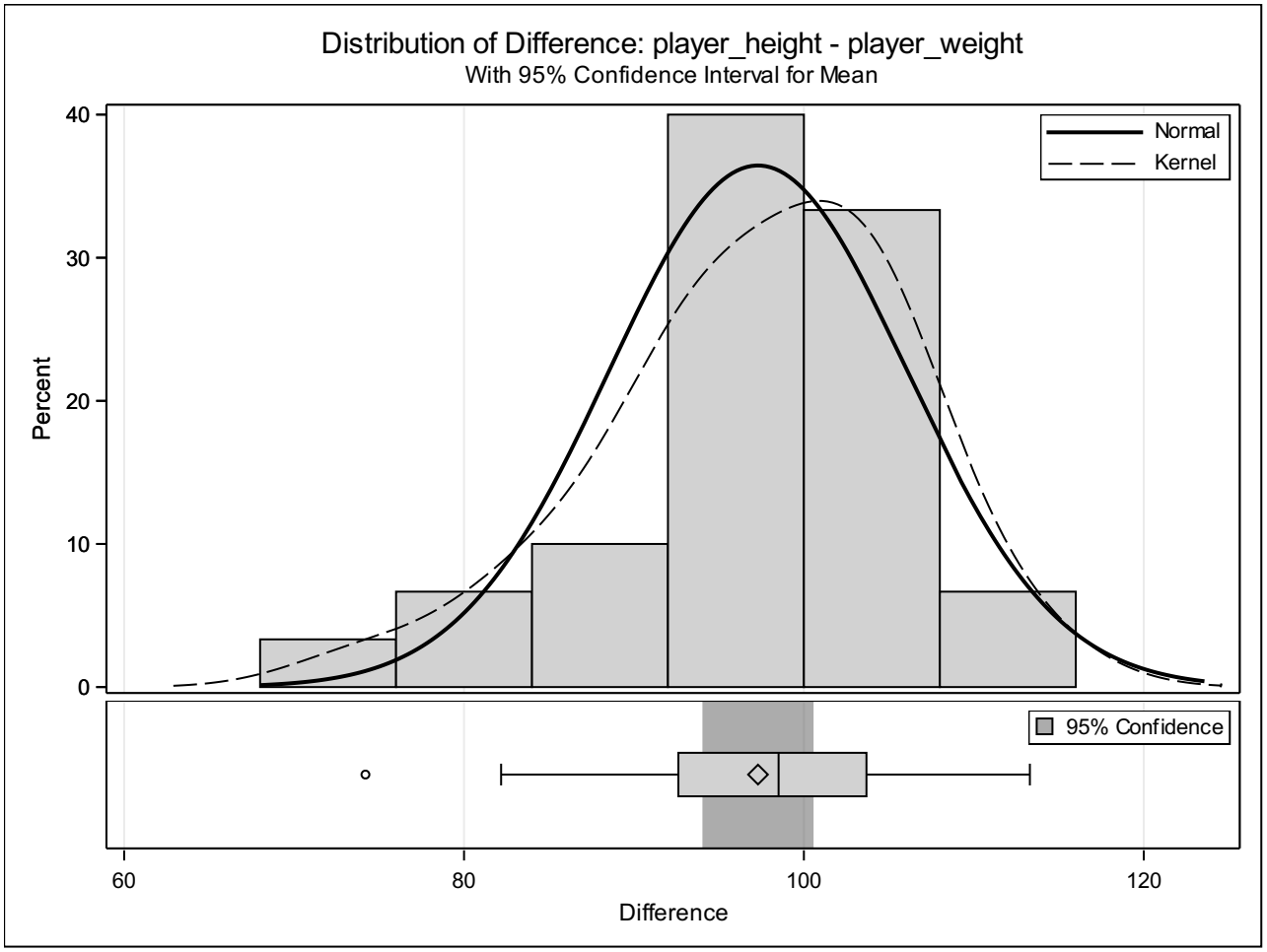
The TTEST Procedure

Difference: player_height - player_weight

N	*Mean*	*Std Dev*	*Std Err*	*Minimum*	*Maximum*
30	97.2985	8.7572	1.5988	74.1983	113.3

Mean	*95% CL Mean*		*Std Dev*	*95% CL Std Dev*	
97.2985	94.0285	100.6	8.7572	6.9743	11.7724

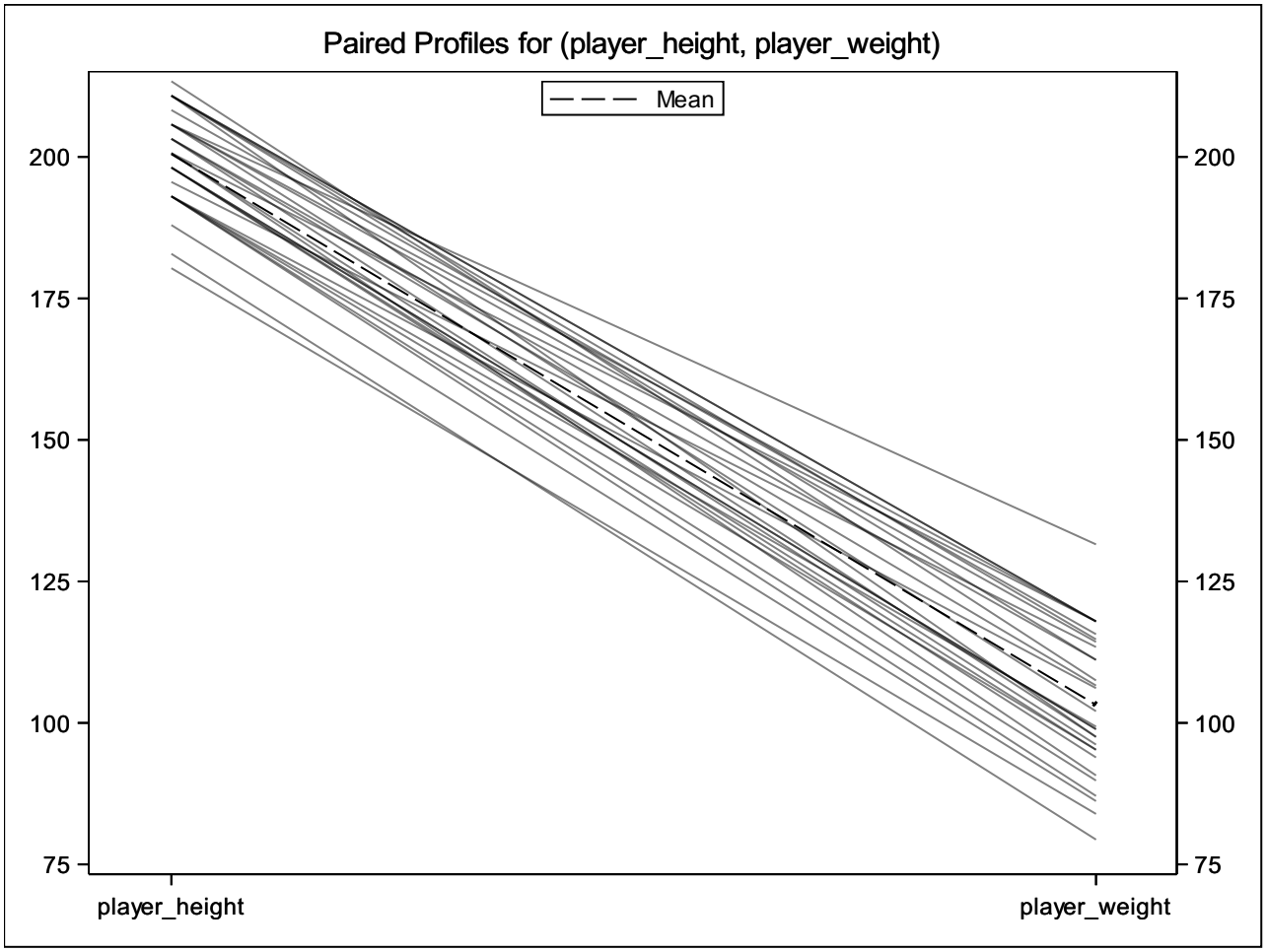
DF	*t Value*	*Pr > |t|*
29	60.86	<.0001



The SAS System

The TTEST Procedure

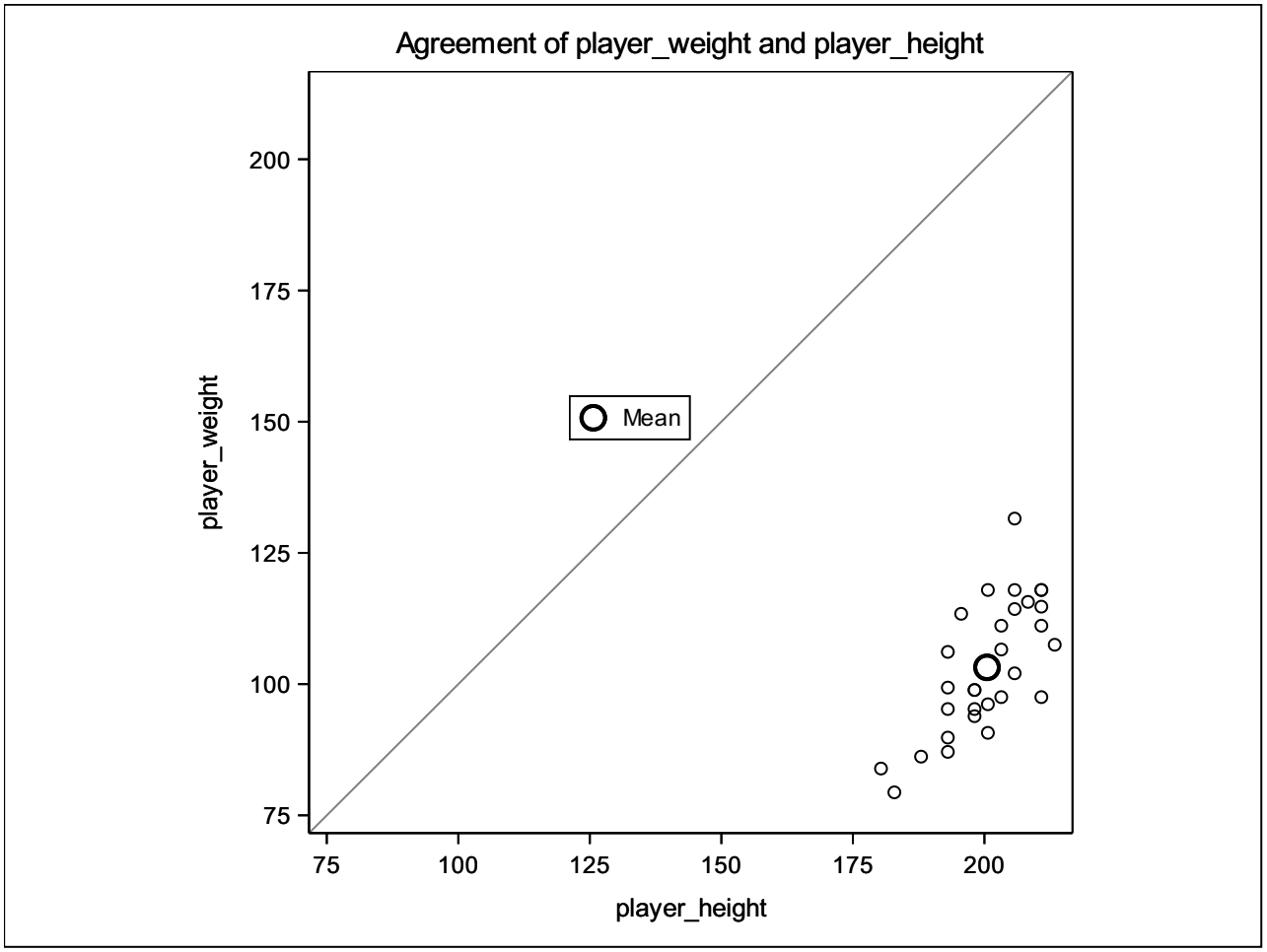
Difference: player_height - player_weight



The SAS System

The TTEST Procedure

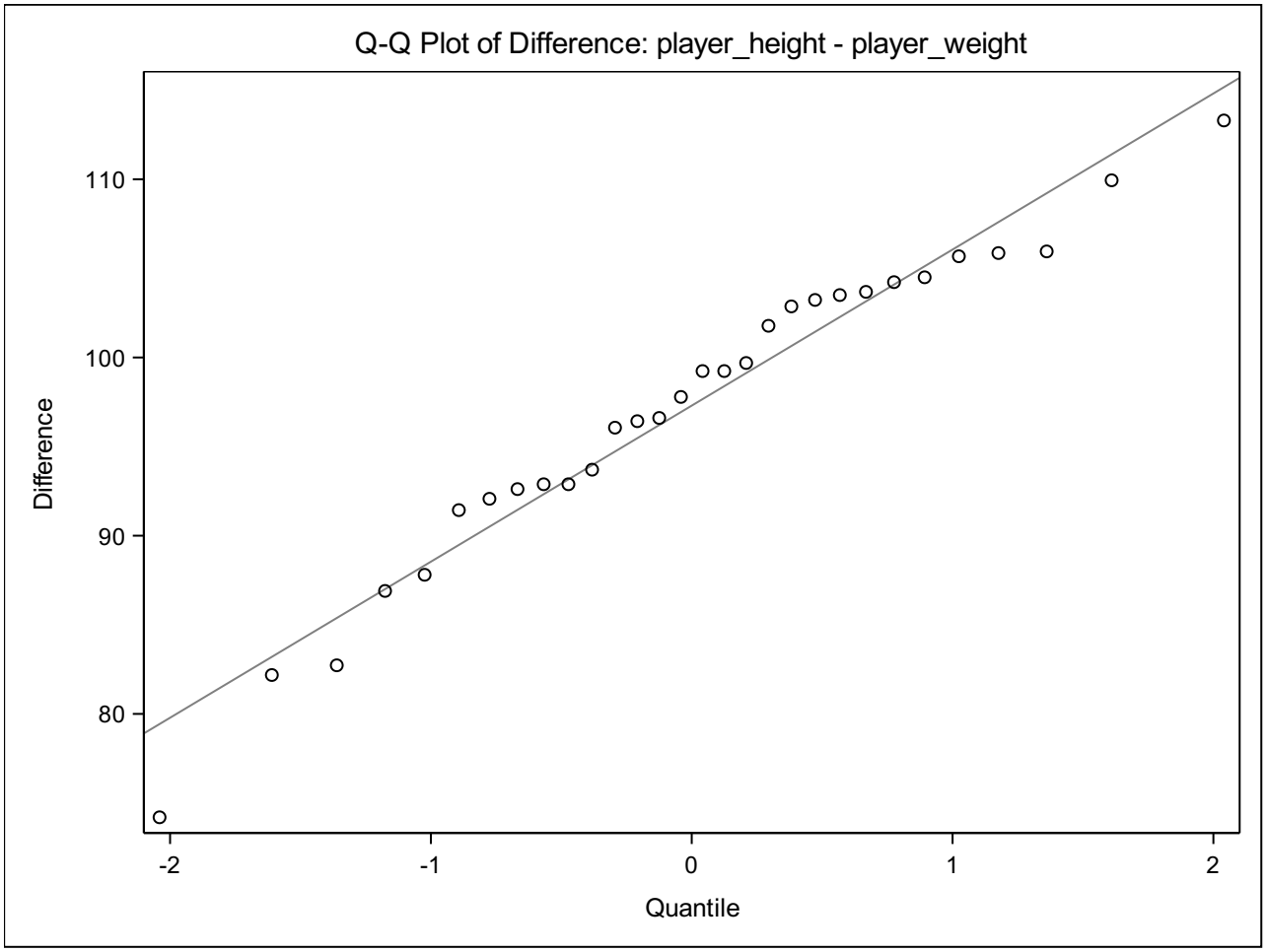
Difference: player_height - player_weight



The SAS System

The TTEST Procedure

Difference: player_height - player_weight



The SAS System

The CORR Procedure

2 Variables:	player_weight 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_weight	*player_height*
player_weight	1.00000	0.71655 <.0001
player_height	0.71655 <.0001	1.00000

The SAS System

The CORR Procedure