NORTHEASTERN UNIVERSITY

360 Huntington Ave., Boston, Massachusetts 02115

COURSE: ENGINEERING PROBABILITY AND STATISTICS (IE6200)

PROJECT TITLE

Statistical analysis of the variation of shooting accuracy of NBA players with their heights

Team members

Anish Kumar Musunuru

Neha Singh

Rashmi Nambiar Pappinisseri Puthenveettil

Data Analytics Engineering, College of Engineering, NEU

Under the guidance of

Nizar Zaarour, PhD

Assistant Teaching Professor

Supply Chain and Information Management Group

D'Amore-McKim School of Business, NEU

Table of Contents

Executive Summary

Basketball has always been considered as a "game of height" and short people have always been discouraged from pursuing the sport. The purpose of this study is to find the relationship between the heights of players in the NBA and their shooting accuracy. As an analyst, the idea is to find out how important height is for succeeding in basketball and to verify the claim that one must be tall in order to do well. However, it must be noted that this study only helps us find out the success in shooting and not in any other aspect of the sport.

The intended audiences for this report are the current/aspiring basketball players, basketball coaches, talent scouts as well as the fans of the sport. The objective of the project is to verify the relationship between the heights of basketball players and their shooting accuracy with the help of Statistical Methods. The sample data used for the project has been taken from the NBA website. The sample size is 45 and is taken based on the minimum number of field goals attempted, in this case 800. So only the players who have attempted to score more than 800 times have been considered for the sample. Only the 2019-20 season statistics have been collected.

For this project, the heights of the players (given in centimeters) are taken as the independent variables (x) with the shooting accuracy (given in percent) as the dependent variables (y). The shooting accuracy is denoted as FG or FG% in this report. The relationship between the two are analyzed using Simple Linear Regression. The analysis shows the presence of a positive non-zero slope indicating the presence of a positive linear relationship of moderate strength. Hypothesis Testing (t-test and f-test) is used to determine whether this relationship is significant. Both the t-test and the f-test showed that the relationship between the heights and shooting accuracy is significant.

In addition to the tests above, a Chi-Squared Goodness-of-Fit test and a Lack of Fit test have been conducted to test how good the model is. Both tests yielded positive results indicating that this model is good. This, however, does not depict the entire story. The confidence interval for this model is good, but the prediction interval is very large. This means that the model is not very good at predicting. The coefficient of determination (R^2) has a low value of 0.29 indicating that there is more unexplained variability than explained variability. The standard error of estimate (SE) has a value 3.39 indicating a moderate accuracy prediction. This further proves that the predictions will be less precise.

Overall, the tests reveal that there is a linear and positive relationship between height and shooting accuracy. However, as mentioned before, it should be noted that shooting accuracy alone does not determine the overall success of a player and someone who is short should not be discouraged. The prediction interval, R^2 , and SE provide further reason to believe this. The major takeaway would be that height does have an impact on one's chances of making a shot, but it is not the only factor that will help in basketball.

Exploratory Data Analysis

The sample consists of 45 NBA players taken from the NBA official website.
Simple Linear Regression analysis is used to determine the relationship between the height of NBA Players in cm and their shooting accuracy percentage (FG). FG is short for "Field Goals" and FG% will be used to represent the shooting accuracy.

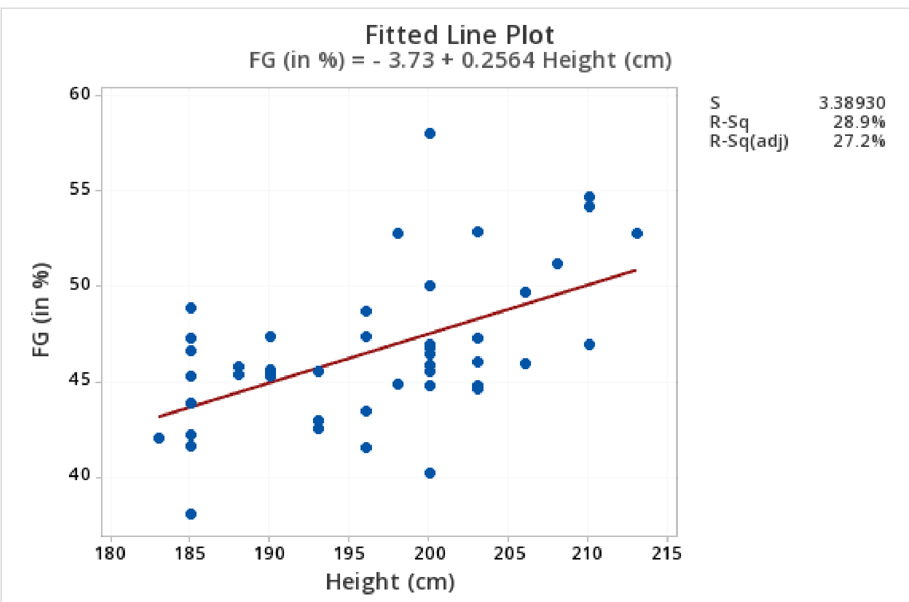
Descriptive statistics

Descriptive statistics are calculated to observe the distribution of the sample data set and summarize the quantitative description of the sample. The detailed descriptive statistics is given in Table A.1 of the Appendix. The standard deviation values for both the variables indicate variation in the data.
The boxplot diagram of height shows no outliers as in figure A.1 and data looks positively skewed as in figure A.3 and boxplot diagram of FG shows four outliers as in figure A.2 and data looks positively skewed as in figure A.4.

Scatter Plot Diagram

Scatterplot diagram identifies the relationships between the two variables, FG (in %) and Height (in cm). Figure 1 shows that the value of FG tends to increase as the value of Height increases. The straight line in red in figure 1 is a trend line that is designed to come as close as possible to all the data points. The trend line has a positive slope, which shows a positive relationship between X and Y.

Figure 1 The scatter plot of variable Y vs X with regression model



Residuals

Residuals are the leftover variations in the data after accounting for the model fit. These residuals can be seen as dots in figure 2:

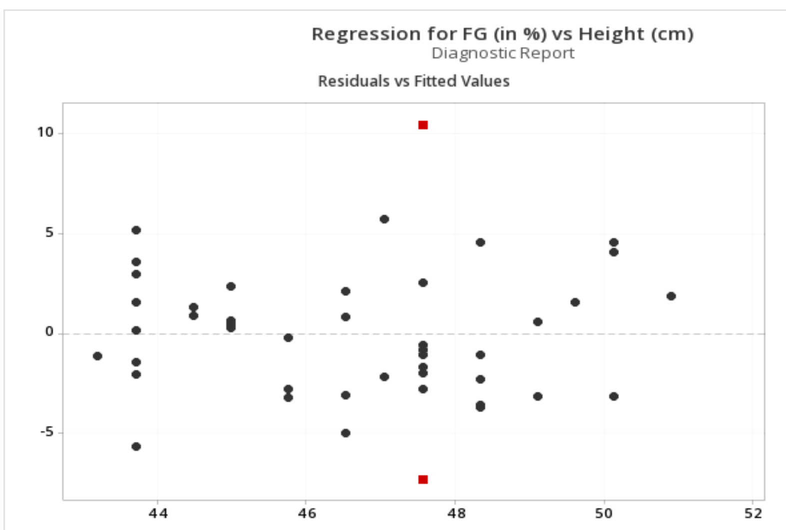
Data = Fit + Residual

Each observation has a residual. If an observation is above the regression line, then it is a positive residual. If the observation is below the line, then it is a negative residual. The value of the residual is the vertical distance from the observation to the line.

Scatter Plot of Residuals vs Fitted Values

The scatter plot with residuals on the *y* axis and the fitted values (estimated responses) on the *x* axis is shown in figure 2. The residual plot shows a fairly random pattern which indicates a linear model is a good fit to the sample data. They are randomly scattered around zero for the entire range of fitted values. Figure A.5 in the appendix verifies that the residuals are normally distributed and that they are close to the fitted straight line. Figure A.6 shows that residuals are not correlated.

Figure 2 Residuals vs Fitted Values



However, the figure 2 above shows two large residuals which are not well fit in the model and these might have an influence on the fitted line. These two large residuals are as follow:

1. Player with height(cm) = 200, Actual FG (in %) = 40.3, Predicted FG (in %) = 47.55 and Residuals = -7.25
2. Player with height(cm) = 200, Actual FG (in %) = 58.0, Predicted FG (in %) = 47.55 and Residuals = 10.45

Analysis of Variance (ANOVA)

Correlation Coefficient (R_{xy})

The value of the correlation coefficient signifies the strength and direction of the linear relationship between the X and Y values and is largely dependent on the number of samples chosen from the population.

In table A.8, it is seen that R_{xy} is 0.54 which signifies a positive correlation between height and accuracy of players considered. This is a significant value as it is not too close to 0 and thus denotes a significant linear relationship.

The regression line can be used to model the linear relationship between X and Y in the population. If the initial hypothesis stating that there is no significant correlation between accuracy and height is considered, it can be rejected with the mentioned value.

Regression Equation

The regression equation is used to represent the estimated relation between X and Y and in this case, it is given as **FG** (%) **= -3.7 + 0.2564*height** (cm). As seen in Table A.2 in the column called **'Coef'**, the value of the coefficients can be given as follows:

b_0 = -3.7 and b_1 = 0.2564 (approximated to 0.26).

b_0 represents the intercept of the regression plot and b_1 indicates the slope. The positive slope is indicative of the fact that there is a positive relationship between accuracy and height, i.e., increase in height shows a certain amount of increase in the shot accuracy in the sample taken.

SSE, SSR and SST

SSE denotes the sum of squared errors or sum of squared residuals. This value is used to represent the discrepancy between the sample data and the estimated model. SSE is 493.96 as shown in Table A.4 in the **'Adj SS'**(adjusted sum of squares) column against the row **'Error'**. This comprises 71.14% of the total dispersion of accuracy value around its mean and is the unexplained variation. A significant deviation is seen from the actual value and a lower SSE value would have given a better fit. Although we see a positive correlation between the variables, it is not significant enough to highlight the existing relation, as the best fit line would require a low SSE.

SSR is the sum of squared regression which indicates the variation in the Y values as indicated by the model with respect to the mean Y value of the sample. The value is seen in Table A.4 as 200.40 in the **'Adj SS'** column against the row '**Regression'**. SSE and SSR add up to give the SST value which is the sum of squared total. This is seen in Table A.4 as 694.35 in the **'Adj SS'** column against the row **'Total'**.

Coefficient of determination (R^2)

R^2 denotes the proportion of variability as explained by the fitted model. It is also given as SSR/SST. From Table A.3, we see that R^2 is 28.86% (approximated to 0.29). This indicates a small value because although there is a variation seen between X and Y values, it is not significant enough to be considered a linear relationship.

This value also indicates that an increase in height by 1 cm causes increase in accuracy only by a factor of 0.29, so there is no best trend in values of height in the sample considered that can give the best accuracy rate in players and that the data points are farther away from fitted values.

A comparatively better R^2 value would be obtained with a greater number of samples; however, it would not have improved the regression fitting considerably. Therefore, the current fit is not a good enough fit to prove the significance of the relationship between height and accuracy rate.

Chi-Square Goodness-of-Fit Test

Hypothesis:

- Null Hypothesis: H_0 : There is no significant difference between observed FG and expected FG values.

- Alternate Hypothesis: H_a : There is a significant difference between observed FG and expected FG values.

At significance level of 5%, $\alpha = 0.05$ and the P-value from Table A.5 in appendix is 1.0.
After comparing the P value with α, it is seen that P > α (1.0 > 0.05)

Decision: Fail to reject the Null Hypothesis H_0.
Conclusion: It is concluded that there is no significant difference between observed FG and expected FG and hence there is insufficient evidence to reject the null hypothesis.

Mean Square Error (MSE)

MSE is the mean of the squared errors and its value is 11.487 as seen in Table A.4 in column **'Adj MS'** against row **'Error'**. It is the square root of SSE and is expected to have a very small value.

Standard error of the estimate (SE Coeff or SE)

SE is the square root of MSE which is a measure of accuracy prediction and is 3.3893, as shown in Table A.3 represented by **'S'**. This value is an estimator of the standard deviation of shot accuracy rate.

Estimated standard deviation of b_1 (S_{b1})

S_{b1} indicates the standard error of regression slope which is the distance that the observed values deviate from the regression line on an average, and its value is 0.0614 as shown in Table A.2 in column **'SE Coef'** against row **'Height(cm)'**.

Hypothesis tests

The sample slope b_1 is the point estimate of unknown population slope β_1.
The hypothesis test is conducted to test whether the linear relationship in the sample is strong enough to model the relationship for the population.

t-Test

Hypothesis:

- Null Hypothesis: H_0 : β_1 = 0
  H_0 : The population slope is equal to zero. Therefore, there is not a significant linear relationship between x and y in the population.

- Alternate Hypothesis: H_a : $\beta_1 \neq 0$
  H_a : The population slope is not equal to zero. Therefore, there is a significant linear relationship between x and y in the population.

At significance level of 5%, $\alpha = 0.05$ for 2 tailed-test, the critical value $t_{\alpha/2}$ using t-distribution table for degrees of freedom(df) = 43 (considering df = n-2, where n is the sample size 45) is given as follows:

$t_{\alpha/2}$ = $t_{0.025}$ = 2.017

Critical region: t > 2.017

Test statistic can be calculated, using the following equation and values given in the table A.8 in appendix, as follows:

$$t = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{b_1}{S_{b_1}} \text{ (according to null hypothesis } \beta_1 = 0 \text{)}$$

$$t = \frac{b_1}{S_{b_1}} = \frac{0.2564}{0.0614} = 4.18$$

After comparing the Critical value with Test Statistics: t > $t_{\alpha/2}$ (4.18 > 2.017)

Decision: Reject the Null Hypothesis H_0.

Conclusion: We conclude that there is sufficient evidence that there is a significant linear relationship between the Height(x) and FG(y) of the population.

F-Test:

Hypothesis:

- Null Hypothesis: H_0 : β_1 = 0
  H_0 : The population slope is equal to zero. Therefore, there is not a significant linear relationship between x and y in the population.

- Alternate Hypothesis: H_a : $\beta_1 \neq 0$
  H_a : The population slope is not equal to zero. Therefore, there is a significant linear relationship between x and y in the population.

At significance level of 5%, $\alpha = 0.05$ for 2 tail-test, the critical value of f_α using f-Distribution table for $v1 = 1$ and $v2$ = 43 (n-2) degree of freedom was $f_{0.05}$(1,43) = 4.067

Critical region: f > 4.067

Test statistic can be calculated using the following equation and values from Table A.8 :

f = $\frac{MSR}{MSE}$ = $\frac{200.4}{11.49}$ = 17.44

After comparing the Critical value with Test Statistics: f > f_α (17.44 > 4.067)

Decision: Reject the Null Hypothesis H_0.

Conclusion: We conclude that there is sufficient evidence that there is a significant linear relationship between the Height(x) and FG(y) of the population. The same conclusion was obtained by the t-test as well.

Confidence and Prediction Intervals for Forecasted Values

Confidence Interval

95% confidence interval means that there is a 95% probability that the linear regression line of the population will lie within the confidence interval of the regression line calculated from sample data. The confidence interval consists of the space between the two green dotted lines given in figure 3 below.
E(y_p) = the mean or expected value of the dependent variable y_p.

And, forecasted value $\widehat{y_p}$ = b_0 + $b_1 x_p$, for random variable x_p = 198 cm from sample is given as $\widehat{y_p}$ = -3.73 + 0.2564*198 = 47.0372

Hence, the 95% confidence interval for the E(y_p) is given as:

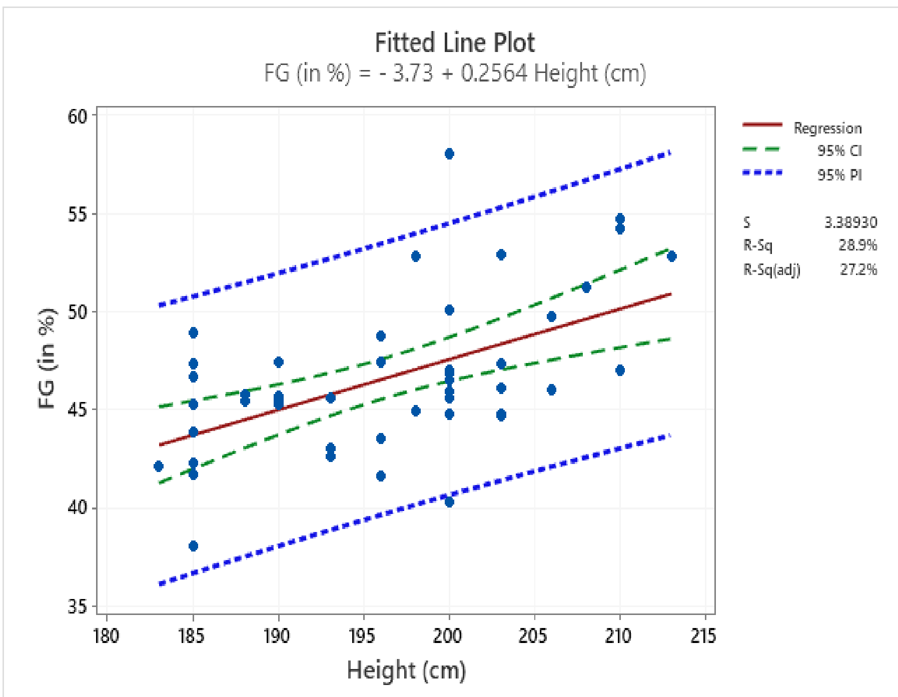
E(y_p) = $\widehat{y_p} \pm (t_{\alpha/2})\,(S_e)(\sqrt{\frac{1}{n} + \frac{(x_p - \underline{x})^2}{\sum \ (x - \underline{x})^2}})$ (using values from table A.8)

= 47.0416 $\pm$ 2.017* 3.39* $\sqrt{\frac{1}{45} + \frac{2.6896}{3048.31}}$

= 47.0416 $\pm$ 1.0391

The 95% confidence interval for the E(y_p), for given x_p(198 cm) is (46.0026,48.0807).

Figure 3. Fitted line Plot



In the figure 3 above, it is seen that the prediction interval (95% PI, in blue dotted line) is wider than the confidence interval (95% CI, in green dotted line).

Prediction interval

95% prediction interval for y_p means that there is a 95% probability that the real value of y (in the population) corresponding to new observation x_p is within this interval.
Here x_p = 198 cm

y_p = $\widehat{y_p} \pm (t_{\alpha/2})$ (S_e)($\sqrt{1 + \frac{1}{n} + \frac{(x_p - \underline{x})^2}{\sum \quad (x - \underline{x})^2}}$) (using values from table A.8)

= 47.0416 $\pm$ 2.017* 3.39* $\sqrt{1 + \frac{1}{45} + \frac{2.6896}{3048.31}}$

= 47.0416 $\pm$ 6.9137.

The 95% prediction interval for the y_p for x_p(198 cm) is (40.1279,53.9553).

Lack of fit Test:

Lack of fit test formally tests how well the model fits the data. It tells us whether there is a difference between random error and the model error. The model error is due to lack of fit of the model. Table A.4 in the appendix shows the pure error, model error, and the difference between them which is the lack of fit. If the majority of the error is due to bad fit and not due to some random error, it suggests that our model has a lack of fit and there is no reasonable relationship in the model.

Hypothesis:

- Null Hypothesis: H_0 : The relationship assumed in the model is reasonable, i.e., there is no lack of fit in the model.

- Alternate Hypothesis: H_a : The relationship assumed in the model is not reasonable, i.e., there is a lack of fit in the model.

At significance level of 5%, $\alpha = 0.05$ and P-Value from table A.4 is 0.936

After comparing the P-Value with α: P > α (0.936 > 0.05)

Decision: Fail to reject the Null Hypothesis H_0.
Conclusion: We conclude that there is not sufficient evidence that there is a lack of fit in the model.

Conclusion

The data collected from the NBA website was analyzed with multiple statistical methods. With height as the independent variable (x) and shooting accuracy as the dependent variable (y), a linear regression analysis was done to find their relationship. The slope was found to be 0.26 which is a positive non-zero value indicating a positive linear relationship between the 2 variables. The regression line is shown in figure 1. The coefficient of determination (R^2) has a low value of 0.29 indicating that there is more unexplained variability than explained variability. This suggests that the predictions will be less precise. A Chi-Squared goodness-of-fit test showed that the model was a good fit at 5% significance level. The hypothesis testing (t-test and F-test) showed that the positive linear relationship between height and shot accuracy discovered in the regression analysis was significant. The prediction interval is large indicating that the model is not ideal for prediction. The lack-of-fit test revealed that there is insufficient evidence to say that the model is a bad fit.

The conclusions made from this data are: there is significant evidence to say that the heights of the players and the shooting accuracy have a positive linear relationship; the model is not great for predictions; although height is a contributing factor for better accuracy, it is however not the only factor. Height does help in basketball but there is no reason to discourage shorter players as height does not guarantee success. The residual plot (figure 2) shows 2 outliers (marked in red). Both points are for players of the same height, which is 200 cm. One player has a shot accuracy of 40.3% while the other has 58%. This gap shows that being tall does not guarantee a better accuracy. Other factors like time spent on practicing shots, experience playing matches, and vision can make a difference on accuracy.

The findings in this project are not very surprising and most of it have been anticipated. The linear relationship between height and shot accuracy is expected and the accuracy being affected by other factors is expected as well. Having a larger sample might have given a higher R^2 value and reduced the overall deviation in the data. The sample data used for this project is limited to players who attempted over 800 shots. This increased the reliability of the shot accuracy stat, but some valuable data may have been lost because of this as larger samples tend to give less deviation. Lower sample size also means lesser points to verify the relationship. The overall shape of the graph can change by introducing more points.

To approach the shot accuracy in a different way, testing the relationship between matches played and shot accuracy will be an interesting way to do it. One would expect a player to get better with every game and it would be interesting to verify that. To achieve this, it would be advisable to consider the matches played in previous seasons. This will help differentiate between experienced players and new players. Experience is the key factor in this study, and it is the games played in their lifetime that counts more than the number of games played in the current season.

APPENDIX

All tables and graphs are as obtained with the sample data on MINITAB tool, unless otherwise specified.

Table A.1 Descriptive statistics of the variables FG% and Height (cm)

Statistics

Variable	Mean	StDev	Variance	CoefVar	Minimum	Q1	Median	Q3	Maximum
FG (in %)	46.620	3.972	15.781	8.52	38.100	44.750	45.900	48.050	58.000
Height (cm)	196.36	8.32	69.28	4.24	183.00	189.00	198.00	203.00	213.00

Variable	IQR	Mode	N for Mode	Skewness
FG (in %)	3.300	44.8, 45.3, 45.6, 47	2	0.70
Height (cm)	14.00	200	9	0.05

The data contain at least five mode values. Only the smallest four are shown.

IQR – Interquartile range; Q1- first quartile; Q3 – third quartile

Table A.2. Coefficients of the simple linear regression equation

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-3.7	12.1	-0.31	0.759	
Height (cm)	0.2564	0.0614	4.18	0.000	1.00

Table A.3. R-squared value

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.38930	28.86%	27.21%	22.53%

Table A.4. ANOVA output table

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	200.40	200.395	17.44	0.000
Height (cm)	1	200.40	200.395	17.44	0.000
Error	43	493.96	11.487		
Lack-of-Fit	11	62.36	5.669	0.42	0.936
Pure Error	32	431.59	13.487		
Total	44	694.35			

SS – Sum of squares; MS – Mean squared value, DF- degrees of freedom

Table A.5. Chi-Square Test Table for FG%

Chi-Square Test

N	DF	Chi-Sq	P-Value
2097.9	44	14.8939	1.000

The above table shows chi-square value, P-value for FG% with degrees of freedom (DF) 44 and sum of all FG(%) values given by 'N'.

Table A.6. Confidence interval (CI) and Prediction Interval (PI) with the best fit value for 95% CI

Prediction

Fit	SE Fit	95% CI	95% PI
47.0416	0.515233	(46.0026, 48.0807)	(40.1279, 53.9553)

Table A.7. Sample data and ANOVA calculation (tool used – MS Excel)

	A	B	C	D	E	F	G	H	I	J	K
1	Height (cm)	FG (in %)	x-xbar	(x-xbar)^2	y-ybar	(y-ybar)^2	(x-xbar)(y-ybar)	yhat	(y-yhat)^2	(yhat-ybar)^2	(y-ybar)^2/ybar
2	185	38.1	-11.36	128.95	-8.52	72.59	96.75	43.71	31.45	8.48	1.56
3	200	40.3	3.64	13.28	-6.32	39.94	-23.03	47.55	52.63	0.87	0.86
4	196	41.6	-0.36	0.13	-5.02	25.20	1.78	46.53	24.29	0.01	0.54
5	185	41.7	-11.36	128.95	-4.92	24.21	55.87	43.71	4.03	8.48	0.52
6	183	42.1	-13.36	178.37	-4.52	20.43	60.37	43.20	1.20	11.73	0.44
7	185	42.3	-11.36	128.95	-4.32	18.66	49.06	43.71	1.98	8.48	0.40
8	193	42.6	-3.36	11.26	-4.02	16.16	13.49	45.76	9.98	0.74	0.35
9	193	43	-3.36	11.26	-3.62	13.10	12.15	45.76	7.62	0.74	0.28
10	196	43.5	-0.36	0.13	-3.12	9.73	1.11	46.53	9.17	0.01	0.21
11	185	43.9	-11.36	128.95	-2.72	7.40	30.89	43.71	0.04	8.48	0.16
12	203	44.7	6.64	44.15	-1.92	3.69	-12.76	48.32	13.13	2.90	0.08
13	203	44.8	6.64	44.15	-1.82	3.31	-12.09	48.32	12.42	2.90	0.07
14	200	44.8	3.64	13.28	-1.82	3.31	-6.63	47.55	7.59	0.87	0.07
15	198	44.9	1.64	2.70	-1.72	2.96	-2.83	47.04	4.59	0.18	0.06
16	185	45.3	-11.36	128.95	-1.32	1.74	14.99	43.71	2.53	8.48	0.04
17	190	45.3	-6.36	40.39	-1.32	1.74	8.39	44.99	0.10	2.66	0.04
18	188	45.4	-8.36	69.82	-1.22	1.49	10.19	44.48	0.85	4.59	0.03
19	190	45.5	-6.36	40.39	-1.12	1.25	7.12	44.99	0.26	2.66	0.03
20	193	45.6	-3.36	11.26	-1.02	1.04	3.42	45.76	0.03	0.74	0.02
21	200	45.6	3.64	13.28	-1.02	1.04	-3.72	47.55	3.82	0.87	0.02
22	190	45.7	-6.36	40.39	-0.92	0.85	5.85	44.99	0.50	2.66	0.02
23	188	45.8	-8.36	69.82	-0.82	0.67	6.85	44.48	1.75	4.59	0.01
24	200	45.9	3.64	13.28	-0.72	0.52	-2.62	47.55	2.74	0.87	0.01
25	206	46	9.64	93.02	-0.62	0.38	-5.98	49.09	9.57	6.11	0.01
26	203	46.1	6.64	44.15	-0.52	0.27	-3.46	48.32	4.94	2.90	0.01
27	200	46.5	3.64	13.28	-0.12	0.01	-0.44	47.55	1.11	0.87	0.00
28	185	46.7	-11.36	128.95	0.08	0.01	-0.91	43.71	8.95	8.48	0.00
29	200	46.8	3.64	13.28	0.18	0.03	0.66	47.55	0.57	0.87	0.00
30	210	47	13.64	186.17	0.38	0.14	5.18	50.12	9.72	12.24	0.00
31	200	47	3.64	13.28	0.38	0.14	1.38	47.55	0.31	0.87	0.00
32	185	47.3	-11.36	128.95	0.68	0.46	-7.72	43.71	12.90	8.48	0.01
33	203	47.3	6.64	44.15	0.68	0.46	4.52	48.32	1.05	2.90	0.01
34	190	47.4	-6.36	40.39	0.78	0.61	-4.96	44.99	5.81	2.66	0.01
35	196	47.4	-0.36	0.13	0.78	0.61	-0.28	46.53	0.76	0.01	0.01
36	196	48.7	-0.36	0.13	2.08	4.33	-0.74	46.53	4.71	0.01	0.09
37	185	48.9	-11.36	128.95	2.28	5.20	-25.89	43.71	26.95	8.48	0.11
38	206	49.7	9.64	93.02	3.08	9.49	29.70	49.09	0.37	6.11	0.20
39	200	50.1	3.64	13.28	3.48	12.11	12.68	47.55	6.48	0.87	0.26
40	208	51.2	11.64	135.59	4.58	20.98	53.33	49.61	2.54	8.91	0.45
41	213	52.8	16.64	277.04	6.18	38.19	102.86	50.89	3.66	18.21	0.82
42	198	52.8	1.64	2.70	6.18	38.19	10.16	47.04	33.16	0.18	0.82
43	203	52.9	6.64	44.15	6.28	39.44	41.73	48.32	20.94	2.90	0.85
44	210	54.2	13.64	186.17	7.58	57.46	103.42	50.12	16.66	12.24	1.23
45	210	54.7	13.64	186.17	8.08	65.29	110.25	50.12	20.99	12.24	1.40
46	200	58	3.64	13.28	11.38	129.50	41.47	47.55	109.11	0.87	2.78

Table A.8. ANOVA calculation

xbar	ybar	Sx	Sy	Sxy	SST	SSE	SSR
420.41261	100.7182	8.32	3.97	17.76	694.35	493.96	200.4
MSE	MSR	Rxy	R squared	b0	b1	SE Coeff (Se)	Sb1
11.49	200.4	0.54	0.29	-3.73	0.26	3.39	0.0614

Figure A.1. Boxplot of Height

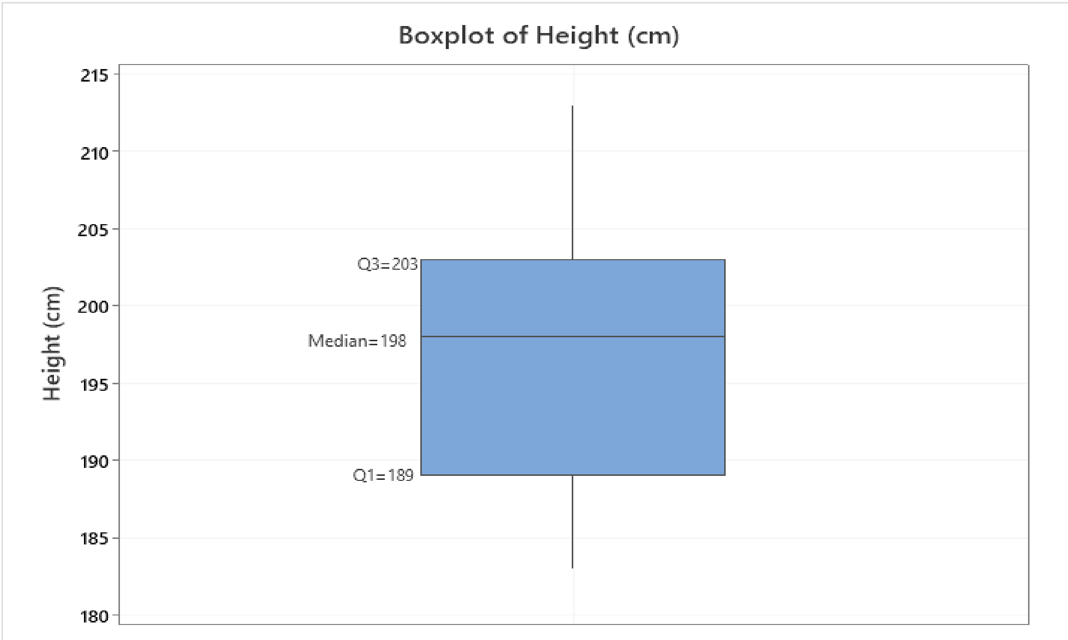


Figure A.2. Boxplot of FG% showing outliers at values 58,54.7 and 54.2

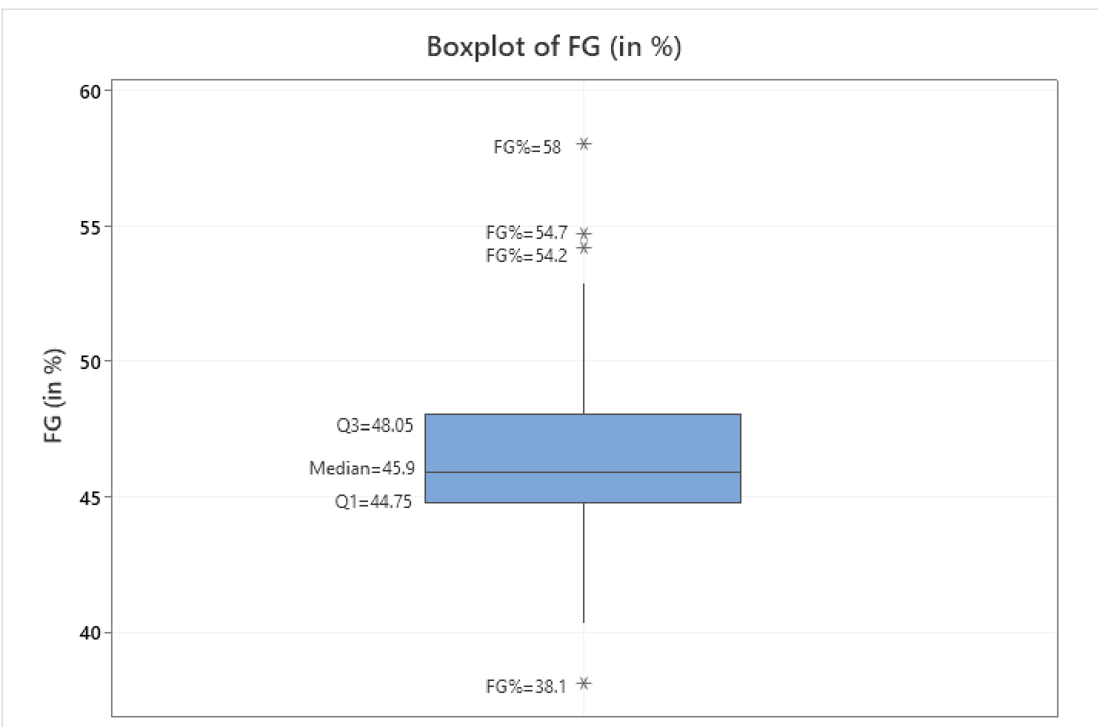


Figure A.3. Histogram of Height

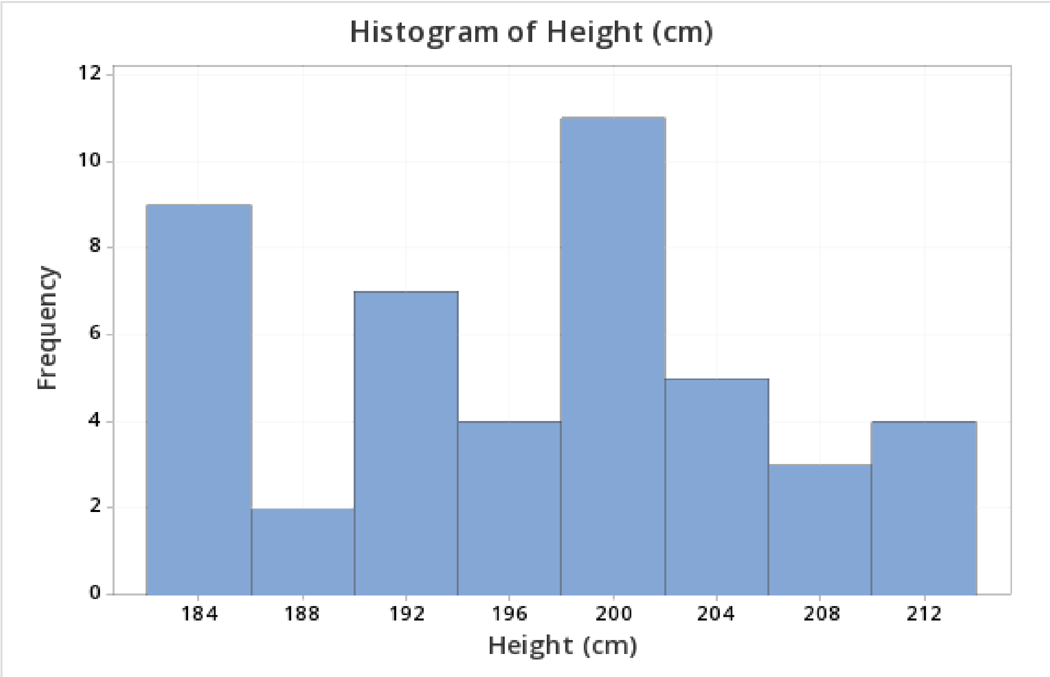


Figure A.4. Histogram of FG

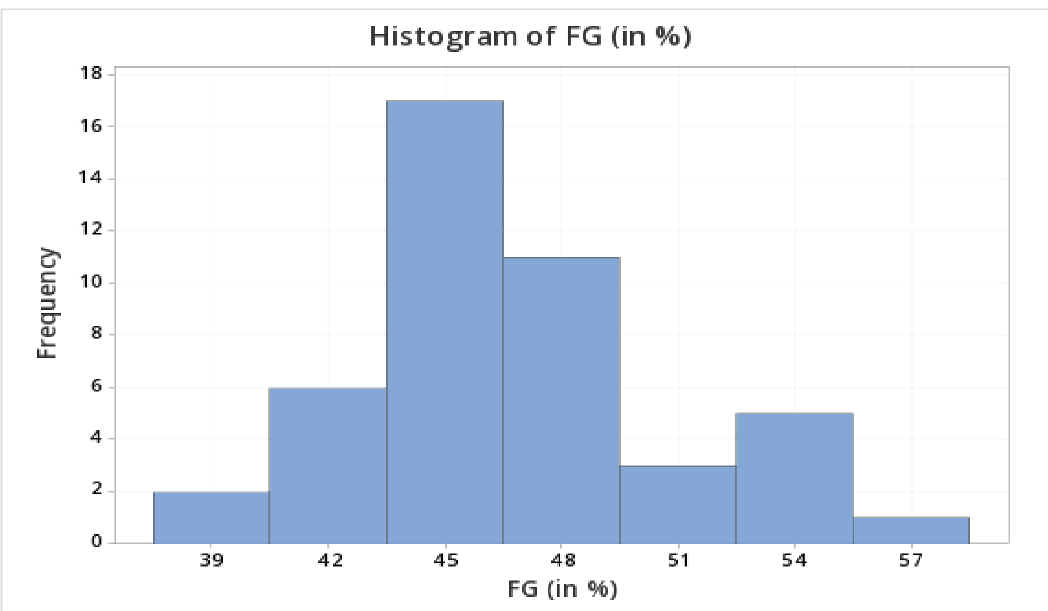


Figure A.5. Normal probability plot of residual

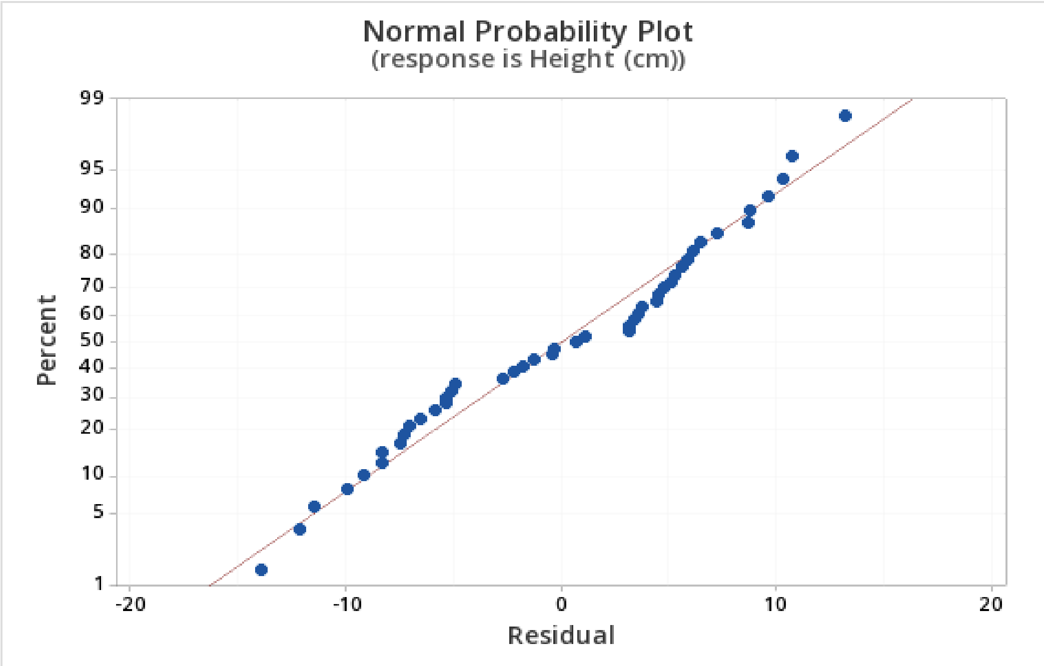
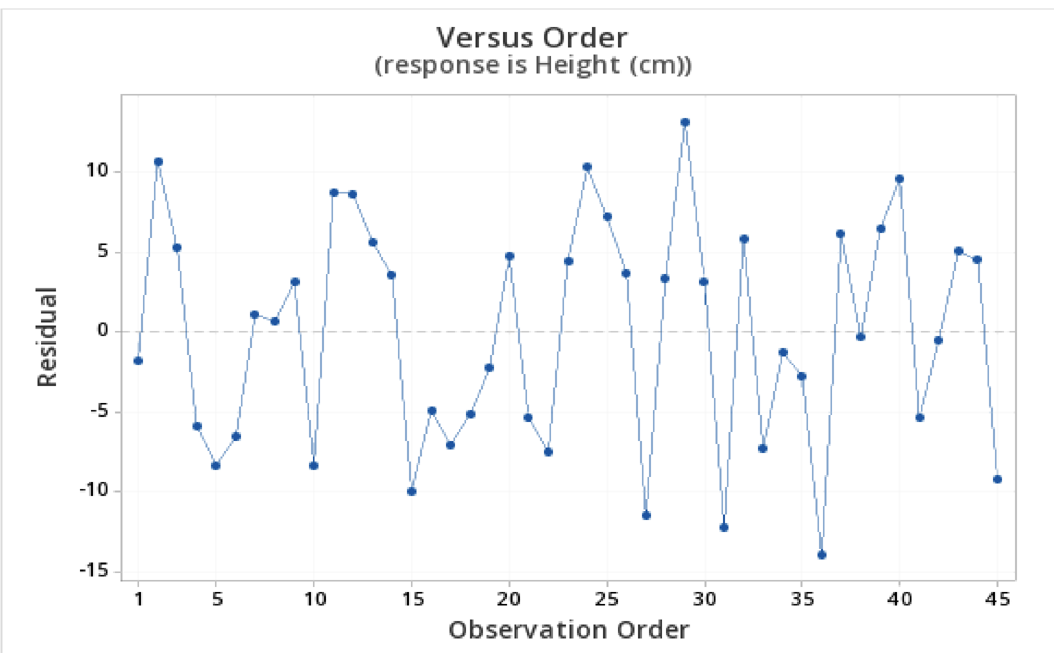


Figure A.6. Residual Vs Order plot



Data Sources

https://stats.nba.com/players/shooting/?Season=2019-20&SeasonType=Regular%20Season&PerMode=Totals

https://stats.nba.com/help/glossary/

https://stats.nba.com/players/bio/