



# American Journal of Biometrics & Biostatistics

## Research Article

# Statistical Analysis on Catheter Length Data Based on Patient Height & Weight without Gender, Age, or Health Condition -

**Darrell Robinson\***

*Department of Biomedical Engineering, Watson School of Engineering & Applied Science, State University of New York at Binghamton, New York*

**\*Address for Correspondence:** Darrell Robinson, Department of Biomedical Engineering, Watson School of Engineering & Applied Science, State University of New York at Binghamton, New York, Tel: +607-297-8198; ORCID: 0000-0003-3454-6741; E-mail: drobin21@binghamton.edu/drobinson289@gmail.com

**Submitted: 29 September 2019; Approved: 13 December 2019; Published: 16 December 2019**

**Cite this article:** Robinson D. Statistical Analysis on Catheter Length Data Based on Patient Height & Weight without Gender, Age, or Health Condition. American J Biom Biostat. 2019;3(1): 001-008.

**Copyright:** © 2019 Robinson D. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### ABSTRACT

A catheter is a thin tube made from medical grade materials that serve a broad range of functions, but mainly catheters are medical devices that can be inserted in the body to treat disease or perform surgical procedures. Catheters have been inserted into body cavities, ducts, or vessels to allow for drainage, administration of therapeutic fluids or gases, operational access for surgery. Catheters help perform tasks in various systems such as cardiovascular, urological, gastrointestinal, neurovascular, and ophthalmic systems. A dataset of 12 patients with varying “weights” and “heights” was recorded along with the lengths of their catheter tubes. This dataset was found from two revered statistical textbooks on linear regression and the Department of Scientific Computing at Florida State University. This dataset was not able to be linked to any particular clinical or experimental research studies, but the dataset can be used to help catheter manufacturers and medical professionals better decide on what particular catheter lengths to use for patients knowing only their height & weight. These research insights could be helpful to healthcare professionals that have patients with incomplete or no healthcare records to decide what catheter length to use. The main investigative inquiry that needed to be answered was how does patient weight & height influence catheter length together and separately? We conducted linear regression and other statistical analysis procedures in R program & Microsoft Excel and discovered that this data exhibited a quality called multicollinearity. With multicollinearity, all predictors (2 or more independent variables) are not significant in an all-encompassing linear aggression, but the predictors might be significant in their own individual linear regressions. Individual linear regression analyses were conducted for both patient height & weight to see how much they both contribute to varying catheter length. Patient weight was found to be more impactful than patient height in relationship to catheter length, even though height and weight are a classical example of multicollinearity predictors.

### INTRODUCTION

A catheter is a thin tube made from medical grade materials serving a broad range of functions. Catheters are medical devices that can be inserted in the body to treat diseases or perform a surgical procedure. By modifying the material or adjusting the way catheters are manufactured, it is possible to tailor catheters for cardiovascular, urological, gastrointestinal, neurovascular, and ophthalmic applications. Catheters can help drain urine from the urinary bladder, administer intravenous fluids & medication, perform angioplasty & angiography, administer oxygen, and even perform embryo transfer by inserting fertilized embryos from in vitro fertilization into the uterus.

A dataset of 12 random patients with varying “weights” and “heights” was recorded along with the lengths of their catheter tubes. This dataset was found on a dataset website of the Department of Scientific Computing at Florida State University [1], but it specifically came from two highly regarded statistics textbooks on linear regression, “Applied Linear Regression: Third Edition” [2] and “Mathematical Algorithms for Linear Regression” [3]. There is research out in the scientific world that uses the lengths of catheters to study and monitor disease, but the predominant variable chosen is patient height.

A great research application of this is, “Influence of fine-bore catheter length on infusion thrombophlebitis in peripheral intravenous nutrition: a randomized controlled trial” [4]. This research describes how previous studies concluded that the risk of thrombophlebitis associated with continuous infusion of Intravenous Nutrition (IVN) by peripheral veins decreases when the catheter length increased to 15 centimeters. The 15 centimeter catheters were used in place of the standard intravenous cannulas to better transport drugs and intravenous fluids to alleviate thrombophlebitis.

There are three questions that we wanted to investigate. How much does patient weight and patient height contribute to the catheter length when combined together? What is the relationship between weight and height in regards to catheter length? How much does weight and height contribute to catheter length individually? We hope to apply the results of this study to better help hospitals study, diagnose, and treat disease using catheters. Everyone is physically different, especially internally, and there are no one-size fits all length

for catheter tubes. It is assumed that the height of a patient is taken into more consideration than the weight of a patient when selecting the length of a catheter. Based on the statistical results, the insights can better help medical professionals determine what catheter length to use for a patient based only on catheter height and weight.

### METHODS

The data in figure 1 was collected from 12 random patients sample size was 12 patients (sample size) and the cardiac catheters were fed from a principal vein into the heart. The genders, age, and health status of the patients were not listed, but we could assume that the patients could possibly have cardiovascular impairments. The standard manufacturing information of the catheters in the dataset was not listed, so we could not figure out the absolute design specifications of the catheters. The original dataset was too small to validate any potential results and we could not find more data based connected to these conditions, so we used R to random generate more data. The data columns included the independent variables of height ( $X_1$ ) and weight ( $X_2$ ) and the dependent variable was catheter length (B).

I = Index/Patient number

A1 = Height (Inches)

A2 = Weight (Pounds)

B = Intercept/Minimal catheter length (Centimeters)

	Patient.Height	Patient.Weight	Catheter.Length
1	42.8	40	37
2	63.5	93.5	50
3	37.5	35.5	34
4	39.5	30	36
5	45.5	52	43
6	38.5	17	28
7	43	38.5	37
8	22.5	8.5	20
9	37	33	34
10	23.5	9.5	30
11	33	21	38
12	58	79	47

Figure 1: Dataset of patient height, patient weight, and catheter length.

C = Total catheter length (Centimeters)

We seek a regression model of the form:  $C = B + A1 * X1 + A2 * X2$

**Scatterplot visualization**

We created scatterplots with the fitted regression lines to see how spread out the dataset was, to see if there was a fitted linear regression line to help us measure any potential error and deviation, and if there were any extreme outliers contributing to the deviation of the data.

**Correlation test & QQ plot visualization**

We performed a correlation test by executing a correlation matrix to see if there were any significant correlation relationships between the variables and to figure out the amount of variability in one variable to another. To compliment the correlation analysis, we created QQ Plots to determine the probability distributions of each of the variables and see if they come from the same distribution.

**ANOVA test**

The One-Way Analysis of Variance (ANOVA) is a statistical test used in determining whether there are any differences that are statistically significant between three or more unrelated groups. This dataset has only has two groups which are really the variables of

patient weight and patient height. This does not meet the requirement of ANOVA needing three or more independent groups in R. We tried to perform a separate ANOVA analysis in R, but the software returned messages that there were not enough arguments to perform ANOVA. However, we were able to perform the ANOVA test in Microsoft Excel.

**Linear regression and multiple linear regression testing**

Figure 2 and figure 3 represent the programming code where we performed two linear regression tests for both patient weight’s separate relationship with catheter length and patient height’s separate relationship with catheter length. We also performed a multiple linear regression test to see how patient weight and patient height combined together influences catheter length. This test was especially needed due to not being able to conduct an ANOVA test.

**RESULTS AND DISCUSSION**

**Numerical summary**

The numerical summary data in figure 4 was obtained as the first step to evaluate the normality of the dataset. All three variables had the 12 data points which matches the number of total patients indicating that this dataset is normal.

```
> library(e1071, pos=15)

> numSummary(Dataset[,c("Patient.Height", "Patient.Weight")],
+ statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd  IQR  0% 25%  50%  75% 100%  n
Patient.Height 40.35833 11.93562  7.625 22.5  36 39.00 43.625 63.5 12
Patient.Weight 38.12500 26.05774 23.000  8.5  20 34.25 43.000 93.5 12

> numSummary(Dataset[, "Catheter.Length"], statistics=c("mean", "sd", "IQR",
+ "quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd  IQR  0% 25%  50%  75% 100%  n
36.16667  8.177945  6.25 20  33 36.5 39.25  50 12
```

Figure 2: Numerical summaries of patient height, weight, and catheter length.

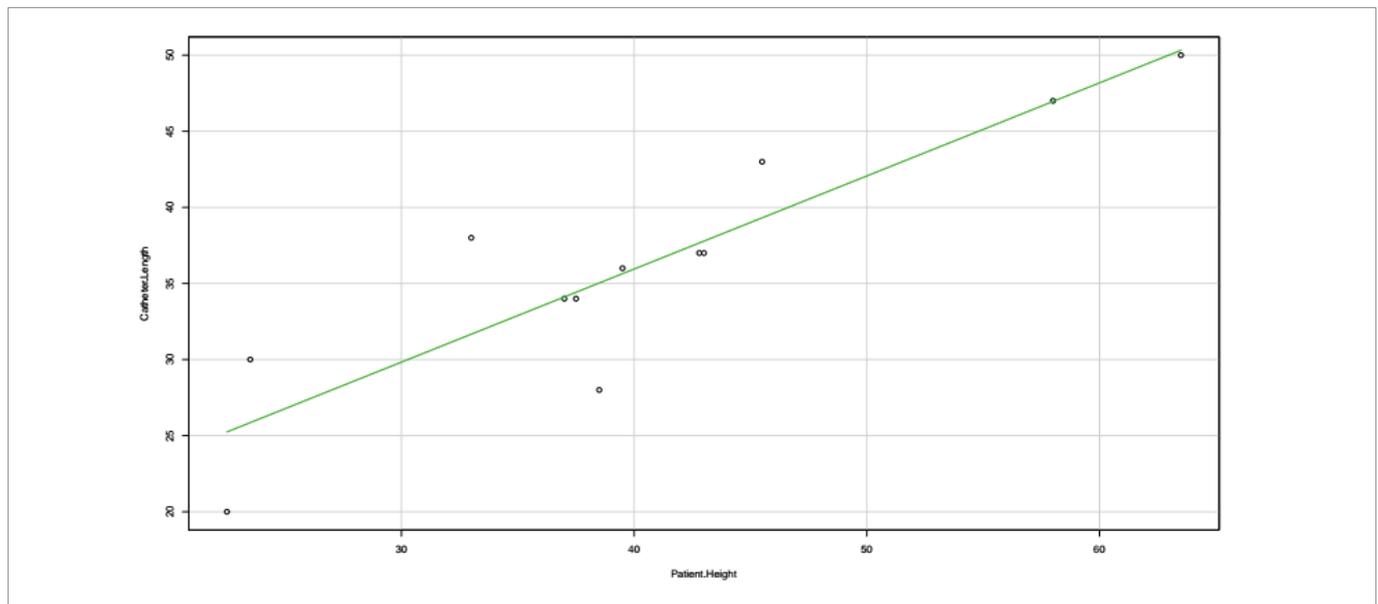


Figure 3: Scatterplot of the patient height & catheter length.

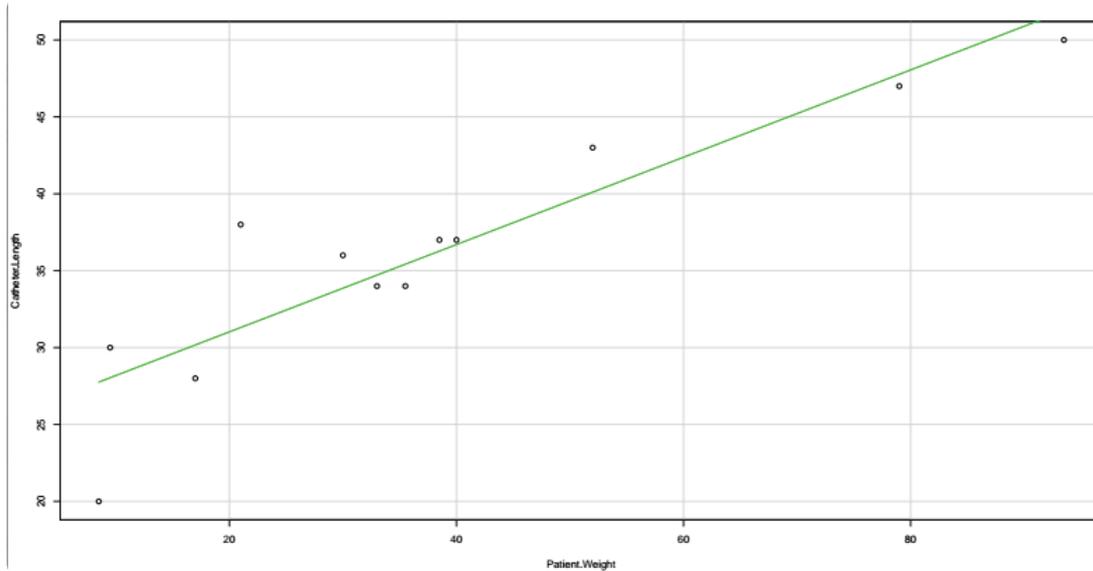


Figure 4: Scatterplot of the patient weight & catheter length.

```

Pearson correlations:
      Catheter.Length Patient.Height Patient.Weight
Catheter.Length      1.0000      0.8928      0.9045
Patient.Height       0.8928      1.0000      0.9611
Patient.Weight       0.9045      0.9611      1.0000

Number of observations: 12

Pairwise two-sided p-values:
      Catheter.Length Patient.Height Patient.Weight
Catheter.Length      <.0001      <.0001
Patient.Height <.0001      <.0001
Patient.Weight <.0001      <.0001

Adjusted p-values (Holm's method)
      Catheter.Length Patient.Height Patient.Weight
Catheter.Length      0.0001      0.0001
Patient.Height 0.0001      <.0001
Patient.Weight 0.0001      <.0001

      Patient.Height Patient.Weight Catheter.Length
Patient.Height      1.0000000      0.9237010      0.7970692
Patient.Weight      0.9237010      1.0000000      0.8181388
Catheter.Length     0.7970692      0.8181388      1.0000000
    
```

Figure 5: Correlation statistics.

**Scatterplot visualization**

According to both of the scatterplots in figure 5 and figure 6, both the correlations of patient height & patient weight with catheter length were slightly normal because of the graph line that fits the data pretty well. Both Patient height and Patient weight had the same number of outliers and errors. There was not much deviation from the regression line.

**ANOVA test**

We were able to perform the ANOVA test in Microsoft Excel and

we retrieved negative results. The sum of squares was 2,107.37. The mean of squares was also 2,107.37, and the degrees of freedom was 1. For the error, the sum of squares was 2881.61, the degrees of freedom was 10, and the mean of squares was 288.16. The total sum of squares was 735.66. The *F*-Test ratio values was quite low at 7.31 indicating that this analysis and model is not an adequate fit for prediction with this dataset.

**Correlation test & QQ plot visualization**

According to the correlation statistics & QQ-Plots presented by figure 7 and figure 8, there was not that much deviation for the

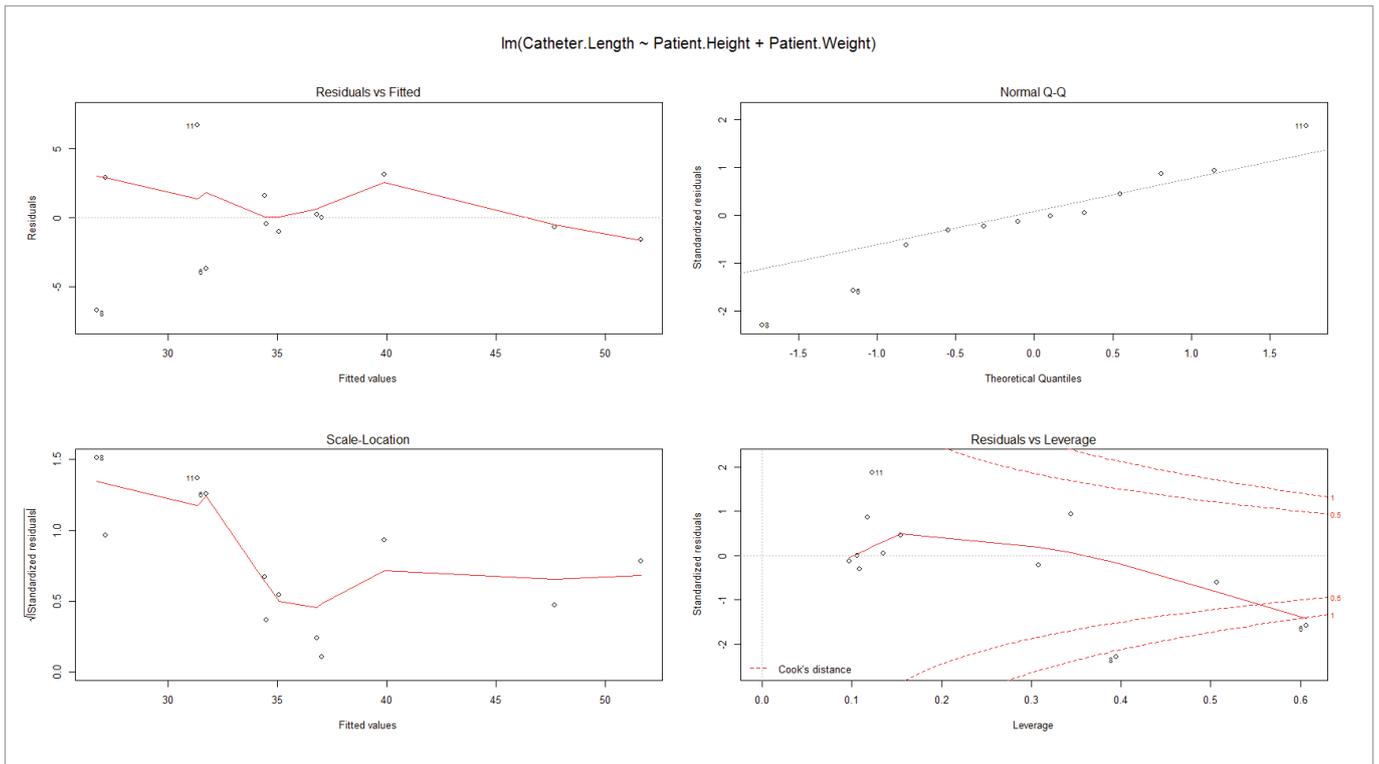


Figure 6: Normal QQ plot for catheter length.

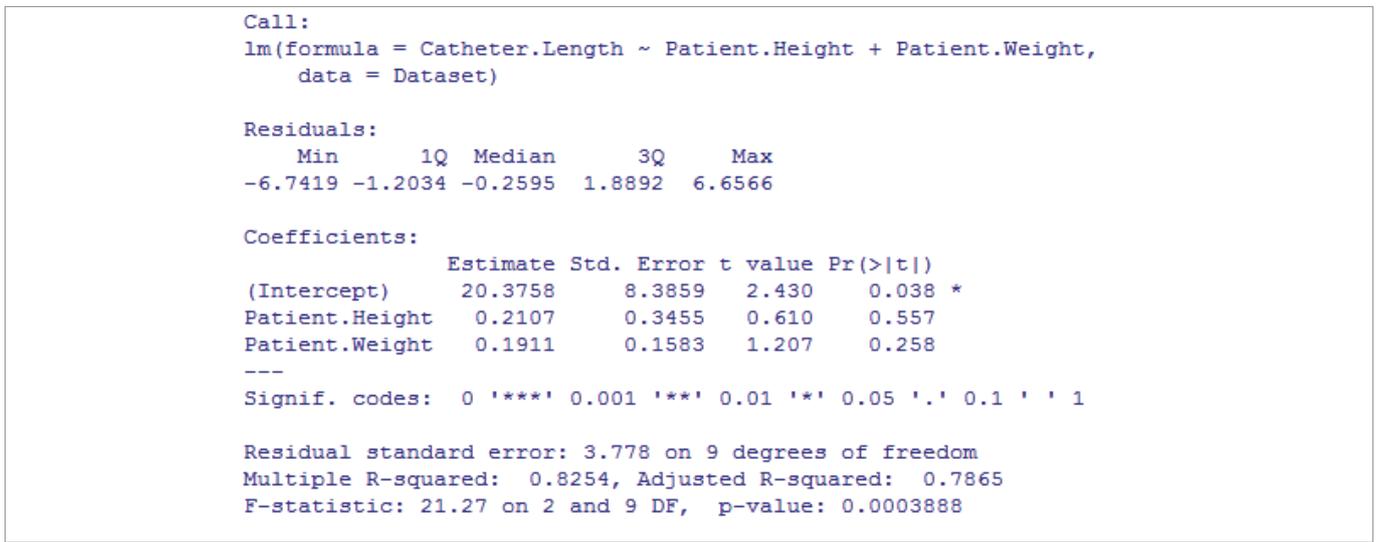


Figure 7: Multiple linear regression statistics - combined patient height & weight effect on catheter length.

correlation line. The main difference between the regression line for patient weight and the regression line for patient height was that the regression line for patient weight started a little higher than the regression line for patient height. According the residuals vs fitted graph, there was not a straight linear line which suggests that this dataset and model is abnormal and irregular. This also supports the result of the *F*-Test value in the ANOVA test.

**Multiple linear regression - patient height & weight combined effect on catheter length**

Regression Line:  $y = 20.3758 + 0.2107x_1 + 0.1911x_2$

Patient height and patient weight seem to work against each

other when combined together from the results of the multiple linear regression analysis displayed in figure 9. Neither patient height or weight were significant, but the intercept was significant. After further investigation, these results were due to a rare statistical quality called “multicollinearity”.

Multicollinearity is a phenomenon in which two or more predictor variables (ex: height & weight) in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a substantial degree of accuracy. In this situation the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. Multicollinearity does not reduce the predictive power of the entire



```
Call:
lm(formula = Catheter.Length ~ Patient.Height, data = Dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-7.0299 -0.6908 -0.2175  1.1908  6.3345

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.47898    4.09405   2.804  0.0187 *
Patient.Height  0.61171    0.09761   6.267  9.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.864 on 10 degrees of freedom
Multiple R-squared:  0.7971, Adjusted R-squared:  0.7768
F-statistic: 39.28 on 1 and 10 DF,  p-value: 9.295e-05
```

Figure 8: Linear regression of patient height's effect on catheter length.

```
Call:
lm(formula = Catheter.Length ~ Patient.Weight, data = Dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-7.7570 -1.5376 -0.2054  2.0043  6.6946

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  25.34409    1.92834  13.143 1.24e-07 ***
Patient.Weight  0.28387    0.04232   6.707 5.32e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.658 on 10 degrees of freedom
Multiple R-squared:  0.8181, Adjusted R-squared:  0.8
F-statistic: 44.99 on 1 and 10 DF,  p-value: 5.317e-05
```

Figure 9: Linear regression on patient weight on catheter length.

```
Call:
lm(formula = Catheter.Length ~ Patient.Height, data = Dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-7.0299 -0.6908 -0.2175  1.1908  6.3345

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.47898    4.09405   2.804  0.0187 *
Patient.Height  0.61171    0.09761   6.267  9.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.864 on 10 degrees of freedom
Multiple R-squared:  0.7971, Adjusted R-squared:  0.7768
F-statistic: 39.28 on 1 and 10 DF,  p-value: 9.295e-05
```

Figure 10: Linear regression of patient height's effect on catheter length.

model, it only affects calculations regarding the predictive power of the individual predictors. With this dataset, multicollinearity boosted the statistical significance of the predictor variables individually. A multiple regression model with correlated predictors can indicate how well the entire bundle of predictors predicts the outcome variable (catheter length in this case), but may not give valid results about any individual predictor, or about which predictors are redundant with respect to others [5-7].

Because of multicollinearity, patient height and patient weight statistically work against each other. When combined together the predictors are statistically insignificant. Neither predictor can contribute any adequate amount of variability. Individually, both predictors become redundant. The *F*-test determines how significant the predictors truly are [7].

According to the correlation test, patient height and patient weight between one another had a correlation coefficient value of 96.11%. Patient height and catheter length's relationship had a correlation coefficient value of 89.28% while patient weight and catheter length's relationship had a correlation coefficient value of 90.45%. According to the correlation coefficient of determination calculations, patient weight specifically accounted for 81.81% of the variability for Catheter Length and Patient height specifically accounted for 79.70% of the variability of the catheter length.

According to the multiple regression line formula, as the height increases by 1 unit, the catheter length increases 0.2107 cm. As the weight increases 1 unit, the catheter length increases by 0.1911 cm.

The multiple R-square value indicates that patient height and weight together account for 82.54% of the variability of the catheter length. Patient height accounts for 25.53% of that variability of catheter length and patient weight accounts for 23.15% of that variability of catheter length. Other possible predictors such as age, blood pressure, cholesterol level, tissue surface area, and body density for example could contribute to the other 33.86% of the variability of the catheter length. More data would need to be acquired in order to statistically analyze these new potential variables. Patient weight and patient height when combined together, within this particular dataset analysis, are not significant even though they still account for a big percentage of the variability of the catheter length.

The *F*-test value for the multiple linear regression model was 21.57 indicating that overall this model is not a good fit for the dataset population and it does not predict the total catheter length very well. The *p*-value that is associated with the *F*-Test value is significant, which says that this model is important for further study and development.

### Linear regression of patient height's effect on catheter length

Regression line of patient height's effect on catheter length:  $y = 11.47898 + 0.61171x$

The results in figure 10 illustrate that individually, patient height is significant. The *t*-value of the patient height (6.267) is less than the *t*-value of the patient weight (6.707). The *t*-value measures the size of the difference relative to the variation in the sample data. The greater the magnitude of the *t*-value, the greater the evidence against the null hypothesis. This provides further support that the null hypothesis of the patient height influencing catheter length more than patient weight in this dataset is incorrect.

The *p*-value of the patient height (9.3e-05) is greater than the *p*-value of the patient weight (5.32e-05) which indicates that the patient height is less significant than the patient weight on an individual basis. The patient height-intercept in this linear regression analysis (11.47898) is at least 2 times less than the patient weight-intercept (25.34409). An inference from this information could be that patient height does not increase the significance of the Intercept more than patient weight.

The multiple R-Square (0.7971) indicates that in this analysis that patient height accounts for 79.71% variability of the catheter length.

The *F*-value (39.28) of patient height indicates that this model is not a great fit to measure the influence of patient height on catheter length. However, the *F*-value of patient height is less than the *F*-value of patient weight also indicating that this model is still a worse fit to measure influence on catheter length.

### Linear regression of patient weight's effect on catheter length

Regression line of patient weight's effect on catheter length:  $y = 25.34409 + 0.28387x$

The multiple R-square (0.8181) indicates that in this analysis that patient weight accounts for 81.81% variability of the catheter length. Other variables such as patient age or patient weight account for the other 19.19% variability of the catheter length.

The difference between the multiple R-square (0.8181) and the Adjusted R-square (0.8000) is 0.0181 which is very small. However, this difference (Multiple R-square-Adjusted R-square) of patient weight is less than the difference (Multiple R-square- Adjusted R-square) of patient height (0.0203). This could be a factor as to why when patient height and patient weight are analyzed together during linear regression that patient height accounts for more of the catheter length than patient weight.

The *p*-value that corresponds to the *F*-value of patient weight is significant (5.317e-05) being less than 0.05 and expressing that this model is a good fit for the dataset.

## CONCLUSION

We cannot fully evaluate catheter length just based on patient height & weight individually. Patient height and patient weight are good predictors when analyzed individually, just based on this dataset, but together, they are redundant due to the quality of multicollinearity. They take away from one another, not leaving much room for the other predictor or variable to influence or impact what type of catheter length to use for a patient.

From the analyses of this small dataset, patient weight is more significant and influential than patient height. More data either needs to be found or randomly generated through computational means to see if this patient weight truly holds more influence on catheter length instead of patient height. Patient weight should be given more strict consideration and focus when creating and selecting catheters to medical operations. Patient weight is a more encompassing variable that can carry other latent variables.

These statistical analysis results can help physicians and hospitals better diagnose and treat disease depending on catheter length and patient weight, especially in making a personalized catheter tube based on the patient's own unique anatomical biometrics. If a patient weighs this much, then this particular catheter length needs to be

this certain measurement. If this particular catheter length needs to be used, then the patient could have this array of diseases or injury. However, more statistical and bioinformatic analysis needs to be done on catheter use and design on patients with particular diseases and injuries to create an effective decision process on what catheter to choose for therapeutic use. This research could provide an avenue of precision medicine decision making and evaluation in the biomedical device field. These results about patient height and patient weight can also help researchers and manufacturers to study the effects of catheter length on the therapeutic treatment of disease.

## REFERENCES

1. Regression linear regression datasets. 2016. <https://fla.st/2NCHhuo>
2. Weisberg S. Applied linear regression third edition. University of Minnesota School of Statistics. John Wiley & Sons Inc Publishing. 2005. <http://bit.ly/2OHSgdb>
3. Spaeth H. Mathematical algorithms for linear regression (Computer Science and Scientific Computing). 1991. ISBN: 0126564604
4. Everitt NJ, McMahon MJ. Influence of fine-bore catheter length on infusion thrombophlebitis in peripheral intravenous nutrition: a randomized controlled trial. *Annals of the royal college of surgeons of England. Ann R Coll Surg Engl.* 1997; 79: 221-4. <http://bit.ly/2O5Hwgy>
5. Gujarati D. Multicollinearity: What happens if the regressors are correlated? (McGraw-Hill). 2003.
6. O'Brien RMA. Caution regarding rules of thumb for variance inflation factors. *Quality & Quantity.* 2007; 41: 673-690. <http://bit.ly/2rSiHNI>
7. Multicollinearity-ISU public homepage server. Model diagnostics-Iowa State University. <http://bit.ly/33JA2pu>