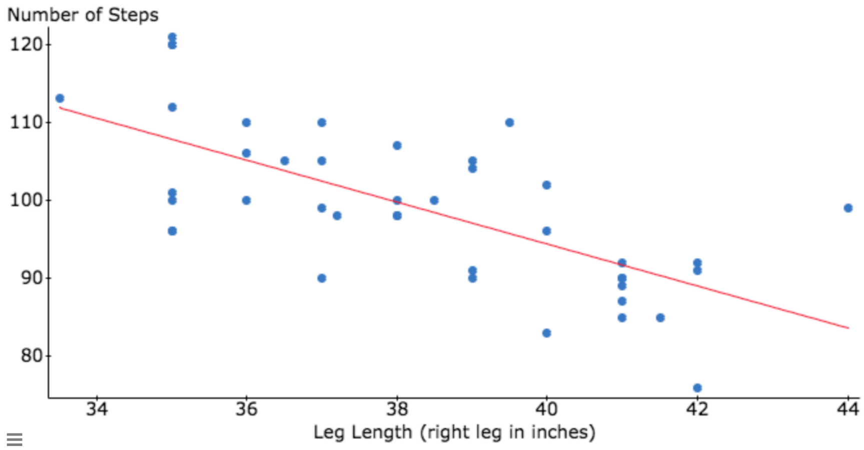
Leg Length and Number of Steps

The following details the results of our class investigation in which we studied the relationship between leg length (measured from hip to foot) and the number us steps it took to walk around a block of classrooms in building 7. Forty-two subjects participated in the study and their data is summarized in the scatter plot below.



Analysis:
The explanatory variable in this study is *leg length* (measured in inches from hip to foot).
The response variable is the *number of steps* required to complete one circuit around a block of classrooms in building 7.

The data showed moderately strong, negative, linear correlation.
The correlation coefficient reflected this with a value of R ≈ -0.7.

The regression equation gave a slope of -2.7 meaning that on average, for every additional inch of leg length, students would take about three fewer steps.

Simple linear regression results:
Dependent Variable: Number of Steps
Independent Variable: Leg Length (right leg in inches)
Number of Steps = 201.95193 - 2.6887291 Leg Length (right leg in inches)
Sample size: 42
R (correlation coefficient) = -0.67809551
R-sq = 0.45981352
Estimate of error standard deviation: 7.6712562

Parameter estimates:

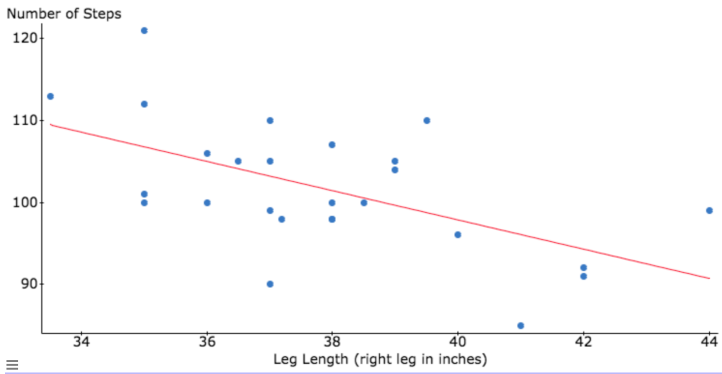
Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	201.95193	17.666908	≠ 0	40	11.431085	<0.0001
Slope	-2.6887291	0.46078469	≠ 0	40	-5.8351095	<0.0001

The coefficient of determination (R^2) has a value of $R^2 \approx 0.46$ means that 46% of the variability in the response variable (Number of Steps) is explained by the *linear relationship* with the explanatory variable (Leg Length).
The **regression equation** (with rounded values):

Number of Steps ≈ 202 – 2.7 × Leg Length
Can be used to predict the number of steps required based on a person's leg length.
For example, a person with a 40" leg is predicted to take 202 – 2.7(40) = 94 steps.

Note that the model predicts a person with a 0" long leg will take 202 steps to complete the circuit and a person with a 75" leg will take 0 steps. These extreme cases show the limitations of the model and the danger of predicting outcomes beyond the range of the actual data points (*extrapolation*).

Comparisons:
Disaggregating (segregating) the data by gender produces the following results. The female set shows a weaker association (than the overall group) that is negatively correlated and linear.



Simple linear regression results for Gender=F:
Dependent Variable: Number of Steps
Independent Variable: Leg Length (right leg in inches)
Number of Steps = 169.35131 - 1.7863585 Leg Length (right leg in inches)
Sample size: 26
R (correlation coefficient) = -0.56630032
R-sq = 0.32069605
Estimate of error standard deviation: 6.6302257

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	169.35131	20.131036	≠ 0	24	8.4124488	<0.0001
Slope	-1.7863585	0.53069899	≠ 0	24	-3.3660484	0.0026

Female results

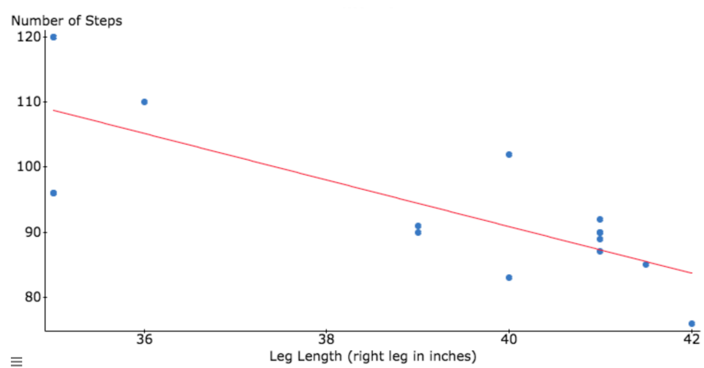
The correlation coefficient reflects the weaker association (R ≈ -0.6) and the R^2 value of 0.3 tells us that only 30% of the variability in number of steps is due to the (linear) relationship with the leg length of the women in the sample. The slope (-1.8 steps per inch) tells us that on average women whose legs are longer by an inch walk two steps fewer than those with shorter legs.

In the male sample we see a stronger negative linear association than among the women (R ≈ -0.8) but also greater variability (possibly due to the smaller sample size). Note the larger Standard Error for both slope and intercept between the two equations. With $R^2 \approx 0.6$ we have 60% of the variability in the number of steps taken by men is explained by linear dependence on leg length. The slope in the equation for males (-3.6steps/inch) tells us that for every inch taller a male in the sample is, they will take about 4 fewer steps to complete the circuit.

Simple linear regression results for Gender=M:
Dependent Variable: Number of Steps
Independent Variable: Leg Length (right leg in inches)
Number of Steps = 234.12857 - 3.5808145 Leg Length (right leg in inches)
Sample size: 16
R (correlation coefficient) = -0.77835057
R-sq = 0.6058296
Estimate of error standard deviation: 8.1002181

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	234.12857	30.101555	≠ 0	14	7.7779558	<0.0001
Slope	-3.5808145	0.77194181	≠ 0	14	-4.6387104	0.0004



Male results

Conclusions:
Overall, males had stronger correlation between leg length and number of steps than females in the study. With twice the R^2 value, we expect estimations for male steps using a linear model to be more accurate than those for women. Taller men tended to take 4 fewer steps for each additional inch of leg length while taller women tended to take only 2 fewer steps. Other factors that might play a role in these results include physiology (muscle composition, body type, skeletal structure), balance, effects of injuries, as well as personal preferences. In general there are a number of other factors that appear to influence the number of steps a person takes making it difficult to use any one (e.g. leg length) as a strong predictor. However, there is sufficient strength in the relationship to allow for general observations- e.g., people with longer legs tend to take fewer steps to complete a journey- on the order of 1 to 4 fewer steps for each additional inch of leg length.