

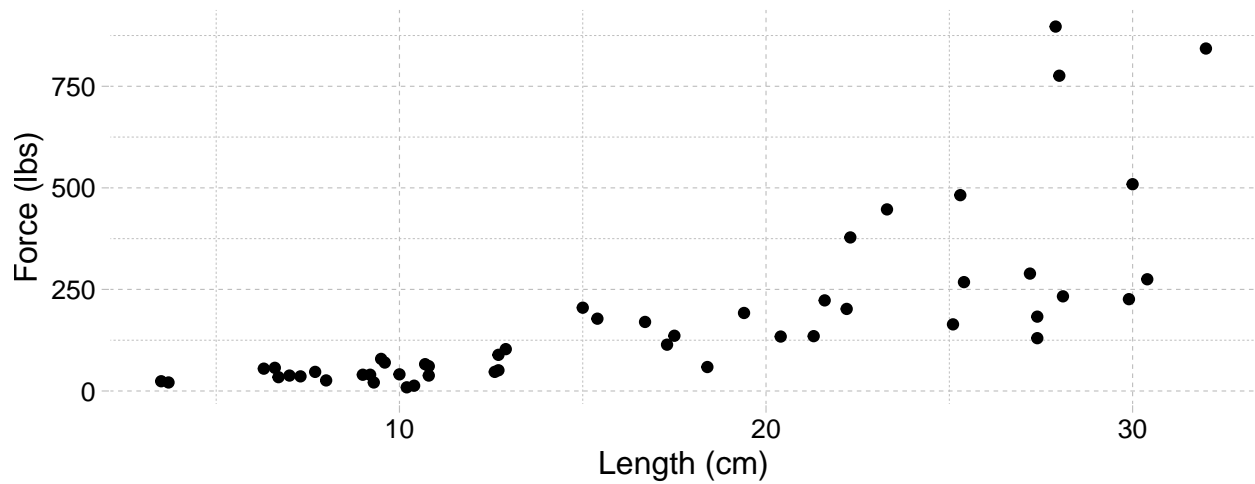
Exam 2

Hayden Atchley

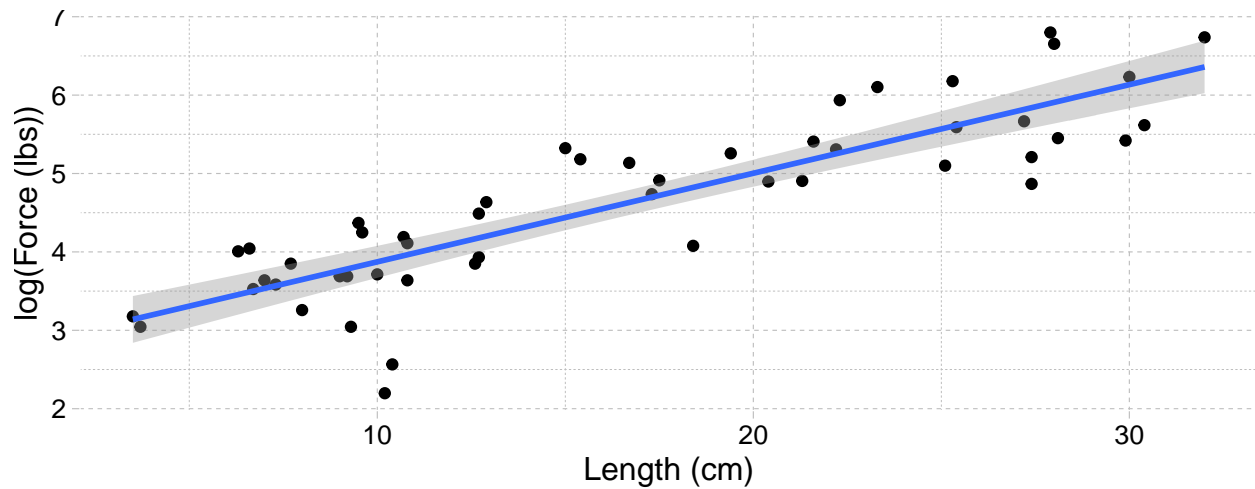
2022-11-01

1

The first thing to do is to look at a scatterplot of the data:



A log transformation of the y-axis results in a more linear relationship:



And a table describing the regression model:

term	estimate	std.error	statistic	p.value
(Intercept)	2.74	0.177	15.46	0
length	0.11	0.010	11.85	0

$R^2 = 0.74$

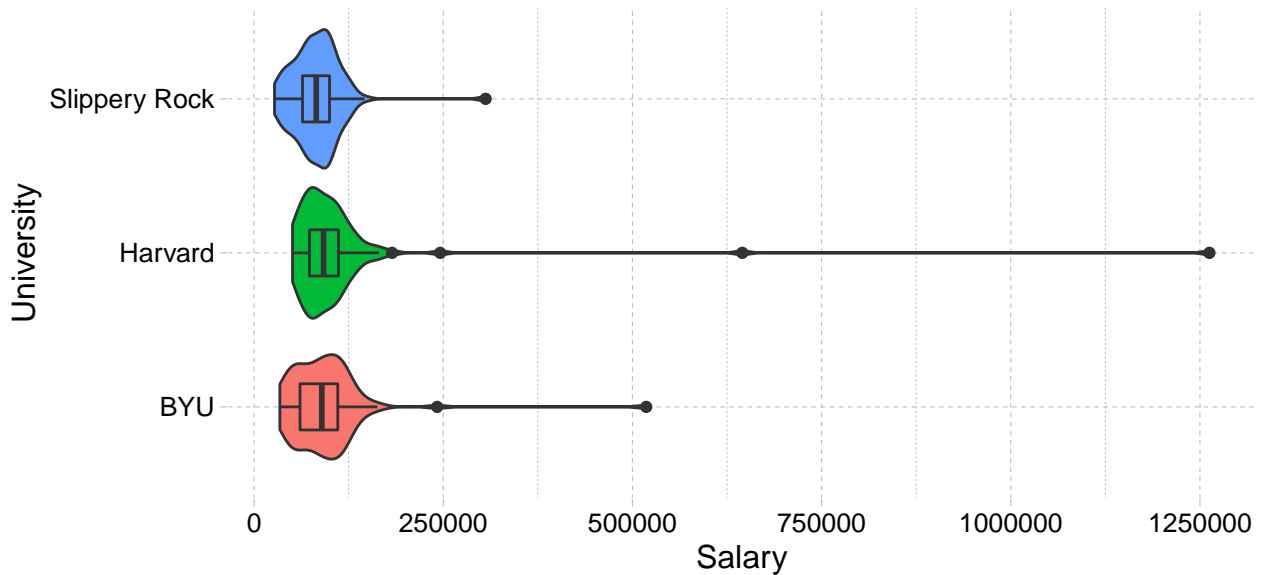
This gives overwhelming evidence of a relationship between the claw length and force applied. With an R-squared value of 0.74, the model also explains a significant amount of the variance in the data. To re-state the model, we found:

$$\begin{aligned}\log(force) &= 2.74 + 0.11 \times length \\ force &= e^{2.74+0.11 \times length} \\ &= 15.54 \times e^{0.11 \times length}.\end{aligned}$$

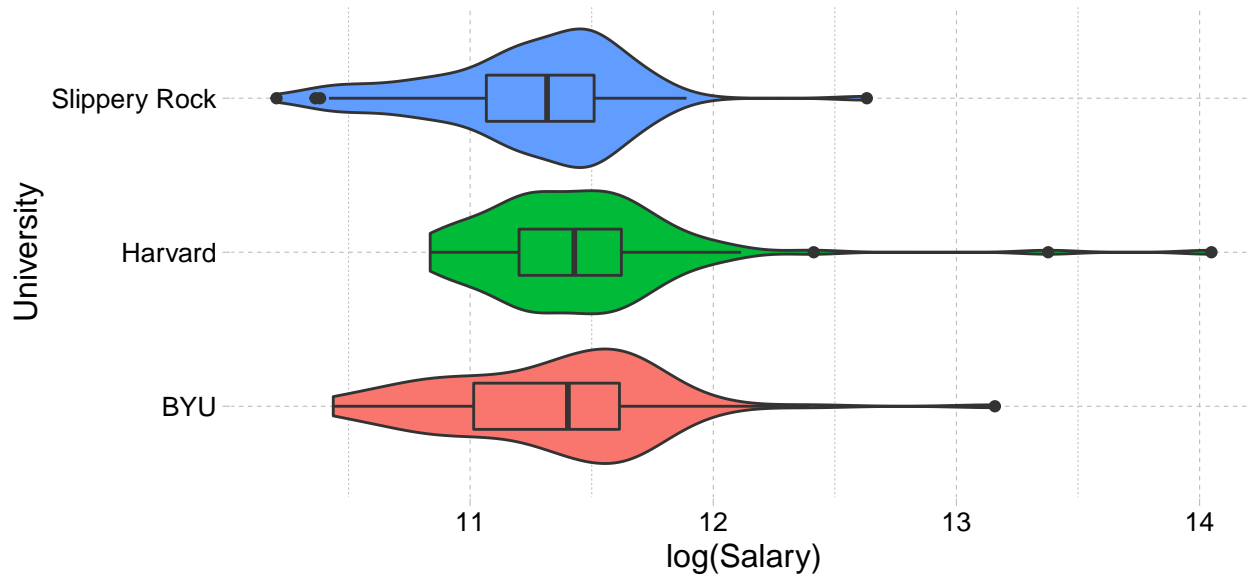
It is important to note that this is an observational experiment, and so while we can conclude that there is an association between claw length and force applied, causality cannot be inferred. The data were obtained from crabs “captured one day”, which is likely a random sample selection (though it is impossible to know for sure without more details regarding the capturing process). Therefore, the results can reasonably be generalized to the crab population in the area.

2

First we look at a plot of the data:



This data could be log-transformed, as incomes are not usually linearly distributed, and several outliers are often present (as can be seen above, especially with the Harvard data). A log-transformed plot is presented below:



Running a one-way ANOVA on the untransformed data gives a p -value of 0.055, which is suggestive of a difference, but inconclusive. An ANOVA test on the log-transformed data gives a p -value of 0.0047, which offers convincing evidence of a difference.

We run a Tukey-Kramer analysis on both the untransformed and log-transformed data. The untransformed analysis is given first:

contrast	estimate	conf.low	conf.high	adj.p.value
Harvard-BYU	18752	-9610	47114	0.266
Slippery Rock-BYU	-9906	-38268	18456	0.689
Slippery Rock-Harvard	-28658	-57020	-296	0.047

And the analysis of the log-transformed data:

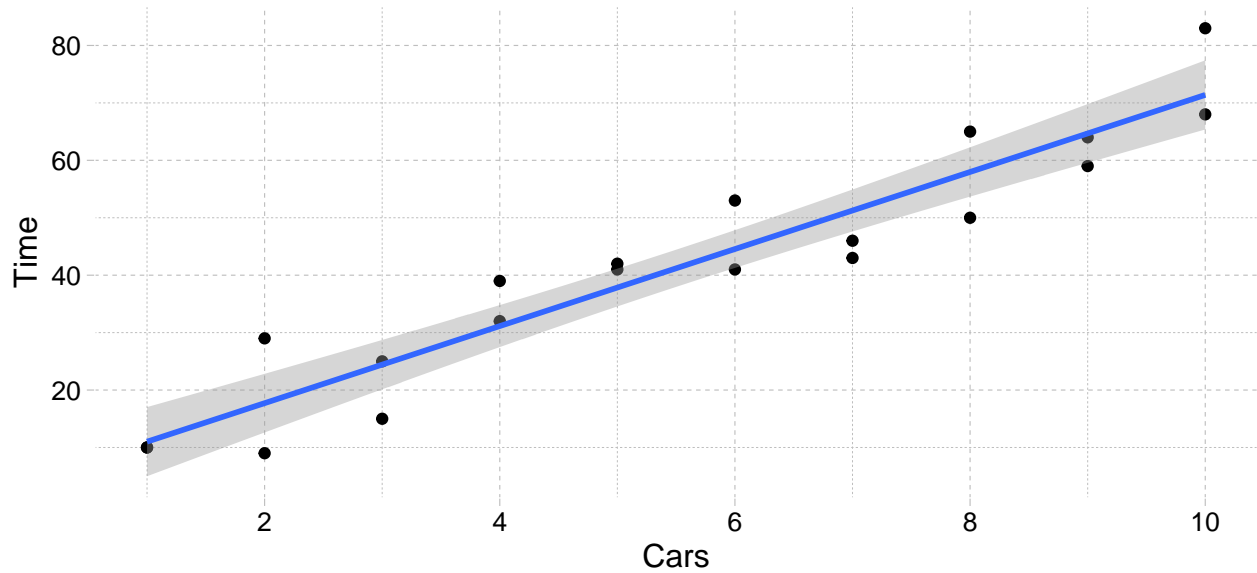
contrast	estimate	conf.low	conf.high	adj.p.value
Harvard-BYU	0.118	-0.025	0.261	0.128
Slippery Rock-BYU	-0.082	-0.224	0.061	0.372
Slippery Rock-Harvard	-0.199	-0.342	-0.057	0.003

In both of these analyses, there is evidence of a difference between Slippery Rock and Harvard salaries, though the evidence is moderate on the original data and much more convincing on the log-transformed data. In any case, it is reasonable to conclude that there is a difference between these two groups.

This result can be generalized to all business graduates from these universities (at least over the time interval that the samples were taken from), as the sample selection was randomized. However, this is an observational study, so while we conclude there is a difference in salaries between Slippery Rock graduates and Harvard graduates, we cannot infer causality.

3

A scatterplot of the data shows that there may be a roughly linear relationship between the number of cars in line and the wait time:



An analysis of the linear model:

term	estimate	std.error	statistic	p.value
(Intercept)	4.30	3.33	1.29	0.213
Cars	6.71	0.54	12.50	0.000

$R^2 = 0.891$

a

The predicted mean wait time with two cars in line is as follows:

fit	lwr	upr
17.7	12.6	22.8

Note that the interval given is the 95% confidence interval for the *mean* wait time, not the 95% prediction interval.

b

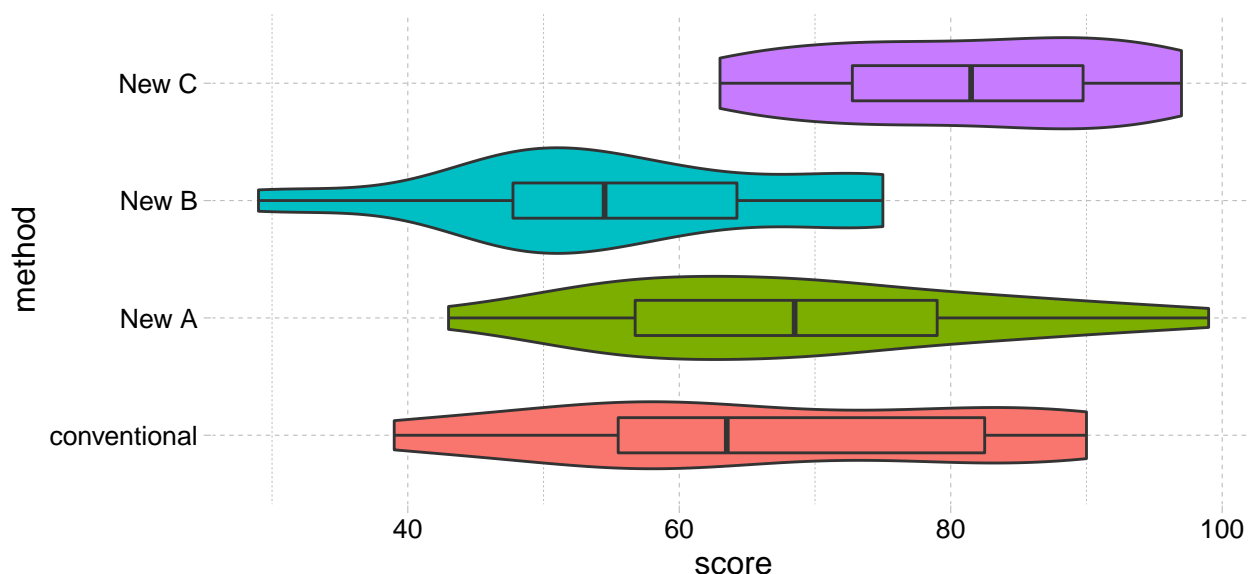
The predicted wait time with 20 cars in line is 138.5 minutes.

c

The value of 20 cars in line is far outside the range of observed data, and while the relationship is approximately linear within the data range, there is no guarantee that that relationship holds outside the data range.

4

A plot of the data:



Since we are only interested in comparing the conventional method to the other groups individually, we use the Dunnett test:

	diff	lwr.ci	upr.ci	pval
New A-conventional	2.0	-8.74	12.74	0.945
New B-conventional	-11.0	-21.74	-0.26	0.043
New C-conventional	14.7	3.96	25.44	0.004

This gives moderate evidence that method B produces lower scores than the conventional method, and strong evidence that method C produces higher scores. There is no evidence that method A differs from the conventional method.

It is unclear how the students were assigned to a teaching method; if it was randomized then we can conclude that the teaching method is responsible for the score difference, but if it was not randomized then no causality can be inferred. It is also unclear how the students were selected; if they were a random sample from a larger group then the results can be generalized to that group, but if not then we cannot generalize.

5

- In performing comparisons between multiple groups, it is good practice to run a t -test for each pair and use those results in the analysis. (T/F)
- The _____ model is obtained by imposing the restrictions of the null hypothesis.
- The difference between the observed value and the estimated mean is called $a(n)$:
 - a. F -value
 - b. Residual
 - c. Standard error
 - d. Sum of squares