Impact of Research on Univerisity Rankings

Thomas Billington, Steven Bottone, Nart Hakij, and Joe Manley

2023-10-23

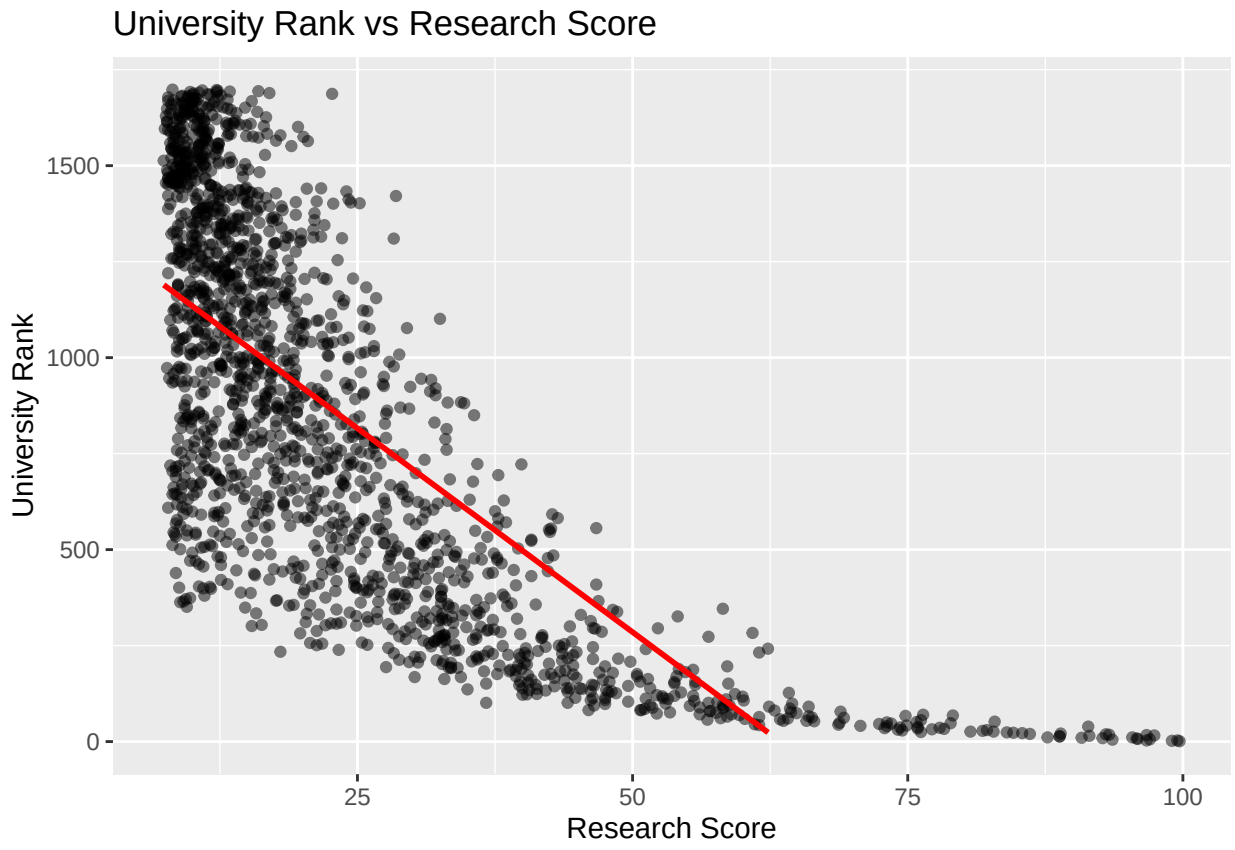
Introduction:

- Many schools either brand themselves as teaching, professional, or research focused universities.
- By measuring the trends of institutions across the globe we can determine if their ability to conduct meaningful research correlate to its statute.
- Goal: Does a university's research efforts have an effect on its global ranking?
- Response variable: Global ranking
- Explanatory variable: Research score (out of 100)
- Observations: 1695 universities

Analysis:

To better understand the relationship, a thorough statistical analysis was conducted.

Plot:



From the graph there appears to be a strong negative linear relationship between research score and university ranking, however this will be validated in later tests. We see that as research score increases, rank decreases(goes towards better rank).

More information can be observed from better understanding the linear model:

Model:

```
Call:
lm(formula = University.Rank ~ Research.Score, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-794.7 -261.5  -14.9  269.9  822.3

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1346.6788    13.7041   98.27   <2e-16 ***
Research.Score  -21.2322     0.4738  -44.81   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 331.8 on 1695 degrees of freedom
Multiple R-squared:  0.5422,    Adjusted R-squared:  0.542
F-statistic:  2008 on 1 and 1695 DF,  p-value: < 2.2e-16
```

From this summary output we can note a few key details about the model.

- $\hat{Y}$ = 1346.679 + (-21.23218)*X
- $\hat{Y}$ is university ranking
- X is research score of university

Statistical Tests and Inference:

Hypothesis Testing:

- H0: There is no association between university ranking and research score
- H1: There is an association between university ranking and research score

The p-value was 2.2eˆ-16 which is much lower than the significance level of 0.05. Therefore, we can reject the null hypothesis, indicating that there is a strong statistical evidence that when university ranking decreases or gets closer to rank 1, research score increases.

Correlation:

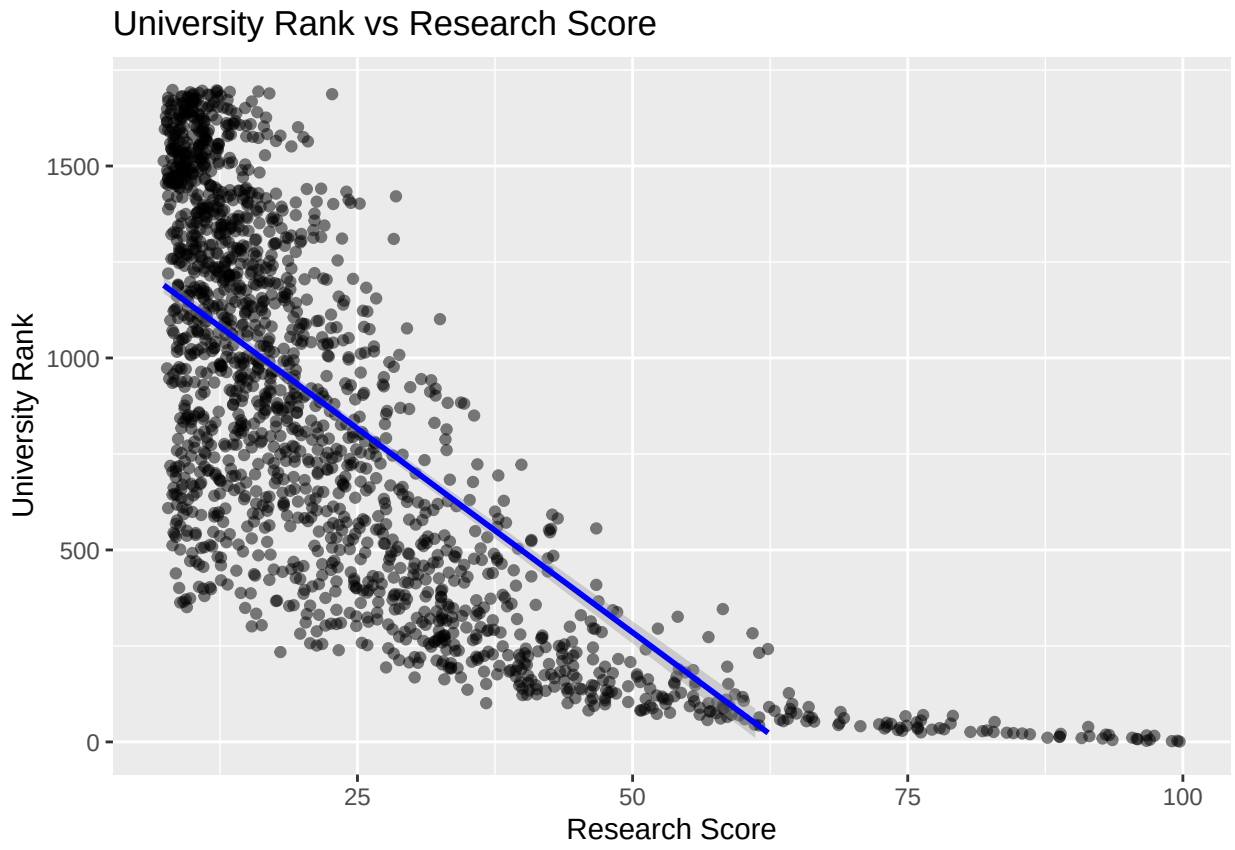
- The Rˆ2 value indicates that 54.2% of the variation in university rank can be explained by the research score.
- The estimated residual standard error is s = 331.8 on 1695 degrees of freedom. This indicates that there is some variability in the models predictions.
- Correlation coefficient between University.Rank and Research.Score: -0.736. This indicates a strong negative correlation, as one variable increases the other decreases.

Confidence Intervals:

```
                   2.5 %    97.5 %
(Intercept)    1319.80011 1373.5575
Research.Score  -22.16156  -20.3028
```

We are 95% confident that as the research score increases by 1%, the overall university rank decreases by 22.2 to 20.3 ranks.

Graphical representation of data with intervals:



Prediction:

Transformation and Comparison:

Due to the shape of the data, a logarithmic transformation was applied.
Graph of transformed data:

Model information:

```
Call:
lm(formula = ln_y ~ Research.Score, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.40523 -0.24076  0.05924  0.26503  1.24660

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     7.6775842  0.0172795  444.32   <2e-16 ***
Research.Score -0.0528821  0.0005975  -88.51   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4183 on 1695 degrees of freedom
Multiple R-squared:  0.8221,    Adjusted R-squared:  0.822
F-statistic:  7834 on 1 and 1695 DF,  p-value: < 2.2e-16
```

From the transformation we can see that the R^2 value of the model is 0.8221. In comparison to the value from the original model (0.542) we can see that the there is a much stronger relationship between the variables when transformed.

Correlation coefficient between log. transformed rank and research score:

```
[1] -0.9067089
```

From this score we see that the data also has a much stronger negative linear correlation with a score of -0.9067089 compared to the -0.736 value for the non-transformed rank.

From these observations, we can choose the logarithmic transformed model as the better model.

Assumptions:

Conclusions: