

# Risk Modeling

Darleen Ziege

## 1. Topic Selection and Introduction

For Mathematical Connection topic selection, I knew I wanted to engage with material in the actuarial science field that was rooted in statistical analysis. Since this is the field I hope to enter after graduation, I knew this would be an interesting deviation from my more theoretical math courses by engaging in a common application within the field. Based on my research, I chose to engage more deeply in current and prospective risk modeling variations, and an evaluation of their relative accuracies based on various goodness of fit models. These models included the R-Squared method, Mean Absolute Error statistics, and Predictive Ratio modeling. I also felt the weight of ethical statistic evaluation, by engaging in a review of Anscombe's Quartet and its warnings about context when evaluating statistics.

## 2. Terminology

Scoring models summarize available, relevant information about consumers and reduce the information into a set of ordered categories that foretell an outcome. For risk scores in particular, we consider these models a normalization of plan premiums or claims experience across a covered population or market to account for differences in risk. Actuaries are business professionals who utilize risk modeling to evaluate the likelihood of potential events and how to manage these risks effectively and efficiently.

Actuarial models can generally be categorized into two main types: prospective and concurrent risk models. Prospective models use information from one year to predict risk expenditures for the following – or some other future – year. Concurrent models on the other hand, use information from one year to predict risk expenditures in that same year. Concurrent models are generally more accurate portrayals than prospective models since they make predictions more closely associated with the period from which the data were drawn.

### 3.1 R-Squared

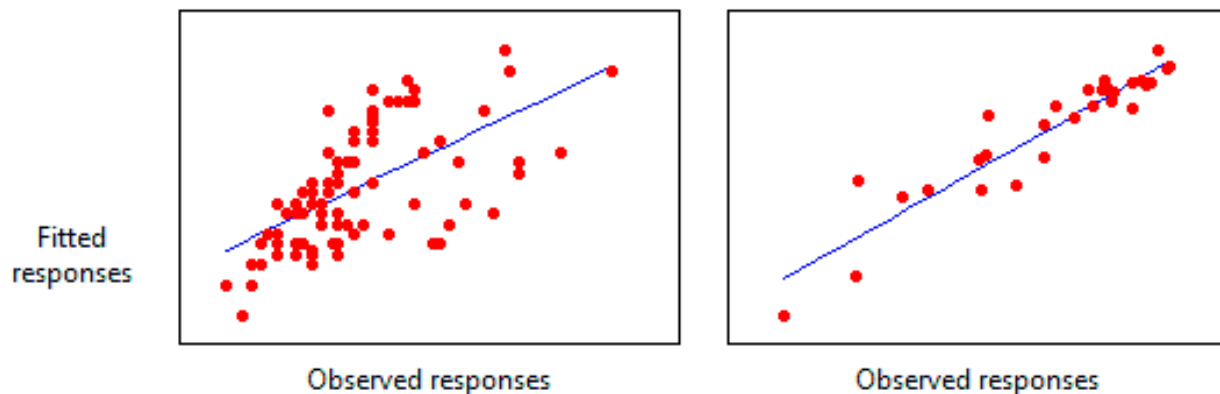
R-Squared statistics ( $R^2$ ) describe goodness of fit distributions based on individual risk levels. It is defined as the percentage of model variation in the dependent variable that is explained by the specified model. R-Squared is calculated as,

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where  $f_i$  is the prediction for observation  $i$ , and  $y_i$  is the actual value for observation  $i$ .

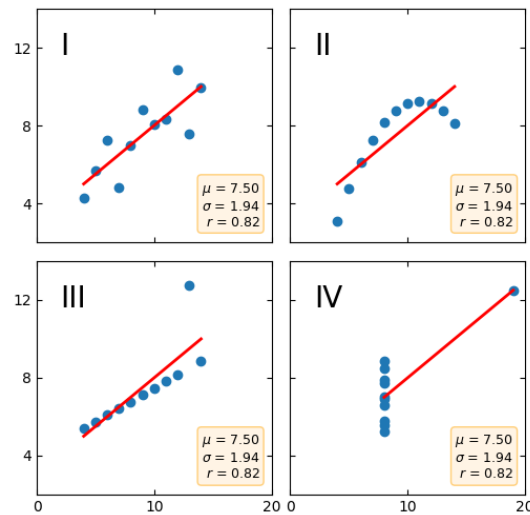
An  $R^2$  value of zero indicates a model with no relation, and a value of one indicates a model perfectly predicts the dependent variable for every observation in the sample. Naturally, higher  $R^2$  indicate a higher risk-scoring model accuracy. This can most easily be seen with a graphical representation, where blue dots are points of data whereas the red line indicates the calculated regression of R-Squared. On the below example, the left example shows scattered data where R-Squared accounts for about 38% data variance, while the example on the right accounts for approximately 87.4%. Note that a higher R-Squared value indicates that data points are more closely aligned with the best fit line of regression.

**Plots of Observed Responses Versus Fitted Responses for Two Regression Models**



## 3.2 Anscombe's Quartet

Francis Anscombe was a French mathematician best known for his criticism of R-Squared values in 1973. Anscombe discovered four data point distributions, shown below, that all display wildly different distributions of data, but have the same statistical analysis values, namely equal means, standard deviations, and R-Squared values. It is unknown how Anscombe came up with his distributions, but they highlight the importance of context when analyzing data points, their regression statistics, and the impact of outliers on impacting distributions.



## 3.3 Mean Absolute Error

Mean Absolute Error (MAE), like the R-Squared Statistic, describe goodness of fit based on individual risk levels. MAE is used to describe how close a prediction is to the actual, observed outcome. Mean Absolute Error is calculated as,

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

where  $|e_i| = |f_i - y_i|$  is the absolute error for observation  $i$ , and  $n$  is the total number of observations.

Unlike  $R^2$ , a lower MAE is preferable for accuracy measures. A MAE of zero indicates a perfectly accurate estimated risk score, since it can be interpreted as an average estimation error of zero. There is also no upper limit on the MAE, which means the accuracy of the mean absolute error is relative to that of similar distributions and must be evaluated as such.

### 3.4 Predictive Ratios

Predictive ratios are the most elusive accuracy evaluation that was considered in this research. Whereas R-Squared and Mean Absolute Error evaluate distributions based on individual data points, predictive ratios provide accuracy based on a snapshot of a specific subgroup of individuals within a population. Despite how these are calculated however, they are generally standardized to their population such that a predictive ratio of 100% indicates that a model, on average, is perfectly unbiased for that sample of individuals. A deviation in the absolute value of a predictive ratio away from 100% indicates possible bias, with a higher chance of bias further removed from 100%.

## 4. Current and Prospective Risk Models

Below are the R-Squared and MAE values for various actuarial risk models, first on an individual level, followed by a group level based on groups of 1,000 and 10,000. The models were also split based on whether they were prospective or concurrent models, since we noted the significant difference earlier in their accuracies, due to their general differences in definition. This study, conducted by the Society of Actuaries, was done to review their respective accuracies and draw comparisons among them to evaluate their quality, as well as how each was impacted by outliers. It is a great example of how these statistical evaluations apply to real actuarial modeling.

On the individual level, we notice the significant difference between uncensored best fit lines and the impact on censoring at \$250 thousand; a point above which values were considered to be significant outliers, since this level was only reached by less than 0.1% of the individuals in any given sample. This is important to note due to the lessons learned from Anscombe's Quartet. Note that MAE scores have all been rescaled to a mean of one which allows them to be expressed as percentage scores.

Table 4.2.2: R-Squared and MAE, Prospective Models

|                               | R-Squared  |                    | MAE        |                    |
|-------------------------------|------------|--------------------|------------|--------------------|
|                               | Uncensored | Censored at \$250k | Uncensored | Censored at \$250k |
| Diagnosis-Only Models         |            |                    |            |                    |
| ACG System                    | 16.2%      | 21.0%              | 100.7%     | 98.7%              |
| CDPS                          | 9.1%       | 11.9%              | 109.2%     | 107.5%             |
| DxCG                          | 18.6%      | 23.8%              | 98.9%      | 96.9%              |
| Impact Pro                    | 18.9%      | 22.8%              | 98.2%      | 96.2%              |
| MARA                          | 20.1%      | 24.9%              | 97.3%      | 95.3%              |
| Truven                        | 20.7%      | 26.4%              | 96.4%      | 94.0%              |
| Wakely                        | 17.0%      | 21.3%              | 100.5%     | 98.6%              |
| Pharmacy-Only Models          |            |                    |            |                    |
| ACG System                    | 11.6%      | 16.5%              | 102.7%     | 100.7%             |
| DxCG                          | 14.8%      | 19.9%              | 100.4%     | 98.4%              |
| Impact Pro                    | 13.7%      | 19.1%              | 101.6%     | 99.6%              |
| MARA                          | 15.1%      | 20.1%              | 99.8%      | 97.8%              |
| MedicaidRx                    | 8.6%       | 12.8%              | 107.6%     | 105.7%             |
| Wakely                        | 9.9%       | 14.9%              | 103.9%     | 101.9%             |
| Diagnosis-and-Pharmacy Models |            |                    |            |                    |
| ACG System                    | 17.2%      | 23.0%              | 97.6%      | 95.5%              |
| CDPS+MRX                      | 10.0%      | 13.3%              | 107.0%     | 105.1%             |
| CRG                           | 17.0%      | 21.7%              | 99.6%      | 97.6%              |
| Impact Pro                    | 20.7%      | 25.8%              | 94.6%      | 92.5%              |
| MARA                          | 22.0%      | 27.7%              | 93.3%      | 91.3%              |
| Wakely                        | 18.5%      | 23.7%              | 97.1%      | 95.1%              |
| Prior Cost Models             |            |                    |            |                    |
| ACG System                    | 17.8%      | 23.7%              | 96.7%      | 94.6%              |
| DxCG                          | 23.8%      | 27.7%              | 91.2%      | 89.1%              |
| MARA                          | 24.8%      | 26.9%              | 91.8%      | 90.1%              |
| SCIO                          | 15.1%      | 22.4%              | 95.8%      | 93.5%              |

Table 4.2.1: R-Squared and MAE, Concurrent Models

|                               | R-Squared  |                    | MAE        |                    |
|-------------------------------|------------|--------------------|------------|--------------------|
|                               | Uncensored | Censored at \$250k | Uncensored | Censored at \$250k |
| Diagnosis-Only Models         |            |                    |            |                    |
| ACG System                    | 44.1%      | 52.4%              | 75.3%      | 73.3%              |
| CDPS                          | 24.2%      | 30.0%              | 92.5%      | 90.6%              |
| DxCG                          | 52.6%      | 61.0%              | 67.6%      | 65.0%              |
| HHS-HCC                       | 41.3%      | 45.2%              | 86.8%      | 85.5%              |
| MARA                          | 52.7%      | 62.6%              | 64.0%      | 61.8%              |
| Truven                        | 52.6%      | 62.7%              | 64.9%      | 61.6%              |
| Wakely                        | 43.2%      | 51.0%              | 76.5%      | 74.3%              |
| Pharmacy-Only Models          |            |                    |            |                    |
| DxCG                          | 29.6%      | 38.4%              | 83.0%      | 80.8%              |
| MARA                          | 30.1%      | 40.1%              | 81.8%      | 79.6%              |
| MedicaidRx                    | 12.9%      | 18.0%              | 100.3%     | 98.3%              |
| Wakely                        | 19.9%      | 28.8%              | 91.4%      | 89.2%              |
| Diagnosis-and-Pharmacy Models |            |                    |            |                    |
| ACG System                    | 45.9%      | 56.4%              | 70.0%      | 67.6%              |
| CDPS-MRx                      | 25.6%      | 32.4%              | 90.0%      | 88.1%              |
| CRG                           | 41.0%      | 49.3%              | 78.2%      | 76.2%              |
| MARA                          | 55.4%      | 66.7%              | 57.9%      | 55.6%              |
| Wakely                        | 44.3%      | 54.2%              | 73.8%      | 71.3%              |

On the group level displayed below, we notice that the error levels are clustered much more tightly than at the individual level. Especially considering the prospective models, the range from best to worst performance of R-Squared and MAE was much smaller. This implies that the difference in predictive power is less relevant for groups than at the individual level.

Table 4.3.2: R-Squared and MAE, Simulated Random Groups, Prospective Models (Uncensored)

|                               | R-Squared |        | MAE   |        | 95th Percentile of Error |        |
|-------------------------------|-----------|--------|-------|--------|--------------------------|--------|
|                               | 1,000     | 10,000 | 1,000 | 10,000 | 1,000                    | 10,000 |
| Diagnosis-Only Models         |           |        |       |        |                          |        |
| ACG System                    | 12.1%     | 16.0%  | 9.4%  | 2.9%   | 22.5%                    | 7.2%   |
| CDPS                          | 7.8%      | 9.6%   | 9.2%  | 3.1%   | 22.5%                    | 7.3%   |
| DxCG                          | 14.5%     | 18.8%  | 9.2%  | 2.9%   | 21.9%                    | 7.2%   |
| Impact Pro                    | 13.1%     | 18.1%  | 9.2%  | 2.9%   | 22.2%                    | 6.9%   |
| MARA                          | 15.2%     | 19.8%  | 9.2%  | 2.9%   | 21.7%                    | 7.1%   |
| Truven                        | 16.6%     | 20.7%  | 9.1%  | 2.8%   | 21.3%                    | 7.3%   |
| Wakely                        | 13.8%     | 17.4%  | 9.2%  | 2.9%   | 21.8%                    | 7.2%   |
| Pharmacy-Only Models          |           |        |       |        |                          |        |
| ACG System                    | 10.6%     | 14.0%  | 9.4%  | 3.0%   | 22.2%                    | 7.1%   |
| DxCG                          | 13.8%     | 14.4%  | 9.2%  | 3.0%   | 22.0%                    | 7.2%   |
| Impact Pro                    | 13.8%     | 13.8%  | 9.2%  | 3.0%   | 21.9%                    | 7.3%   |
| MARA                          | 14.0%     | 15.3%  | 9.2%  | 2.9%   | 22.1%                    | 7.1%   |
| MedicaidRx                    | 8.8%      | 8.4%   | 9.5%  | 3.1%   | 22.6%                    | 7.3%   |
| Wakely                        | 10.3%     | 9.7%   | 9.4%  | 3.0%   | 22.9%                    | 7.5%   |
| Diagnosis-and-Pharmacy Models |           |        |       |        |                          |        |
| ACG System                    | 13.9%     | 18.5%  | 9.4%  | 2.9%   | 22.3%                    | 7.1%   |
| CDPS+MRX                      | 9.7%      | 10.7%  | 9.5%  | 3.1%   | 22.6%                    | 7.2%   |
| CRG                           | 11.8%     | 12.1%  | 9.3%  | 3.0%   | 22.2%                    | 7.4%   |
| Impact Pro                    | 15.8%     | 20.6%  | 9.1%  | 2.9%   | 22.0%                    | 6.8%   |
| MARA                          | 18.5%     | 21.6%  | 8.9%  | 2.8%   | 21.7%                    | 7.0%   |
| Wakely                        | 16.0%     | 18.5%  | 9.1%  | 2.9%   | 21.4%                    | 7.1%   |
| Prior Cost Models             |           |        |       |        |                          |        |
| ACG System                    | 14.5%     | 20.0%  | 9.2%  | 2.9%   | 21.9%                    | 6.9%   |
| DxCG                          | 22.1%     | 24.3%  | 8.8%  | 2.8%   | 23.0%                    | 7.1%   |
| MARA                          | 22.4%     | 24.9%  | 8.7%  | 2.8%   | 21.9%                    | 6.8%   |
| SCIO                          | 14.3%     | 15.8%  | 9.2%  | 2.9%   | 22.1%                    | 7.2%   |

Table 4.3.1: R-Squared and MAE, Simulated Random Groups, Concurrent Models (Uncensored)

|                               | R-Squared |        | MAE   |        | 95th Percentile of Error |        |
|-------------------------------|-----------|--------|-------|--------|--------------------------|--------|
|                               | 1,000     | 10,000 | 1,000 | 10,000 | 1,000                    | 10,000 |
| Diagnosis-Only Models         |           |        |       |        |                          |        |
| ACG System                    | 43.7%     | 48.7%  | 7.3%  | 2.3%   | 17.2%                    | 5.6%   |
| CDPS                          | 26.9%     | 30.5%  | 8.6%  | 2.7%   | 21.0%                    | 6.5%   |
| DxCG                          | 49.4%     | 56.2%  | 6.9%  | 2.1%   | 16.2%                    | 5.4%   |
| HHS-HCC                       | 40.2%     | 46.0%  | 7.7%  | 2.3%   | 18.8%                    | 5.9%   |
| MARA                          | 51.0%     | 57.8%  | 6.8%  | 2.1%   | 16.0%                    | 5.4%   |
| Truven                        | 49.6%     | 55.7%  | 6.8%  | 2.1%   | 16.8%                    | 5.1%   |
| Wakely                        | 41.8%     | 48.5%  | 7.5%  | 2.3%   | 17.7%                    | 6.0%   |
| Pharmacy-Only Models          |           |        |       |        |                          |        |
| DxCG                          | 28.6%     | 33.3%  | 8.3%  | 2.6%   | 20.7%                    | 6.4%   |
| MARA                          | 28.5%     | 32.8%  | 8.3%  | 2.6%   | 19.5%                    | 6.5%   |
| MedicaidRx                    | 14.4%     | 15.4%  | 9.3%  | 3.0%   | 22.0%                    | 7.1%   |
| Wakely                        | 19.2%     | 22.1%  | 9.0%  | 2.8%   | 21.5%                    | 7.0%   |
| Diagnosis-and-Pharmacy Models |           |        |       |        |                          |        |
| ACG System                    | 46.4%     | 51.9%  | 7.1%  | 2.2%   | 17.1%                    | 5.6%   |
| CDPS-MRx                      | 27.8%     | 32.3%  | 9.5%  | 2.7%   | 21.1%                    | 6.5%   |
| CRG                           | 40.5%     | 42.8%  | 7.5%  | 2.4%   | 18.2%                    | 5.8%   |
| MARA                          | 53.9%     | 60.0%  | 6.6%  | 2.0%   | 15.5%                    | 5.2%   |
| Wakely                        | 43.0%     | 49.1%  | 7.6%  | 2.4%   | 17.4%                    | 6.0%   |

We can also consider the predictive ratios for both current and prospective models, calculated based on the mean risk score divided by the mean actual cost for a subgroup of individuals from the sample population. Populations for these particular ratios were based on sex and age, with a total of six sample groups. Like the MAE scores above, these values were rescaled to one over the population.

There are significant differences in the predictive abilities of these ratios based on populations. We note that this is due to how various models are calculated; which provides context for these statistics. The HHS-HCC model for example is calibrated to predict liability rather than cost, and since children have the lowest average cost of healthcare expenditures, the model will indicate under-predicting the risk associated with children. We can do this type of analysis for each individual model type; for further research engage with Hileman & Steele (2016) cited in references.

Table 4.4.5: Predictive Ratios by Age-Sex, Prospective Models (Offered Weights; \$250,000 Censoring)

|                               | Children, 0-6 | Children, Age 7-18 | Males, Age 19-44 | Males, Age 45-64 | Females, Age 19-44 | Females, Age 45-64 |
|-------------------------------|---------------|--------------------|------------------|------------------|--------------------|--------------------|
| Diagnosis-Only Models         |               |                    |                  |                  |                    |                    |
| ACG System                    | 99.8%         | 98.7%              | 108.7%           | 100.6%           | 88.7%              | 104.0%             |
| CDPS                          | 212.4%        | 226.0%             | 119.0%           | 76.3%            | 93.3%              | 75.9%              |
| DxCG                          | 87.1%         | 88.7%              | 99.8%            | 104.3%           | 96.4%              | 102.7%             |
| Impact Pro                    | 116.0%        | 107.9%             | 102.1%           | 99.5%            | 101.6%             | 95.4%              |
| MARA                          | 105.9%        | 96.7%              | 103.7%           | 102.7%           | 96.7%              | 99.1%              |
| Truven                        | 96.3%         | 99.2%              | 101.0%           | 100.4%           | 99.5%              | 100.2%             |
| Wakely                        | 102.0%        | 100.3%             | 103.9%           | 100.4%           | 99.1%              | 98.7%              |
| Pharmacy-Only Models          |               |                    |                  |                  |                    |                    |
| ACG System                    | 106.8%        | 100.6%             | 111.0%           | 98.5%            | 91.0%              | 102.6%             |
| DxCG                          | 86.1%         | 90.3%              | 103.1%           | 104.6%           | 100.3%             | 98.8%              |
| Impact Pro                    | 111.2%        | 92.9%              | 99.0%            | 100.9%           | 100.0%             | 100.5%             |
| MARA                          | 101.1%        | 95.4%              | 106.8%           | 101.0%           | 101.2%             | 97.4%              |
| MedicaidRx                    | 147.0%        | 134.0%             | 116.3%           | 91.6%            | 100.5%             | 88.8%              |
| Wakely                        | 94.2%         | 101.0%             | 105.8%           | 100.6%           | 99.4%              | 98.3%              |
| Diagnosis-and-Pharmacy Models |               |                    |                  |                  |                    |                    |
| ACG System                    | 99.7%         | 97.3%              | 108.3%           | 99.2%            | 93.3%              | 102.8%             |
| CDPS+MRx                      | 199.5%        | 219.0%             | 117.5%           | 78.2%            | 93.5%              | 77.5%              |
| CRG                           | 139.1%        | 113.2%             | 73.9%            | 99.0%            | 112.4%             | 94.7%              |
| Impact Pro                    | 100.7%        | 99.9%              | 101.8%           | 100.1%           | 102.4%             | 97.8%              |
| MARA                          | 106.2%        | 98.2%              | 104.1%           | 102.2%           | 98.2%              | 98.1%              |
| Wakely                        | 98.8%         | 100.1%             | 103.9%           | 100.9%           | 99.1%              | 98.7%              |

Table 4.4.4: Predictive Ratios by Age-Sex, Concurrent Models (Offered Weights; \$250,000 Censoring)

|                               | Children, 0-6 | Children, Age 7-18 | Males, Age 19-44 | Males, Age 45-64 | Females, Age 19-44 | Females, Age 45-64 |
|-------------------------------|---------------|--------------------|------------------|------------------|--------------------|--------------------|
| Diagnosis-Only Models         |               |                    |                  |                  |                    |                    |
| ACG System                    | 104.5%        | 90.0%              | 97.9%            | 102.9%           | 93.9%              | 104.2%             |
| CDPS                          | 219.1%        | 216.9%             | 120.5%           | 80.7%            | 88.6%              | 76.5%              |
| DxCG                          | 88.0%         | 89.4%              | 100.4%           | 103.8%           | 94.9%              | 103.6%             |
| HHS-HCC                       | 88.1%         | 88.8%              | 100.1%           | 105.8%           | 94.6%              | 102.4%             |
| MARA                          | 115.9%        | 104.5%             | 100.8%           | 98.7%            | 97.3%              | 100.0%             |
| Truven                        | 94.0%         | 91.9%              | 98.5%            | 99.8%            | 101.0%             | 102.6%             |
| Wakely                        | 108.7%        | 100.4%             | 101.0%           | 100.0%           | 98.2%              | 100.0%             |
| Pharmacy-Only Models          |               |                    |                  |                  |                    |                    |
| DxCG                          | 92.2%         | 91.6%              | 102.0%           | 104.4%           | 99.3%              | 99.1%              |
| MARA                          | 105.5%        | 99.7%              | 103.8%           | 99.8%            | 100.2%             | 98.5%              |
| MedicaidRx                    | 229.9%        | 220.3%             | 123.5%           | 79.5%            | 94.2%              | 71.4%              |
| Wakely                        | 97.7%         | 101.9%             | 106.0%           | 98.8%            | 101.0%             | 98.2%              |
| Diagnosis-and-Pharmacy Models |               |                    |                  |                  |                    |                    |
| ACG System                    | 108.6%        | 99.3%              | 101.3%           | 98.9%            | 99.7%              | 100.1%             |
| CDPS+MRx                      | 202.2%        | 209.1%             | 119.7%           | 82.4%            | 89.6%              | 78.2%              |
| CRG                           | 122.2%        | 109.1%             | 95.7%            | 104.4%           | 91.4%              | 98.9%              |
| MARA                          | 113.9%        | 103.9%             | 101.6%           | 98.6%            | 98.7%              | 99.3%              |
| Wakely                        | 103.0%        | 98.8%              | 104.2%           | 97.8%            | 99.6%              | 100.8%             |

## 5. Connections and Applications

The above example of analysis is naturally rooted in actuarial healthcare and insurance research, but this can naturally be extended to various fields. Financial sectors utilize risk modeling to extend loans, bonds, and many other financing devices to individuals and companies alike. In addition, R-Squared and MAE accuracy can be utilized to evaluate relevance in many fields of research such as biology, economics, and machine learning. Many business analysts also utilize predictive ratios to evaluate the state of their business operations.

## References

- Parikh, Ravi. "Anscombe's Quartet, and Why Summary Statistics Don't Tell the Whole Story." *Heap*. April 21, 2014. Retrieved September 2020.
- "Anscombe's Quartet." *Matplotlib.org*. April 8, 2020. Retrieved September 2020.
- "What is an Actuary?" *beanactuary.org*. 2020. Retrieved September 2020.
- Ogee, A. & Ellis, M. & et.al. "Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit?" *The Minitab Blog*. May 30, 2013. Retrieved September 2020.
- Hileman, G. & Steele, S. "Accuracy of Claims-Based Risk Scoring Models." *Society of Actuaries*. October 2016. Retrieved September 2020.