

The Missing Statistic in the Decile Table: The Confidence Interval

Bruce Ratner, Ph.D.

The decile table has become for most modelers a universal tabular display of model performance. The decile table, by definition, has the point estimate of the statistic Cum Lift, which indicates how much better is a given model than the chance model (random selection of individuals). The confidence interval that furnishes the precision of the model is considered necessary to complete the decile-table model assessment. The purpose of this article is to augment the decile table with the missing the confidence interval estimate. I review briefly the basics of statistical inference, and then outline the bootstrap method, which is exercised in the construction of the confidence interval.

Statistical Inference

Statistical inference refers to drawing conclusions about a population parameter using sample information. To estimate population parameters there are two approaches, requiring clarifying terminology, which I make readily available.

1. Estimator vs. Estimate: Consider, statisticians refer to the sample mean as the *estimator* of the population mean, and the value of a sample mean is an *estimate* of the population mean. Analogously, statisticians focus on the sample proportion.
2. Sampling Error: A point estimate (namely, a sample statistic) of population parameter is prone to sampling error (variation), and is not likely to equal the population parameter in a given sample. Statisticians are more interested in the range, in which the population parameter will lie, not in a point estimate. Confidence intervals are preferred to point estimates, because confidence intervals indicate the *precision* (togetherness) of the estimates: *The missing confidence interval in the decile table has baffled me.*
3. Point Estimation: Obtain a single-value estimate for the population parameter.
4. Confidence Interval Estimation: Calculate an interval {i.e., $(a < X < b)$, and X is the sample statistic} within which the population parameter lies. The interval means that the population parameter is greater than a , and less than b .
5. Margin of Error: The range of values both above and below the sample statistic is called the margin of error. The margin of error = critical value x standard error of the statistic. The critical value is typically the familiar "z-score," and the standard error is the *standard deviation* of the sampling distribution. (The standard error confuses many a student.)
6. Confidence Interval Estimation Formula: Sample Statistic \pm Margin of Error
7. Confidence Level: The probability that the population parameter will lie within a confidence interval is called the confidence level, denoted by $1 - \alpha$, where common α

values are 0.05, and 0.10. (The assignment of these always-used values was flippantly uttered by the one of the fathers of statistics.) [Who?](#)

8. Confidence Interval Interpretation: A 95% confidence interval for a population parameter means that, in repeated sampling, say, a 100, 95% of the 100 confidence intervals estimated will include the population parameter; and 5 of the 100 confidence intervals estimated will not include the population parameter.
9. Question: What is the probability of any one of the 100-sample point estimates fall within a confidence interval? [Answer](#).

Decile Table

Historians trace the first use of the decile table, originally called a *gains chart* with roots in the *direct mail* business, circa wee 1950s. [1] The gains chart is hallmarked by solicitations found inside the covers of matchbooks. More recently, the decile table has transcended the origin of the gains chart toward a generalized measure of model performance. The term *decile* was first used by Galton in 1882. [2]

The decile table is a tabular display of model performance. It has become for most modelers a universal measure of model performance, for a either binary or continuous dependent variable. The decile table is widely used for today's big data. I illustrate the construction and interpretation of the binary response (yes=1, no =0) decile table found in Slides 4 and 5, below. The response model, on which the decile table is based, is not shown. Keep in mind: The eight-step construction detailed below for the binary dependent variable is identical (except for Prob_est (Probability_estimate) is replaced by Y_pred(iction)) for a continuous dependent variable Y, such as profit, sales, or write-offs.

Construction of the Response Decile Table

1. Score the validation sample or fresh file using the response model under consideration. Every individual receives a model score, Prob_est, the model's estimated probability of response.
2. Rank the scored file, in descending order by Prob_est.
3. Divide the ranked and scored file into ten *equal* groups. The Decile variable is created, which takes on ten ordered 'values': top (1), 2, 3, 4, 5, 6, 7, 8, 9, and bottom (10). The 'top' decile consists of the best 10% of individuals most likely to respond; decile 2 consists of the next 10% of individuals most likely to respond. And so on, for the remaining deciles. Accordingly, Decile separates and orders the individuals on an ordinal scale ranging from most to least likely to respond.
4. Number of Individuals is the number of individuals in each decile, 10% of the total size of the sample/file.
5. Number of Responses is the actual - not predicted - number of responses in each decile. The model identifies 865 actual responders in the top decile. In decile 2, the model identifies 382 actual responders. And so on, for the remaining deciles.

6. Decile Response Rate is the actual response rate for each decile group. It is Number of Responses divided by Number of Individuals for each decile group. For the top decile, the response rate is 18.7% ($=865/4,617$). For the second decile, the response rate is 8.3% ($=382/4,617$). And so forth, for the remaining deciles.
7. Cumulative Response Rate for a given depth-of-file (the aggregated or cumulative deciles) is the response rate among the individuals in the cumulative deciles. For example, the cumulative response rate for the top decile (10% depth-of-file) is 18.7% ($=865/4,617$). For the top two deciles (20% depth-of-file), the cumulative response rate is 13.5% ($=(865+382)/(4,617+4,617)$). Et cetera, for the remaining deciles.
8. Cum Lift - for a given depth-of-file - is the Cumulative Response Rate divided by the overall response rate of the sample/file (4.6%), multiplied by 100. It measures how much better one can expect to do with the model than without a model. For example, a Cum Lift of 411 for the top decile means that when soliciting to the top 10% of the file based on the model, one can expect 4.11 times the total number of responders found by randomly selecting 10%-of-file. The Cum Lift of 296 for top two deciles means that when selecting to 20% of the file based on the model, one can expect 2.96 times the total number of responders found by soliciting 20%-of-file without a model. And so and so, for the remaining deciles.

Slide-show Construction of a Response Decile Table

What is the Decile Table?

Model Performance Criterion

... is the DECILE ANALYSIS.

1. **Apply** model to the file (score the file).
2. **Rank** the scored file, in descending order.
3. **Divide** the ranked file into 10 equal groups.
4. **Calculate** Cum Lift.
5. **Assess** model performance.
The **best** model identifies the **most** response in the **upper deciles**.

Target variable can be binary Response, or any continuous performance measure, e.g., Profit.

Slide 1: What is the Decile Table?

What is the Decile Table?

RESPONSE Model Criterion

- How well the model **correctly classifies** response in the **upper** deciles.
- Perfect Response Model
 - among 100 individuals
 - 40 responders
 - 60 nonresponders.

| Decile | Number of Individuals | Number of Responses |
|--------|-----------------------|---------------------|
| 1 | 10 | 10 |
| 2 | 10 | 10 |
| 3 | 10 | 10 |
| 4 | 10 | 10 |
| 5 | 10 | 0 |
| 6 | 10 | 0 |
| 7 | 10 | 0 |
| 8 | 10 | 0 |
| 9 | 10 | 0 |
| bottom | 10 | 0 |
| Total | 100 | 40 |

Slide 2: Response Model Criterion

What is the Decile Table?

RESPONSE Model Goal

- We seek a model that identifies the **maximum** responses in the **upper** deciles.

| Decile | Total Response |
|--------|----------------|
| top | max |
| 2 | max |
| 3 | max |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| bottom | |

Slide 3: Response Model Goal

What is the Decile

Response Decile Analysis

One can expect 4.11 times the responders obtained using no model, targeting the "top" decile.

| Decile | Number of Customers | Number of Responses | Decile Response Rate | Cum Response Rate | Cum Response Lift |
|--------|---------------------|---------------------|----------------------|-------------------|-------------------|
| 1 | 4,617 | 865 | 18.7% | 18.7% | 411 |
| 2 | 4,617 | 382 | 8.3% | 13.5% | 296 |
| 3 | 4,617 | 290 | 6.3% | 11.1% | 244 |
| 4 | 4,617 | 128 | 2.8% | 9.0% | 198 |
| 5 | 4,617 | 97 | 2.1% | 7.6% | 167 |
| 6 | 4,617 | 81 | 1.8% | 6.7% | 146 |
| 7 | 4,617 | 79 | 1.7% | 5.9% | 130 |
| 8 | 4,617 | 72 | 1.6% | 5.4% | 118 |
| 9 | 4,617 | 67 | 1.5% | 5.0% | 109 |
| bottom | 4,617 | 43 | 0.9% | 4.6% | 100 |
| TOTAL | 46,170 | 2,104 | 4.6% | | |

Slide 4: Response Decile Analysis Top/(1) Decile Cum Lift

What is the Decile

Response Decile Analysis

One can expect **2.96** times the responders obtained using no model targeting the "top 2" deciles.

| Decile | Number of Customers | Number of Responses | Decile Response Rate | Cum Response Rate | Cum Response Lift |
|--------------|---------------------|---------------------|----------------------|-------------------|-------------------|
| 1 | 9,234 | 1,247 | 18.7% | 18.7% | 411 |
| 2 | | | 8.3% | 13.5% | 296 |
| 3 | 4,617 | 290 | 6.3% | 11.1% | 244 |
| 4 | 4,617 | 128 | 2.8% | 9.0% | 198 |
| 5 | 4,617 | 97 | 2.1% | 7.6% | 167 |
| 6 | 4,617 | 81 | 1.8% | 6.7% | 146 |
| 7 | 4,617 | 79 | 1.7% | 5.9% | 130 |
| 8 | 4,617 | 72 | 1.6% | 5.4% | 118 |
| 9 | 4,617 | 67 | 1.5% | 5.0% | 109 |
| bottom | 4,617 | 43 | 0.9% | 4.6% | 100 |
| TOTAL | 46,170 | 2,104 | 4.6% | | |

Slide 5: Response Decile Analysis Top-two/(1+2) Deciles Cum Lift

The Bootstrap

Bootstrapping alludes to a German legend about Baron Munchhausen, who was able to lift himself out of a swamp by pulling himself up by his own hair. In later versions, he was using his own boot straps to pull himself out of the sea that gave rise to the term bootstrapping. A bootstrap was a loop of leather sewn onto the back of each boot to hold onto when pulling boots onto ones feet. Bootstraps were still being used on leather boots during the early 20th century. In popular fiction when a poor boy became wealthy through his own efforts, he was said to have "pulled himself up by his own bootstraps". This metaphor continued into business financing where a highly profitable business might grow rapidly without external financing. [3]

In statistics, the *bootstrap* is a method to determine the trustworthiness of a statistic, like the standard deviation, a measure of variability, or Cum Lift, a measure of model predictiveness of identifying the upper performing individuals often displayed in a decile table. In other words, the *bootstrap* method is a justly procedure to determine the dependability of any statistic.

The bootstrap is a computer-intensive approach to statistical inference. [4] It is the most popular resampling method, using the computer to extensively resample the sample at hand. [5, 6] By random selection with replacement from the sample, some individuals occur more than once in a *bootstrap* sample, and some individuals occur not at all. Each same-size bootstrap sample will be slightly different from one another. This variation makes it possible to induce an empirical sampling distribution of the desired statistic, from which estimates of bias and variability are determined. [7]

The bootstrap is a flexible technique for assessing the accuracy (closeness-to-true value) and precision of any statistic. For everyday statistics, such as the mean, the standard deviation, regression coefficients and R-squared, the bootstrap provides an alternative to traditional parametric methods. For statistics with unknown properties, such as the median and Cum Lift, traditional parametric methods do not exist. Thus, the bootstrap provides a viable alternative over the inappropriate use of traditional methods, which yield risky results.

The bootstrap falls also into the class of non-parametric procedures. It does not rely on unrealistic parametric assumptions. Consider testing the significance of a variable in a regression model built using ordinary least-squares estimation. [8] Say, the error terms are not normally distributed, a clear violation of the least-squares assumptions. [9] The significance testing may yield inaccurate results due to the model assumption not being met. In this situation, the bootstrap is a feasible approach in determining the significance of the coefficient without concern of any assumptions. As a non-parametric method the bootstrap does not rely on theoretical derivations required in traditional parametric methods.

How To Bootstrap [10]

The key assumption of the bootstrap is that the sample is the best estimate of the unknown population. [11] Treating the sample as the population, the analyst repeatedly draws same-size random samples with replacement from the original sample. The analyst estimates the desired statistic's sampling distribution from the many bootstrap samples, and is able to calculate a bias-reduced bootstrap estimate of the statistic, and a bootstrap estimate of the standard error of the statistic. The bootstrap procedure is listed neatly in ten steps in Slide 7, below. I provide a simple illustration of bootstrapping the familiar standard deviation in Slides 8 and 9, below. The computer programming for doing the bootstrap requires minimal coding skills. [Interested in my little app for bootstrapping?](#)

HOW TO BOOTSTRAP

1. State desired calculation, say Y .
2. Treat sample as population.
3. Calculate Y on `sam_pop`; call it `SAM_EST`.
4. Draw a bootstrap sample from the population, i.e., an independent random selection with replacement (of size n).
5. Calculate Y on the bootstrap sample to produce a pseudo-value; call it BS_1 .
6. Repeat steps 4 and 5 ' m ' times.
7. We have: $BS_1, BS_2, BS_3, \dots, BS_m$.
8. $BSest(Y) = 2 \times SAM_EST - mean(BS)$
9. $SE_{BS} = StdDev(BS)$
10. $95\%CI = BSest(Y) \pm |z_{.025}| SE_{BS}$

Slide 7: How To Bootstrap

Simple Illustration

Consider a sample from a distribution
(no reason to assume it's normal),
which produced the following eleven values:

.1 .1 .1 .4 .5 1.0 1.1 1.3 1.9 1.9 4.7

Set confidence limits on the standard deviation.

Solution

Bootstrap methods are not performed by hand.
They require computer power.

Slide 8: Simple Illustration of a Bootstrapped Standard Deviation

RESAMPLING STATS OUTPUT - Simple Illustration

Sample size = 10
of simulations = 100

SAM_STDEV = 1.3435

B =

| | | | | | |
|---------|---------|---------|---------|---------|---------|
| 1.3431 | 0.60332 | 1.7076 | 0.6603 | 1.4614 | 1.43 |
| 1.4312 | 1.2965 | 1.9242 | 0.71063 | 1.3841 | 1.7656 |
| 0.73151 | 0.70404 | 0.643 | 1.6366 | 1.8288 | 1.6313 |
| 0.61427 | 0.76485 | 1.3417 | 0.69474 | 2.2153 | 1.2581 |
| 1.4533 | 1.353 | 0.479 | 0.62902 | 2.1982 | 0.73666 |
| 0.66341 | 1.4098 | 1.8892 | 2.0633 | 0.73756 | 0.69121 |
| 1.2893 | 0.67032 | 1.7316 | 0.60083 | 1.4493 | 1.437 |
| 1.3671 | 0.49677 | 0.70309 | 0.51897 | 0.65701 | 0.59898 |
| 1.3784 | 0.7181 | 1.4312 | 1.3985 | 0.6603 | 1.2972 |
| 0.47854 | 2.0658 | 1.7825 | 0.63281 | 1.8755 | 0.39384 |
| 0.69194 | 0.6343 | 1.31 | 1.3491 | 0.70079 | 1.3754 |
| 0.29609 | 1.5522 | 0.62048 | 1.8657 | 1.3919 | 1.6596 |
| 1.3726 | 2.0877 | 1.6659 | 1.4372 | 0.72111 | 1.4356 |
| 0.83327 | 1.4056 | 1.7404 | 1.796 | 1.7957 | 1.3994 |
| 1.399 | 1.3653 | 1.3726 | 1.2665 | 1.2874 | 1.8172 |
| 1.322 | 0.56569 | 0.74863 | 1.4085 | 1.6363 | 1.3802 |
| 0.60194 | 1.9938 | 0.57937 | 0.74117 | | |

Mean_BS = 1.2034

SE_boot = 0.50084

Thus, BOOTSTRAP estimate of the standard deviation is

$$2 * 1.3435 - 1.2034 = 1.4836$$

The 95% confidence interval for the standard deviation:

$$1.4836 \pm |z_{.025}| * 0.50084$$

$$1.4836 \pm 1.96 * 0.50084$$

$$1.4836 \pm 0.9816$$

$$(0.502 , 2.465)$$

Slide 9: Result of Illustration

I demonstrate the bootstrap as a nonparametric alternative technique of estimating the standard deviation, and provide a parametric vs. bootstrap comparison in Slide 10, below. (Of course, you suspected that the sample was drawn from a normal distribution.) Examination of the two methods to establish similarities and dissimilarities are left to the reader. Notation: Rows N and BS represent the CI estimates for the normal and bootstrap, respectively.

| | |
|---|---------------------------------|
| <i>Estimate Comparison</i> | |
| (Sample was from a Normal distribution.) | |
| N: | .93xxxxxxxxxxxxxxxxx2.35 |
| BS: | .50xxxxxxxxxxxxxxxxx2.48 |

Slide 10: Parametric vs. Bootstrap Estimates Comparison

The Decile Table with the Missing the Confidence Interval

Following the steps in Slide 7, the data analyst can obtain the missing confidence interval (CI) estimates in the decile table in Slide 11, below. Interpretation of the decile-level CI estimates warrants no modification of the CI estimates meaning as presented in Statistical Inference, page 1, item 8. Notation used is (Lower, Upper), which represents the lower and upper limits of the CI estimates.

Response Decile Analysis with Confidence Interval

| Decile | Number of Customers | Number of Responses | Decile Response Rate | Cum Response Rate | Cum Response Lift | 95% CI (Lower, Upper) |
|--------------|---------------------|---------------------|----------------------|-------------------|-------------------|-----------------------|
| 1 | 4,617 | 865 | 18.7% | 18.7% | 411 | (399, 423) |
| 2 | 4,617 | 382 | 8.3% | 13.5% | 296 | (286, 304) |
| 3 | 4,617 | 290 | 6.3% | 11.1% | 244 | (238, 350) |
| 4 | 4,617 | 128 | 2.8% | 9.0% | 198 | (193, 203) |
| 5 | 4,617 | 97 | 2.1% | 7.6% | 167 | (163, 171) |
| 6 | 4,617 | 81 | 1.8% | 6.7% | 146 | (142, 150) |
| 7 | 4,617 | 79 | 1.7% | 5.9% | 130 | (127, 133) |
| 8 | 4,617 | 72 | 1.6% | 5.4% | 118 | (115, 121) |
| 9 | 4,617 | 67 | 1.5% | 5.0% | 109 | (106, 111) |
| bottom | 4,617 | 43 | 0.9% | 4.6% | 100 | (100,100) |
| TOTAL | 46,170 | 2,104 | 4.6% | | | |

Slide 11: Decile Table with the Missing the Confidence Interval

Conclusion:

The decile table has become for most modelers a universal tabular display of model performance. The decile table includes the point estimate of the statistic Cum Lift, which indicates how much better is a given model than the chance model. The confidence interval that furnishes the precision of the model is needed to complete the decile-table model assessment. I review briefly the basics of statistical inference, outline the bootstrap method, provide a comparison, at least for the standard deviation, of how the bootstrap fares against the parametric method. I set out successfully to augment the decile table with the missing confidence interval in the decile table. I open to discussion the value of the confidence interval estimates added to the standard decile table, as well as, the bootstrap route for estimation of any statistic, not only the Cum Lift.

Reference:

- 1 - The decile table has ten rows of equal number of individuals, irrespectively of model score. There can be individuals with the same model score in adjacent deciles. In a gains chart, there are as many rows as there are distinct model scores. Thus, there are no individuals with the same model score across gains-chart rows.
- 2 - Galton, F., "Report of the Anthropometric Committee," in *Report of the 51st Meeting of the British Association for the Advancement of Science*, 1882, pp. 245-260.
- 3 - Wikipedia

- 4 - Noreen, E. W., *Computer Intensive Methods for Testing Hypotheses*, John Wiley & Sons, Inc., 1989.
- 5 - Other resampling methods include the jackknife, infinitesimal jackknife, delta method, influence function, and random subsampling.
- 6 - Efron, B., *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, 1982.
Accuracy includes bias, variance, and error.
- 7 - A sampling distribution can be considered as the frequency curve of a sample statistic from an infinite number of samples.
- 8 - Is the coefficient equal to zero?
- 9 - Draper, N. R. and Smith, H., *Applied Regression Analysis*, John Wiley & Sons, Inc., 1966
- 10 - This bootstrap method is the normal approximation. Others are Percentile, B-C Percentile, and Percentile-t
- 11- Actually, the sample distribution function is the nonparametric maximum likelihood estimate of the population distribution function.