Quality Auditor Consistency

Introduction

Auditing calls is a critical element of quality monitoring. Many call centers have established quality-monitoring guidelines, which independent auditors are supposed to follow when evaluating calls.

Many hours of planning and calibration can go into designing a quality monitoring system and then training an auditing team to implement the system. In an ideal world, every auditor would interpret the scoring guidelines the same way, thereby scoring each call exactly the same way.

Of course, we do not live in an ideal world. Even the most specific of guidelines can be open to interpretation when exposed to real-world calls.

Still, we need auditors to follow the guidelines to a certain degree of consistency or quality scoring loses its credibility. How can we assess whether scoring differences by auditors are due to randomness, bias, or other factors?

Fortunately, statistical tools are available to help make such determinations. One such tool is the chi-square test of independence. This test can be used to determine whether there is evidence of auditor bias or not.

This test and corresponding data visualization are freely available from an open source statistical software program called R.

The plots and tests in this article are based on functions from a contributed R package called *vcd*, which stands for "Visualizing Categorical Data." The authors of this package are David Meyer, Achim Zeileis, and Kurt Hornik[1] .

What is a chi-square test for independence?

The chi-square test for independence tests whether two categorical variables are associated or not. One way to visualize these associations is through association plots, which are graphical ways of displaying variances between factors.

---

[1] David Meyer, Achim Zeileis, and Kurt Hornik, *The Strucplot Framework: Visualizing Multi-way Contingency Tables with vcd*. http://cran.r-project.org/web/packages/vcd/vignettes/strucplot.pdf

Quality Auditor Consistency

The example below shows an association plot of the distribution of hair and eye color of 592 statistics students from the University of Delaware in 1974. This tests whether the eye color and hair color are independent for this population.

Figure 1 shows the result of a chi-square test for independence between hair and eye color. We are testing whether hair and eye color combinations are due to randomness, or if some hair and eye color combinations tend to occur together.

The *p-value* of the test indicates whether there is evidence of independence or not. If the p-value is greater than 0.05, then we say that the variables are independent. If the p-value is less than 0.05, then the variables are dependent.

The p-value for this test (bottom right of plot) is much less than 0.05. This indicates that hair and eye color are dependent on each other. For instance, people with black hair are more likely to have brown eyes than blue eyes.

In addition to the p-value from the test, we have boxes of various heights and colors for each combination of hair and eye color:

1. The width of each rectangle is associated with the number of data points in that category. The wider the box, the higher the number of data points.

2. The heights and colors of each rectangle are related to how different the percentages are to what is expected of independence:

    a. Deep red – the percentage is much lower than expected

    b. Light red – the percentage is slightly lower than expected

    c. Gray – the percentage is about expected

    d. Light blue – the percentage is slightly higher than expected

    e. Deep blue – the percentage is much higher than expected.

In this example, the deep blue bar for Brown eyes and Black hair indicates that this combination occurs together too frequently to be random. In addition, the deep red bar for Brown eyes and Blond hair shows that this combination occurs together too infrequently to be random.

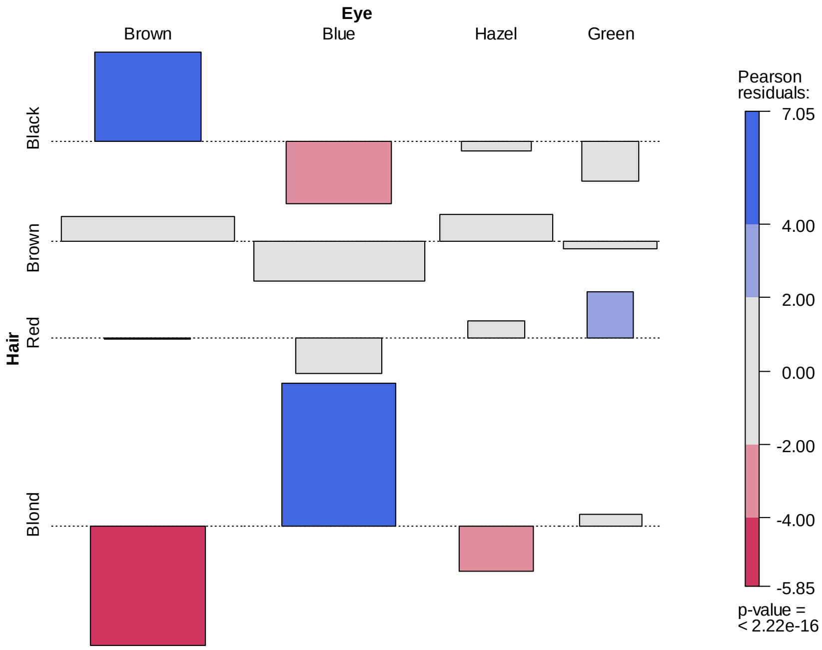

Figure 1. Association plot of hair and eye color

<u>How can these tests and plots help assess auditor consistency?</u>

If each auditor is assigned a random set of calls, then the scoring tendencies should be the same for each auditor. In auditor consistency, we test whether call scoring is independent of the auditor or not. An example best illustrates this method.

The tables below show summarized example data for a team of eight reviewers assessing calls by an attribute called “Politeness” by month. The possible ratings for Politeness are Poor, Average, and Superior.

Quality Auditor Consistency

June

Politeness Rating	A.G.	B.R.	D.T.	E.H.	E.F.	G.T.	L.W.	R.V.
Poor	2	0	4	1	7	6	2	1
Average	27	17	37	32	21	36	34	87
Superior	4	1	16	2	0	3	2	4

July

Politeness Rating	A.G.	B.R.	D.T.	E.H.	E.F.	G.T.	L.W.	R.V.
Poor	1	0	2	0	5	2	0	2
Average	46	13	25	34	43	16	20	66
Superior	5	4	24	3	4	5	1	13

August

Politeness Rating	A.G.	B.R.	D.T.	E.H.	E.F.	G.T.	L.W.	R.V.
Poor	0	2	5	0	3	7	2	5
Average	11	20	28	20	26	46	17	55
Superior	3	1	25	3	1	4	2	11

September

Politeness Rating	A.G.	B.R.	D.T.	E.H.	E.F.	G.T.	L.W.	R.V.
Poor	9	1	5	1	5	3	0	13
Average	35	27	50	13	41	28	28	105
Superior	4	2	7	2	0	4	2	9

The differing numbers of call reviews by auditor makes interpreting data in this form a challenge. It is difficult to interpret auditor-scoring tendencies directly from this table. Association plots can make this task easier as well as provide a statistical basis of comparison.

Figure 2 shows the trend for this group on Politeness based on the data above. Each panel shows an association plot for a particular month from June through September. The initials at the top of each panel (A.G., B.R., D.T., E.H., E.F., G.T., L.W., and R.V.) are the initials for the auditors. The ratings on the left side of the panel are Poor, Average, and Superior.

The June plot in the upper left shows a vast majority of gray bars and three colored bars. The Superior rating for auditor D.T. is a deep blue, indicating that D.T. tended to score Politeness as “Superior” far more frequently than the other auditors did. Meanwhile, auditor “E.F.” has a light blue bar for “Poor,” indicating that E.F. tended to score Politeness as Poor slightly more frequently than the other auditors did. R.V. has a light red bar for “Poor,” indicating the R.V. tended to score Politeness as Poor slightly less frequently.

The June data indicates that auditors D.T., E.F., and R.V. have some scoring proportions that are significantly different from the rest of the team. Either the auditors are showing some bias in scoring this category, or other factors in the call samples are driving the scoring differences.

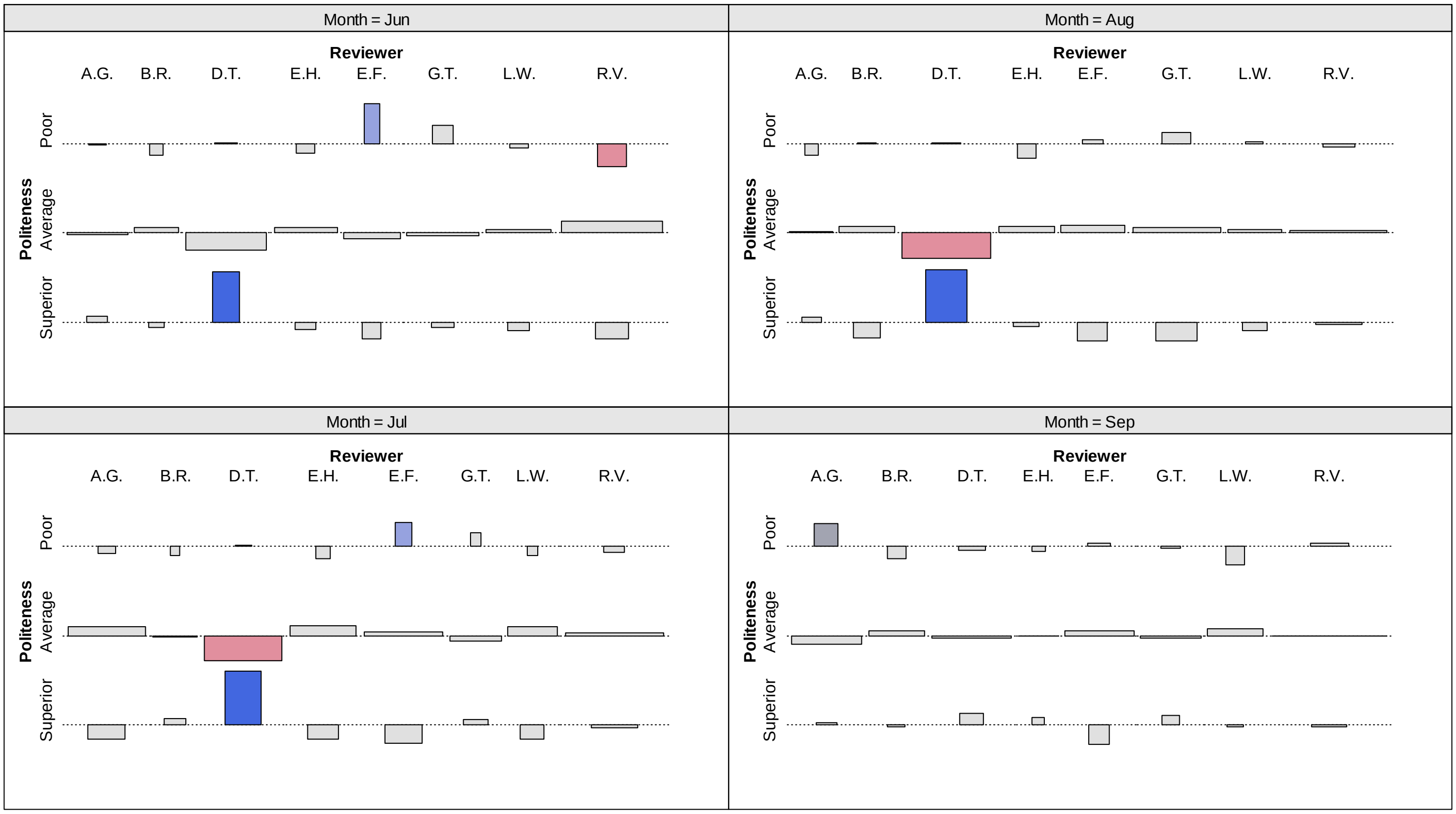

Figure 2. Association plots of Politeness ratings by month for a team of eight auditors

In the next month, July, auditor D.T. still shows a tendency to score Politeness as Superior more often than others do. We also see a light red bar for D.T. – Average. This indicates that D.T. may be scoring some calls as Superior that others score as Average. E.F. still has the light blue for Poor, which is a slight indication of potential bias, since this occurred two months in a row.

In August, D.T. persists as the only auditor with scores that are out of calibration. The three consecutive months of scoring differences indicate a trend. Either D.T.'s call samples have Superior Politeness ratings due to an outside factor (e.g. perhaps a particular site tends to emphasize Polite calls); or, D.T. has a tendency to interpret some calls as Superior that others would interpret as Average.

Finally, September's data shows gray bars almost universally across the board. All of the bars are now gray, indicating improved auditor consistency.

<u>Conclusion</u>

Association plots are powerful tools for revealing scoring tendencies and sources of bias. How to handle these biases is the next task, which can be the greater challenge. Coaching and calibration sessions are the preferred methods of addressing auditor bias. Association plots can be part of such sessions, but the emphasis should be on following scoring guidelines in a consistent manner rather than the raw numbers.

Sam Thomas is a Principal Consultant and Data Analyst for Revelant Technologies, LLC. In his role, Sam applies advanced statistical methods to call center operations to identify improvement opportunities. Sam has a Master of Science in Mathematics and is an Associate of the Society of Actuaries. Contact Sam at sam.thomas@revelanttech.com or 317.696.9214.