The Odds of Dark Skin Toned Players of Receiving Red Cards Are Higher than the Odds of Light Skin Players of Receiving Red Cards[1]

Ismael Flores Cervantes[1,2]*

Affiliations

[1]Senior Statistician, Westat

[2]PhD student at the Joint Program of Survey Methodology at the University of Maryland

*Correspondence to: ismaelflorescervantes@westat.com

Abstract

A logistic analysis with repeated measure shows that the odds of a dark skin toned player receiving a red card are 1.39 (1.10, 1.75) times higher than the odds for a light skin toned player receiving it. The results lend some support to research on assimilation to stereotypes in social perception and cultural preferences for light skin which predicts an association between darker skin tone and receiving more red cards. The analysis is not without problems due to a large percentage of missing values in the data and auxiliary variables with measurement errors not accounted in the analysis.

Summary

The odds of a dark skin toned player (scale=1) receiving a red card are 1.39 times higher than the odds for a light skin toned player (scale=0) receiving a red card. The 95% confidence interval of the odd ratio is (1.10, 1.75).

[1] The analysis is based on logistic regression. There is confusion on how the odd ratios should be reported because most researcher and readers misinterpret odds ratios as relative risks. In this submission, the odd ratios are spelled out to avoid this situation. However, the title becomes cumbersome.

Approach

In order to explain our approach we borrow concepts from clinical trial and observational studies and use the analogy of patient subject to different treatments. In this case, the player is the "patient" which receives different "treatment" or encounters with a referee. The treatments are different (different referees with different implicit and explicit cultural preferences for light skin tone), the results of the treatment are red card or not[2]. There are auxiliary variables defined at the player/patient level such as age, weight, height, position, and skin tone. Also a treatment, which can be applied more than once, is represented for multiple games with the same referee. Since it is the same player (i.e., "patient"), the observation and treatments are clustered within player. In other words, the data are correlated. Any analysis that do not account for the correlated data (i.e., assuming independence) will underestimate the standard errors and incorrect inferences may be made. Since we are modeling getting a red card or not, the dependent variable has a binary outcome that is analyzed using generalized linear model with *logit* as the link function. This type of analysis corresponds to logistic regression. We use SUDAAN to do the analysis. SUDAAN offers design corrected analysis of clinical trial data and experiments or observational studies. We use SAS to clean the data.

Problems with implemented approach

Logistic regression is based on generalized linear models (GLM). In linear models, auxiliary variables with measurement errors can produce biased estimates (Fuller, 1987). The data have continuous auxiliary variables with measurement errors for MEANIAT and MEANEXP and standard errors are provided (variables SEIAT and SEEXP). The variables RATER1 and RATER2 are measuring a latent variable for skin tone (that is no necessarily on the scale for RATER). These two categorical measurements can be seen as multiple measurement of this variable. The differences between the two assessments can be modeled as measurement errors reflected as misclassification between RATER1 and RATER2. Although there is software for regression with corrections of estimates taking into account categorical and continuous auxiliary variables with measurement errors, no function is available to account for the effect due to clustering of players and generalized models (i.e., logistic regression). At least on sensitivity analysis need to be done for accounting for these errors. Otherwise, there will be some reluctance to accept these results. Programing new functions in R can solve this problem but it requires more time than the one allocated for this analysis.

The second problem is the large percentage of missing data. The analysis assumes a listwise approach, where observations with missing values (skin tone, and the measures for preferences for light skin) are excluded from the model. Basically we assume that data missing is at random (MAR). This assumption is not likely to hold. Imputation of missing values is a possible alternative but there are not many auxiliary variables to carry out the imputation. Another approach is to repeat the analysis observing how the estimates change when all missing values are assumed either dark or light skin or when all missing measures for preferences for light skin

[2] Since when one player received a red card, there player is expel of the game. In other words, a player cannot receive more than one red card per game. As a result, we considered the "event" of getting a red card ad a success.

are either high or low. This would help to examine the extreme cases and determine if the conclusion are robust. Such analysis is not included in the implemented approach.

Data file

The analysis was based on a file provided by the *Crowdstorming Research: Many analysts, one dataset Research Protocol* coordinated by Raphael Silberzahn, Eric Uhlmann, Dan Martin, and Brian Nosek.

The variables shown in Table 1 were created in this analysis:

Table 1. Derived Variables

Variable	Definition	Comments
PV	VICTORIES/sum(TIES, VICTORIES, DEFEATS);	Proportion of victories for the player/referee dyad
AGE	Based on BIRTHDAY	Age of player
N_POSITION	Sequential values of ordered POSITION	Numeric values of player position (software only uses numeric values)
N_LEGUE	Sequential values of ordered LEAGUECOUNTRY	Numeric values of league (software only uses numeric values)
PAYERID	Sequential values of ordered PLAYER	Numeric values of PLAYER (software only uses numeric values)
N_CLUB	Sequential values of ordered CLUB	Numeric values of PLAYER (software only uses numeric values)
CMEANEXP	Quartiles of MEANEXP	Sequential values 1 to 4 for quartiles of MEANEXP
CMEANIAT	Quartiles of MEANIAT	Sequential values 1 to 4 for quartiles of MEANIAT

Initial Approach

The initial approach used LEAGUE as an independent variable. Although we observed that there variables had the same value for a single player, we did not realize that there were games where the player belonged to another league. In other words, the variable did not represent the league the player belonged when the game was played. Unfortunately we had found a large interaction between league and skin tone. Those initial results were barely significant without that interaction term. Once the interaction was include in the model, the results become significant. The reason is that the OR varied by league and was not significant in some of the leagues (being most significant in one league).

After the feedback we realized that we could not justify the use of LEAGUE because it reflected either the current league of the player or the league the player has played most games (i.e., modal league).

Also based on the feedback, we decided to do a descriptive analysis (i.e. table) to test if there was an effect due to skin tone before including any modeling. SUDAAN provides a procedure CROSSTAB that can be used to test the association of categorical variables such as received a red card and skin tone (accounting for clustered data with repeated values). We used the Cochran–Mantel–Haenszel (CMH) test, which allows the comparison of two categorical variables. The test is often used in observational studies where random assignment of subjects to different treatments cannot be controlled.

Final Approach

The final approach uses two analyses. The first analysis is descriptive where we test the association between getting a red card and skin tone. The second analysis uses logistic regression. All analyses reflect the effect of clustered and correlated data.

Descriptive analysis and Chi-square tests

1. Table TABLE Y *RATER1 where Y=1 (red card), 0 (no red card)

The proportions of players-games with red cards are shown in Figure 1. The figure shows that the proportion of player-games pairs with red cards increases from 0.38, 0.44, 0.45, 0.47 and 0.52 as the value or rater increases from 0, 0.25, 0.5, 0.75 and 1.

The Cochran–Mantel–Haenszel (CMH) test of association, shown in Figure 2 is barely significant (0.059). However, is we use the CMH test for trend (that is more strict than the CMH for general association) the test shows that there is a trend between the proportions of red cards and the ordinal values of RATER1 that is significant (0.0047).

Figure 1. Proportion of players-games pairs with a red flag by skin tone

Y		RATER1					
		Total	0	0.250	0.500	0.750	1
1	Sample Size	1589	567	595	175	130	122
	Weighted Size	1589.00	567.00	595.00	175.00	130.00	122.00
	SE Weighted	56.57	38.33	38.17	22.55	18.22	17.46
	Col Percent	0.43	0.38	0.44	0.45	0.47	0.52
	SE Col Percent	0.01	0.02	0.02	0.04	0.05	0.06
	Lower 95% Limit COLPER	0.40	0.34	0.40	0.38	0.38	0.42
	Upper 95% Limit COLPER	0.45	0.43	0.49	0.54	0.57	0.64

Figure 2. Cochran–Mantel–Haenszel (CMH) test for Y *RATER1

```
Hypothesis Test
   Test Statistic              Test
                         DF   Value     P-Value
-----------------------------------------------
CMH General
  Association
   Wald chi-square        4     9.08     0.0591
   Wald-F                 4     2.27     0.0595
   Adj Wald F             4     2.27     0.0598
CMH Trend
   Wald chi-square        1     7.99     0.0047
   Wald-F                 1     7.99     0.0047
   Adj Wald F             1     7.99     0.0047
-----------------------------------------------
```

2. Table TABLE   Y *RATER1 controlling by N_POSITION

In the second table, we controlled by position. There are no much differences between these reulsts and the previous table.

Figure 3. Cochran–Mantel–Haenszel (CMH) test for Y *RATER1 controlling by N_POSITION

```
  Test Statistics for Stratum-Adjusted Hypotheses
  Variable Y by Variable RATER1
  Controlling for: Variable N_POSITION
by: Hypothesis Test, Test Statistic.

-------------------------------------------------
Hypothesis Test               Test
   Test Statistic        DF   Value       P-Value
-------------------------------------------------
CMH General
  Association
   Wald chi-square        4     9.7046     0.0457
   Wald-F                 4     2.4262     0.0461
   Adj Wald F             4     2.4226     0.0463
CMH Trend
   Wald chi-square        1     8.7217     0.0031
   Wald-F                 1     8.7217     0.0032
   Adj Wald F             1     8.7217     0.0032
-------------------------------------------------
```

3. Table TABLE   Y *RATER1 controlling by N_POSITION and CMEANEXP (quartiles of CMEANEXP)

After controlling by N_POSITION and CMEANEXP, Figure 4 shows the CMH test of association is significant after controlling for position and CMEANEXP. The next step is to quantify these relationships.

Figure 4. Cochran–Mantel–Haenszel (CMH) test for Y *RATER1 controlling for: Variable N_POSITION and CMEANEXP.

```
Variance Estimation Method: Taylor Series (WR)
  Test Statistics for Stratum-Adjusted Hypotheses
  Variable Y by Variable RATER1
  Controlling for: Variable N_POSITION and CMEANEXP
by: Hypothesis Test, Test Statistic.

-------------------------------------------------
Hypothesis Test              Test
   Test Statistic        DF   Value       P-Value
-------------------------------------------------
CMH General
  Association
   Wald chi-square        4    11.2026     0.0244
   Wald-F                 4     2.8007     0.0246
   Adj Wald F             4     2.7966     0.0248
CMH Trend
   Wald chi-square        1    10.9317     0.0009
   Wald-F                 1    10.9317     0.0010
   Adj Wald F             1    10.9317     0.0010
-------------------------------------------------
```

Logistic Regression

Once we have determined that different skin tone player receive different red cards, we proceed to quantify these relationships.
We use logistic regression accounting for multiple measures and clustered data. The most parsimonious models are

MODEL Y= RATER1 MEANIAT PV N_POSITION * MEANIAT ;

And

MODEL Y= RATER1 MEANEXP PV N_POSITION * MEANEXP ;

for the two measures of cultural preference for lighter skin tone. Notice that there is an interaction term for N_POSITION and MEANEXP (or MEANIAT). Although interaction terms in logistic regression are not easy to interpret, the overall significance of RATER1 depends on the inclusion of this term. Including the interaction tem made consistent the results between the logistic regression and the Chis-squared tests in the first part of the analysis.

The odd ratios (OR) for skin tone (lighter skin tone as the reference) for the first model are shown in Figure 5.
The odds for dark skinned toned players (RATER1=1) to receive a red card are 1.39 higher than the odds of light skin players (RATER1=0) to receive a red card[3]. Although it is significant

[3] Notices that we are not describing this result as dark skin toned players are 1.39 times more likely to receive a red card than a light skin toned player. The previous sentence describes the relative risk (p1/p2) and not the odd ratio (p1/(1-p1))/(p2/(1-p2)). The correct interpretation of the OR of 1.39 is that for every dark skin toned player who does not receive a red card, 1.39 times as many dark skin toned players will receive a red card than the number of light skinned tone players receiving a red card for every light skin player who does not receive one (Osborne 2006).

(0.5%), is still close to 1. Note the progressive odds ratios (1.14, 1.22, 1.29, and 1.36) observed from the previous analysis (i.e., trend). However, some of these are not statistically different from RATE=0.

Figure 5. Odd ratio for RATER and MEANIAT

Independent Variables and Effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
Intercept	0.00	0.00	0.00
RATER1			
0	1.00	1.00	1.00
0.250	1.14	0.99	1.31
0.500	1.22	1.00	1.50
0.750	1.29	1.03	1.62
1	1.36	1.08	1.72
MEANIAT	30.89	4.05	235.59

We do not get very different results if we use MEANEXP as shown in the Figure 6.

Figure 6. Odd ratio for RATER and MEANEXP

Independent Variables and Effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
Intercept	0.00	0.00	0.01
RATER1			
0	1.00	1.00	1.00
0.250	1.14	0.99	1.31
0.500	1.21	0.99	1.49
0.750	1.27	1.01	1.59
1	1.36	1.08	1.71
MEANEXP	0.63	0.34	1.18

References

Osborne JW (2006) Bringing balance and technical accuracy to reporting odds ratios and the results of logistic regression analyses. *Practical Assessment, Research & Evaluation* 11(7):1-6.

Research Triangle Institute (2012). *SUDAAN Language Manual, Volumes 1 and 2, Release 11*. Research Triangle Park, NC: Research Triangle Institute.

SAS Institute Inc. (2012). *SAS/STAT ® 9.3.* Cary, NC: SAS Institute, Inc.

Fuller WA. (1987). Measurement Error models. New York: John Wiley & Sons.