Zimmerman, Dale L.; Tang, Jun; Huang, Rui
Outline analyses of the called strike zone in major league Baseball. (English) [Zbl 1435.62445]

Summary: We extend statistical shape analytic methods known as outline analysis for application to the
strike zone, a central feature of the game of baseball. Although the strike zone is rigorously defined by
Major League Baseball’s official rules, umpires make mistakes in calling pitches as strikes (and balls) and
may even adhere to a strike zone somewhat different than that prescribed by the rule book. Our methods
yield inference on geometric attributes (centroid, dimensions, orientation and shape) of this “called strike
zone” (CSZ) and on the effects that years, umpires, player attributes, game situation factors and their
interactions have on those attributes. The methodology consists of first using kernel discriminant analysis
to determine a noisy outline representing the CSZ corresponding to each factor combination, then fitting
existing elliptic Fourier and new generalized superelliptic models for closed curves to that outline and
finally analyzing the fitted model coefficients using standard methods of regression analysis, factorial
analysis of variance and variance component estimation. We apply these methods to PITCHf/x data
comprising more than three million called pitches from the 2008–2016 Major League Baseball seasons to
address numerous questions about the CSZ. We find that all geometric attributes of the CSZ, except its
size, became significantly more like those of the rule-book strike zone from 2008-2016 and that several
player attribute/game situation factors had statistically and practically significant effects on many of
them. We also establish that the variation in the horizontal center, width and area of an individual
umpire’s CSZ from pitch to pitch is smaller than their variation among CSZs from different umpires.

MSC:

62P99 Applications of statistics
62H30 Classification and discrimination; cluster analysis (statistical aspects)
62H25 Factor analysis and principal components; correspondence analysis

Keywords:
elliptic Fourier model; kernel discriminant analysis; morphometrics; orthogonal distance fitting; shape
analysis; superellipse

Software:
Momocs; XML2R; pitchRx; R

Full Text: DOI Euclid

References:


