World-Record Progress in Track Events

Potentially Novel Means to Detect Doping

Danny Bubb

NY Data Science Academy

January 31st, 2018

Why care about doping?

- modern Olympiad is worth saving
- doping is dangerous and causes health problems
- competition is vital to human endeavor
  - countries that "play" together tend not to war (like trade)

Types of Doping

- Steroids
  - reduce inflammation - cortico
  - build muscle - anabolic
- Blood doping
  - transfusions - both autologous and homologous
  - synthetic oxygen carriers and hormones (EPO, CERA)
- Insulin - increases glucose storage and bio-availability
- hGH - oh to be young again!!!

Track Events and Energy Systems

- $100m$- purely anaerobic
- $1500m$ - anywhere from $80/20 \rightarrow 84/16$
- $10000m$ - $\frac{95}{5}$

Transition between anaerobic and aerobic respiration occurs at about $15 - 30s$

How to Get Caught

- Be either stupid or unfortunate and test positive
- Get caught with drugs - i.e. Festina affair
- Biological passport
  - establishes permissable limits and tracks athlete's blood values and tests over time to detect malfeasance
- non-analytical positive -> Data Science??!!!!

Statistical Tests Performed Here

- As the density plots show, the distribution of times are not normal as they are highly skewed and kurtotic
- Levene tests tell you whether the variances are changing (homo- versus heteroskedastic)
- Wilcoxon Rank Sum Test
  - H_0 : the medians of both distributions are the same
  - H_1 : the medians differ
- Kruskal-Wallis Rank Sum Test
  - extension of Anova to non-parametric statistics

Analysis of Bookends

- $100m$ - the most telling result is that there is not an appreciable difference in the distribution of times for known dopers (n= 121) and the all-time performance list (sample $n = 121$)
  - suggests either doping is ineffective (!?) or that the two groups are really the same (everyone dopes!)
- $10000m$ - especially when considering elite times, the performances in the "EPO" era stand out
  - after an effective test for EPO is implemented, times come back down from the stratosphere
  - Despite the fact that nearly twice as many sub-27 performances come from the "post" era, no one has come within $25s$ of the record since 2011
  - Does this shed any light on the effectiveness of "microdosing"?

Analysis of 1500

- No obvious trends wrt to "eras"
  - Does this suggest that there is no EPO influence or rather that microdosing is just as effective for this event?
- Looking at WR progression through the $50's$ suggests that steroids may have had a strong influence on times
- Future analysis will focus on the number of performers who are elite in both the $1500m$ and $5000m$ races.