Variability Study Paper - Statistical Methodology

Kiegan Rice

1 Introduction

1.1 Intro to Gauge R&R

- Response measurement type
  - Example: measuring length of a screw (y)

- Utility of Gauge R&R, consider distance between two measured objects
  - Example: difference in length between two screws: $y_a - y_b = d$

- Traditional model applied is mixed-effects, emphasizes variance components for random effects

1.2 Intro to complex data structures

- Explosion of data collection and growing complexity of data structures
  - Example: considering multiple measurements on a screw
  - Background info: multivariate Gauge R&R approaches (include citations)
- Data science leverages complex data structures
  - Beginning with complex structure, apply actions to structure to achieve goal: some quantitative result
    - *figure*: pipeline sketched out with overview images of data structure at each point

  - Point: need to adapt definition of distance between two objects with complex structure and how we apply statistical models to quantify measurement variability

1.3 Example of a data science pipeline with complex data structures is automated forensic firearms comparison

- firearm barrels engrave striation patterns on fired bullets

- forensic question: same source or different source?

- data science approach translates complex striation pattern on surface into measured data

- reproducibility of measurement system applied to forensic evidence important to establish (wording needs work)

1.4 End of introduction

- using forensic firearms analysis as a motivating example
- propose approaches to adapting gauge R&R framework to two complex data structures within a data science pipeline

2 Methodology

Note: I am torn about the order of these next two sections. Do I formally introduce the R&R structure and model first, and follow that with our data structures? Or introduce the complex structures in our data science pipeline first (along with data structure), then introduce R&R framework, then come back to data structure?

2.1 defining complex data structures in forensic firearms analysis

- answering the forensic question using data science
  - first, measure objects
  - second, compare two objects using a similarity metric
- measurement process 1: translation of physical object to complex data structure
  - *figure*: resulting data structure (one signature)
  - *details*: data structure in vector notation
- measurement process 2: measurement of similarity between paired objects
  - *figure*: same-source signatures, different-source signatures
  - *details*: data structure, range of similarity scores

2.2 Traditional three-factor gauge R&R model

- define parts, operators, devices
- define model
  - $j = 1, \ldots, n_p$; $k = 1, \ldots, n_o$; $m = 1, \ldots, n_d$; $n = 1, \ldots, n_r$
  - $y_{jkmn} = \mu + p_j + o_k + d_m + po_{jk} + pd_{jm} + od_{km} + pod_{jkm} + e_{jkmn}$
  - fixed, unknown measurement mean μ and random effects...
- model assumptions
- model outcomes

2.3 Reframing three-factor R&R model for structured signature data

- structure of peaks and valleys violates assumption of measurement independence
- fixed effects structure
  - need to figure out a good way to present this

- subsampling
  - maintains model assumption of independence
  - subsampling indices
  - *figure*: autocorrelation
  - *figure*: subsampling indices
- model and model assumptions
  - $y_{ijkmn} = \mu_i + p_{ij} + o_{ik} + d_{im} + po_{ijk} + pd_{ijm} + od_{ikm} + pod_{ijkm} + e_{ijkmn}$
  - independent random variables
  - variance components
- multiple phases
  - *figure*: phased approach

2.4 Reframing three-factor R&R model for paired response data

- response is univariate measurement
- study factors are not single-factor (levels do not map one-to-one to response levels)

3 Data Collected

still need to decide on the scope we are presenting here - I think one barrel type, maybe Orange?

3.1 Study design

- three bullets (parts), two machines, eight operators
- three to five repetitions and repetition definition
- Total number of resulting signatures

3.2 separating out LEAs

- six separate LEA patterns results in six individual models, one for each LEA
- each barrel-land engraves on three LEAs, one per bullet

4 Results

4.1 LEA signatures

- results table

- summary table with $\sigma_{repeat}, \sigma_{reprod}$?
- results figure

4.2 pairwise similarity scores

- change in data structure changes results structure as well

- results table

- results figure

5 Conclusions

- operator effect shows most when we have bullets with damage
  - without maintaining the data format, units (i.e. not using Fourier analysis) we wouldn't have this interpretability

- preserve model assumptions and structure
- estimate things relevant to data we are working with

- can get actionable items for a process