

Technical Paper: Balancing Type I and Type II Errors in a  
Recommendation System for Dispatching an Ambulance Based on  
Automated Crash Reports from Cell Phones

Brad Burkman

Updated 9 November 2022



# Contents

<b>Index</b>	<b>4</b>
<b>Acronyms</b>	<b>5</b>
<b>0 Notes to Self (Temp)</b>	<b>7</b>
0.1 To Do . . . . .	8
<b>1 CRSS Data</b>	<b>9</b>
1.1 3 March 2022 File Update . . . . .	10
1.2 Derived Data . . . . .	10
<b>2 Imputing Missing Data</b>	<b>11</b>
2.1 Missing Data in the CRSS Data Set . . . . .	12
<b>Bibliography</b>	<b>14</b>



# Acronyms

**CRSS** Crash Report Sampling System. 10, 12

**IVEware** Imputation and Variance Estimation Software. 12



## Chapter 0

# Notes to Self (Temp)

0.1	To Do . . . . .	8
-----	-----------------	---

## 0.1 To Do

- Finish Imputation
- Finalize Binning
- Focus: False Positives and Negatives
- Run experiments
- Figure out model



# Chapter 1

## CRSS Data

1.1	3 March 2022 File Update . . . . .	10
1.2	Derived Data . . . . .	10

## 1.1 3 March 2022 File Update

On 3 March 2022, Crash Report Sampling System (CRSS) posted a file, `_READ ME Before Downloading Files.pdf` that they had updated the CRSS files. We downloaded new copies and compared.

On March 2, 2022, at about 11 a.m. ET, NHTSA updated data files for 2020 FARS and CRSS, including files from previous years.

[https://static.nhtsa.gov/nhtsa/downloads/CRSS/0\\_READ%20ME%20Before%20Downloading%20Files/\\_READ%20ME%20Before%20Downloading%20Files.pdf](https://static.nhtsa.gov/nhtsa/downloads/CRSS/0_READ%20ME%20Before%20Downloading%20Files/_READ%20ME%20Before%20Downloading%20Files.pdf)

## 1.2 Derived Data

Many of the features in the big data files (Accident, Vehicle, Person) are derived from other information, either as a count or as a summary.

URBANICITY does not actually tell us whether the crash occurred in town or out; it just tells us whether the PSU county had a population of 250,000 or more. Most counties have both urban and rural areas, so the distinction may not be as useful as one would think.

## Chapter 2

# Imputing Missing Data

2.1	Missing Data in the CRSS Data Set . . . . .	12
-----	---	----

## 2.1 Missing Data in the CRSS Data Set

The Crash Report Sampling System (CRSS) data set is well curated and in most features has categories signifying unknown or unreported values. The data set does not have nonsensical or misspelled values. It probably does have values that, due to errors in observation or transcription, are incorrect, but we will not address that problem.

In the WEATHER feature, for instance, ten values signify a specific kind of weather, one value indicates “other” kind of weather, one signifies Unreported, and one Reported as Unknown.

Feature	Code	Meaning	% of Samples	% Hospitalized
WEATHER	1	Clear	70.02	16.39
WEATHER	2	Rain	8.82	16.05
WEATHER	3	Sleet or Hail	0.11	18.23
WEATHER	4	Snow	1.47	12.48
WEATHER	5	Fog, Smog, Smoke	0.33	22.50
WEATHER	6	Severe Crosswinds	0.05	15.79
WEATHER	7	Blowing Sand, Soil, Dirt	0.02	12.50
WEATHER	8	Other	0.06	14.93
WEATHER	10	Cloudy	14.35	15.77
WEATHER	11	Blowing Snow	0.04	13.67
WEATHER	12	Freezing Rain or Drizzle	0.02	13.12
WEATHER	98	Not Reported	4.54	13.20
WEATHER	99	Reported as Unknown	0.16	10.77

The CRSS authors used Imputation and Variance Estimation Software (IVEware) [1] to impute values for codes 98 and 99, making the feature `WEATHR_IM`.

Code	Meaning	% of Samples (Original)	% of Samples (Imputed)	Difference	% of Missing Values	% Hospitalized
1	Clear	70.0227	73.3489	3.3262	70.8335	16.2245
2	Rain	8.8223	9.2981	0.4758	10.1325	15.9803
3	Sleet or Hail	0.1141	0.1206	0.0065	0.1384	18.0180
4	Snow	1.4716	1.5444	0.0728	1.5503	12.3618
5	Fog, Smog, Smoke	0.3332	0.3548	0.0216	0.46	21.6973
6	Severe Crosswinds	0.0531	0.0602	0.0071	0.1512	14.1753
7	Blowing Sand, Soil, Dirt	0.0161	0.0169	0.0008	0.017	11.9266
8	Other	0.0551	0.0573	0.0022	0.0469	15.1762
10	Cloudy	14.3481	15.1257	0.7776	16.5595	15.7146
11	Blowing Snow	0.0431	0.0469	0.0038	0.0809	12.5828
12	Freezing Rain or Drizzle	0.0248	0.0262	0.0014	0.0298	13.6095



# Bibliography

- [1] T. Raghunathan et al. *IVEware: Imputation and Variation Estimation Software*. Version 0.3. 2016. URL: <https://www.src.isr.umich.edu/software/iveware/>.