

Data Documentation – Final Project

Project # 1

Dataset: `gavotte`

In the 2000 United States Presidential election, there was a lot of controversy especially in Florida where the voting machines were problematic. The Data set `gavotte` in the **package** “`faraway`” contains a sample of 159 individual in the counties of Georgia with 10 variables.

Use `help(gavotte)` in R to learn more about the variable names and dataset.

Issue: A potential voter goes to the polling station where it is determined whether he or she is registered to vote. If so, a ballot is issued. However, a vote is not recorded if the person fails to vote for President, votes for more than one candidate or the equipment fails to record the vote. For example, we can see that in Appling county, $6617 - 6099 = 518$ ballots did not result in voted for President. This is called the undercount.

We wish to determine the factors that significantly affects the undercount.

Obtain `summary(gavotte)` to get summary information about the variables.

Obtain the relative undercount:

```
gavotte$undercount <- ( ballots – votes)/ballots
```

Now, fit appropriate models for undercount as response variable with all possible predictors in the dataset as described in the “Final Project Instruction” document.

Project # 2

Dataset: urine

The Data set urine in the **package** “boot” has 79 rows and 7 columns.

79 urine specimens were analyzed in an effort to determine if certain physical characteristics of the urine might be related to the formation of calcium oxalate crystals.

Use `help(urine)` in R to learn more about the variable names and dataset.

Fit appropriate models for an indicator of the presence of calcium oxalate crystals as response variable with all possible predictors in the dataset as described in the “Final Project Instruction” document.

Project # 3:

Dataset: bacteria

The Data set bacteria in the **package** “MASS” has 220 rows and 6 columns.

Dr A. Leach tested the effects of a drug on 50 children with a history of otitis media in the Northern Territory of Australia. The children were randomized to the drug or the a placebo, and also to receive active encouragement to comply with taking the drug.

The presence of *H. influenzae* was checked at weeks 0, 2, 4, 6 and 11: 30 of the checks were missing and are not included in this data frame.

Use `help(bacteria)` in R to learn more about the variable names and dataset.

Fit appropriate models for an indicator of the presence or absence of bacteria as response variable with all possible predictors in the dataset as described in the “Final Project Instruction” document.