

Lab #4

Biostatistics 210

0. BACKGROUND

This lab is meant to give you an introduction to the concepts of missing data in predictors in regression. The data is patterned after a cohort study of patients diagnosed with primary biliary cirrhosis with the aim of finding predictors of survival

1. DOWNLOAD

• Download the model building dataset `lab4.dta`. The data here is faked so you can have the experience of look at different scenarios for missing data and seeing how different patterns of missingness affect the complete case analysis

• `dead` (1=died, 0 = survived)

• `copper_50` (serum cooper in 50 mg/dL units)

• `bilir_2` (bilirubin in 2 mg/dL units)

• `hepatomegaly` (1=present, 0 = absent)

• `spiders` (1=present, 0 = absent)

Think of the variable `spiders` as potential missing and based on this, I've created 3 scenarios with `spiders` missing. For each scenario, I created

`spiders_miss`k`` the available `spiders` values for scenario `k`

`miss`k`` (1 = spider data is missing, 0 = not missing)

2. EXPLORING MISSINGNESS

Take some time to explore the missingness in each scenario. Consider:

1. What fraction of spider values are missing in each scenario?
2. What fraction of non-missing data has spiders present: how does that compare with the actual spiders data
3. Can you see any association between other data points and whether the spiders is missing?

In scenario 2, look closely at copper. In scenario 3, look at hepatomegaly and mortality

3. LOGISTIC REGRESSION MODEL

The data has been generated to fit a logistic regression model with OR: 1.5 for copper_50, 1.5 for bilir_2, and 2.0 for spiders and hepatomegaly. Verify this on the full data

```
logistic dead spiders hepat bilir copper
```

Examine the missingness by fitting logistic regression models for the 3 missingness scenarios:

```
logistic dead spider_miss1 hepat bilir copper
```

```
logistic dead spider_miss2 hepat bilir copper
```

```
logistic dead spider_miss3 hepat bilir copper
```

How do the results of each of these compare to the true parameters. Can you link this to the pattern of missingness?