Non-parametric Statistics: Notes 1

Gregory Matthews [1]

[1]Department of Mathematics
Loyola University Chicago

Spring 2016

Outline

Introduction

- Big question: What is the difference between parametric and non-parametric analysis?
- With parametric statistics we make assumptions about the probability distributions of variables. The assumptions often involve parametric distributions. (OR if we have a large sample the Central Limit Theorem (CLT) can help us out).
- Non-parametric statistics does not make these distributional assumptions about the data. (May make other assumptions though!)
- Parametric: More assumptions, more power
- Non-parametric: Less assumptions, less power (usually, but not always)

- Hypothesis tests
  - Sign test
  - Wilcoxon test
  - Permutation tests
- Non-parametric Survival techniques
- Boostrap techniques
- Smoothing techniques
- Robust fitting
- Classification and Regression Trees (Maybe)

Sign Test

- Say we want to test $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$ where θ is the median.
- If we let p be the probability that any x is larger than 0 we can re-write the null hypothesis as....
- $H_0 : p = 0.5$ vs $H_1 : p \neq 0.5$
- **Assumptions**: The data are independent and identically distribution (iid). (No claims about WHAT distribution. Just SOME distribution.)

Sign Test

- The test statistic here is $S_T = \Sigma_{i=1}^{n} I(x_i \leq \theta_0)$
- But this is just a fancy way of counting how many observations are smaller than θ_0.
- Question we need to answer: If the null hypothesis is true (i.e. $p = 0.5$), how often will we observe a S_T that is as or more extreme than what we observe? (What am I getting at here?)

```r
dat
```

```
##  [1] -1.9141315  1.0548585  2.6688824 -4.1913954  1.3582494  1.5121118
##  [7] -0.6494799 -0.5932637 -0.6289040 -1.2800757 -0.4543854 -1.4967729
```

```r
table(sign(dat))
```

```
##
## -1  1
##  8  4
```

```r
#Prob binom is greater than or equal to 8
1-pbinom(7,12,0.5)
```

```
## [1] 0.1938477
```

```r
#Prob binom is less than or equal to 4
  pbinom(4,12,0.5)
```

```
## [1] 0.1938477
```

```r
dat
```

```
##  [1] -1.9141315  1.0548585  2.6688824 -4.1913954  1.3582494  1.5121118
##  [7] -0.6494799 -0.5932637 -0.6289040 -1.2800757 -0.4543854 -1.4967729
```

```r
table(sign(dat))
```

```
##
## -1  1
##  8  4
```

```r
#R has a built in function for this test
binom.test(8,12)
```

```
##
##  Exact binomial test
##
## data:  8 and 12
## number of successes = 8, number of trials = 12, p-value = 0.3877
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3488755 0.9007539
## sample estimates:
## probability of success
##              0.6666667
```

```
dat
```

```
##  [1]  0.7927644  0.8547330  0.4453483  0.2732514  0.8029437 -0.4089780
##  [7]  0.8150493  0.3619079  0.3579201  0.0403390  0.4418761  1.4086560
```

```
table(sign(dat))
```

```
##
## -1  1
##  1 11
```

```
pbinom(1,12,.5)+1-pbinom(10,12,.5)
```

```
## [1] 0.006347656
```

```
binom.test(1,12)
```

```
##
##  Exact binomial test
##
## data:  1 and 12
## number of successes = 1, number of trials = 12, p-value = 0.006348
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.002107593 0.384796165
## sample estimates:
## probability of success
##             0.08333333
```