

1/12/2010

36-402/608 ADA-II
Breakout #1 Results

H. Seltman

```
# Analysis of Sparrow data from Statistical Sleuth, chapter 2.  
# Full data from: http://students.clarku.edu/~rking/bumpusI.pdf  
# Note: "case0201.csv" data on book CD has only has adult male humerus data.
```

```
### Data setup and initial examination ###  
sparrow = read.table("fullBumpus.txt", header=TRUE)  
dim(sparrow) # [1] 136 13
```

```
# Check which columns were converted to factors  
sapply(sparrow, class)
```

```
#      Bird      Female      Adult      Survive      Length      Alar      Weight      SkullLen  
# "integer" "integer" "integer" "integer" "integer" "integer" "numeric" "numeric"  
#      Humerus      Femur      TibTars      SkullWid      Sternum  
# "numeric" "numeric" "numeric" "numeric" "numeric"
```

```
summary(sparrow[,1:10])
```

```
#      Bird      Female      Adult      Survive      Length  
# Min.   : 1.00   Min.   :0.0000   Min.   : 0.0000   Min.   :0.0000   Min.   :152.0  
# 1st Qu.:17.75   1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:157.0  
# Median :34.50   Median :0.0000   Median : 1.0000   Median :1.0000   Median :160.0  
# Mean   :34.62   Mean    :0.3603   Mean    : 0.6782   Mean    :0.5294   Mean    :159.5  
# 3rd Qu.:51.25   3rd Qu.:1.0000   3rd Qu.: 1.0000   3rd Qu.:1.0000   3rd Qu.:162.0  
# Max.   :72.00   Max.    :1.0000   Max.    : 1.0000   Max.    :1.0000   Max.    :167.0  
#  
#      NA's      :49.0000  
#      Alar      Weight      SkullLen      Humerus      Femur  
# Min.   :230.0   Min.   :22.60   Min.   :29.80   Min.   :0.6590   Min.   :0.6530  
# 1st Qu.:242.0   1st Qu.:24.57   1st Qu.:31.10   1st Qu.:0.7180   1st Qu.:0.7007  
# Median :246.0   Median :25.50   Median :31.60   Median :0.7330   Median :0.7130  
# Mean   :245.3   Mean    :25.52   Mean    :31.57   Mean    :0.7324   Mean    :0.7130  
# 3rd Qu.:249.0   3rd Qu.:26.50   3rd Qu.:32.02   3rd Qu.:0.7495   3rd Qu.:0.7312  
# Max.   :256.0   Max.    :31.00   Max.    :33.40   Max.    :0.7800   Max.    :0.7670
```

```
# Make factors from categorical data (based on "indicator" naming)  
sparrow$Survive = factor(sparrow$Survive, labels=c("Perished","Survived"))  
sparrow$Female = factor(sparrow$Female, labels=c("Male","Female"))  
sparrow$Adult = factor(sparrow$Adult, labels=c("Juvenile","Adult"))
```

```

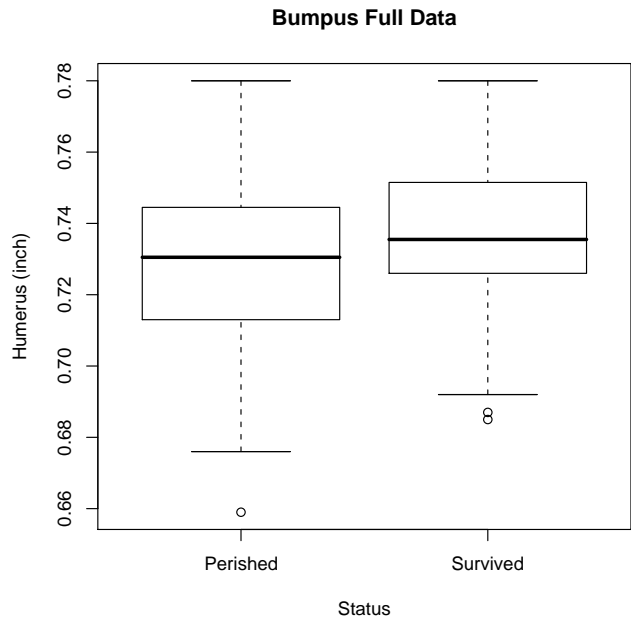
### Examine data collection structure ###
with(sparrow, table(Survive, Female, Adult))
# , , Adult = Juvenile
#           Female
# Survive   Male Female
# Perished  12      0
# Survived  16      0
# , , Adult = Adult
#           Female
# Survive   Male Female
# Perished  24      0
# Survived  35      0
#
with(sparrow, table(Survive, Female, Adult, useNA="ifany"))
#, , Adult = Juvenile
#           Female
# Survive   Male Female
# Perished  12      0
# Survived  16      0
# , , Adult = Adult
#           Female
# Survive   Male Female
# Perished  24      0
# Survived  35      0
# , , Adult = NA
#           Female
# Survive   Male Female
# Perished  0      28
# Survived  0      21

# Create a grouping variable to match data collection
# (Does not work if attach() was used.)
sparrow$Group = 1
sparrow$Group[sparrow$Female=="Male"] = 2
sparrow$Group[sparrow$Female=="Male" & sparrow$Adult=="Adult"] = 3
sparrow$Group = factor(sparrow$Group, labels=c("Female",
  "JuvMale", "AdultMale"))
table(sparrow$Group, useNA="ifany")
# Female   JuvMale AdultMale
#    49         28         59

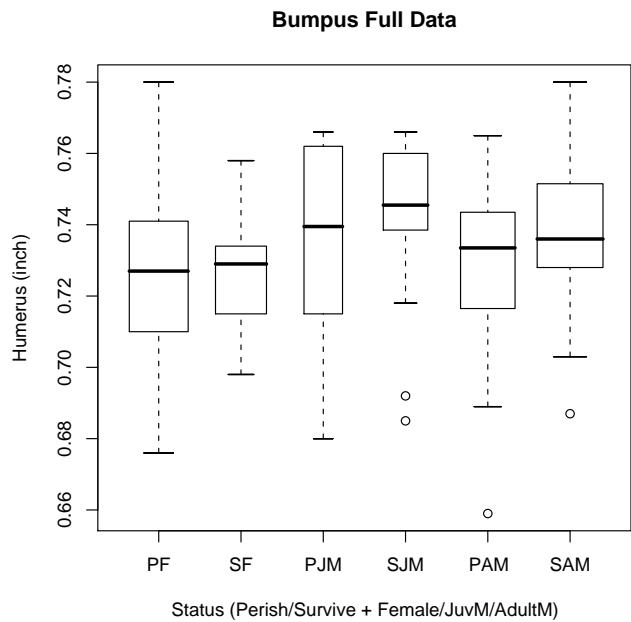
```

Graphical EDA

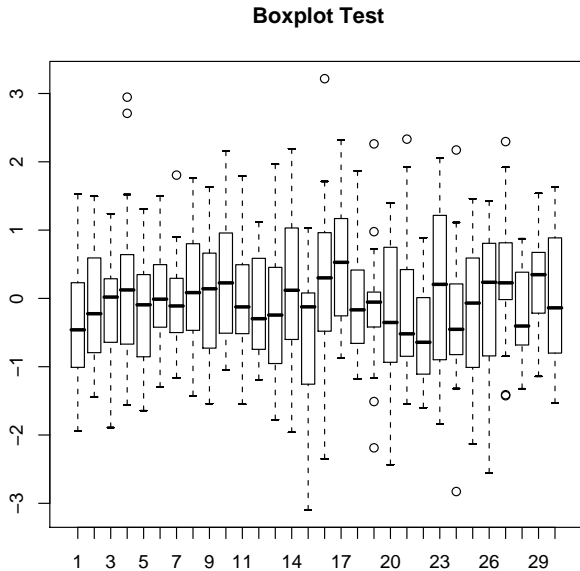
```
boxplot(Humerus~Survive, data=sparrow, xlab="Status",  
        ylab="Humerus (inch)", main="Bumpus Full Data")
```



```
boxplot(Humerus~Survive+Group, data=sparrow,  
        xlab="Status (Perish/Survive + Female/JuvM/AdultM)",  
        ylab="Humerus (inch)", main="Bumpus Full Data", names=  
        c("PF", "SF", "PJM", "SJM", "PAM", "SAM"), varwidth=T)
```



```
# Use simulation to see how variable similar IID Normal boxplots would be (with n=20).
boxplot(split(rnorm(20*30),rep(1:30,each=20)), main="Boxplot Test")
```



```
### Non-graphical EDA ###
```

```
with(sparrow, aggregate(Humerus, list(status=Survive, group=Group), mean))
```

```
# status group x
#1 Perished Female 0.7260357
#2 Survived Female 0.7283333
#3 Perished JuvMale 0.7347500
#4 Survived JuvMale 0.7414375
#5 Perished AdultMale 0.7279167
#6 Survived AdultMale 0.7380000
```

```
round(with(sparrow, aggregate(Humerus, list(status=Survive, group=Group), sd))$x, 3)
# [1] 0.026 0.016 0.029 0.024 0.024 0.020
```

```
### Formal testing: t-tests ###
```

```
Sel = sparrow$Group=="AdultMale"
```

```
with(sparrow[Sel,], t.test(Humerus~Survive, var.equal=TRUE))
```

```
# t = -1.777, df = 57, p-value = 0.0809
```

```
# Use an "anonymous (nameless & temporary) function to compare means:
```

```
sapply(split(sparrow,sparrow$Group),
```

```
function(x) t.test(x$Humerus~x$Survive, var.equal=TRUE)$conf.int)
```

```
# Female JuvMale AdultMale
# [1,] -0.01528490 -0.02750834 -0.021446053
# [2,] 0.01068967 0.01413334 0.001279386
```