Booklet of Code and Output
for
STAC32 Midterm Exam

October 26, 2015

List of Figures in this document by page:

List of Figures

```
league homeruns
American 122
American 103
American 100
American 96
American 93
American 86
American 84
American 80
American 74
American 73
American 71
American 64
American 64
American 57
National 110
National 106
National 94
National 90
National 89
National 86
National 80
National 77
National 77
National 76
National 74
National 70
National 65
National 63
National 60
National 48
```

Figure 1: Home run data for baseball teams

```
SAS> data wines;
SAS>   infile 'wines.txt' expandtabs firstobs=2;
SAS>   input caseprice location $;
SAS>
SAS> proc sgplot;
SAS>   vbox caseprice / category=location;
```

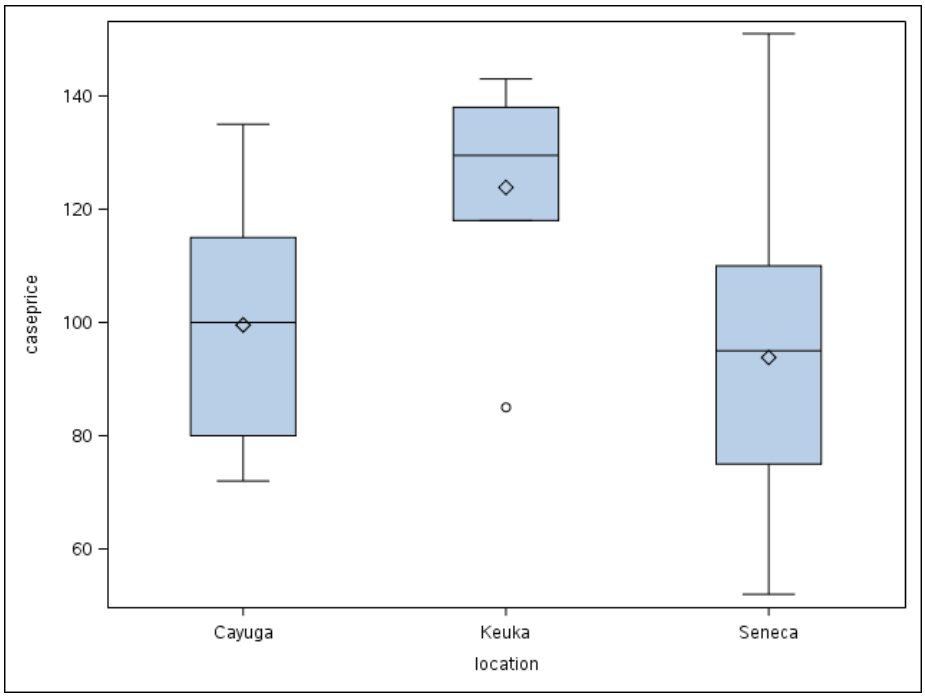

Figure 2: Finger Lakes wine case prices

```
R>   head(boys,n=20)

 [1] 1 3 3 2 3 2 3 3 2 1 2 3 3 2 1 2 2 3 1 2

R>   table(boys)

boys
  0   1   2   3   4
 10  53 108  90  21
```

Figure 3: ADHD boys' task ratings: first 20 observations and table counting observed ratings

```
R> t.test(boys,conf.level=0.90,mu=2.35)

        One Sample t-test

data:  boys
t = -2.4834, df = 281, p-value = 0.0136
alternative hypothesis: true mean is not equal to 2.35
90 percent confidence interval:
 2.115666 2.302773
sample estimates:
mean of x
  2.20922
```

Figure 4: Confidence interval and hypothesis test for ADHD boys data

F26 sas

	A	B	C	D	E	F	G
1	Product	Calories	Fat	Carbs	Fibre	Protein	
2	Caffe Latte	190	7	18	0	12	
3	Caffe Mocha	260	8	41	2	13	
4	Cappuccino	120	4	12	0	8	
5	Caramel Macchiato	240	7	34	0	10	
6	Cinnamon Dolce Latte	260	6	40	0	11	
7	Flavoured Latte	250	6	36	0	12	
8	Iced Caffe Latte	130	4.5	13	0	8	
9	Iced Caffe Mocha	200	6	35	2	9	
10	Iced Caramel Macchiato	230	6	33	0	10	
11	Iced Cinnamon Dolce Latte	200	4	34	0	7	
12	Iced Flavoured Latte	250	6	36	0	12	
13	Iced Peppermint Mocha	260	6	52	2	8	
14	Iced Peppermint White Chocolate Mocha	400	9	72	0	10	
15	Iced Pumpkin Spice Latte	250	4	44	0	10	
16	Iced Skinny Flavoured Latte	110	4	12	0	7	
17	Iced Toffee Mocha	280	3.5	51	2	12	
18	Iced White Chocolate Mocha	340	9	55	0	10	
19	Peppermint Mocha	330	8	57	2	12	
20	Peppermint White Chocolate Mocha	470	12	78	0	14	

Figure 5: Coffee spreadsheet

```
timeofday seconds
early 68
early 138
early 75
early 186
early 68
early 217
early 93
early 90
early 71
early 154
early 166
early 130
early 72
early 81
early 76
early 129
evening 299
evening 367
evening 331
evening 257
evening 260
evening 269
evening 252
evening 200
evening 296
evening 204
evening 190
evening 240
evening 350
evening 256
evening 282
evening 320
late 216
late 175
late 274
late 171
late 187
late 213
late 221
late 139
late 226
late 128
late 236
late 128
late 217
late 196
late 201
late 161
```

Figure 6: File download data

```
            Df Sum Sq Mean Sq F value   Pr(>F)
timeofday    2 204641  102320   46.03 1.31e-11 ***
Residuals   45 100020    2223
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 7: R analysis for downloads data, part 1

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = seconds ~ timeofday, data = dl)

$timeofday
                   diff        lwr       upr    p adj
evening-early 159.9375  119.53988 200.33512 0.00e+00
late-early     79.6875   39.28988 120.08512 5.57e-05
late-evening  -80.2500 -120.64762 -39.85238 4.99e-05
```

Figure 8: R analysis for downloads data, part 2

```
R> boxplot(seconds~timeofday,data=dl)
```

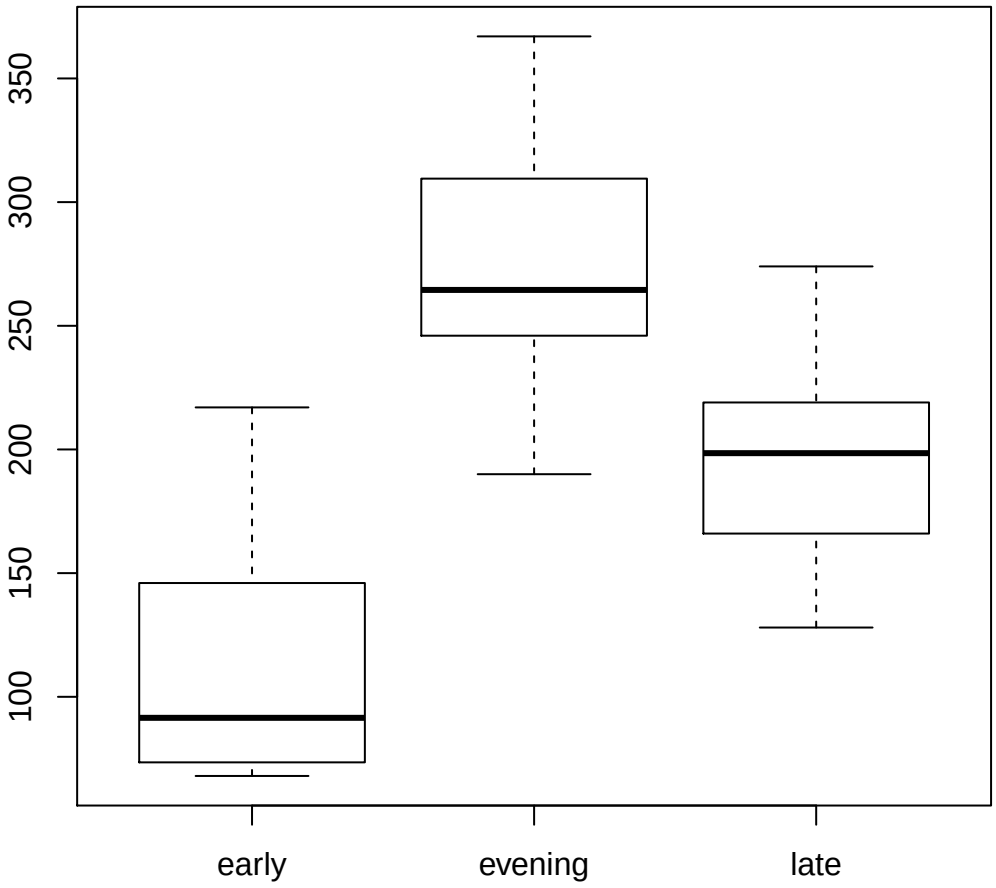

Figure 9: Boxplots of download times by time of day

```
0.119
0.118
0.099
0.118
0.275
0.358
0.080
0.158
0.310
0.105
0.073
0.832
0.517
0.851
0.269
0.433
0.141
0.135
0.175
```

Figure 10: Arsenic concentrations in toenail clippings in New Hampshire

```
R> arsenic=read.table("arsenic.txt",header=F)
R> x=arsenic$V1
R> boxplot(x)
```

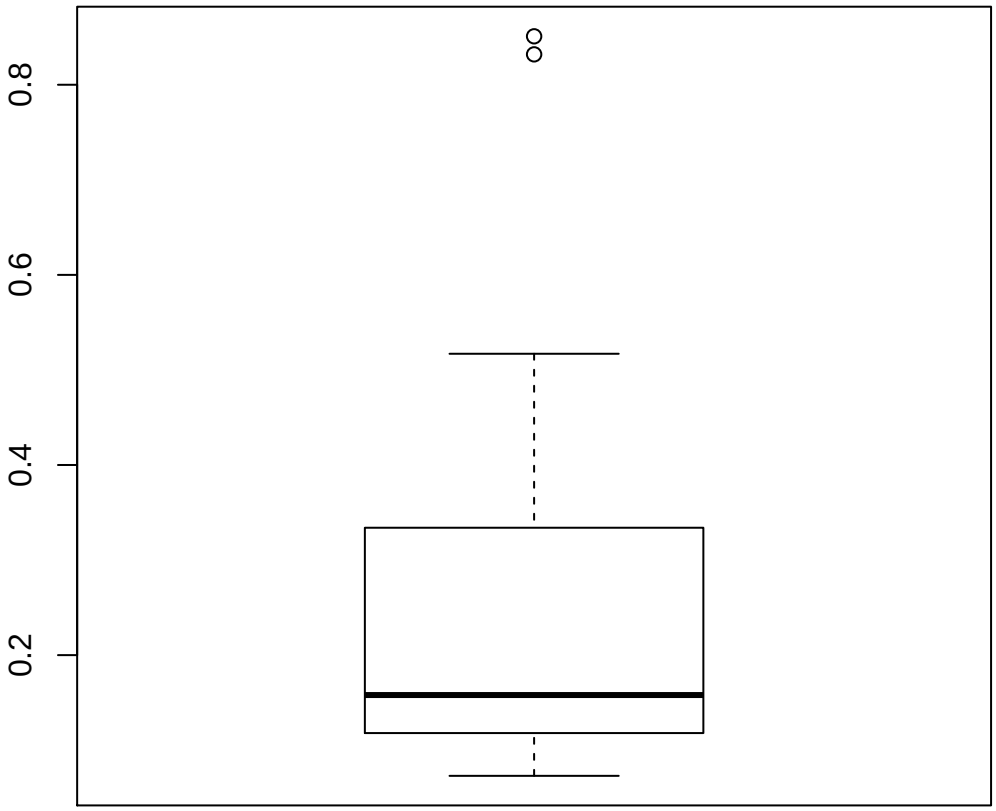

Figure 11: Boxplot of arsenic levels

```
R> sign.test=function(med,mydata) {
R>   n=length(mydata)
R>   tab=table(mydata<med)
R>   stat=min(tab)
R>   pval=2*pbinom(stat,n,0.5)
R>   return(pval)
R> }
```

Figure 12: R function to implement two-sided sign test

medians	pvals
0.10	0.004425049
0.11	0.019210815
0.12	0.359283447
0.13	0.359283447
0.14	0.647605896
0.15	1.000000000
0.16	1.000000000
0.17	1.000000000
0.18	0.647605896
0.19	0.647605896
0.20	0.647605896
0.21	0.647605896
0.22	0.647605896
0.23	0.647605896
0.24	0.647605896
0.25	0.647605896
0.26	0.647605896
0.27	0.359283447
0.28	0.167068481
0.29	0.167068481
0.30	0.167068481
0.31	0.167068481
0.32	0.063568115
0.33	0.063568115
0.34	0.063568115
0.35	0.063568115
0.36	0.019210815
0.37	0.019210815
0.38	0.019210815
0.39	0.019210815
0.40	0.019210815
0.41	0.019210815
0.42	0.019210815
0.43	0.019210815
0.44	0.004425049
0.45	0.004425049
0.46	0.004425049
0.47	0.004425049
0.48	0.004425049
0.49	0.004425049
0.50	0.004425049

Figure 13: P-values (second column) for running the sign test on various different hypothesized population medians (first column)

```
id group sex beats
1 Control Female 159
2 Control Female 183
3 Control Female 140
4 Control Female 140
5 Control Female 125
6 Control Female 155
7 Control Female 148
8 Control Female 132
9 Control Female 158
10 Control Female 136
201 Control Male 127
202 Control Male 99
203 Control Male 157
204 Control Male 102
205 Control Male 97
206 Control Male 122
207 Control Male 128
208 Control Male 136
209 Control Male 142
210 Control Male 127
401 Runners Female 119
402 Runners Female 84
403 Runners Female 89
404 Runners Female 119
405 Runners Female 127
406 Runners Female 111
407 Runners Female 115
408 Runners Female 109
409 Runners Female 111
410 Runners Female 120
601 Runners Male 100
602 Runners Male 120
603 Runners Male 93
604 Runners Male 107
605 Runners Male 138
606 Runners Male 96
607 Runners Male 107
608 Runners Male 119
609 Runners Male 99
610 Runners Male 102
```

Figure 14: Some of the heart beats data for active and sedentary people

```
R>   borneo=read.table("borneo.txt",header=T)
R>   boxplot(species~status,data=borneo)
```

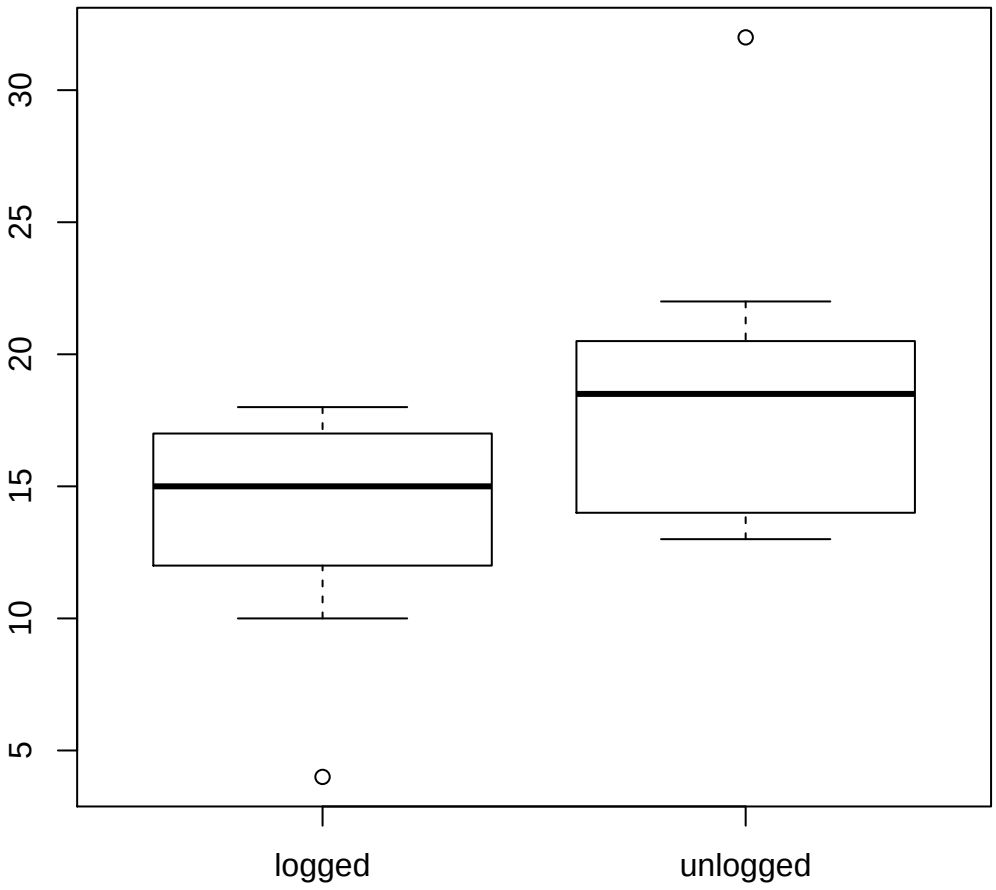

Figure 15: Logged vs. unlogged plots in Borneo

```
R>   means.obs=aggregate(species~status,borneo,mean)
R>   means.diff.obs=means.obs$species[1]-means.obs$species[2]
R>
R>   shuff=function(mydata) {
R>     attach(mydata)
R>     shuff.status=sample(status)
R>     means=aggregate(species~shuff.status,mydata,mean)
R>     detach(mydata)
R>     return(means$species[1]-means$species[2])
R>   }
```

Figure 16: Randomization test part 2

```
R> rand.dist=replicate(1000,shuff(borneo))
R> tab=table(rand.dist<=means.diff.obs)
R> tab

FALSE  TRUE
  982    18
```

Figure 17: Randomization test part 3

```
R> hist(rand.dist)
R> abline(v=means.diff.obs,lty="dashed")
```

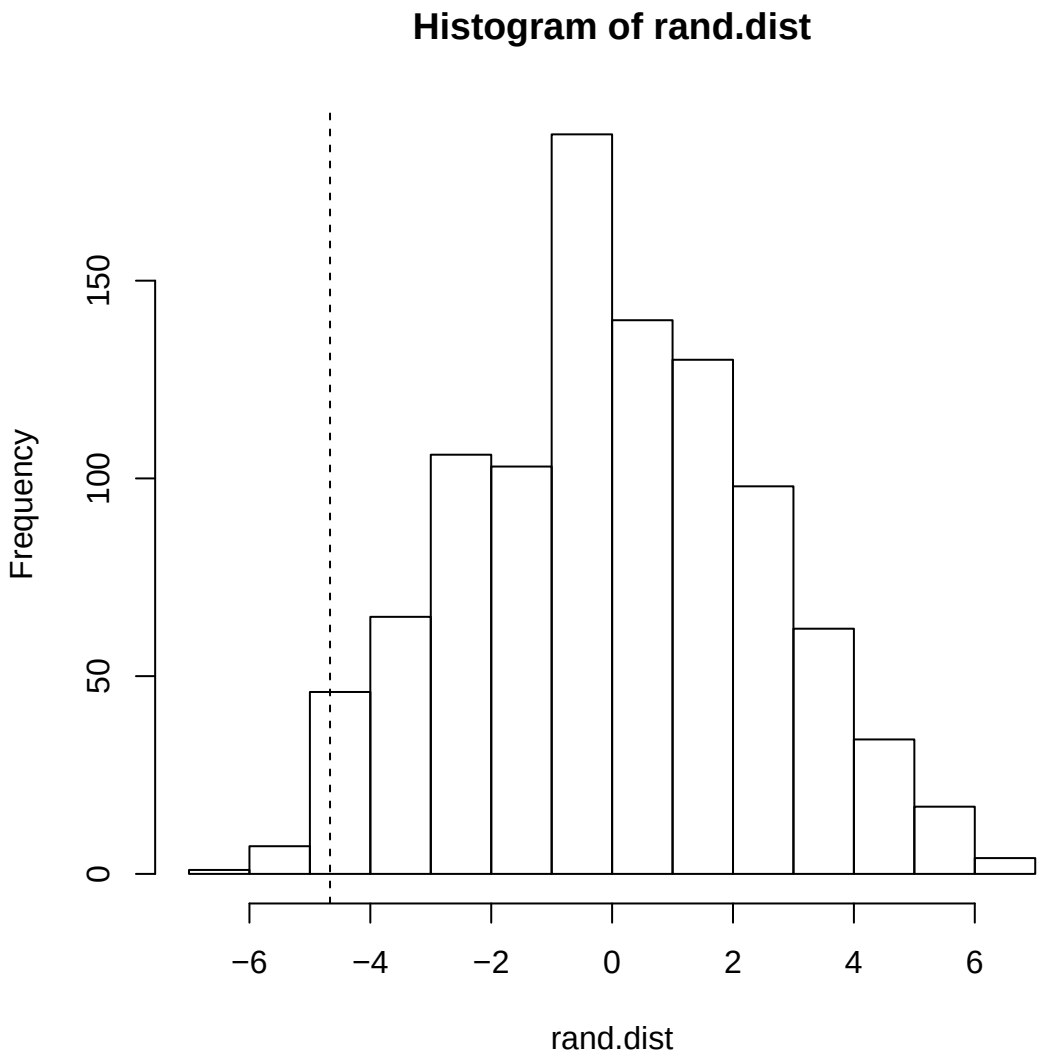

Figure 18: Randomization test part 4