Booklet of Code and Output
for
STAC32 Midterm Exam

October 13, 2018

```
grade boehm
kindergarten 17
kindergarten 20
kindergarten 24
kindergarten 34
kindergarten 34
kindergarten 38
first 23
first 25
first 27
first 34
first 38
first 47
second 22
second 23
second 26
second 32
second 34
second 34
second 36
second 38
second 38
second 42
second 48
second 50
```

Figure 1: Data for Boehm Basic Concepts Test for hard-of-hearing children

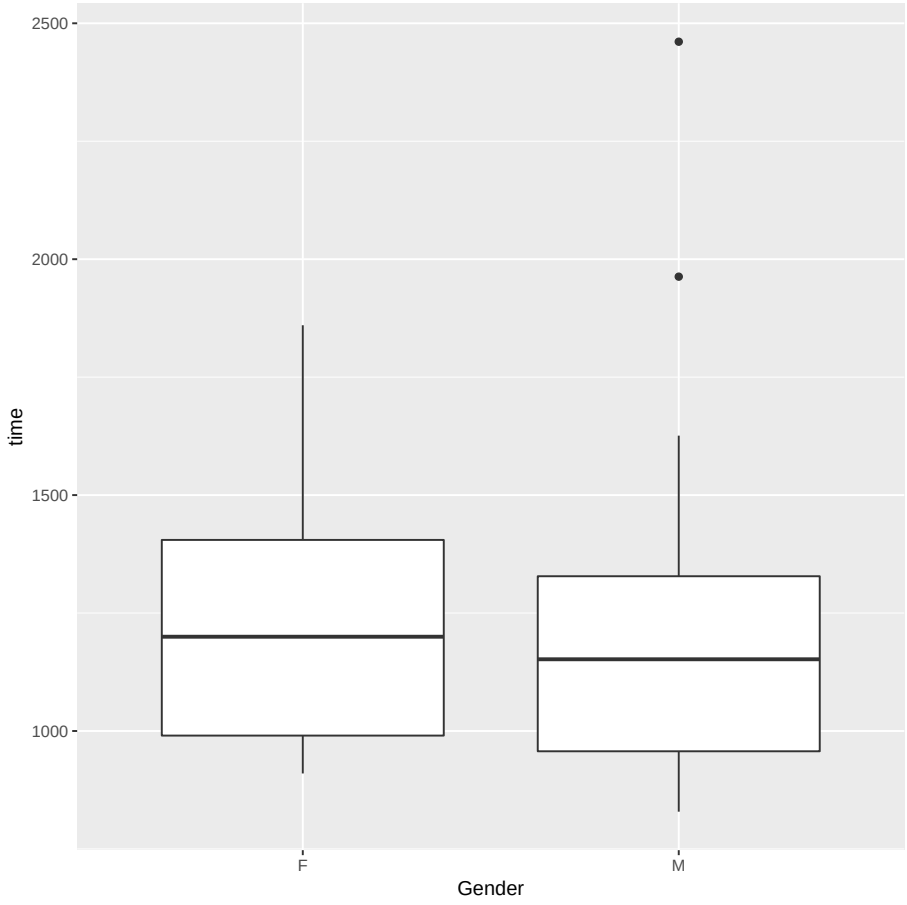

Figure 2: Lake Ontario swimming boxplot for males and females

```
## # A tibble: 2 x 6
##   Gender     n  mean median    sd   iqr
##   <chr>  <int> <dbl>  <dbl> <dbl> <dbl>
## 1 F         33 1228.  1200.  243.  415.
## 2 M         25 1226.  1152   368.  371
```

Figure 3: Summary of swimming times for Lake Ontario swims, for males and females

store	promotion	prom_sales	prev_sales
1	sampling	38	21
2	sampling	39	26
3	sampling	36	22
4	sampling	45	28
5	sampling	33	19
6	shelf_regular	43	34
7	shelf_regular	38	26
8	shelf_regular	38	29
9	shelf_regular	27	18
10	shelf_regular	34	25
11	shelf_display	24	23
12	shelf_display	32	29
13	shelf_display	31	30
14	shelf_display	21	16
15	shelf_display	28	29

Figure 4: Cracker promotion data

```
ggplot(pulserates,aes(x=Pulse))+geom_histogram(bins=8)
```

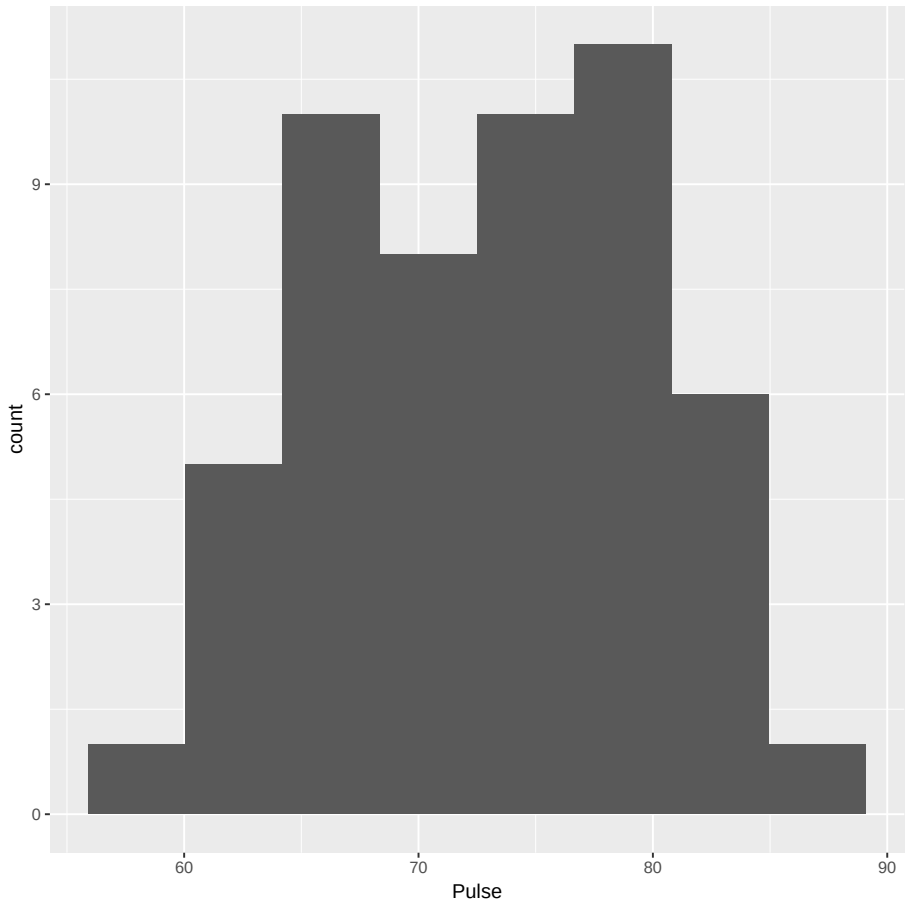

Figure 5: Pulse rate data histogram

```
with(pulserates, t.test(Pulse, conf.level=0.90))
##
##  One Sample t-test
##
## data:  Pulse
## t = 80.864, df = 51, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  71.18632 74.19829
## sample estimates:
## mean of x
##  72.69231
```

Figure 6: Pulse rate data t-test (1)

```
##
##  One Sample t-test
##
## data:  Pulse
## t = 1.8826, df = 51, p-value = 0.03274
## alternative hypothesis: true mean is greater than 71
## 95 percent confidence interval:
##  71.18632      Inf
## sample estimates:
## mean of x
##  72.69231
```

Figure 7: Pulse rate data t-test (2)

```
## [1] 0.2786014
```

Figure 8: Power calculation result for newborn girls in remote region, $n = 25$

```
## # A tibble: 2 x 2
##   `pvals <= 0.05`     n
##   <lgl>           <int>
## 1 FALSE             175
## 2 TRUE              825
```

Figure 9: Power simulation result for newborn girls in remote region, $n = 100$

```
## Parsed with column specification:
## cols(
##  player = col_integer(),
##  salary = col_double()
## )
## # A tibble: 25 x 2
##    player salary
##     <int>  <dbl>
##  1      1  2.5
##  2      2  0.925
##  3      3  3.25
##  4      4  0.6
##  5      5  1.75
##  6      6  2.5
##  7      7  1.6
##  8      8  1.85
##  9      9  1.4
## 10     10  3
## # ... with 15 more rows
```

Figure 10: NHL salaries (some)

```
sign_test(nhl,salary,1.0)
## $above_below
## below above
##     7    18
##
## $p_values
##   alternative    p_value
## 1       lower 0.99268335
## 2       upper 0.02164263
## 3   two-sided 0.04328525
```

Figure 11: Sign test output for NHL player salaries

```
subject before after
1 28 28
2 35 15
3 14 2
4 20 20
5 25 31
6 40 19
7 18 6
8 15 17
9 21 1
10 19 5
11 32 12
12 42 20
13 26 30
14 37 19
15 19 0
16 38 16
17 23 4
18 24 19
```

Figure 12: Data for posthypnotic suggestion study on smokers

```
ggplot(smoking, aes(sample=before))+stat_qq()+stat_qq_line()
```

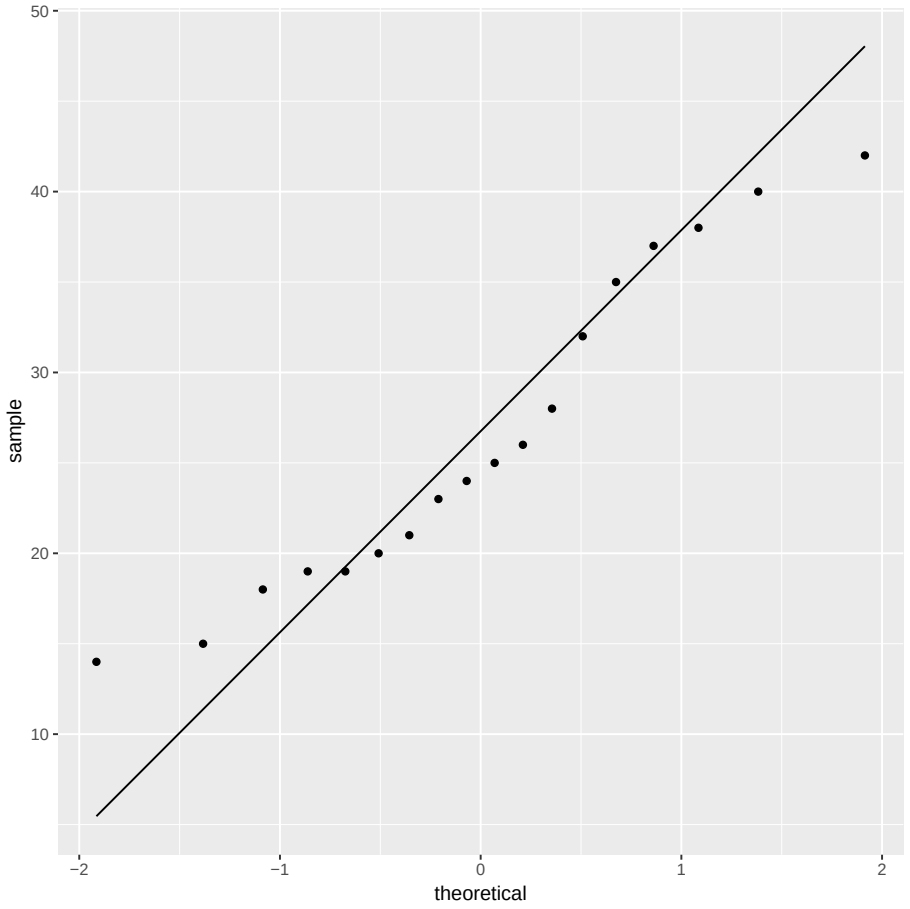

Figure 13: Normal quantile plots for smokers data part 1

```
ggplot(smoking, aes(sample=after))+stat_qq()+stat_qq_line()
```

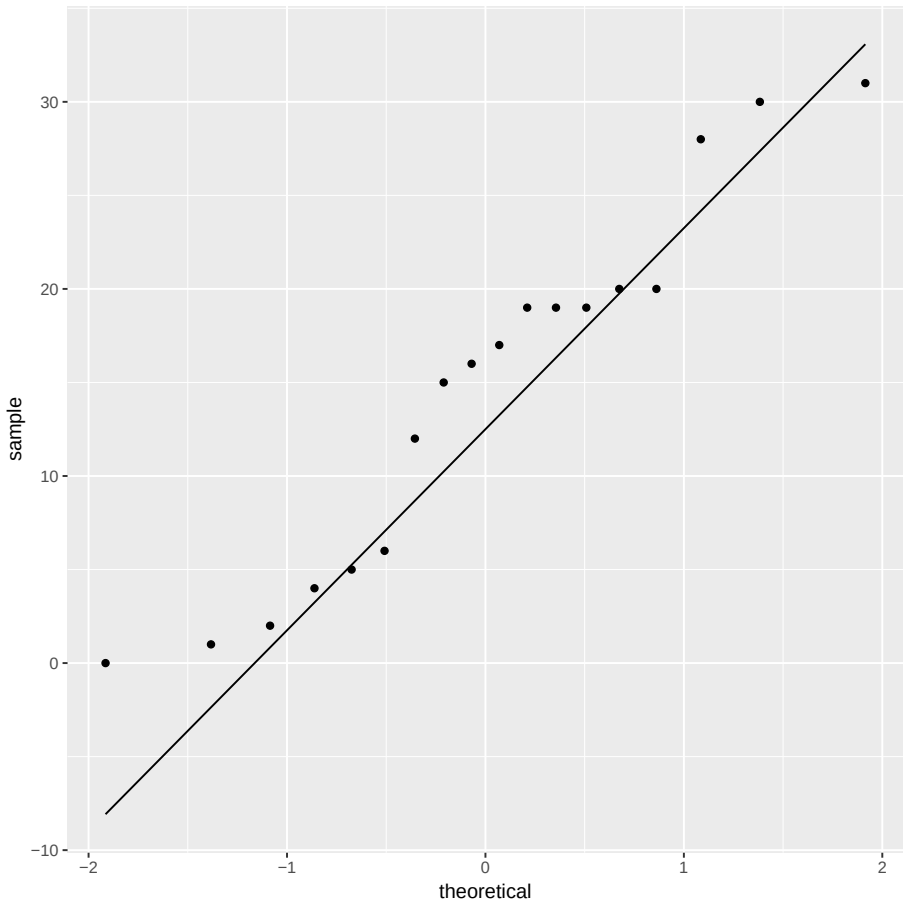

Figure 14: Normal quantile plots for smokers data part 2

```
smoking %>%
    mutate(difference=before-after) -> smoking2
ggplot(smoking2, aes(sample=difference))+stat_qq()+stat_qq_line()
```

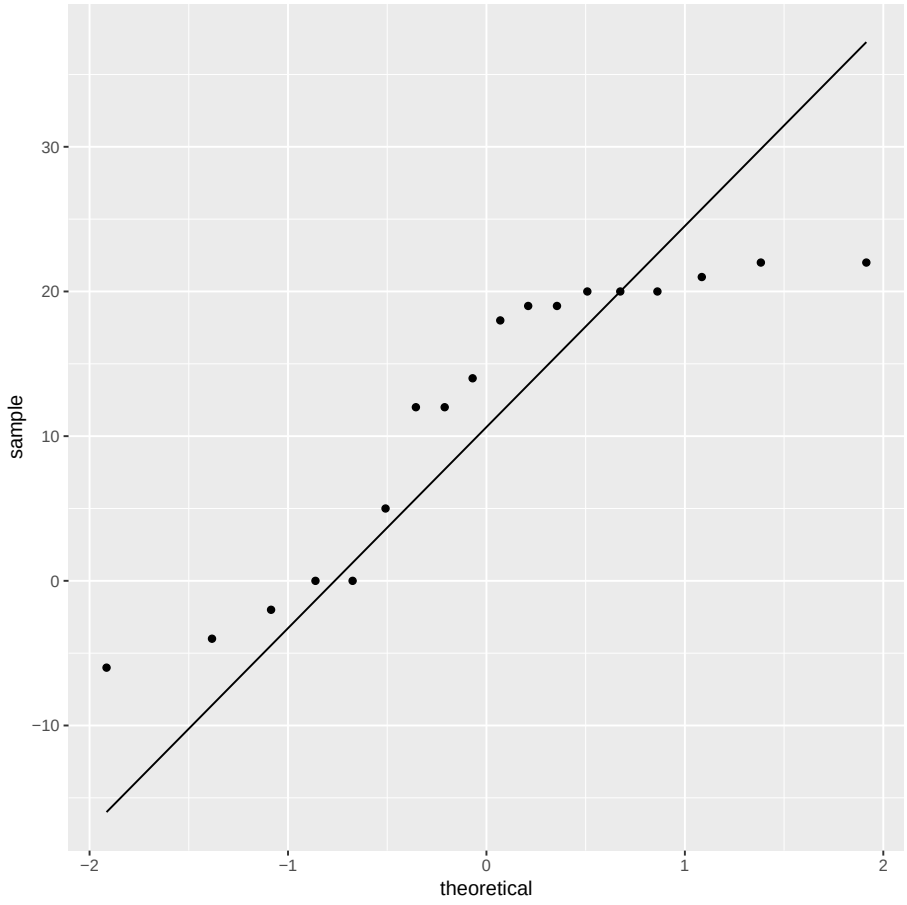

Figure 15: Normal quantile plots for smokers data part 3

```
with(smoking, t.test(before, after, alternative="less", paired=T))
##
##  Paired t-test
##
## data:  before and after
## t = 4.9633, df = 17, p-value = 0.9999
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 15.90585
## sample estimates:
## mean of the differences
##                11.77778
```

Figure 16: Analysis 1 for smokers data

```
with(smoking, t.test(before, after, alternative="greater", paired=T))
##
##  Paired t-test
##
## data:  before and after
## t = 4.9633, df = 17, p-value = 5.917e-05
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  7.649708      Inf
## sample estimates:
## mean of the differences
##                11.77778
```

Figure 17: Analysis 2 for smokers data

```
with(smoking, t.test(before, after, alternative="greater"))
##
##  Welch Two Sample t-test
##
## data:  before and after
## t = 3.7688, df = 33.625, p-value = 0.0003162
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.491845      Inf
## sample estimates:
## mean of x mean of y
##  26.44444  14.66667
```

Figure 18: Analysis 3 for smokers data

```
sign_test(smoking2, difference, 0)
## $above_below
## below above
##     3    13
##
## $p_values
##   alternative    p_value
## 1       lower 0.99790955
## 2       upper 0.01063538
## 3   two-sided 0.02127075
```

Figure 19: Analysis 4 for smokers data

```
## # A tibble: 48 x 2
##    side  height
##    <chr>  <dbl>
##  1 north    7.1
##  2 north    7.2
##  3 north    7.4
##  4 north    7.6
##  5 north    7.6
##  6 north    7.7
##  7 north    7.7
##  8 north    7.9
##  9 north    8.1
## 10 north    8.4
## 11 north    8.5
## 12 north    8.8
## 13 east     6.9
## 14 east     7
## 15 east     7.1
## 16 east     7.2
## 17 east     7.3
## 18 east     7.3
## 19 east     7.4
## 20 east     7.6
## # ... with 28 more rows
```

Figure 20: Elm tree heights data (some)

```
ggplot(trees, aes(sample=height))+
    stat_qq()+stat_qq_line()+facet_wrap(~side)
```

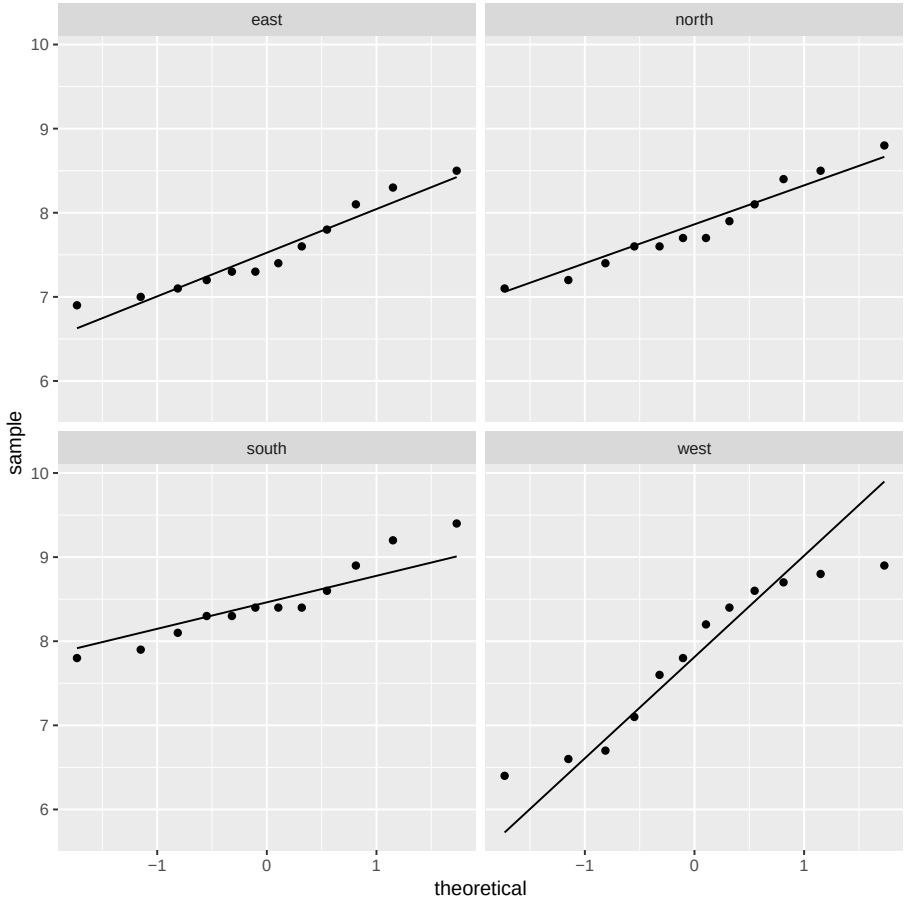

Figure 21: Elm tree heights normal quantile plots

```
trees.1=aov(height~side, data=trees)
summary(trees.1)
##             Df Sum Sq Mean Sq F value  Pr(>F)
## side         3  5.632  1.8772   4.595 0.00697 **
## Residuals   44 17.975  0.4085
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 22: Elm trees ANOVA

```
median_test(trees, height, side)
## $table
##        above
## group   above below
##   east      3     9
##   north     4     7
##   south    10     1
##   west      6     6
##
## $test
##        what       value
## 1 statistic 11.18181818
## 2        df  3.00000000
## 3   P-value  0.01078227
```

Figure 23: Elm trees Mood median test

```
## # A tibble: 4 x 3
##   side  median  mean
##   <chr>  <dbl> <dbl>
## 1 east    7.35  7.54
## 2 north   7.7   7.83
## 3 south   8.4   8.48
## 4 west    8     7.82
```

Figure 24: Elm trees table of means and medians

```
TukeyHSD(trees.1)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = height ~ side, data = trees)
##
## $side
##                    diff        lwr        upr     p adj
## north-east   0.29166667 -0.4050315 0.98836483 0.6806460
## south-east   0.93333333  0.2366352 1.63003150 0.0046083
## west-east    0.27500000 -0.4216982 0.97169817 0.7189018
## south-north  0.64166667 -0.0550315 1.33836483 0.0808319
## west-north  -0.01666667 -0.7133648 0.68003150 0.9999050
## west-south  -0.65833333 -1.3550315 0.03836483 0.0701302
```

Figure 25: Elm trees Tukey analysis

```
pairwise_median_test(trees, height, side)
## # A tibble: 6 x 4
##   g1    g2    p_value adj_p_value
##   <chr> <chr>   <dbl>       <dbl>
## 1 east  north 0.123        0.739
## 2 east  south 0.00284      0.0170
## 3 east  west  0.201        1.20
## 4 north south 0.0143       0.0858
## 5 north west  0.414        2.49
## 6 south west  0.414        2.49
```

Figure 26: Elm trees pairwise Mood tests