Booklet of Code and Output
for
STAC32 Midterm Exam

October 21, 2017

```
time_of_day	download_time
early	69
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-night	216
late-night	175
late-night	274
late-night	171
late-night	187
late-night	213
late-night	221
late-night	139
late-night	226
late-night	128
late-night	236
late-night	128
late-night	217
late-night	196
late-night	201
late-night	161

-:@---  download.txt   All L32    (Text Fill)
```

Figure 1: File download data (screenshot)

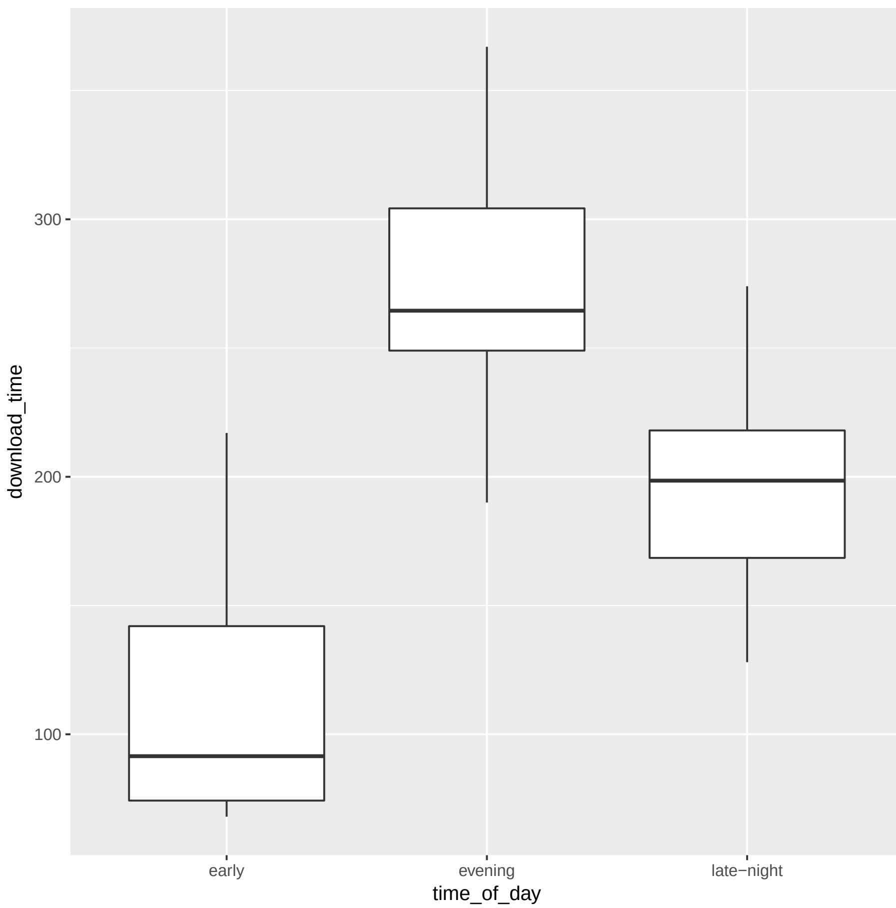


Figure 2: Boxplot for downloading data

```
## # A tibble: 47 x 3
##     Year Gender Height
##    <int>  <chr>  <dbl>
##  1  1896    Men  1.810
##  2  1900    Men  1.900
##  3  1904    Men  1.800
##  4  1908    Men  1.905
##  5  1912    Men  1.930
##  6  1920    Men  1.935
##  7  1924    Men  1.980
##  8  1928    Men  1.940
##  9  1932    Men  1.970
## 10  1936    Men  2.030
## 11  1948    Men  1.980
## 12  1952    Men  2.040
## 13  1956    Men  2.120
## 14  1960    Men  2.160
## 15  1964    Men  2.160
## 16  1968    Men  2.240
## 17  1972    Men  2.230
## 18  1976    Men  2.250
## 19  1980    Men  2.360
## 20  1984    Men  2.350
## 21  1988    Men  2.360
## 22  1992    Men  2.340
## 23  1996    Men  2.390
## 24  2000    Men  2.360
## 25  2004    Men  2.360
## 26  2008    Men  2.360
## 27  2012    Men  2.380
## 28  1928  Women  1.590
## 29  1932  Women  1.657
## 30  1936  Women  1.600
## 31  1948  Women  1.680
## 32  1952  Women  1.670
## 33  1956  Women  1.760
## 34  1960  Women  1.850
## 35  1964  Women  1.900
## 36  1968  Women  1.820
## 37  1972  Women  1.920
## 38  1976  Women  1.930
## 39  1980  Women  1.970
## 40  1984  Women  2.020
## 41  1988  Women  2.030
## 42  1992  Women  2.020
## 43  1996  Women  2.050
## 44  2000  Women  2.010
## 45  2004  Women  2.060
## 46  2008  Women  2.050
## 47  2012  Women  2.050
```

Figure 3: The high-jump data

```
subject hypnotized score
1 yes 8.5
2 yes 9.6
3 yes 10.0
4 yes 9.2
5 yes 8.9
6 yes 10.8
7 no 12.6
8 no 13.8
9 no 11.6
10 no 12.2
11 no 12.1
12 no 13.0
```

Figure 4: The Stroop test data

hypnotized	N	Mean	Std Dev	Std Err	Minimum	Maximum
no	6	12.5500	0.7740	0.3160	11.6000	13.8000
yes	6	9.5000	0.8246	0.3367	8.5000	10.8000
Diff (1-2)		3.0500	0.7997	0.4617		

hypnotized	Method	Mean	95% CL Mean		Std Dev
no		12.5500	11.7378	13.3622	0.7740
yes		9.5000	8.6346	10.3654	0.8246
Diff (1-2)	Pooled	3.0500	2.2132	Infty	0.7997
Diff (1-2)	Satterthwaite	3.0500	2.2128	Infty	

hypnotized	Method	95% CL Std Dev	
no		0.4831	1.8982
yes		0.5147	2.0225
Diff (1-2)	Pooled	0.5588	1.4034
Diff (1-2)	Satterthwaite		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	10	6.61	<.0001
Satterthwaite	Unequal	9.9601	6.61	<.0001

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	5	5	1.14	0.8927

Figure 5: Stroop data t-test

```
## # A tibble: 202 x 13
##       Sex   Sport   RCC   WCC    Hc    Hg  Ferr   BMI   SSF `%Bfat`   LBM
##     <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <int> <dbl> <dbl>   <dbl> <dbl>
##  1 female Netball  4.56  13.3  42.2  13.6    20 19.16  49.0   11.29 53.14
##  2 female Netball  4.15   6.0  38.0  12.7    59 21.15 110.2   25.26 47.09
##  3 female Netball  4.16   7.6  37.5  12.3    22 21.40  89.0   19.39 53.44
##  4 female Netball  4.32   6.4  37.7  12.3    30 21.03  98.3   19.63 48.78
##  5 female Netball  4.06   5.8  38.7  12.8    78 21.77 122.1   23.11 56.05
##  6 female Netball  4.12   6.1  36.6  11.8    21 21.38  90.4   16.86 56.45
##  7 female Netball  4.17   5.0  37.4  12.7   109 21.47 106.9   21.32 53.11
##  8 female Netball  3.80   6.6  36.5  12.4   102 24.45 156.6   26.57 54.41
##  9 female Netball  3.96   5.5  36.3  12.4    71 22.63 101.1   17.93 55.97
## 10 female Netball  4.44   9.7  41.4  14.1    64 22.80 126.4   24.97 51.62
## 11 female Netball  4.27  10.6  37.7  12.5    68 23.58 114.0   22.62 58.27
## 12 female Netball  3.90   6.3  35.9  12.1    78 20.06  70.0   15.01 57.28
## 13 female Netball  4.02   9.1  37.7  12.7   107 23.01  77.0   18.14 57.30
## 14 female Netball  4.39   9.6  38.3  12.5    39 24.64 148.9   26.78 54.18
## 15 female Netball  4.52   5.1  38.8  13.1    58 18.26  80.1   17.22 42.96
## 16 female Netball  4.25  10.7  39.5  13.2   127 24.47 156.6   26.50 54.46
## 17 female Netball  4.46  10.9  39.7  13.7   102 23.99 115.9   23.01 57.20
## 18 female Netball  4.40   9.3  40.4  13.6    86 26.24 181.7   30.10 54.38
## 19 female Netball  4.83   8.4  41.8  13.4    40 20.04  71.6   13.93 57.58
## 20 female Netball  4.23   6.9  38.3  12.6    50 25.72 143.5   26.65 61.46
## 21 female Netball  4.24   8.4  37.6  12.5    58 25.64 200.8   35.52 53.46
## 22 female Netball  3.95   6.6  38.4  12.8    33 19.87  68.9   15.59 54.11
## 23 female Netball  4.03   8.5  37.7  13.0    51 23.35 103.6   19.61 55.35
## 24 female   BBall  3.96   7.5  37.5  12.3    60 20.56 109.1   19.75 63.32
## 25 female   BBall  4.41   8.3  38.2  12.7    68 20.67 102.8   21.30 58.55
## 26 female   BBall  4.14   5.0  36.4  11.6    21 21.86 104.6   19.88 55.36
## 27 female   BBall  4.11   5.3  37.3  12.6    69 21.88 126.4   23.66 57.18
## 28 female   BBall  4.45   6.8  41.5  14.0    29 18.96  80.3   17.64 53.20
## 29 female   BBall  4.10   4.4  37.4  12.5    42 21.04  75.2   15.58 53.77
## 30 female   BBall  4.31   5.3  39.6  12.8    73 21.69  87.2   19.99 60.17
## # ... with 172 more rows, and 2 more variables: Ht <dbl>, Wt <dbl>
```

Figure 6: Australian athletes data (some)

```r
ggplot(athletes,aes(x=BMI))+geom_histogram(bins=8)
```

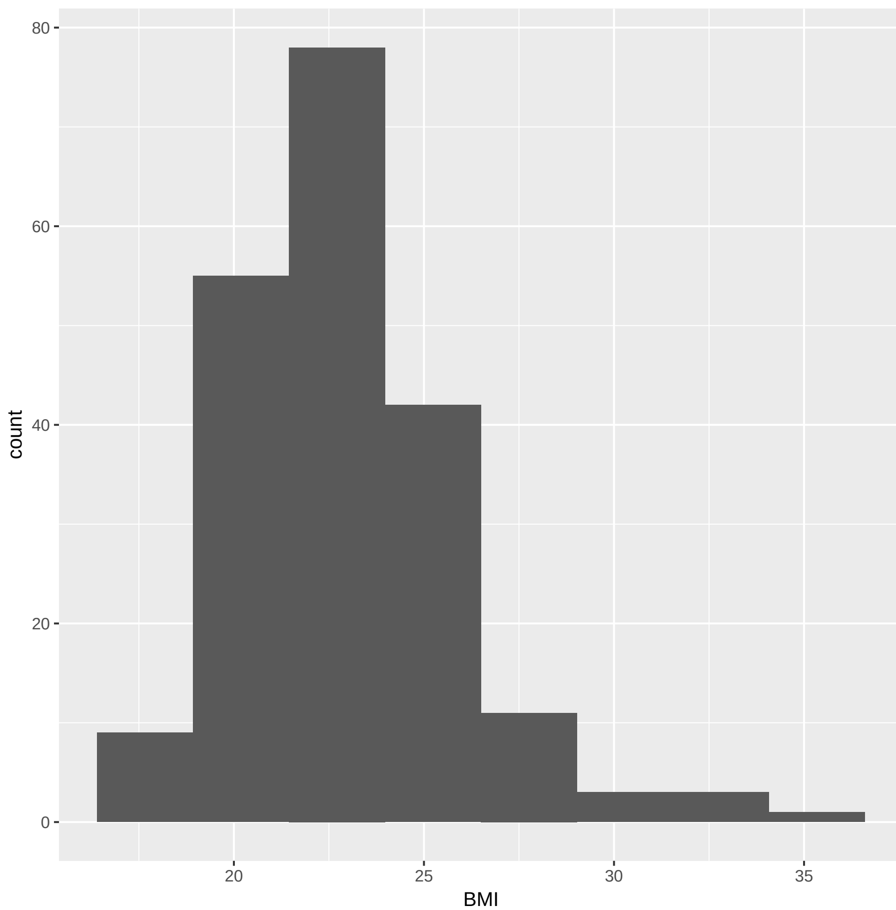


Figure 7: Australian athletes BMI histogram

```
samp=function(true_1,true_2,sd_1,sd_2,n_1,n_2) {
    r1=rnorm(n_1,true_1,sd_1)
    r2=rnorm(n_2,true_2,sd_2)
    g1=rep("a",n_1)
    g2=rep("b",n_2)
    d=tibble(value=c(r1,r2),group=c(g1,g2))
    ans=t.test(value~group,data=d)
    ans$p.value
}
pp=replicate(1000,samp(60,55,8,10,25,22))
tibble(pp) %>% count(pp<=0.05)
## # A tibble: 2 x 2
##   `pp <= 0.05`     n
##          <lgl> <int>
## 1        FALSE   543
## 2         TRUE   457
```

Figure 8: Some R code

```
proc import
  datafile='/home/ken/diabetes.txt'
  out=diabetes
  dbms=dlm
  replace;
  getnames=yes;
  delimiter=' ';

proc print;
```

Obs	age
1	35.5
2	44.5
3	39.8
4	33.3
5	51.4
6	51.3
7	30.5
8	48.9
9	42.1
10	40.3
11	46.8
12	38
13	40.1
14	36.8
15	39.3
16	71.1
17	73.4
18	65.4
19	42.6
20	42.8
21	59.8
22	52.4
23	26.2
24	60.9
25	45.6
26	27.1
27	47.3
28	36.6
29	23.2
30	55.6
31	45.1
32	52.2
33	43.5

Figure 9: Diabetes patients data

```
proc sgplot;
  histogram age;
```

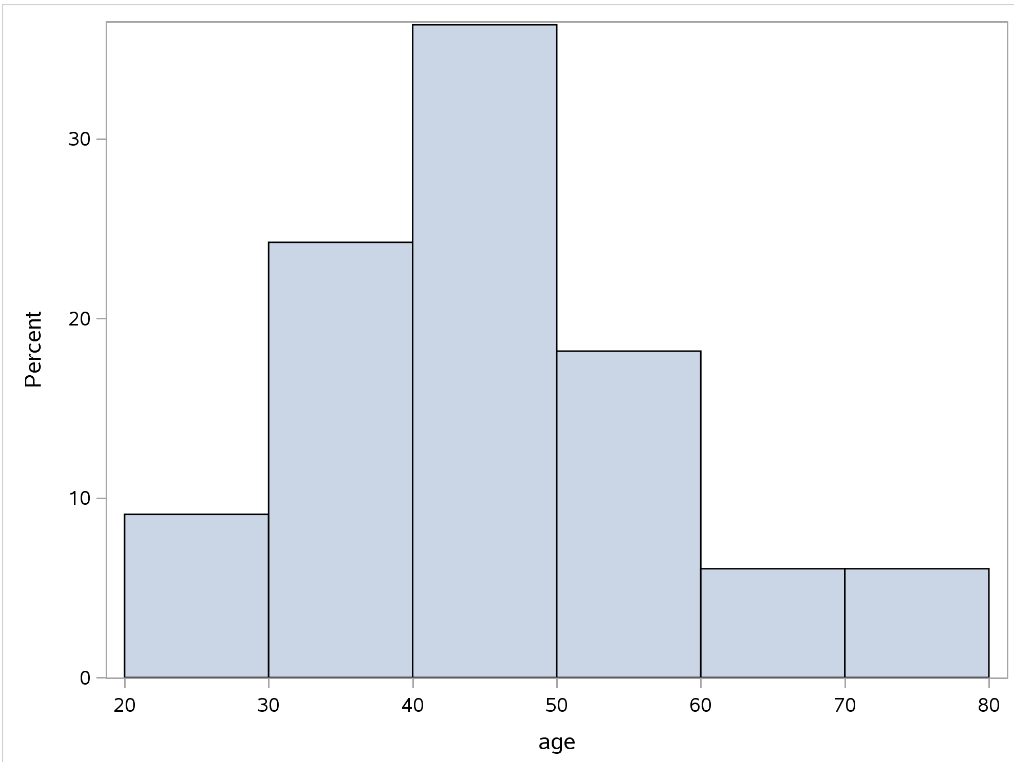


Figure 10: Histogram of diabetes data

```
                    The UNIVARIATE Procedure
                         Variable:  age

                 Tests for Location: Mu0=37

     Test           -Statistic-    -----p Value------

     Student's t    t  3.897204    Pr > |t|    0.0005
     Sign           M       8.5    Pr >= |M|   0.0046
     Signed Rank    S       188    Pr >= |S|   0.0003
```

Figure 11: Diabetes tests for location

```
diabetes=read_delim("diabetes.txt"," ")
## Parsed with column specification:
## cols(
##   age = col_double()
## )
diabetes %>% count(age<37)
## # A tibble: 2 x 2
##   `age < 37`     n
##        <lgl> <int>
## 1      FALSE    25
## 2       TRUE     8
succ=20:33
tibble(succ,prob=dbinom(succ,33,0.5)) %>% print(n=Inf)
## # A tibble: 14 x 2
##     succ         prob
##    <int>        <dbl>
##  1    20 6.672536e-02
##  2    21 4.130617e-02
##  3    22 2.253064e-02
##  4    23 1.077552e-02
##  5    24 4.489801e-03
##  6    25 1.616328e-03
##  7    26 4.973318e-04
##  8    27 1.289379e-04
##  9    28 2.762955e-05
## 10    29 4.763715e-06
## 11    30 6.351620e-07
## 12    31 6.146729e-08
## 13    32 3.841706e-09
## 14    33 1.164153e-10
```

Figure 12: R output for diabetes data analysis

```
## Parsed with column specification:
## cols(
##   bottom = col_double(),
##   surface = col_double()
## )
## # A tibble: 10 x 2
##     bottom surface
##      <dbl>   <dbl>
##  1  0.430   0.415
##  2  0.266   0.238
##  3  0.567   0.390
##  4  0.531   0.410
##  5  0.707   0.605
##  6  0.716   0.609
##  7  0.651   0.632
##  8  0.589   0.523
##  9  0.469   0.411
## 10  0.723   0.612
```

Figure 13: Zinc concentration data

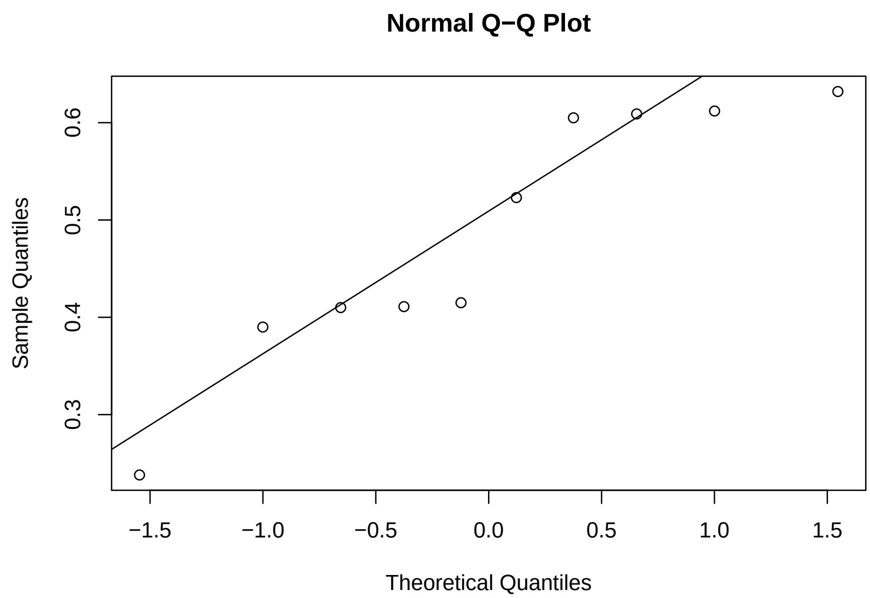
```
with(zinc,t.test(surface,bottom,alternative="greater"))
##
##  Welch Two Sample t-test
##
## data:  surface and bottom
## t = -1.2913, df = 17.779, p-value = 0.8934
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.1884396        Inf
## sample estimates:
## mean of x mean of y
##    0.4845    0.5649
```

Figure 14: Zinc *t*-test 1

```
with(zinc,t.test(bottom,surface,paired=T,alternative="greater"))
##
##  Paired t-test
##
## data:  bottom and surface
## t = 4.8638, df = 9, p-value = 0.0004456
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.0500982       Inf
## sample estimates:
## mean of the differences
##                  0.0804
```

Figure 15: Zinc t-test 2

```r
qqnorm(zinc$surface) ; qqline(zinc$surface)
```




```r
qqnorm(zinc$bottom) ; qqline(zinc$bottom)
```

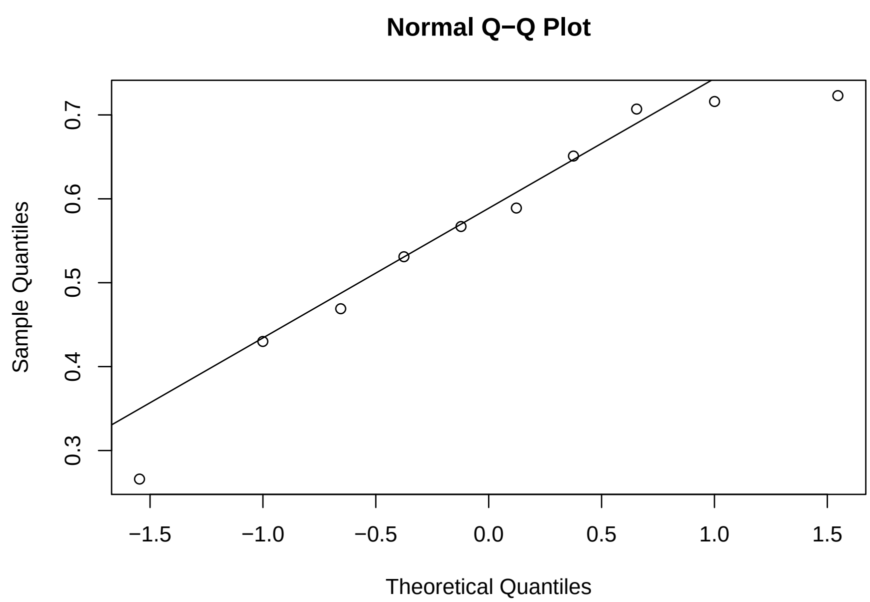


Figure 16: Zinc normal quantile plots of bottom and surface measurements

```
diff=zinc$bottom-zinc$surface
qqnorm(diff)
qqline(diff)
```

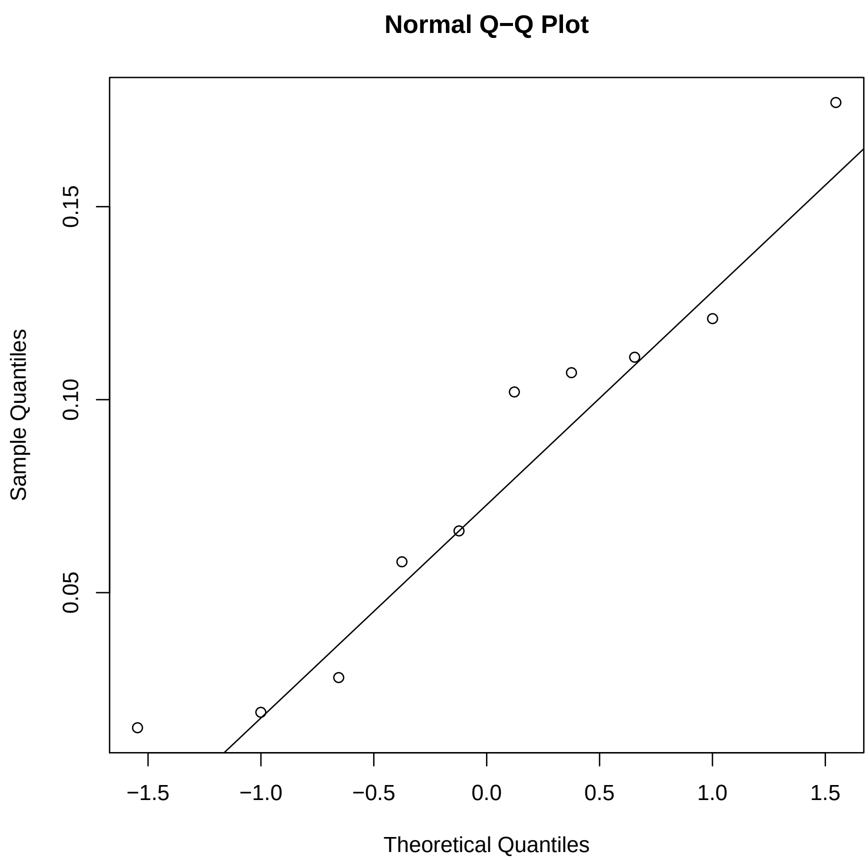


Figure 17: Zinc normal quantile plot of differences

```r
prenatal=read_csv("prenatal.csv")
## Parsed with column specification:
## cols(
##   care = col_character(),
##   apgar = col_integer()
## )
prenatal
## # A tibble: 15 x 2
##      care apgar
##     <chr> <int>
##  1  usual     8
##  2  usual     7
##  3  usual     6
##  4  usual     5
##  5  usual     2
##  6  usual     8
##  7  usual     7
##  8  usual     3
##  9 visits     9
## 10 visits     9
## 11 visits     7
## 12 visits     8
## 13 visits    10
## 14 visits     9
## 15 visits     6
```

Figure 18: Prenatal care data

```
prenatal %>% summarize(med=median(apgar))
## # A tibble: 1 x 1
##     med
##   <int>
## 1     7
library(smmr)
median_test(prenatal,apgar,care)
## $table
##         above
## group    above below
##   usual      2     4
##   visits     5     1
##
## $test
##        what      value
## 1 statistic 3.08571429
## 2        df 1.00000000
## 3   P-value 0.07898258
```

Figure 19: Hypothesis test for prenatal care data