Booklet of Code and Output
for
STAC32 Midterm Exam

October 24, 2016

```
cup tempdiff
SIGG 12
SIGG 16
SIGG 9
SIGG 23
SIGG 11
SIGG 20.5
SIGG 12.5
SIGG 20.5
SIGG 24.5
Starbucks 13
Starbucks 7
Starbucks 7
Starbucks 17.5
Starbucks 10
Starbucks 15.5
Starbucks 6
Starbucks 6
CUPPS 6
CUPPS 6
CUPPS 18.5
CUPPS 10
CUPPS 17.5
CUPPS 11
CUPPS 6.5
Nissan 2
Nissan 1.5
Nissan 2
Nissan 3
Nissan 0
Nissan 7
Nissan 0.5
Nissan 6
```

Figure 1: Data to be read in

```
carmpg=read_delim("carmpg.txt"," ")
## Parsed with column specification:
## cols(
##   row_number = col_integer(),
##   country = col_character(),
##   mpg = col_integer()
## )
carmpg
## # A tibble: 328 x 3
##    row_number country   mpg
##         <int>   <chr> <int>
##  1          1      us    18
##  2          2      us    15
##  3          3      us    18
##  4          4      us    16
##  5          5      us    17
##  6          6      us    15
##  7          7      us    14
##  8          8      us    14
##  9          9      us    14
## 10         10      us    15
## # ... with 318 more rows
carmpg %>% count(country)
## # A tibble: 2 x 2
##   country     n
##     <chr> <int>
## 1   japan    79
## 2      us   249
```

Figure 2: Summary of car gas mileage data

```
##
##  Welch Two Sample t-test
##
## data:  mpg by country
## t = 12.946, df = 136.87, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  9.014218      Inf
## sample estimates:
## mean in group japan    mean in group us
##            30.48101            20.14458
```

Figure 3: Results of t-test for car gas mileage data

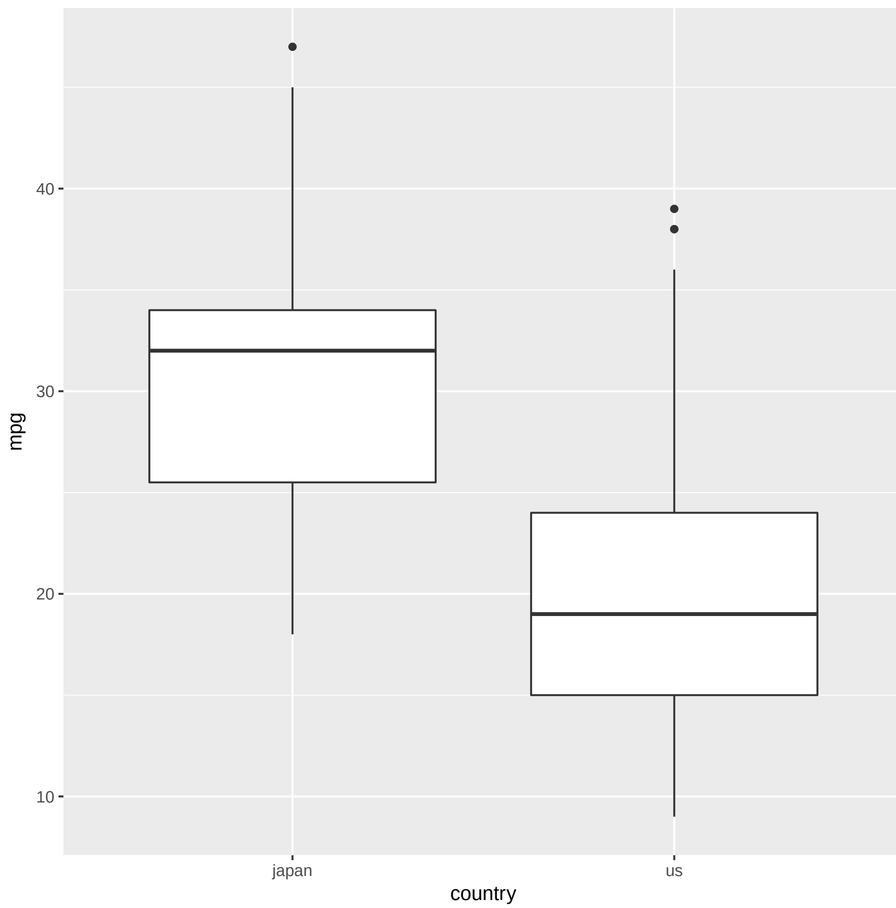

Figure 4: Boxplots of gas mileages for US and Japanese cars

```
after before
6 6
8 5
6 6
7 5
9 7
8 5
9 4
6 6
7 7
5 8
9 4
8 5
6 4
8 6
6 7
```

Figure 5: Police trainees' recollection of licence plates, before and after memory training

N	Mean	Std Dev	Std Err	Minimum	Maximum
15	1.5333	2.1996	0.5679	-3.0000	5.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1.5333	0.5330	Infty	2.1996	1.6104	3.4689

DF	t Value	Pr > t
14	2.70	0.0086

Figure 6: Analysis of police trainees data

```
proc univariate noprint;
  qqplot before / normal(mu=est sigma=est);
```

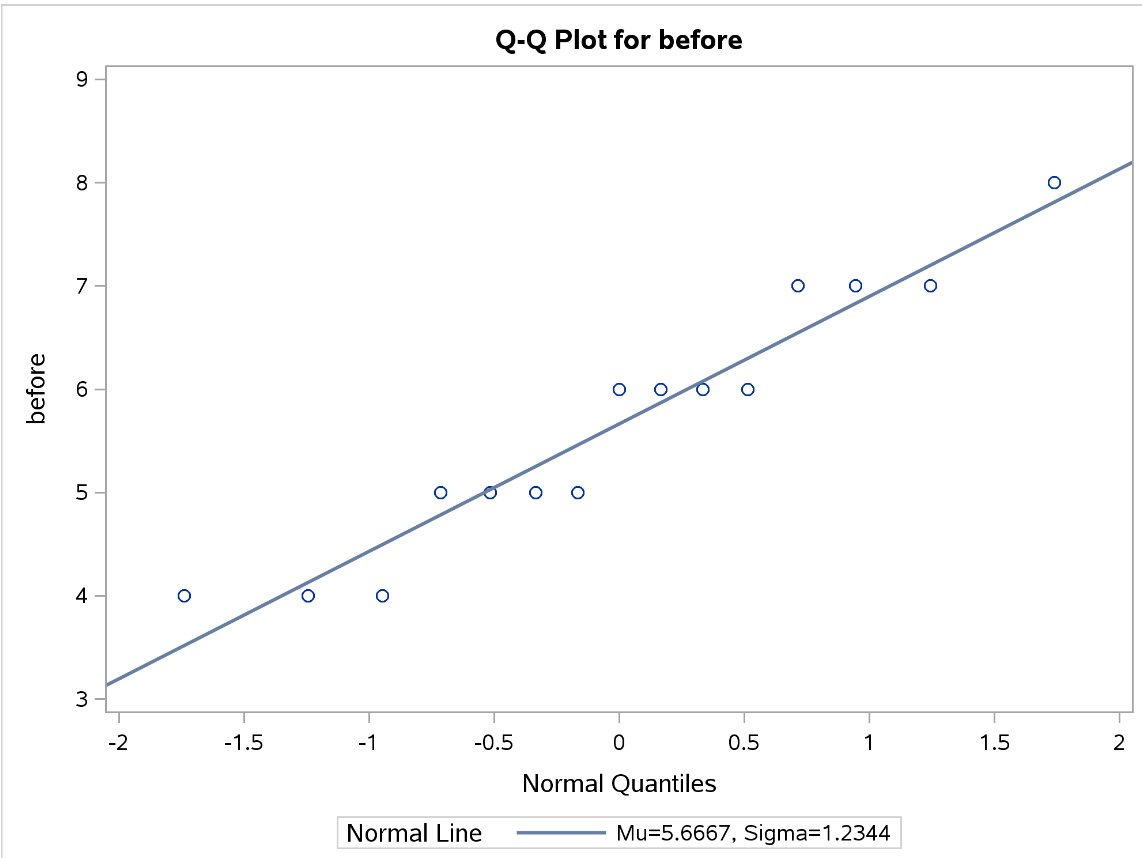

Figure 7: Normal quantile plot of before scores for police data

```
proc univariate noprint;
  qqplot after / normal(mu=est sigma=est);
```

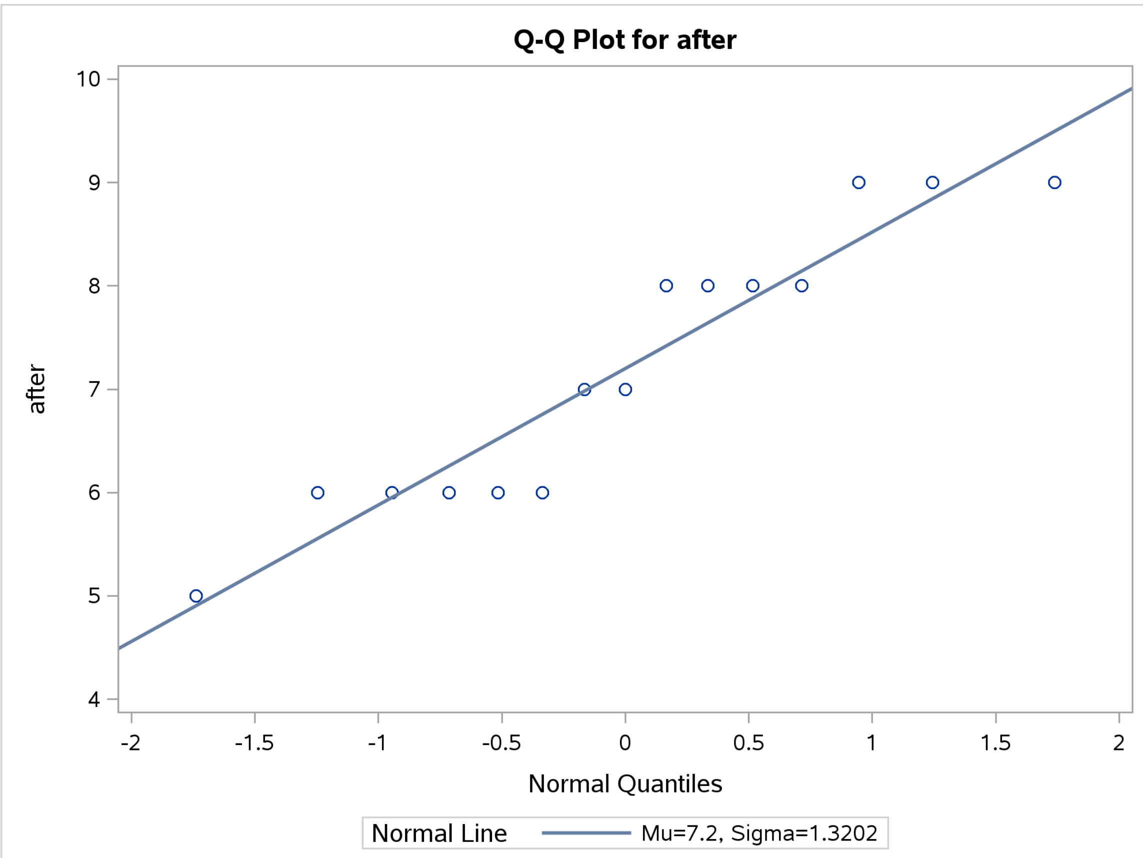

Figure 8: Normal quantile plot of after scores for police data

```
  data police2;
    set police;
    diff=after-before;

proc univariate noprint;
  qqplot diff / normal(mu=est sigma=est);
```

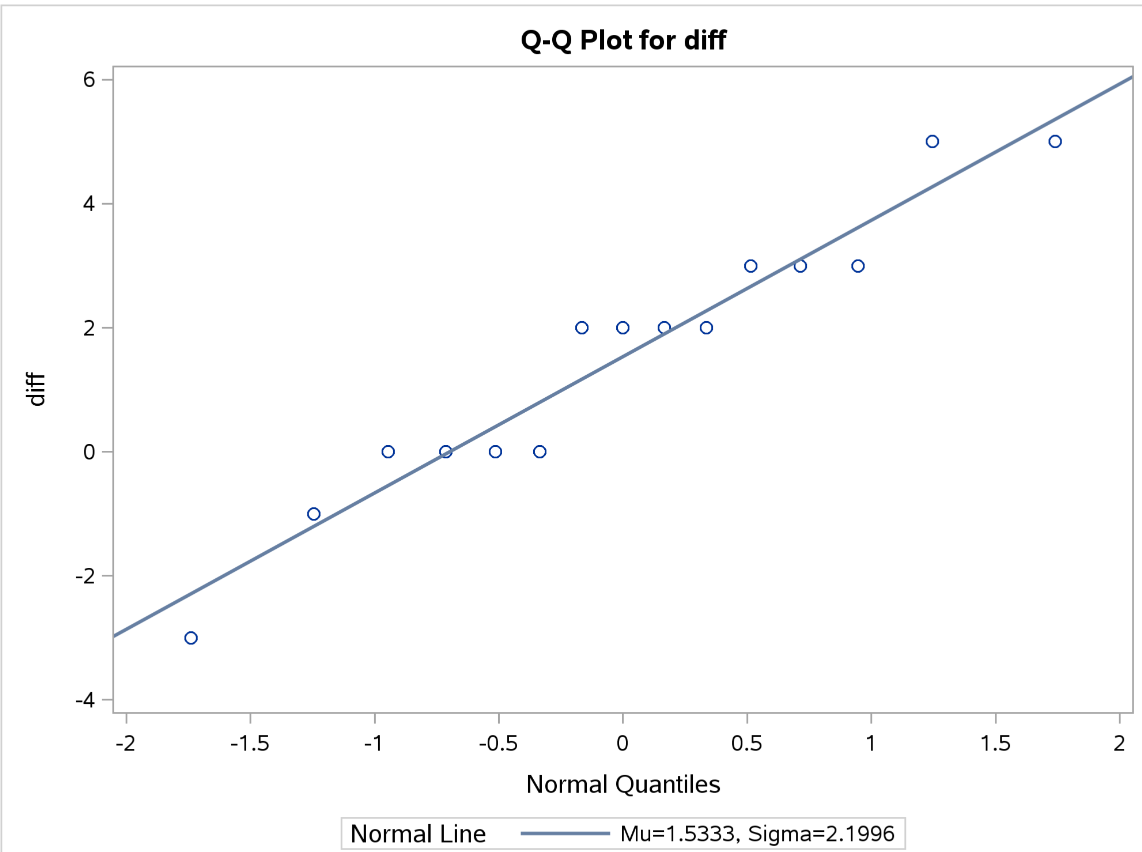

Figure 9: Normal quantile plot of scores differences for police data

```
Time
96
327
250
157
17
262
93
49
36
325
512
67
145
115
237
120
240
422
171
253
264
203
113
116
126
92
189
315
19
181
```

Figure 10: Data on time taken to complete IRS forms

```r
irs=read.table("irs.txt",header=T)
ggplot(irs,aes(x=Time))+geom_histogram(bins=10)
```

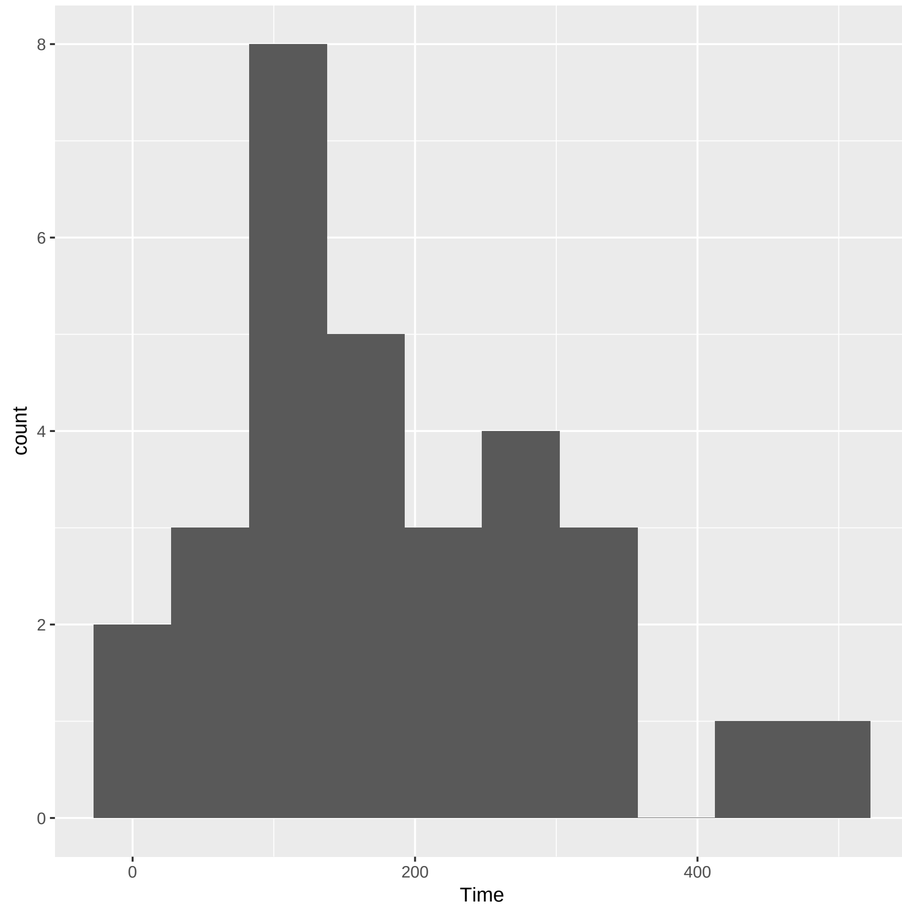

Figure 11: Histogram of IRS form completion times

```r
table(irs$Time>150)
##
## FALSE  TRUE
##    14    16
```

Figure 12: Table for sign test

```r
success=0:30
d=data.frame(success,prob=pbinom(success,30,0.5))
print(d,row.names=F)
##  success         prob
##        0 9.313226e-10
##        1 2.887100e-08
##        2 4.339963e-07
##        3 4.215166e-06
##        4 2.973806e-05
##        5 1.624571e-04
##        6 7.154532e-04
##        7 2.611440e-03
##        8 8.062401e-03
##        9 2.138697e-02
##       10 4.936857e-02
##       11 1.002442e-01
##       12 1.807973e-01
##       13 2.923324e-01
##       14 4.277678e-01
##       15 5.722322e-01
##       16 7.076676e-01
##       17 8.192027e-01
##       18 8.997558e-01
##       19 9.506314e-01
##       20 9.786130e-01
##       21 9.919376e-01
##       22 9.973886e-01
##       23 9.992845e-01
##       24 9.998375e-01
##       25 9.999703e-01
##       26 9.999958e-01
##       27 9.999996e-01
##       28 1.000000e+00
##       29 1.000000e+00
##       30 1.000000e+00
```

Figure 13: Table of binomial distribution for $n = 30, p = 0.5$

```
sgn=function(med,z) {
  n=length(z)
  tab=table(z>med)
  small=min(tab)
  p=pbinom(small,n,0.5)
  return(2*p)
}
```

Figure 14: Function to obtain P-value for sign test

```r
meds=seq(100,300,5)
pvals=sapply(meds,sgn,irs$Time)
data.frame(meds,pvals)
##    meds        pvals
## 1   100 0.0161248017
## 2   105 0.0161248017
## 3   110 0.0161248017
## 4   115 0.0987371467
## 5   120 0.3615946081
## 6   125 0.3615946081
## 7   130 0.5846647117
## 8   135 0.5846647117
## 9   140 0.5846647117
## 10  145 0.8555355519
## 11  150 0.8555355519
## 12  155 0.8555355519
## 13  160 1.1444644481
## 14  165 1.1444644481
## 15  170 1.1444644481
## 16  175 0.8555355519
## 17  180 0.8555355519
## 18  185 0.5846647117
## 19  190 0.3615946081
## 20  195 0.3615946081
## 21  200 0.3615946081
## 22  205 0.2004884221
## 23  210 0.2004884221
## 24  215 0.2004884221
## 25  220 0.2004884221
## 26  225 0.2004884221
## 27  230 0.2004884221
## 28  235 0.2004884221
## 29  240 0.0427739453
## 30  245 0.0427739453
## 31  250 0.0161248017
## 32  255 0.0052228794
## 33  260 0.0052228794
## 34  265 0.0003249142
## 35  270 0.0003249142
## 36  275 0.0003249142
## 37  280 0.0003249142
## 38  285 0.0003249142
## 39  290 0.0003249142
## 40  295 0.0003249142
## 41  300 0.0003249142
```

Figure 15: Calculations for confidence interval

```
patterns setting
9 praise
8 praise
8 praise
9 praise
7 praise
2 criticism
5 criticism
4 criticism
3 criticism
9 interest
3 interest
7 interest
8 interest
5 interest
6 interest
5 silence
7 silence
3 silence
6 silence
7 silence
```

Figure 16: Data for pattern recognition experiment

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

proc means;
  var patterns;
  class setting;
```

```
                                The MEANS Procedure

                           Analysis Variable : patterns

               N
setting      Obs     N            Mean         Std Dev         Minimum         Maximum
--------------------------------------------------------------------------------
criticism      4     4       3.5000000       1.2909944       2.0000000       5.0000000

interest       6     6       6.3333333       2.1602469       3.0000000       9.0000000

praise         5     5       8.2000000       0.8366600       7.0000000       9.0000000

silence        5     5       5.6000000       1.6733201       3.0000000       7.0000000
--------------------------------------------------------------------------------
```

Figure 17: Reading in pattern-recognition data and showing means

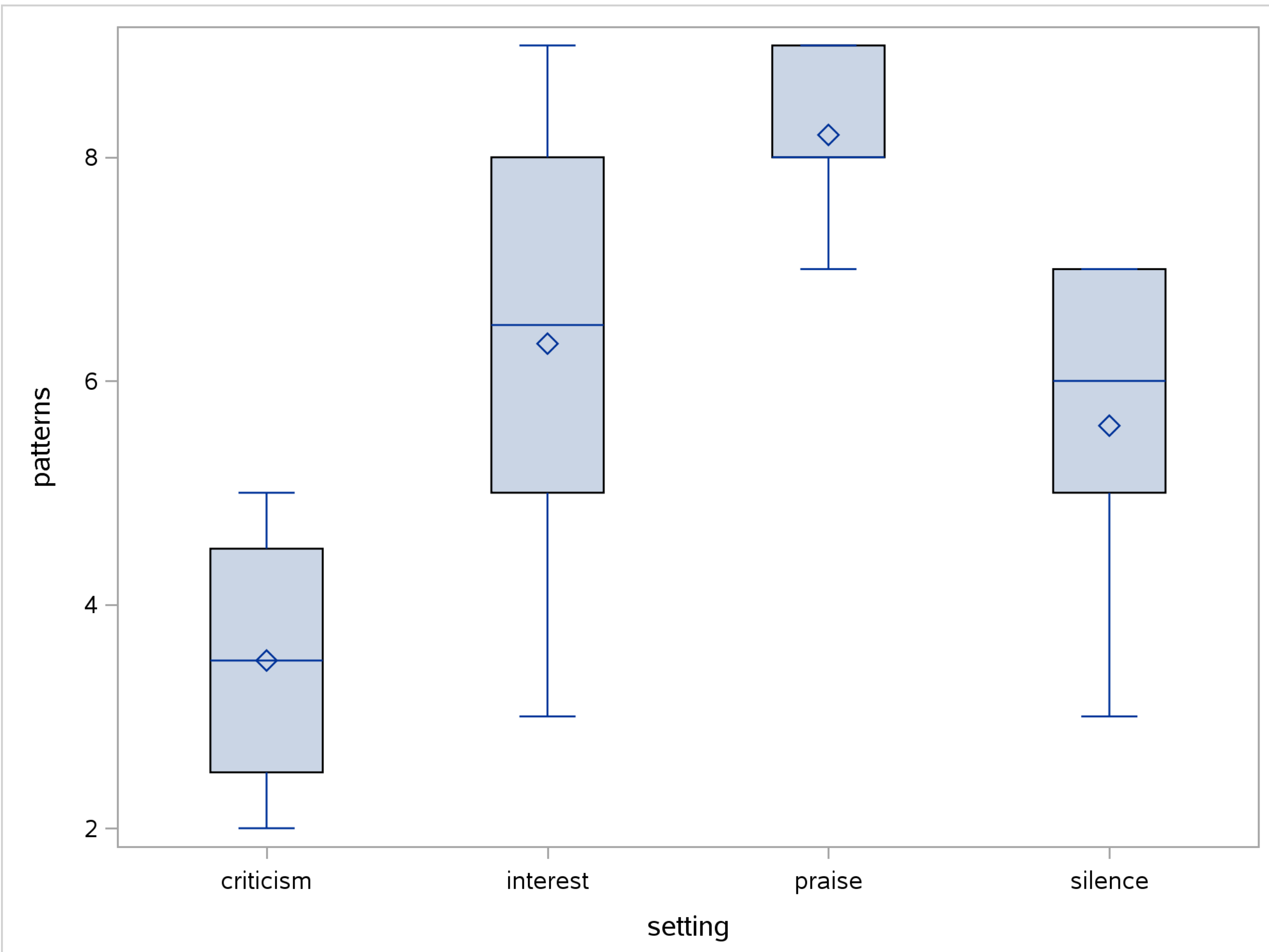

Figure 18: Boxplots of pattern-recognition data

```
proc anova;
  class setting;
  model patterns=setting;
  means setting / tukey;
```

```
                              The ANOVA Procedure

                        Dependent Variable: patterns

                                        Sum of
 Source                      DF        Squares    Mean Square   F Value   Pr > F

 Model                        3    50.61666667    16.87222222      6.38   0.0048

 Error                       16    42.33333333     2.64583333

 Corrected Total             19    92.95000000
             R-Square     Coeff Var      Root MSE    patterns Mean

             0.544558      26.88598      1.626602         6.050000
 Source                      DF       Anova SS    Mean Square   F Value   Pr > F

 setting                      3    50.61666667    16.87222222      6.38   0.0048
```

Figure 19: Analysis of variance for pattern-recognition data

```
                    The ANOVA Procedure

      Tukey's Studentized Range (HSD) Test for patterns

NOTE: This test controls the Type I experimentwise error rate.


         Alpha                                     0.05
         Error Degrees of Freedom                    16
         Error Mean Square                     2.645833
         Critical Value of Studentized Range    4.04606
Comparisons significant at the 0.05 level are indicated by ***.


                           Difference
       setting                Between     Simultaneous 95%
     Comparison                 Means    Confidence Limits

praise     - interest          1.8667     -0.9513   4.6846
praise     - silence           2.6000     -0.3433   5.5433
praise     - criticism         4.7000      1.5782   7.8218  ***
interest   - praise           -1.8667     -4.6846   0.9513
interest   - silence           0.7333     -2.0846   3.5513
interest   - criticism         2.8333     -0.1706   5.8373
silence    - praise           -2.6000     -5.5433   0.3433
silence    - interest         -0.7333     -3.5513   2.0846
silence    - criticism         2.1000     -1.0218   5.2218
criticism  - praise           -4.7000     -7.8218  -1.5782  ***
criticism  - interest         -2.8333     -5.8373   0.1706
criticism  - silence          -2.1000     -5.2218   1.0218
```

Figure 20: Tukey's method for pattern-recognition data