Booklet of Code and Output
for
STAC32 Final Exam

December 7, 2017

Figure captions are *below* the Figures they refer to.

```
LowCalorie LowFat LowCarbo Control
8 2 3 2
9 4 5 2
6 3 4 -1
7 5 2 0
3 1 3 3
```

Figure 1: Weight loss data

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## diet         3  75.75   25.25   8.559 0.00128 **
## Residuals   16  47.20    2.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

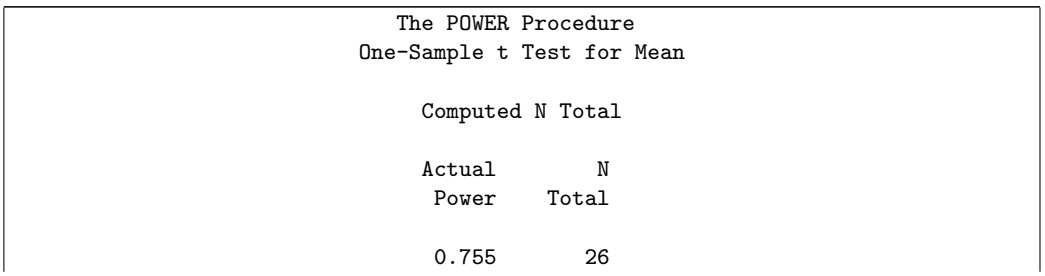
Figure 2: Weight loss ANOVA

```
##                      diff       lwr         upr        p adj
## LowCalorie-Control   5.4  2.292137  8.50786303 0.0007219131
## LowCarbo-Control     2.2 -0.907863  5.30786303 0.2199271176
## LowFat-Control       1.8 -1.307863  4.90786303 0.3769278810
## LowCarbo-LowCalorie -3.2 -6.307863 -0.09213697 0.0424599998
## LowFat-LowCalorie   -3.6 -6.707863 -0.49213697 0.0205480632
## LowFat-LowCarbo     -0.4 -3.507863  2.70786303 0.9823292501
## # A tibble: 4 x 2
##         diet weightloss_mean
##        <chr>           <dbl>
## 1 LowCalorie             6.6
## 2   LowCarbo             3.4
## 3     LowFat             3.0
## 4    Control             1.2
```

Figure 3: Weight loss Tukey, and table of weight loss mean by diet

```
New,Old
42.1,42.7
41,43.6
41.3,43.8
41.8,43.3
42.4,42.5
42.8,43.5
43.2,43.1
42.3,41.7
41.8,44
42.7,44.1
```

Figure 4: Packing machine data



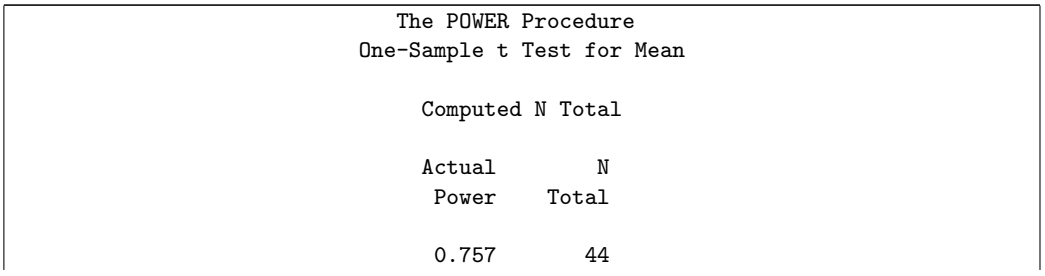
```
The POWER Procedure
One-Sample t Test for Mean

Computed N Total

Actual        N
 Power    Total

 0.755       26
```

Figure 5: Output of sample size calculation



```
The POWER Procedure
One-Sample t Test for Mean

Computed N Total

Actual        N
 Power    Total

 0.757       44
```

Figure 6: Output of a second sample size calculation

```
proc print data=smelling(obs=40);
```

Obs	agegroup	smell
1	1	1.381
2	1	1.322
3	1	1.162
4	1	1.275
5	1	1.381
6	1	1.275
7	1	1.322
8	1	1.492
9	1	1.322
10	1	1.381
11	1	1.162
12	1	1.013
13	1	1.322
14	1	1.322
15	1	1.275
16	1	1.492
17	1	1.322
18	1	1.322
19	1	1.492
20	1	1.322
21	1	1.381
22	1	1.234
23	1	1.162
24	1	1.381
25	1	1.381
26	1	1.381
27	1	1.322
28	1	1.381
29	1	1.322
30	1	1.381
31	1	1.275
32	1	1.492
33	1	1.275
34	1	1.322
35	1	1.275
36	1	1.381
37	1	1.234
38	1	1.105
39	2	1.234
40	2	1.234

Figure 7: Smell score data (some)

```
proc sgplot;
  vbox smell / category=agegroup;
```

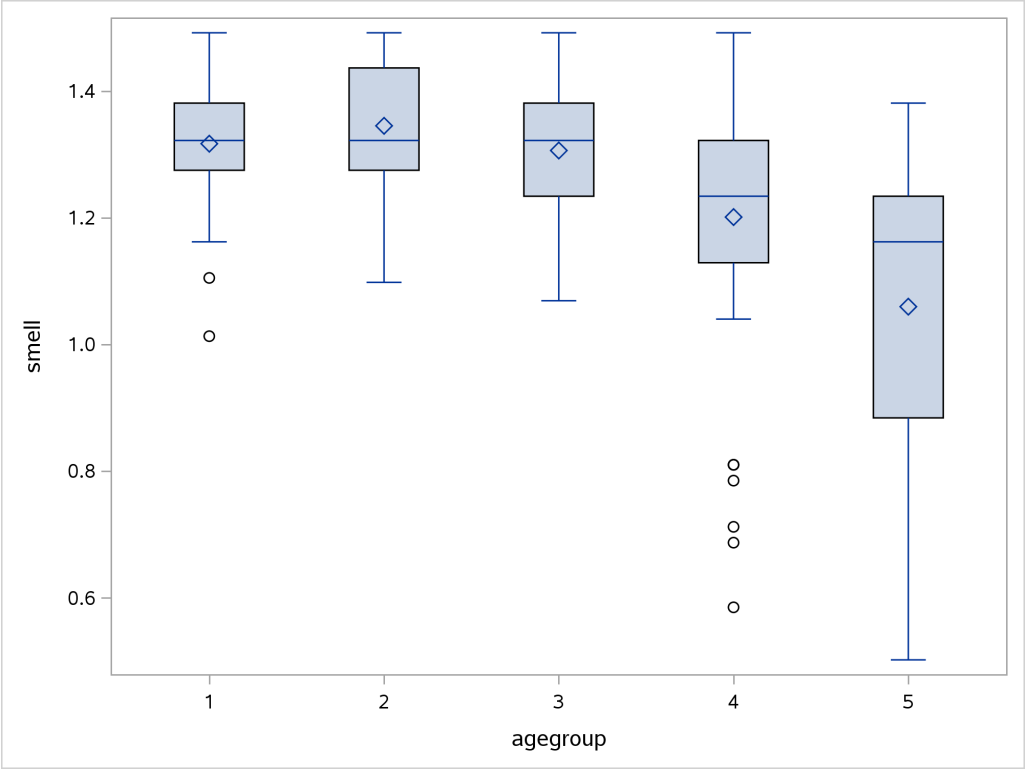


Figure 8: Smell score boxplots by age group

```
                          The ANOVA Procedure

                        Class Level Information

                    Class         Levels    Values

                    agegroup           5    1 2 3 4 5
                  Number of Observations Read         180
                  Number of Observations Used         180
                           The ANOVA Procedure

                      Dependent Variable: smell

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

Model                        4     2.13878141     0.53469535     16.65   <.0001

Error                      175     5.61970399     0.03211259

Corrected Total            179     7.75848539
              R-Square     Coeff Var      Root MSE    smell Mean

              0.275670      14.52664      0.179200      1.233594
Source                      DF       Anova SS    Mean Square   F Value   Pr > F

agegroup                     4     2.13878141     0.53469535     16.65   <.0001
                          The ANOVA Procedure

               Levene's Test for Homogeneity of smell Variance
                ANOVA of Squared Deviations from Group Means

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

       agegroup         4      0.0799      0.0200       6.35    <.0001
       Error          175      0.5503     0.00314
                        Welch's ANOVA for smell

                  Source            DF    F Value    Pr > F

                  agegroup      4.0000      13.72    <.0001
                  Error        78.7489
               Level of            ------------smell------------
               agegroup      N             Mean          Std Dev

               1            38       1.31689474       0.10365373
               2            36       1.34513889       0.10574232
               3            21       1.30614286       0.12748933
               4            43       1.20109302       0.21961425
               5            42       1.05961905       0.24594635
```

Figure 9: Smell score Welch ANOVA results

```
## # A tibble: 2,700 x 9
##       X1 region population   sex   age education income statusquo  vote
##    <int>  <chr>      <int> <chr> <int>     <chr>  <int>     <dbl> <chr>
##  1     1      N     175000     M    65         P  35000   1.00820     Y
##  2     2      N     175000     M    29        PS   7500  -1.29617     N
##  3     3      N     175000     F    38         P  15000   1.23072     Y
##  4     4      N     175000     F    49         P  35000  -1.03163     N
##  5     5      N     175000     F    23         S  35000  -1.10496     N
##  6     6      N     175000     F    28         P   7500  -1.04685     N
##  7     7      N     175000     M    26        PS  35000  -0.78626     N
##  8     8      N     175000     F    24         S  15000  -1.11348     N
##  9     9      N     175000     F    41         P  15000  -1.01292     U
## 10    10      N     175000     M    41         P  15000  -1.29617     N
## 11    11      N     175000     M    64         P  15000   1.36566     Y
## 12    12      N     175000     M    19         S  35000   1.02791     U
## 13    13      N     175000     F    27        PS     NA   1.43448     Y
## 14    14      N     175000     F    46         S  75000   1.50684     Y
## 15    15      N     175000     M    36        PS  35000   1.49026  <NA>
## 16    16      N     175000     M    22         S  15000  -1.14189     A
## 17    17      N     175000     F    20        PS  15000  -0.85579     N
## 18    18      N     175000     M    30         S  35000  -0.90689     U
## 19    19      N     175000     M    67         P  75000   1.32279     Y
## 20    20      N     175000     F    50         S   2500  -1.05805     U
## 21    21      N     175000     F    38         S  35000   1.38534     Y
## 22    22      N     175000     F    55        PS  35000   1.45602     Y
## 23    23      N     175000     M    18         S  75000   1.54808     Y
## 24    24      N     175000     F    24        PS  35000   1.61471     Y
## 25    25      N     175000     M    58         P  35000   1.52601     Y
## 26    26      N     175000     F    28         S  35000   0.55328     U
## 27    27      N     175000     F    34         P   7500   1.33619     A
## 28    28      N     175000     F    43         P     NA   0.15489     A
## 29    29      N     175000     M    20        PS  35000  -0.89281     N
## 30    30      N     175000     M    53        PS  35000  -1.29617     A
## 31    31      N     175000     F    32        PS  35000   1.50986     Y
## 32    32      N     175000     F    37        PS 200000   0.96525     Y
## 33    33      N     175000     M    36        PS   7500   0.15489     U
## 34    34      N     175000     M    44        PS  35000   1.16851     N
## 35    35      N     175000     M    20        PS  15000  -1.29617     N
## 36    36      N     175000     F    36         P  35000  -0.35147     Y
## 37    37      N     175000     M    37         P   7500   1.43203     Y
## 38    38      N     175000     M    30         S  35000   0.26360     A
## 39    39      N     175000     M    30         P  15000   0.11067     Y
## 40    40      N     175000     F    19         P  15000   0.72102     A
## # ... with 2,660 more rows
```

Figure 10: Chile plebiscite data (some)

Obs	year	sales
1	0	98
2	1	135
3	2	162
4	3	178
5	4	221
6	5	232
7	6	283
8	7	300
9	8	374
10	9	395

Figure 11: Annual sales data

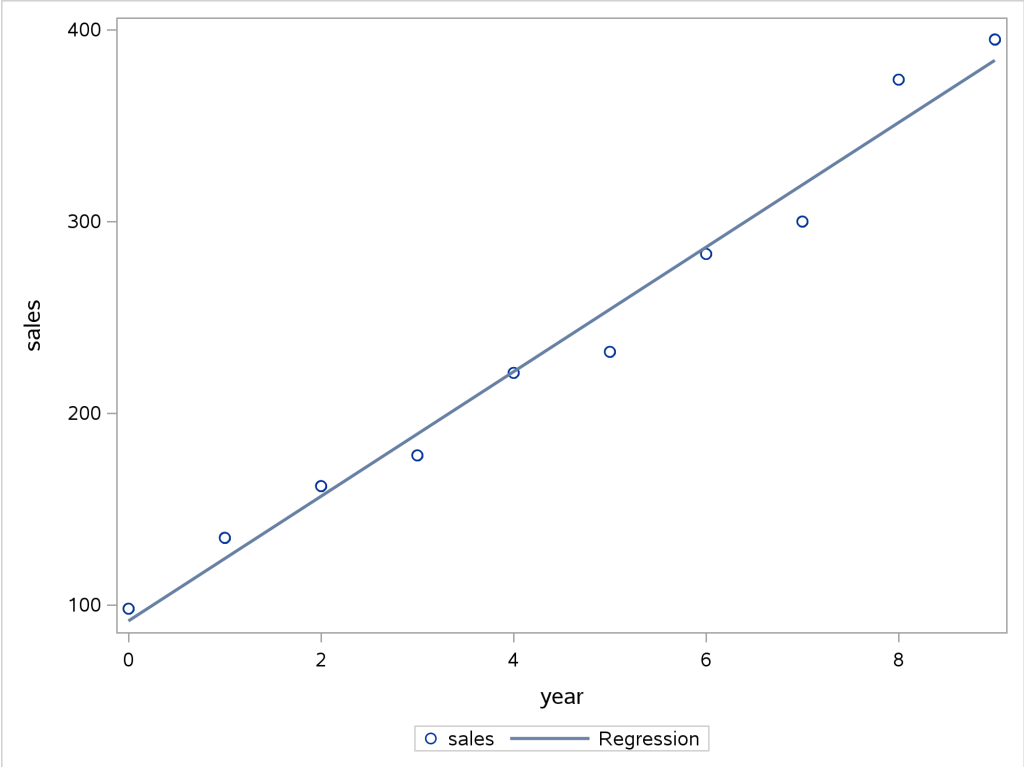


Figure 12: Scatterplot of annual sales data

```
                          The REG Procedure
                            Model: MODEL1
                       Dependent Variable: sales

                 Number of Observations Read          10
                 Number of Observations Used          10
                          Analysis of Variance

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

Model                     1          87124          87124     387.39    <.0001
Error                     8     1799.22424      224.90303
Corrected Total           9          88924

            Root MSE             14.99677    R-Square     0.9798
            Dependent Mean      237.80000    Adj R-Sq     0.9772
            Coeff Var             6.30646
                          Parameter Estimates

                         Parameter       Standard
     Variable     DF       Estimate          Error    t Value    Pr > |t|

     Intercept     1       91.56364        8.81441      10.39      <.0001
     year          1       32.49697        1.65109      19.68      <.0001
```

Figure 13: Regression text output for sales data

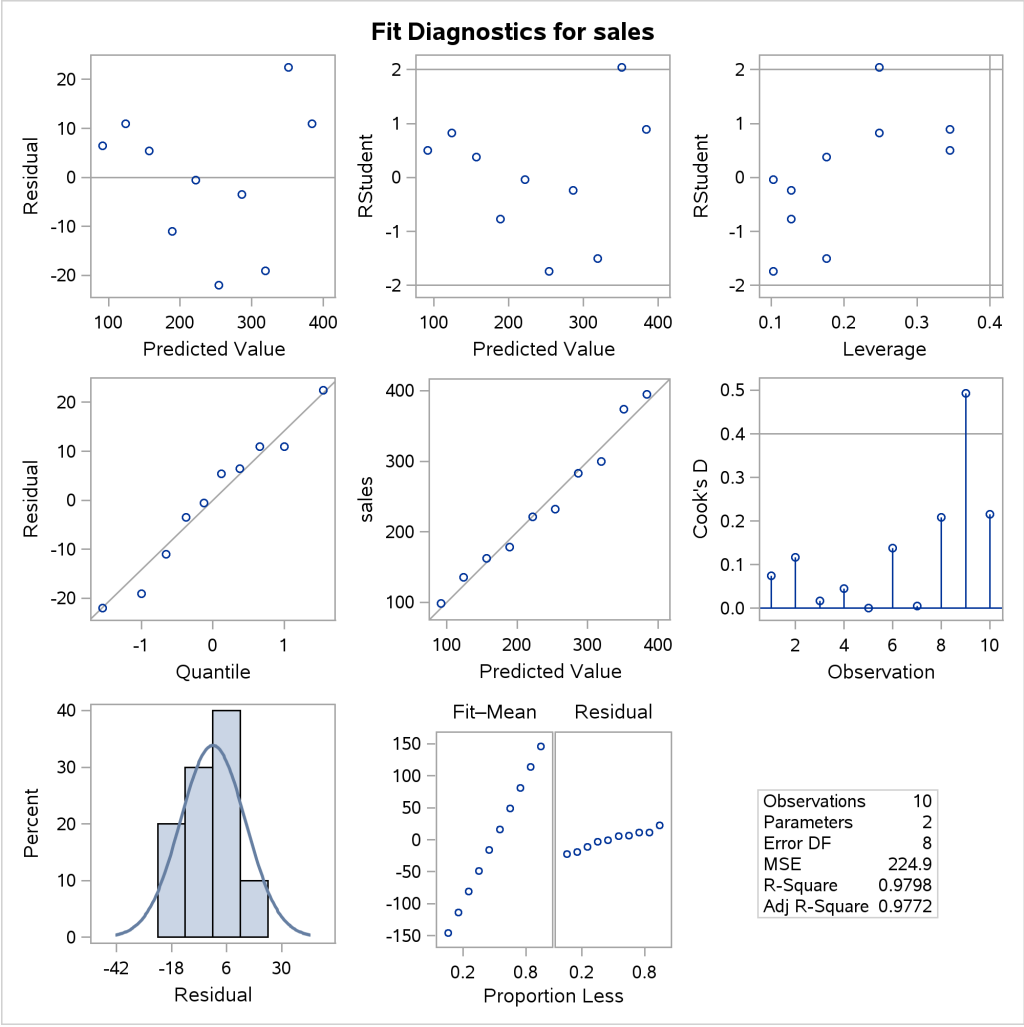


Figure 14: Regression graphics output for sales data

```
proc transreg;
  model boxcox(sales)=identity(year);
```

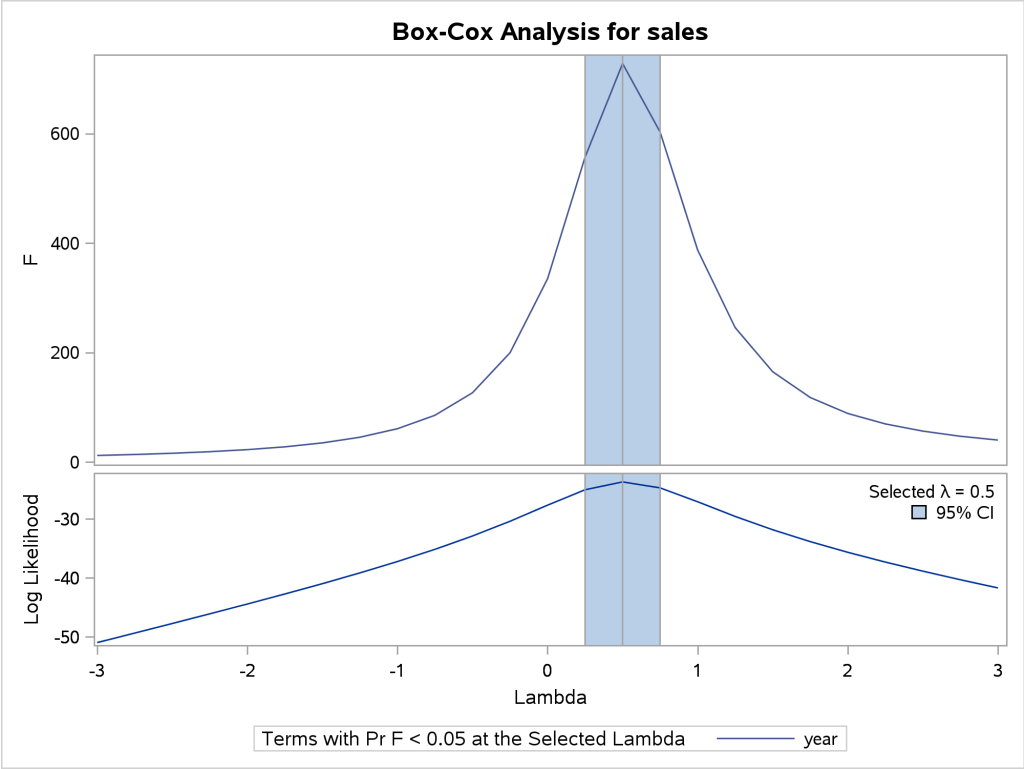


Figure 15: "Proc transreg" output for sales data

```
## Parsed with column specification:
## cols(
##  PIQ = col_integer(),
##  Brain = col_double(),
##  Height = col_double(),
##  Weight = col_integer()
## )
## # A tibble: 38 x 4
##      PIQ  Brain Height Weight
##    <int>  <dbl>  <dbl>  <int>
##  1   124  81.69   64.5    118
##  2   150 103.84   73.3    143
##  3   128  96.54   68.8    172
##  4   134  95.15   65.0    147
##  5   110  92.88   69.0    146
##  6   131  99.13   64.5    138
##  7    98  85.43   66.0    175
##  8    84  90.49   66.3    134
##  9   147  95.55   68.8    172
## 10   124  83.39   64.5    118
## # ... with 28 more rows
```

Figure 16: Performance IQ data (some)

```r
piq.1=lm(PIQ~Brain+Height+Weight,data=perf)
summary(piq.1)
##
## Call:
## lm(formula = PIQ ~ Brain + Height + Weight, data = perf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32.74 -12.09  -3.84  14.17  51.69
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.114e+02  6.297e+01   1.768 0.085979 .
## Brain        2.060e+00  5.634e-01   3.657 0.000856 ***
## Height      -2.732e+00  1.229e+00  -2.222 0.033034 *
## Weight       5.599e-04  1.971e-01   0.003 0.997750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.79 on 34 degrees of freedom
## Multiple R-squared:  0.2949,Adjusted R-squared:  0.2327
## F-statistic: 4.741 on 3 and 34 DF,  p-value: 0.007215
```

Figure 17: Performance IQ regression 1

```
piq.2=lm(PIQ~Brain,data=perf)
summary(piq.2)
##
## Call:
## lm(formula = PIQ ~ Brain, data = perf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.077 -17.508  -2.095  17.097  41.571
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.6519    43.7118   0.106   0.9158
## Brain         1.1766     0.4806   2.448   0.0194 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.21 on 36 degrees of freedom
## Multiple R-squared:  0.1427,Adjusted R-squared:  0.1189
## F-statistic: 5.994 on 1 and 36 DF,  p-value: 0.01935
```

Figure 18: Performance IQ regression 2

```
anova(piq.2,piq.1)
## Analysis of Variance Table
##
## Model 1: PIQ ~ Brain
## Model 2: PIQ ~ Brain + Height + Weight
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1     36 16198
## 2     34 13322  2    2875.7 3.6696 0.03606 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 19: Performance IQ `anova`

```
## # A tibble: 397 x 6
##         rank discipline yrs.since.phd yrs.service    sex salary
##        <chr>      <chr>          <int>       <int>  <chr>  <int>
##  1      Prof          B             19          18   Male 139750
##  2      Prof          B             20          16   Male 173200
##  3  AsstProf          B              4           3   Male  79750
##  4      Prof          B             45          39   Male 115000
##  5      Prof          B             40          41   Male 141500
##  6 AssocProf          B              6           6   Male  97000
##  7      Prof          B             30          23   Male 175000
##  8      Prof          B             45          45   Male 147765
##  9      Prof          B             21          20   Male 119250
## 10      Prof          B             18          18 Female 129000
## 11 AssocProf          B             12           8   Male 119800
## 12  AsstProf          B              7           2   Male  79800
## 13  AsstProf          B              1           1   Male  77700
## 14  AsstProf          B              2           0   Male  78000
## 15      Prof          B             20          18   Male 104800
## 16      Prof          B             12           3   Male 117150
## 17      Prof          B             19          20   Male 101000
## 18      Prof          A             38          34   Male 103450
## 19      Prof          A             37          23   Male 124750
## 20      Prof          A             39          36 Female 137000
## 21      Prof          A             31          26   Male  89565
## 22      Prof          A             36          31   Male 102580
## 23      Prof          A             34          30   Male  93904
## 24      Prof          A             24          19   Male 113068
## 25 AssocProf          A             13           8 Female  74830
## 26      Prof          A             21           8   Male 106294
## 27      Prof          A             35          23   Male 134885
## 28  AsstProf          B              5           3   Male  82379
## 29  AsstProf          B             11           0   Male  77000
## 30      Prof          B             12           8   Male 118223
## # ... with 367 more rows
```

Figure 20: Faculty salary data (some)

```
t.test(salary~sex,data=salaries,alternative="less")
##
##  Welch Two Sample t-test
##
## data:  salary by sex
## t = -3.1615, df = 50.122, p-value = 0.001332
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6620.263
## sample estimates:
## mean in group Female   mean in group Male
##             101002.4             115090.4
```

Figure 21: T-test to compare salaries for male and female faculty members

```
salary.1=lm(salary~rank+discipline+yrs.since.phd+yrs.service+sex,
            data=salaries)
summary(salary.1)
##
## Call:
## lm(formula = salary ~ rank + discipline + yrs.since.phd + yrs.service +
##     sex, data = salaries)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -65248 -13211  -1775  10384  99592
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    78862.8     4990.3  15.803  < 2e-16 ***
## rankAsstProf  -12907.6     4145.3  -3.114  0.00198 **
## rankProf       32158.4     3540.6   9.083  < 2e-16 ***
## disciplineB    14417.6     2342.9   6.154 1.88e-09 ***
## yrs.since.phd    535.1      241.0   2.220  0.02698 *
## yrs.service     -489.5      211.9  -2.310  0.02143 *
## sexMale         4783.5     3858.7   1.240  0.21584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22540 on 390 degrees of freedom
## Multiple R-squared:  0.4547,Adjusted R-squared:  0.4463
## F-statistic:  54.2 on 6 and 390 DF,  p-value: < 2.2e-16
```

Figure 22: Regression to predict salaries

Obs	month	year	comp	indproc
1	2	2012	102.9	2.052
2	3	2012	101.5	2.026
3	4	2012	100.8	2.002
4	5	2012	98	1.949
5	6	2012	97.3	1.942
6	7	2012	93.5	1.887
7	8	2012	97.5	1.986
8	9	2012	102.2	2.053
9	10	2012	105	2.102
10	11	2012	107.2	2.113
11	12	2012	105.1	2.058
12	1	2013	103.9	2.06
13	2	2013	103	2.035
14	3	2013	104.8	2.08
15	4	2013	105	2.102
16	5	2013	107.2	2.15

Figure 23: Microcomputer component data