Booklet of Code and Output
for
STAC32 Final Exam

December 14, 2018

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

```
Region Modern Historic
1 1610 1590
2 2230 2360
3 5270 5161
4 6990 7170
5 2010 1920
6 4560 4760
7 780 660
8 6510 6320
9 2850 2920
10 3550 2440
11 1710 1340
12 2050 2180
13 2750 3110
14 2550 2070
15 6750 7330
16 3670 2980
```

Figure 1: Grain yields in modern and historic times

```
systolic  age  weight
     132   52     173
     143   59     184
     153   67     194
     162   73     211
     154   64     196
     168   74     220
     137   54     180
     149   61     188
     159   65     207
     128   46     167
     166   72     217
```

Figure 2: Blood pressure data

```
##               Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 22.5126162  9.8476336 2.286094 0.05157837
## age          0.5795478  0.2290198 2.530557 0.03522485
## weight       0.4703846  0.1172728 4.011030 0.00388947
```

Figure 3: Regression for blood pressure data

Obs	child	walked
1	1	14.2
2	2	12.3
3	3	12.7
4	4	12.3
5	5	13.1
6	6	13.5
7	7	12
8	8	13.5
9	9	12.9
10	10	13.8
11	11	11.6
12	12	11.9
13	13	13.9
14	14	13.6
15	15	12.3
16	16	12.9
17	17	14.1
18	18	12.8

Figure 4: Walking data

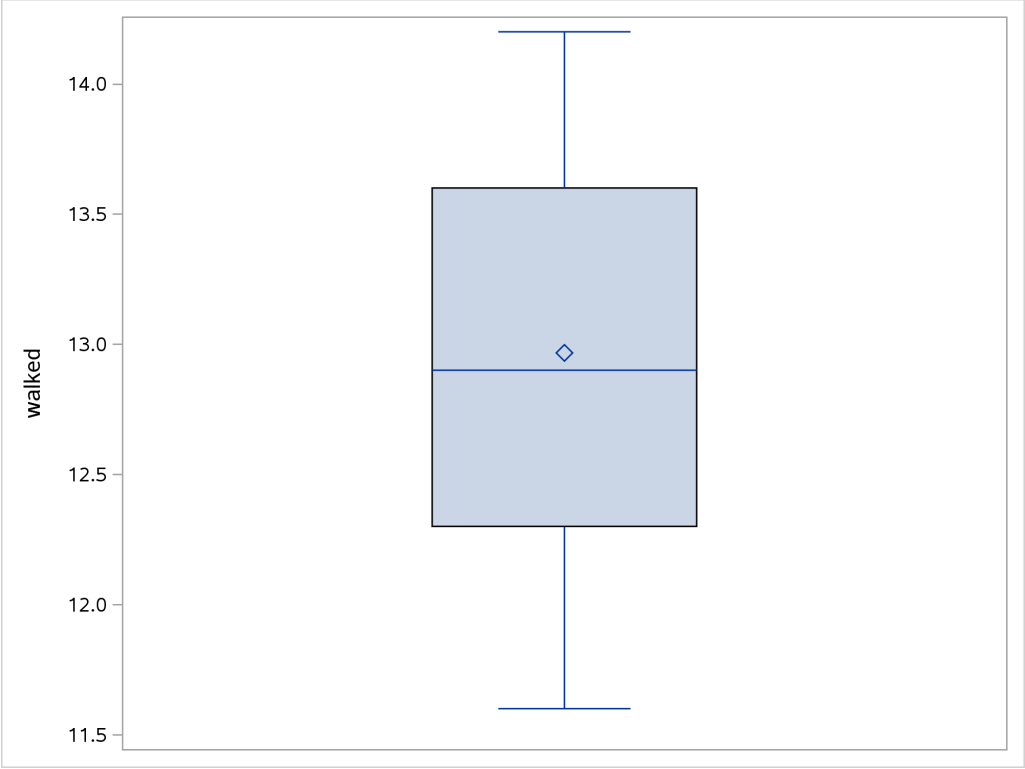

Figure 5: Boxplot of walking ages

```
proc univariate location=12.5;
  var walked;
```

```
                    The UNIVARIATE Procedure
                       Variable:  walked

                            Moments

N                          18    Sum Weights                18
Mean                12.9666667    Sum Observations         233.4
Std Deviation       0.79483628    Variance            0.63176471
Skewness            -0.0394481    Kurtosis            -1.1379773
Uncorrected SS         3037.16    Corrected SS             10.74
Coeff Variation     6.12984275    Std Error Mean      0.18734471

                  Basic Statistical Measures

          Location                    Variability

      Mean     12.96667     Std Deviation            0.79484
      Median   12.90000     Variance                 0.63176
      Mode     12.30000     Range                    2.60000
                            Interquartile Range      1.30000

                Tests for Location: Mu0=12.5

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

         Student's t    t  2.490952    Pr > |t|    0.0234
         Sign           M         3    Pr >= |M|   0.2379
         Signed Rank    S      49.5    Pr >= |S|   0.0293

                    Quantiles (Definition 5)

                     Level         Quantile

                     100% Max          14.2
                     99%               14.2
                     95%               14.2
                     90%               14.1
                     75% Q3            13.6
                     50% Median        12.9
                     25% Q1            12.3
                     10%               11.9
                     5%                11.6
                     1%                11.6
                     0% Min            11.6
```

Figure 6: Output for walking ages

Obs	region	cases
1	A	1
2	A	8
3	A	8
4	A	8
5	A	7
6	A	8
7	A	8
8	A	1
9	A	3
10	A	3
11	A	3
12	A	2
13	A	5
14	A	1
15	A	4
16	A	6
17	B	1
18	B	1
19	B	3
20	B	1
21	B	4
22	B	8
23	B	5
24	B	4
25	B	4
26	B	4
27	B	2
28	B	2
29	B	5
30	B	6
31	B	9

Figure 7: Fox rabies data

region	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	16	4.7500	2.8166	0.7042	1.0000	8.0000
B	15	3.9333	2.4339	0.6284	1.0000	9.0000
Diff (1-2)		0.8167	2.6388	0.9484		

region	Method	Mean	95% CL Mean		Std Dev
A		4.7500	3.2491	6.2509	2.8166
B		3.9333	2.5855	5.2812	2.4339
Diff (1-2)	Pooled	0.8167	-1.1230	2.7563	2.6388
Diff (1-2)	Satterthwaite	0.8167	-1.1141	2.7475	

region	Method	95% CL Std Dev	
A		2.0806	4.3593
B		1.7819	3.8385
Diff (1-2)	Pooled	2.1016	3.5474
Diff (1-2)	Satterthwaite		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	29	0.86	0.3962
Satterthwaite	Unequal	28.821	0.87	0.3940

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	15	14	1.34	0.5903

Figure 8: T-test for fox rabies data

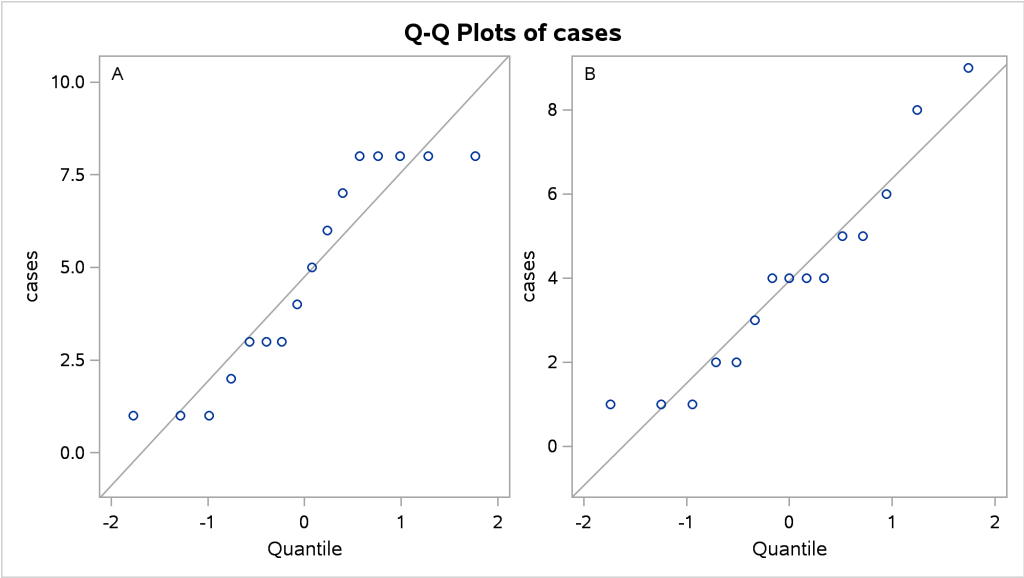

Figure 9: Graphical output from t-test for fox rabies data

Obs	test1	test2	test3	test4	proficiency
1	88	86	110	100	87
2	80	62	97	99	100
3	96	110	107	103	103
4	76	101	117	93	95
5	80	100	101	95	88
6	73	78	85	95	84
7	58	120	77	80	74
8	116	105	122	116	102
9	104	112	119	106	105
10	99	120	89	105	97
11	64	87	81	90	88
12	126	133	120	113	108
13	94	140	121	96	89
14	71	84	113	98	78
15	111	106	102	109	109
16	109	109	129	102	108
17	100	104	83	100	102
18	127	150	118	107	110
19	99	98	125	108	95
20	82	120	94	95	90
21	67	74	121	91	85
22	109	96	114	114	103
23	78	104	73	93	80
24	115	94	121	115	104
25	83	91	129	97	83

Figure 10: Job proficiency data

The REG Procedure
Model: MODEL1
Dependent Variable: proficiency

Number of Observations Read 25
Number of Observations Used 25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2192.13011	548.03253	20.17	<.0001
Error	20	543.30989	27.16549		
Corrected Total	24	2735.44000			

Root MSE	5.21205	R-Square	0.8014
Dependent Mean	94.68000	Adj R-Sq	0.7617
Coeff Var	5.50491		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	100.88142	30.03086	3.36	0.0031
test1	1	0.84060	0.21337	3.94	0.0008
test2	1	-0.19182	0.09142	-2.10	0.0488
test3	1	-0.04574	0.07258	-0.63	0.5357
test4	1	-0.58529	0.40537	-1.44	0.1643

Figure 11: Job proficiency regression, text output

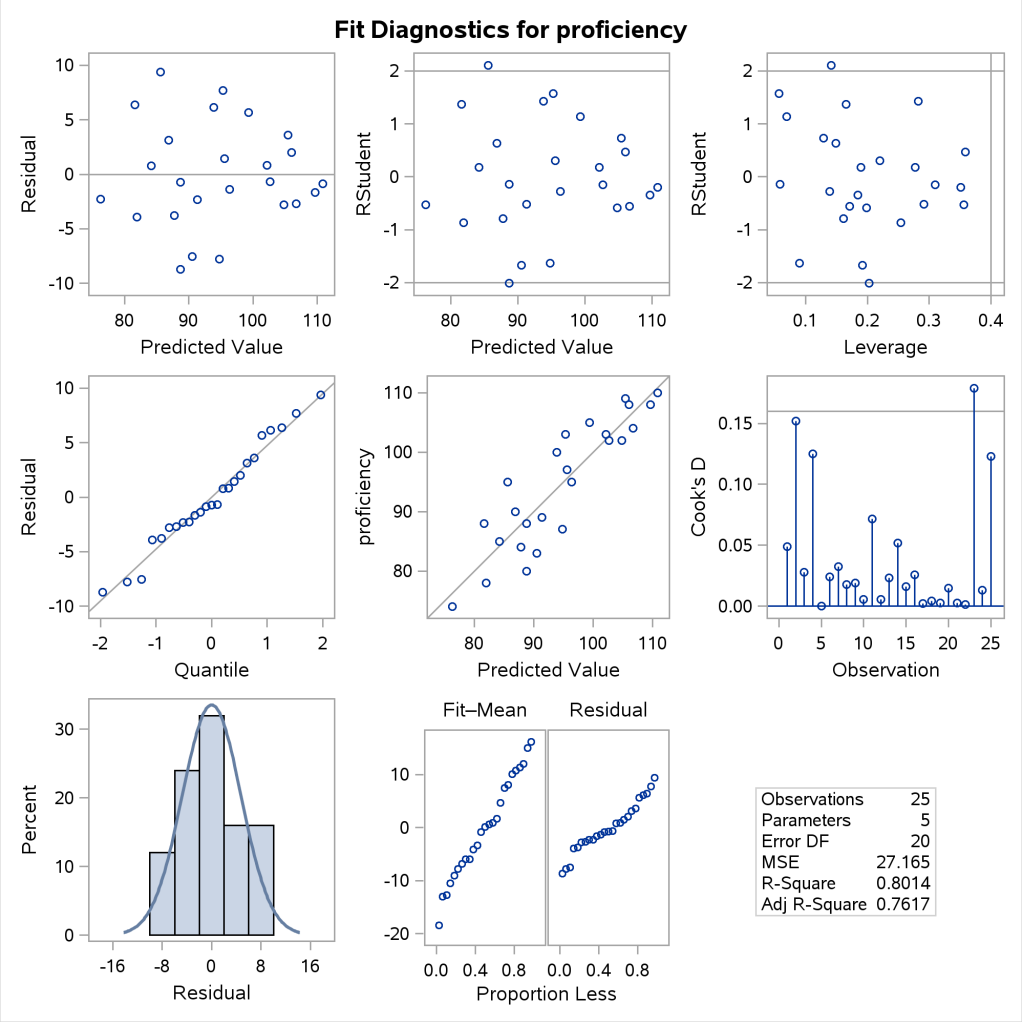

Figure 12: Job proficiency regression, graphics output part 1

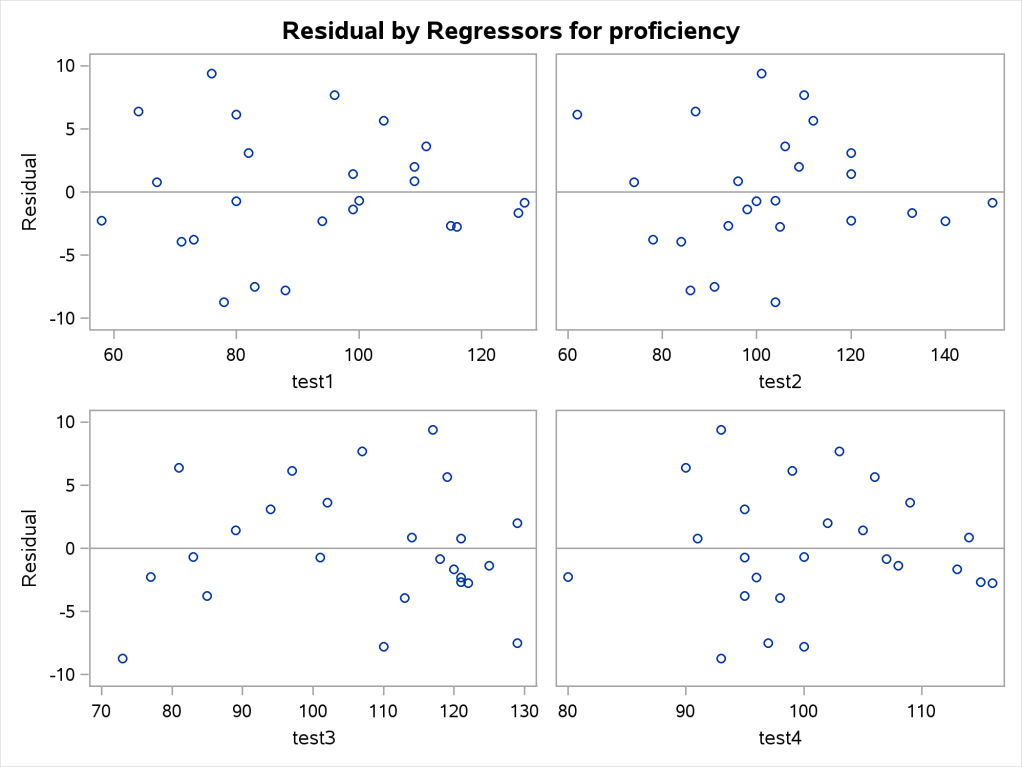

Figure 13: Job proficiency regression, graphics output part 2

```
## # A tibble: 7 x 3
##     row    xx    yy
##   <dbl> <dbl> <dbl>
## 1     1    11    10
## 2     2    12    12
## 3     3    13    13
## 4     4    14    15
## 5     5    15    22
## 6     6    16    18
## 7     7    17    20
```

Figure 14: Regression data

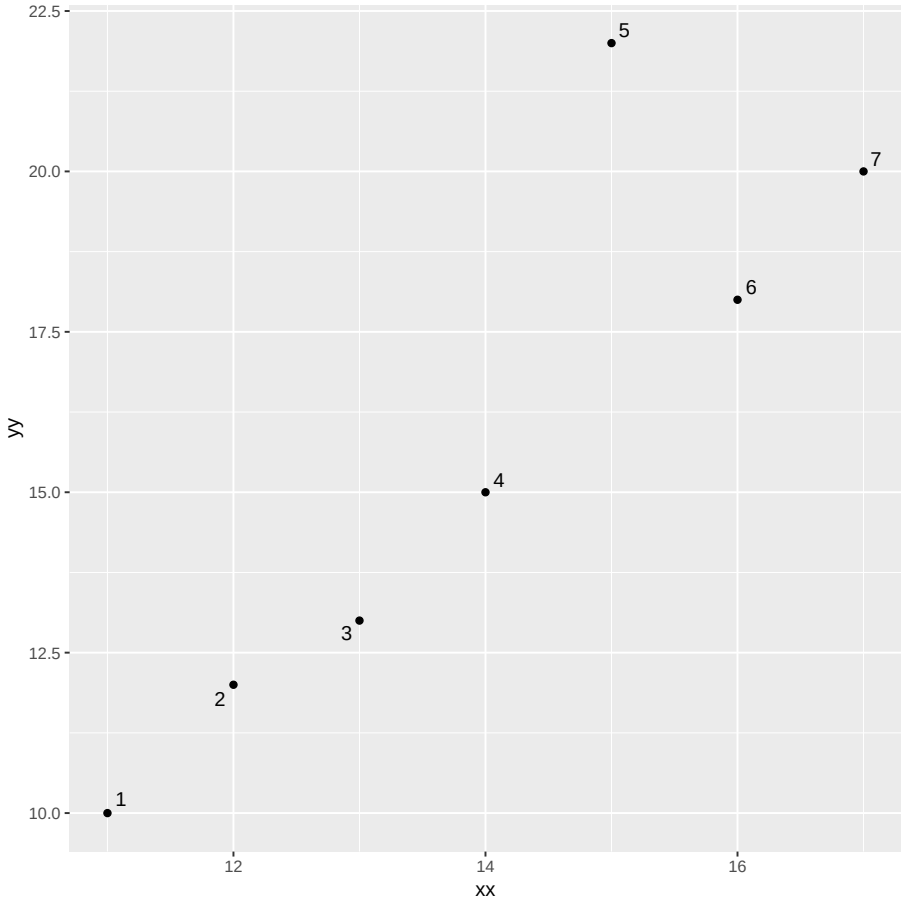

Figure 15: Scatter plot of regression data

```
rsq=function(d) {
  yy.1=lm(yy~xx, data=d)
  summary(yy.1)$r.squared
}
```

Figure 16: Function to fit a regression and return its R-squared

```
omit1=function(d,i) {
  d %>% slice(-i) %>% rsq()
}
```

Figure 17: Using the previous function to fit a regression with one row omitted, and to return its R-square

```
## # A tibble: 7 x 4
##     row    xx    yy   rsq
##   <dbl> <dbl> <dbl> <dbl>
## 1     1    11    10 0.692
## 2     2    12    12 0.758
## 3     3    13    13 0.783
## 4     4    14    15 0.795
## 5     5    15    22 0.996
## 6     6    16    18 0.803
## 7     7    17    20 0.771
```

Figure 18: Output from running `omit1` on data frame `dd`

```
## # A tibble: 20 x 2
##    smile     leniency
##    <chr>        <dbl>
##  1 neutral        2
##  2 false          8
##  3 false          7.5
##  4 miserable      3.5
##  5 felt           5
##  6 miserable      6
##  7 false          6
##  8 false          4.5
##  9 neutral        2.5
## 10 neutral        3
## 11 miserable      5
## 12 felt           5
## 13 false          6.5
## 14 felt           4
## 15 neutral        2.5
## 16 miserable      5
## 17 felt           3.5
## 18 felt           3
## 19 miserable      5.5
## 20 felt           3.5
```

Figure 19: Leniency data (20 randomly chosen rows out of 136)

```
smile_leniency %>%
    group_by(smile) %>%
    summarize(n=n(), mean=mean(leniency), med=median(leniency))
## # A tibble: 4 x 4
##   smile         n  mean   med
##   <chr>     <int> <dbl> <dbl>
## 1 false        34  5.37  5.5
## 2 felt         34  4.91  4.75
## 3 miserable    34  4.91  4.75
## 4 neutral      34  4.12  4
```

Figure 20: Leniency data sample sizes, means, and medians

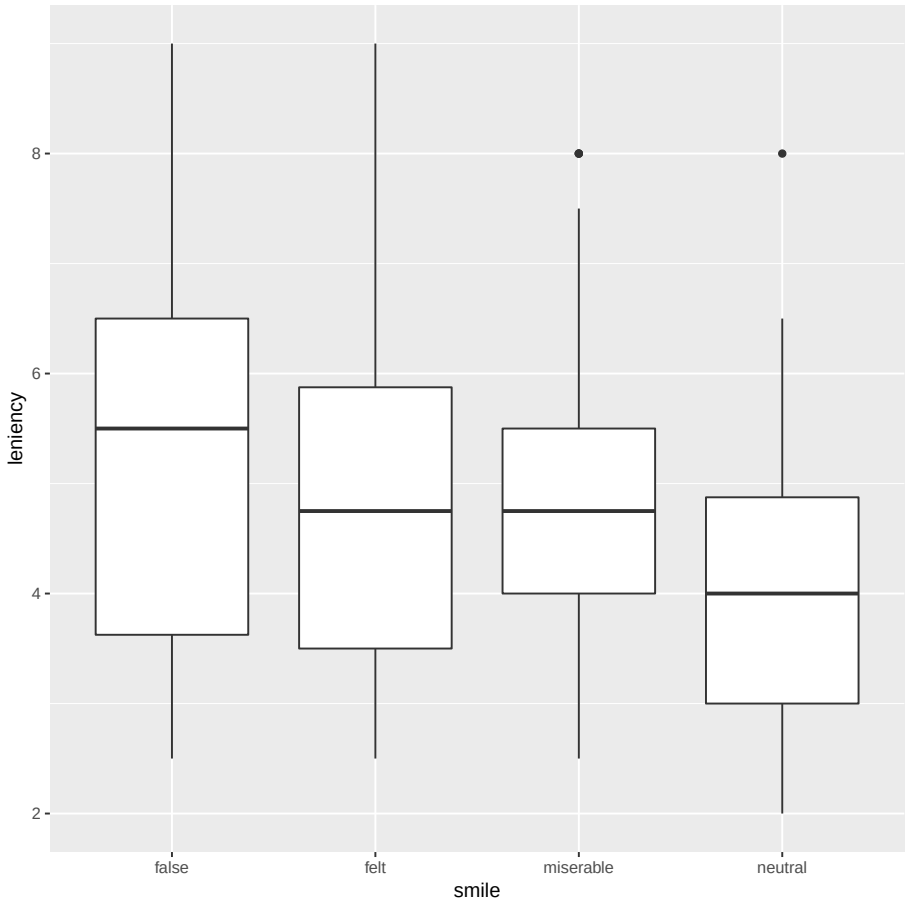

Figure 21: Leniency boxplots

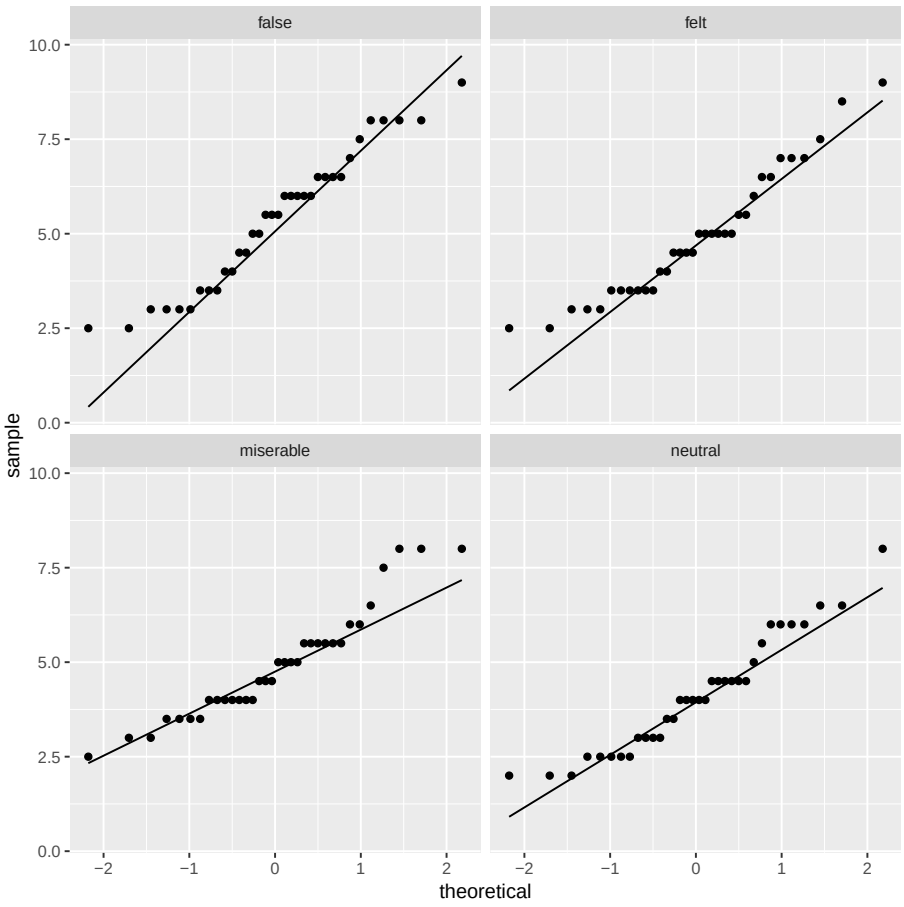

Figure 22: Leniency normal quantile plots

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## smile         3   27.5   9.178   3.465 0.0182 *
## Residuals   132  349.7   2.649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 23: Leniency data ANOVA

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  leniency and smile
## F = 3.4169, num df = 3.000, denom df = 73.091, p-value = 0.02172
```

Figure 24: Leniency data Welch ANOVA

```
## $table
##            above
## group       above below
##   false        21    11
##   felt         17    13
##   miserable    17    14
##   neutral       9    19
##
## $test
##        what      value
## 1 statistic 7.13901843
## 2        df 3.00000000
## 3   P-value 0.06759634
```

Figure 25: Leniency data Mood's median test

```
## # A tibble: 6 x 4
##   g1        g2        p_value adj_p_value
##   <chr>     <chr>       <dbl>       <dbl>
## 1 false     felt      0.121        0.723
## 2 false     miserable 0.207        1.24
## 3 false     neutral   0.00966      0.0580
## 4 felt      miserable 1            6
## 5 felt      neutral   0.0606       0.363
## 6 miserable neutral   0.0795       0.477
```

Figure 26: Leniency data: pairwise median tests

```
##
## Pairwise comparisons using Games-Howell test
## data:  leniency by factor(smile)
##           false felt  miserable
## felt      0.708 -     -
## miserable 0.667 1.000 -
## neutral   0.016 0.184 0.134
##
## P value adjustment method:  none
## alternative hypothesis:  two.sided
```

Figure 27: Leniency data: Games-Howell

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = leniency ~ smile, data = smile_leniency)
##
## $smile
##                         diff       lwr        upr     p adj
## felt-false        -0.4558824 -1.483012  0.5712478 0.6562329
## miserable-false   -0.4558824 -1.483012  0.5712478 0.6562329
## neutral-false     -1.2500000 -2.277130 -0.2228699 0.0102192
## miserable-felt     0.0000000 -1.027130  1.0271301 1.0000000
## neutral-felt      -0.7941176 -1.821248  0.2330125 0.1888804
## neutral-miserable -0.7941176 -1.821248  0.2330125 0.1888804
```

Figure 28: Leniency data: Tukey

```
## Parsed with column specification:
## cols(
##  Treatment = col_character(),
##  sales1 = col_integer(),
##  sales2 = col_integer()
## )
## # A tibble: 15 x 3
##    Treatment            sales1 sales2
##    <chr>                 <int>  <int>
##  1 athlete                  92     69
##  2 athlete                  68     44
##  3 athlete                  74     58
##  4 athlete                  52     38
##  5 athlete                  65     54
##  6 physician-stationery     77     74
##  7 physician-stationery     80     75
##  8 physician-stationery     70     73
##  9 physician-stationery     73     78
## 10 physician-stationery     79     82
## 11 physician-checkout       64     66
## 12 physician-checkout       43     49
## 13 physician-checkout       81     84
## 14 physician-checkout       68     75
## 15 physician-checkout       71     77
```

Figure 29: Marker sales data

```
drop1(markers.1, test="F")
## Single term deletions
##
## Model:
## sales2 ~ sales1 + Treatment
##           Df Sum of Sq     RSS    AIC F value    Pr(>F)
## <none>                  176.53 44.982
## sales1     1    1190.7 1367.20 73.687  74.194 3.214e-06 ***
## Treatment  2    1397.3 1573.81 73.798  43.534 5.947e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 30: Marker sales regression output 1

```
summary(markers.1)
##
## Call:
## lm(formula = sales2 ~ sales1 + Treatment, data = markers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7636 -2.7666  0.7781  2.4288  5.7406
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -5.99860    7.03500  -0.853    0.412
## sales1                          0.83474    0.09691   8.614 3.21e-06 ***
## Treatmentphysician-checkout    21.60674    2.57598   8.388 4.15e-06 ***
## Treatmentphysician-stationery 19.12547    2.59110   7.381 1.39e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.006 on 11 degrees of freedom
## Multiple R-squared:  0.939,Adjusted R-squared:  0.9223
## F-statistic: 56.39 on 3 and 11 DF,  p-value: 5.758e-07
```

Figure 31: Marker sales regression output 2

```
proc print;
```

Obs	x
1	0
2	26
3	30
4	33
5	34
6	35
7	37
8	39
9	41
10	44
11	48
12	104

Figure 32: Data for estimating `sigma`

Obs	x16	x50	x84
1	26	36	48

Figure 33: 16th, 50th (median) and 84th percentiles of column `x` in Figure 32

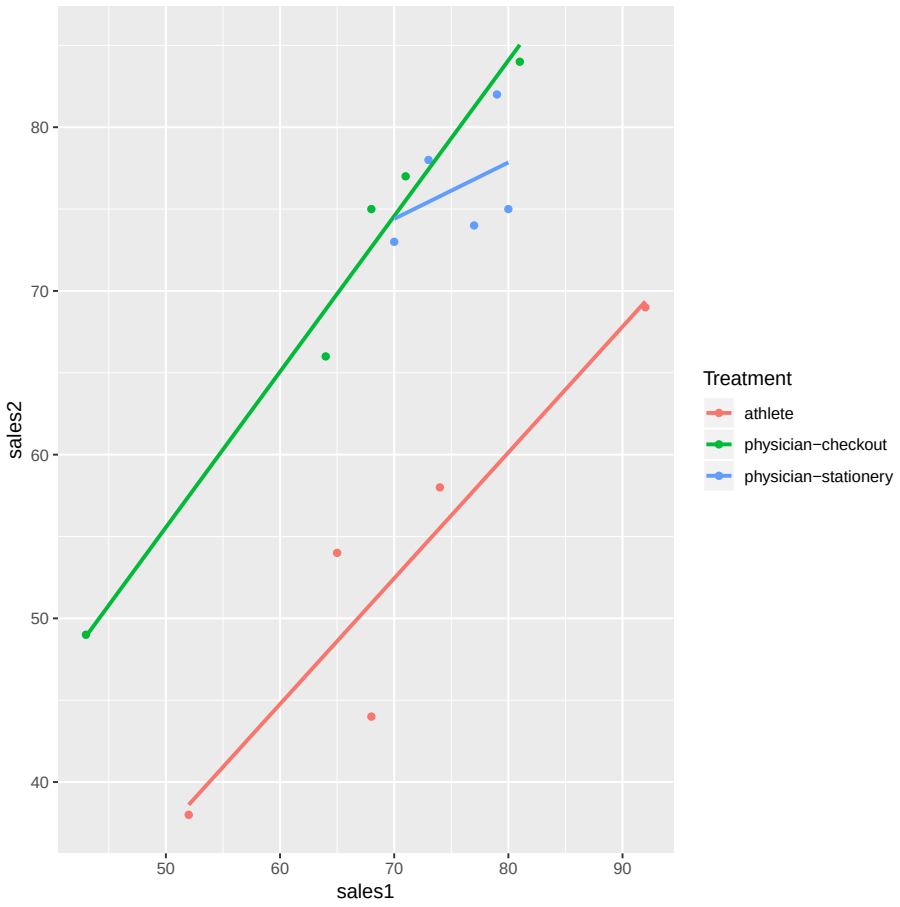

Figure 34: Marker sales plot