STA442s04 Overheads Set Five:
More Dummy Variable Regression

```
/******************** smile1.sas **********************/
options linesize=79 pagesize=100 noovp formdlim='_';
title 'Beekman smile-gaze data:  Conversation One';
title2 'DV is smile time of person one';

proc format;
     value sexfmt 1 = 'Female' 2 = 'Male';

data smile;
     infile 'smilegz1.dat' firstobs=2; /* Skip the first line */
     input id sex1  sex2  trntime1  smtime1  gztime1  smtime2  gztime2;
/* Comment the labels out for now
     label id       = 'Conversation Number'
           sex1     = 'Sex of Person 1'
           sex2     = 'Sex of Person 2'
           trntime1 = 'Speaking Time of Person One in sec'
           smtime1  = 'Smiling Time of Person One in sec'
           gztime1  = 'Gazing Time of Person One in sec'
           smtime2  = 'Smiling Time of Person Two in sec'
           gztime2  = 'Gazing Time of Person Two in sec';
*/
     smiledif = smtime1 - smtime2;
     smileav = (smtime1+smtime2)/2;
     /* Combination and dummy vars */
     combo = 10*sex1 + sex2;
     if combo = 11 then ff=1; else ff=0;
     if combo = 12 then fm=1; else fm=0;
     if combo = 21 then mf=1; else mf=0;
     if combo = 22 then mm=1; else mm=0;
     /* Difference between smiling of Female and Male, where defined  */
     if combo=12 then FminusM = smtime1 - smtime2;
        else if combo = 21 then FminusM = smtime2 - smtime1;
        else FminusM = .;
     label FminusM = 'Smile time of Female minus Male';
     format sex1 sex2 sexfmt.;

proc freq;
     table sex1*sex2 / norow nocol nopercent missing;
     table (ff--mm) * combo / norow nocol nopercent missing;
proc corr;
     var trntime1  smtime1  gztime1  smtime2  gztime2;
```

```
/* Now I want the variable labels back. Here is how you can have data steps
   after proc steps, if you really want to. */

data labsmile;
     set smile; /* Bring in the data set smile created above. */
     label id       = 'Conversation Number'
           sex1     = 'Sex of Person 1'
           sex2     = 'Sex of Person 2'
           trntime1 = 'Speaking Time of Person One in sec'
           smtime1  = 'Smiling Time of Person One in sec'
           gztime1  = 'Gazing Time of Person One in sec'
           smtime2  = 'Smiling Time of Person Two in sec'
           gztime2  = 'Gazing Time of Person Two in sec';

/* Now take a look at the means in a nice tabular format. */

proc tabulate;
     class sex1 sex2;
     var smtime1;
     table (sex1 all),(sex2 all) * (mean*smtime1);

/* Basic one-way ANOVA -- well, fairly basic */
proc glm;
     class combo;
     model smtime1 = combo;
     means combo;
     means combo / bon tukey scheffe;
     estimate 'FvsM_Beh' combo 1 1 -1 -1 / divisor = 2;

/* Now oneway using proc reg and dummy variables.
   First with intercept */

proc reg;
     model smtime1 = ff fm mf;
     sexdif: test ff+fm = mf; /* Av smiling of Females vs Males */

/* Special tests are easier with cell means coding: No intercept => No algebra
   Here are some questions we want to ask.

   1. Do females smile more at males, or at other females?
   2. Do males smile more at females, or at other males?
   3. Do females reveive more smiling from males, or from other females?
   4. Do males reveive more smiling from females, or from other males?
   5. Who smiles more on average, males or females?
   6. Who receives more smiling on average, males or females?
   7. Does the sex difference in mean smiling time depend on sex of partner?
*/
```

```
proc reg;
     model smtime1 = ff fm mf mm / noint;
     sexdif:  test ff+fm = mf+mm; /* Compare M and F again; it's easier */
     alleq:   test ff=fm=mf=mm; /*  The one-way ANOVA test again */
     q1:      test ff=fm;
     q2:      test mf=mm;
     q3:      test ff=mf;
     q4:      test fm=mm;
     q5:      test ff+fm = mf+mm; /* Same as sexdiff */
     q6:      test ff+mf = fm+mm;
     q7:      test ff-fm = mf-mm;

/* Actually it's a two-way ANOVA. q5 and q6 are about main effects,
   and q7 is about the interaction. */

proc glm;
     class sex1 sex2;
     model smtime1 = sex1|sex2;

/*   The model statement could have been
     model smtime1 = sex1 sex2 sex1*sex2;   */

/* One problem with this analysis is that it throws away information about
   person 2. So let's ask:

9.  In conversations between males and females, who smiles more?

10. Is there more average smiling in conversations between women than in
    conversations between men?

11. Does average amount of smiling in a conversation depend on sex composition
    of the dyad?

12. Is average amount of smiling just a linear function of the number of
    women?

*/

/* q9 two ways: first the fancy way */

proc reg;
     model smiledif = ff fm mf mm / noint;
     q9:   test fm=mf=0;
proc univariate normal plot;
     var FminusM;

proc reg;
     model smileav = ff fm mf mm / noint;
     q10:   test ff=mm;
     q11:   test 2*ff = fm+mf = 2*mm;
     q12:   test 2*ff - fm+mf = fm+mf - 2*mm;
```

```
proc freq;
     table sex1*sex2 / norow nocol nopercent missing;
     table (ff--mm) * combo / norow nocol nopercent missing;
```

TABLE OF SEX1 BY SEX2

SEX1 \ SEX2	Female	Male	Total
Female	11	11	22
Male	11	11	22
Total	22	22	44

TABLE OF FF BY COMBO

FF \ COMBO	11	12	21	22	Total
0	0	11	11	11	33
1	11	0	0	0	11
Total	11	11	11	11	44

TABLE OF FM BY COMBO

FM \ COMBO	11	12	21	22	Total
0	11	0	11	11	33
1	0	11	0	0	11
Total	11	11	11	11	44

TABLE OF MF BY COMBO

MF COMBO

Frequency	11	12	21	22	Total
0	11	11	0	11	33
1	0	0	11	0	11
Total	11	11	11	11	44

TABLE OF MM BY COMBO

MM COMBO

Frequency	11	12	21	22	Total
0	11	11	11	0	33
1	0	0	0	11	11
Total	11	11	11	11	44

```
proc corr;
    var trntime1  smtime1  gztime1  smtime2  gztime2;
```

Correlation Analysis

5 'VAR' Variables: TRNTIME1 SMTIME1 GZTIME1 SMTIME2 GZTIME2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
TRNTIME1	44	157.2045	46.1144	6917	58.0000	263.0000
SMTIME1	44	34.1136	26.5271	1501	1.0000	134.0000
GZTIME1	44	212.3864	43.0015	9345	118.0000	297.0000
SMTIME2	44	34.2045	34.5803	1505	0	138.0000
GZTIME2	44	225.1591	43.1595	9907	58.0000	291.0000

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 44

	TRNTIME1	SMTIME1	GZTIME1	SMTIME2	GZTIME2
TRNTIME1	1.00000	0.01314	-0.56791	0.25519	0.34910
	0.0	0.9326	0.0001	0.0946	0.0202
SMTIME1	0.01314	1.00000	-0.08263	0.59727	0.08130
	0.9326	0.0	0.5939	0.0001	0.5999
GZTIME1	-0.56791	-0.08263	1.00000	-0.21012	-0.12699
	0.0001	0.5939	0.0	0.1710	0.4114
SMTIME2	0.25519	0.59727	-0.21012	1.00000	0.30215
	0.0946	0.0001	0.1710	0.0	0.0462
GZTIME2	0.34910	0.08130	-0.12699	0.30215	1.00000
	0.0202	0.5999	0.4114	0.0462	0.0

```
/* Now take a look at the means in a nice tabular format. */

proc tabulate;
     class sex1 sex2;
     var smtime1;
     table (sex1 all),(sex2 all) * (mean*smtime1);
```

	Sex of Person 2		
	Female	Male	ALL
	MEAN	MEAN	MEAN
	Smiling Time of Person One in sec	Smiling Time of Person One in sec	Smiling Time of Person One in sec
Sex of Person 1			
Female	47.27	39.55	43.41
Male	32.09	17.55	24.82
ALL	39.68	28.55	34.11

```
/* Basic one-way ANOVA -- well, fairly basic */
proc glm;
     class combo;
     model smtime1 = combo;
     means combo;
     means combo / bon tukey scheffe;
     estimate 'FvsM_Beh' combo 1 1 -1 -1 / divisor = 2;
```

```
                  Beekman smile-gaze data:  Conversation One                 4
                        DV is smile time of person one
                                        12:55 Wednesday, February 18, 2004

                        General Linear Models Procedure
                            Class Level Information

                        Class    Levels    Values

                        COMBO         4    11 12 21 22


                    Number of observations in data set = 44
```

```
                  Beekman smile-gaze data:  Conversation One                 5
                        DV is smile time of person one
                                        12:55 Wednesday, February 18, 2004

                        General Linear Models Procedure

Dependent Variable: SMTIME1   Smiling Time of Person One in sec
                                      Sum of            Mean
Source                   DF          Squares          Square  F Value    Pr > F

Model                     3     5293.8863636    1764.6287879     2.83    0.0506

Error                    40    24964.5454545     624.1136364

Corrected Total          43    30258.4318182

                  R-Square             C.V.        Root MSE       SMTIME1 Mean

                  0.174956         73.23249       24.982266          34.113636
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
COMBO	3	5293.8863636	1764.6287879	2.83	0.0506

Source	DF	Type III SS	Mean Square	F Value	Pr > F
COMBO	3	5293.8863636	1764.6287879	2.83	0.0506

General Linear Models Procedure

Level of COMBO	N	SMTIME1 Mean	SMTIME1 SD
11	11	47.2727273	33.8676569
12	11	39.5454545	24.2213279
21	11	32.0909091	24.0518380
22	11	17.5454545	13.5747091

General Linear Models Procedure

Tukey's Studentized Range (HSD) Test for variable: SMTIME1

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 40 MSE= 624.1136
Critical Value of Studentized Range= 3.791
Minimum Significant Difference= 28.553

Means with the same letter are not significantly different.

Tukey Grouping		Mean	N	COMBO
	A	47.27	11	11
	A			
B	A	39.55	11	12
B	A			
B	A	32.09	11	21
B				
B		17.55	11	22

General Linear Models Procedure

Bonferroni (Dunn) T tests for variable: SMTIME1

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 40 MSE= 624.1136
Critical Value of T= 2.78
Minimum Significant Difference= 29.57

Means with the same letter are not significantly different.

Bon Grouping		Mean	N	COMBO
	A	47.27	11	11
	A			
B	A	39.55	11	12
B	A			
B	A	32.09	11	21
B				
B		17.55	11	22

General Linear Models Procedure

Scheffe's test for variable: SMTIME1

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 40 MSE= 624.1136
Critical Value of F= 2.83875
Minimum Significant Difference= 31.087

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	COMBO
A	47.27	11	11
A			
A	39.55	11	12
A			
A	32.09	11	21
A			
A	17.55	11	22

```
estimate 'FvsM_Beh' combo 1 1 -1 -1 / divisor = 2;
```


General Linear Models Procedure

Dependent Variable: SMTIME1 Smiling Time of Person One in sec

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
FvsM_Beh	18.5909091	2.47	0.0180	7.53243674

```
/* Now oneway using proc reg and dummy variables.
   First with intercept */

proc reg;
     model smtime1 = ff fm mf;
     sexdif: test ff+fm = mf; /* Av smiling of Females vs Males */
```

Model: MODEL1
Dependent Variable: SMTIME1 Smiling Time of Person One in sec

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	5293.88636	1764.62879	2.827	0.0506
Error	40	24964.54545	624.11364		
C Total	43	30258.43182			

Root MSE	24.98227	R-square	0.1750
Dep Mean	34.11364	Adj R-sq	0.1131
C.V.	73.23249		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	17.545455	7.53243674	2.329	0.0250
FF	1	29.727273	10.65247420	2.791	0.0080
FM	1	22.000000	10.65247420	2.065	0.0454
MF	1	14.545455	10.65247420	1.365	0.1797

Dependent Variable: SMTIME1
Test: SEXDIF Numerator: 3801.8409 DF: 1 F value: 6.0916
              Denominator: 624.1136 DF: 40 Prob>F: 0.0180

Repeating the table for reference ...

	Sex of Person 2		
	Female	Male	ALL
	MEAN	MEAN	MEAN
	Smiling Time of Person One in sec	Smiling Time of Person One in sec	Smiling Time of Person One in sec
Sex of Person 1			
Female	47.27	39.55	43.41
Male	32.09	17.55	24.82
ALL	39.68	28.55	34.11

```
 /* Special tests are easier with cell means coding: No intercept => No algebra
   Here are some questions we want to ask.

   1. Do females smile more at males, or at other females?
   2. Do males smile more at females, or at other males?
   3. Do females reveive more smiling from males, or from other females?
   4. Do males reveive more smiling from females, or from other males?
   5. Who smiles more on average, males or females?
   6. Who receives more smiling on average, males or females?
   7. Does the sex difference in mean smiling time depend on sex of partner?
*/

proc reg;
     model smtime1 = ff fm mf mm / noint;
     sexdif:  test ff+fm = mf+mm; /* Compare M and F again; it's easier */
     alleq:   test ff=fm=mf=mm; /*  The one-way ANOVA test again */
     q1:      test ff=fm;
     q2:      test mf=mm;
     q3:      test ff=mf;
     q4:      test fm=mm;
     q5:      test ff+fm = mf+mm; /* Same as sexdiff */
     q6:      test ff+mf = fm+mm;
     q7:      test ff-fm = mf-mm;
```

Model: MODEL1
NOTE: No intercept in model. R-square is redefined.
Dependent Variable: SMTIME1 Smiling Time of Person One in sec

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	56498.45455	14124.61364	22.631	0.0001
Error	40	24964.54545	624.11364		
U Total	44	81463.00000			

Root MSE	24.98227	R-square	0.6935
Dep Mean	34.11364	Adj R-sq	0.6629
C.V.	73.23249		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
FF	1	47.272727	7.53243674	6.276	0.0001
FM	1	39.545455	7.53243674	5.250	0.0001
MF	1	32.090909	7.53243674	4.260	0.0001
MM	1	17.545455	7.53243674	2.329	0.0250

Variable	DF	Variable Label
FF	1	
FM	1	
MF	1	
MM	1	

1. Do females smile more at males, or at other females?
2. Do males smile more at females, or at other males?
3. Do females reveive more smiling from males, or from other females?
4. Do males reveive more smiling from females, or from other males?
5. Who smiles more on average, males or females?
6. Who receives more smiling on average, males or females?
7. Does the sex difference in mean smiling time depend on sex of partner?

```
Dependent Variable: SMTIME1
Test: SEXDIF   Numerator:   3801.8409  DF:    1   F value:   6.0916
               Denominator:  624.1136  DF:   40   Prob>F:    0.0180


Dependent Variable: SMTIME1
Test: ALLEQ    Numerator:   1764.6288  DF:    3   F value:   2.8274
               Denominator:  624.1136  DF:   40   Prob>F:    0.0506


Dependent Variable: SMTIME1
Test: Q1       Numerator:    328.4091  DF:    1   F value:   0.5262
               Denominator:  624.1136  DF:   40   Prob>F:    0.4724


Dependent Variable: SMTIME1
Test: Q2       Numerator:   1163.6364  DF:    1   F value:   1.8645
               Denominator:  624.1136  DF:   40   Prob>F:    0.1797


Dependent Variable: SMTIME1
Test: Q3       Numerator:   1267.6818  DF:    1   F value:   2.0312
               Denominator:  624.1136  DF:   40   Prob>F:    0.1619


Dependent Variable: SMTIME1
Test: Q4       Numerator:   2662.0000  DF:    1   F value:   4.2652
               Denominator:  624.1136  DF:   40   Prob>F:    0.0454


Dependent Variable: SMTIME1
Test: Q5       Numerator:   3801.8409  DF:    1   F value:   6.0916
               Denominator:  624.1136  DF:   40   Prob>F:    0.0180


Dependent Variable: SMTIME1
Test: Q6       Numerator:   1364.2045  DF:    1   F value:   2.1858
               Denominator:  624.1136  DF:   40   Prob>F:    0.1471


Dependent Variable: SMTIME1
Test: Q7       Numerator:    127.8409  DF:    1   F value:   0.2048
               Denominator:  624.1136  DF:   40   Prob>F:    0.6533
```

```
/* Actually it's a two-way ANOVA. q5 and q6 are about main effects,
   and q7 is about the interaction. */

proc glm;
     class sex1 sex2;
     model smtime1 = sex1|sex2;

/*   The model statement could have been
     model smtime1 = sex1 sex2 sex1*sex2;   */
```

General Linear Models Procedure
Class Level Information

Class	Levels	Values
SEX1	2	Female Male
SEX2	2	Female Male

Number of observations in data set = 44

General Linear Models Procedure

Dependent Variable: SMTIME1 Smiling Time of Person One in sec

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5293.8863636	1764.6287879	2.83	0.0506
Error	40	24964.5454545	624.1136364		
Corrected Total	43	30258.4318182			

R-Square	C.V.	Root MSE	SMTIME1 Mean
0.174956	73.23249	24.982266	34.113636

```
proc glm;
     class sex1 sex2;
     model smtime1 = sex1|sex2;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEX1	1	3801.8409091	3801.8409091	6.09	0.0180
SEX2	1	1364.2045455	1364.2045455	2.19	0.1471
SEX1*SEX2	1	127.8409091	127.8409091	0.20	0.6533

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX1	1	3801.8409091	3801.8409091	6.09	0.0180
SEX2	1	1364.2045455	1364.2045455	2.19	0.1471
SEX1*SEX2	1	127.8409091	127.8409091	0.20	0.6533

```
/* One problem with this analysis is that it throws away information about
   person 2. So let's ask:

9.  In conversations between males and females, who smiles more?

10. Is there more average smiling in conversations between women than in
    conversations between men?

11. Does average amount of smiling in a conversation depend on sex composition
    of the dyad?

12. Is average amount of smiling just a linear function of the number of
    women?

*/

/* q9 two ways: first the fancy way */

proc reg;
     model smiledif = ff fm mf mm / noint;
     q9:   test fm=mf=0;
proc univariate normal plot;
     var FminusM;
```

Model: MODEL1
NOTE: No intercept in model. R-square is redefined.
Dependent Variable: SMILEDIF

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	1867.09091	466.77273	0.571	0.6851
Error	40	32692.90909	817.32273		
U Total	44	34560.00000			

Root MSE	28.58886	R-square	0.0540
Dep Mean	-0.09091	Adj R-sq	-0.0406
C.V.	-31447.74237		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
FF	1	-1.818182	8.61986462	-0.211	0.8340
FM	1	10.272727	8.61986462	1.192	0.2404
MF	1	-7.727273	8.61986462	-0.896	0.3754
MM	1	-1.090909	8.61986462	-0.127	0.8999

---

9. In conversations between males and females, who smiles more?

```
q9:   test fm=mf=0;
```


Dependent Variable: SMILEDIF

Test: Q9	Numerator:	908.8182	DF: 2	F value:	1.1119
	Denominator:	817.3227	DF: 40	Prob>F:	0.3389

```
     if combo=12 then FminusM = smtime1 - smtime2;
        else if combo = 21 then FminusM = smtime2 - smtime1;
        else FminusM = .;
     label FminusM = 'Smile time of Female minus Male';

proc univariate normal plot;
     var FminusM;
```

```
                  Beekman smile-gaze data:  Conversation One                19
                        DV is smile time of person one
                                        12:55 Wednesday, February 18, 2004

                              Univariate Procedure

Variable=FMINUSM       Smile time of Female minus Male

                                   Moments

                N                 22  Sum Wgts          22
                Mean               9  Sum              198
                Std Dev     30.73543  Variance    944.6667
                Skewness    1.318085  Kurtosis     3.12146
                USS            21620  CSS            19838
                CV          341.5048  Std Mean    6.552816
                T:Mean=0    1.373455  Pr>|T|        0.1841
                Num ^= 0          21  Num > 0           12
                M(Sign)          1.5  Pr>=|M|       0.6636
                Sgn Rank        34.5  Pr>=|S|       0.2392
                W:Normal    0.903231  Pr<W          0.0322


                            Quantiles(Def=5)

                 100% Max       104       99%        104
                  75% Q3         25       95%         48
                  50% Med         5       90%         36
                  25% Q1         -8       10%        -24
                   0% Min       -32        5%        -27
                                           1%        -32
                 Range          136
                 Q3-Q1           33
                 Mode           -24
```

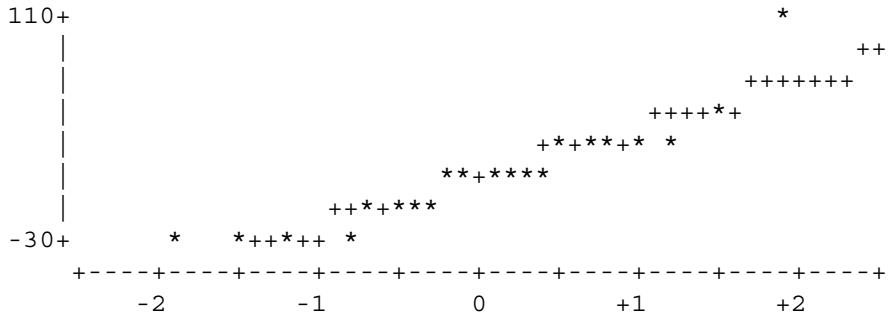
Extremes

Lowest	Obs	Highest	Obs
-32(	41)	29(	20)
-27(	17)	31(	38)
-24(	43)	36(	32)
-24(	16)	48(	14)
-23(	28)	104(	22)

Missing Value	.
Count	22
% Count/Nobs	50.00

```
 Stem Leaf                     #             Boxplot
   10 4                        1                0
    8                                           
    6                                           
    4 8                         1                |
    2 45916                     5             +-----+
    0 037929                    6             *--+--*
   -0 8731                      4             +-----+
   -2 27443                     5                |
      ----+----+----+----+
  Multiply Stem.Leaf by 10**+1
```

Normal Probability Plot



10. Is there more average smiling in conversations between women than in conversations between men?

11. Does average amount of smiling in a conversation depend on sex composition of the dyad?

12. Is average amount of smiling just a linear function of the number of women?

```
proc reg;
     model smileav = ff fm mf mm / noint;
     q10:   test ff=mm;
     q11:   test 2*ff = fm+mf = 2*mm;
     q12:   test 2*ff - fm+mf = fm+mf - 2*mm;
```

Skipping the ANOVA summary table ...

```
                Beekman smile-gaze data:  Conversation One                21
                      DV is smile time of person one
                                          12:55 Wednesday, February 18, 2004

Dependent Variable: SMILEAV
Test: Q10      Numerator:   4980.0455  DF:    1   F value:   7.3345
               Denominator:   678.992  DF:   40   Prob>F:    0.0099


Dependent Variable: SMILEAV
Test: Q11      Numerator:   2513.0341  DF:    2   F value:   3.7011
               Denominator:   678.992  DF:   40   Prob>F:    0.0335


Dependent Variable: SMILEAV
Test: Q12      Numerator:   3722.7348  DF:    1   F value:   5.4827
               Denominator:   678.992  DF:   40   Prob>F:    0.0243
```