

Factorial ANOVA

```
/* potato.sas */
options linesize=79 noovp formdlim='_';
title 'Rotten potatoes';
title2 'Two-factor ANOVA several different ways';

proc format;
  value tfmt 1 = 'Cool' 2 = 'Warm';

data spud;
  infile 'potato2.data' firstobs=2; /* Skip the first line that R uses */
  input id bact temp rot;
  /* Cell means coding for all 6 treatment combinations */
  if temp=1 and bact=1 then mu11=1; else mu11=0;
  if temp=1 and bact=2 then mu12=1; else mu12=0;
  if temp=1 and bact=3 then mu13=1; else mu13=0;
  if temp=2 and bact=1 then mu21=1; else mu21=0;
  if temp=2 and bact=2 then mu22=1; else mu22=0;
  if temp=2 and bact=3 then mu23=1; else mu23=0;
  combo = 10*temp+bact;
  format temp tfmt.;

proc means;
  class bact temp;
  var rot;

/* Better looking output from proc tabulate */

proc tabulate;
  class bact temp;
  var rot;
  table (temp all),(bact all) * (mean*rot);

proc glm;
  title3 'Standard 2-way ANOVA with proc glm';
  class bact temp;
  model rot=temp|bact; /* Could have said bact temp bact*temp */
  means temp|bact;

/* Need to plot it; SAS is not the tool. */
```

```
/* Now generate the tests for main effects and interaction using cell means coding.
```

	BACTERIA TYPE		
TEMP	1	2	3
Cool	mu11	mu12	mu13
Warm	mu21	mu22	mu23

*/

```
/* The test statement of proc reg uses variable names to stand for the corresponding regression coefficients. By naming the effect cell mean coding dummy variables the same as the population cell means, I can just state the null hypothesis. Isn't this a cute SAS trick? */
```

```
proc reg;
  title3 'Using the proc reg test statement and cell means coding';
  model rot = mu11--mu23 / noint;
  Overall:      test mu11=mu12=mu13=mu21=mu22=mu23;
  Temperature: test mu11+mu12+mu13 = mu21+mu22+mu23;
  Bacteria:     test mu11+mu21 = mu12+mu22 = mu13+mu23;
  Bact_by_Temp1: test mu11-mu21 = mu12-mu22 = mu13-mu23;
  Bact_by_Temp2: test mu12-mu11 = mu22-mu21,
                    mu13-mu12 = mu23-mu22;
```

```
/* Bact_by_Temp1 checks equality of temperature effects.
   Bact_by_Temp2 checks parallel line segments. They are equivalent. */
```

```
proc glm;
  title3 'Proc glm: Using contrasts on the combination variable';
  class combo;      /* 11 12 13  21 22 23 */
  model rot=combo;
  contrast 'Main Effect for Temperature'
    combo 1 1 1 -1 -1 -1;
  contrast 'Main Effect for Bacteria'
    combo 1 -1 0  1 -1 0,
    combo 0 1 -1  0  1 -1;
  contrast 'Temperature by Bacteria Interaction'
    combo 1 -1 0 -1 1 0,
    combo 0 1 -1  0 -1 1;
```

```
/* Illustrate effect coding */
```

```
data mashed;
  set spud;
  /* Effect coding, with interactions */
  if bact = 1 then b1 = 1;
  else if bact = 2 then b1 = 0;
  else if bact = 3 then b1 = -1;
  if bact = 1 then b2 = 0;
  else if bact = 2 then b2 = 1;
  else if bact = 3 then b2 = -1;
  if temp = 1 then t = 1;
  else if temp = 2 then t = -1;
  /* Interaction terms */
  tb1 = t*b1; tb2 = t*b2;
```

```

proc reg;
  title3 'Effect coding';
  model rot = b1 b2 t tb1 tb2;
  Temperature: test t=0;
  Bacteria: test b1=b2=0;
  Bact_by_Temp: test tb1=tb2=0;

/* Do some exploration to follow up the interaction. The regression
with cell means coding is easiest. The final product of several runs
is shown below. For reference, here is the table of population means again.

```

TEMP	BACTERIA TYPE		
	1	2	3
Cool	mu11	mu12	mu13
Warm	mu21	mu22	mu23

```

proc reg;
  title3 'Further exploration using cell means coding';
  model rot = mu11--mu23 / noint;
  /* Pairwise comparisons of marginal means for Bacteria Type */
  Bact1vs2: test mu11+mu21=mu12+mu22;
  Bact1vs3: test mu11+mu21=mu13+mu23;
  Bact2vs3: test mu12+mu22=mu13+mu23;
  /* Effect of temperature for each bacteria type */
  Temp_for_Bac1: test mu11=mu21;
  Temp_for_Bac2: test mu12=mu22;
  Temp_for_Bac3: test mu13=mu23;
  /* Effect of bacteria type for each temperature */
  Bact_for_CoolTemp: test mu11=mu12=mu13;
  Bact_for_WarmTemp: test mu21=mu22=mu23;
  /* Pairwise comparisons of bacteria types at warm temperature */
  Bact1vs2_for_WarmTemp: test mu21=mu22;
  Bact1vs3_for_WarmTemp: test mu21=mu23;
  Bact2vs3_for_WarmTemp: test mu22=mu23;

```

```

/* We have done a lot of tests. Concerned about buildup of Type I
error? We can make ALL the tests into Scheffe follow-ups of the
initial test for equality of the 6 cell means. The Scheffe family
includes all COLLECTIONS of contrasts, not just all contrasts. */

```

```

proc iml;
  title3 'Table of critical values for all possible Scheffe tests';
  numdf = 5; /* Numerator degrees of freedom for initial test */
  dendf = 48; /* Denominator degrees of freedom for initial test */
  alpha = 0.05;
  critval = finv(1-alpha,numdf,dendf);
  zero = {0 0}; S_table = repeat(zero,numdf,1); /* Make empty matrix */
  /* Label the columns */
  namz = {"Number of Contrasts in followup test"
         " Scheffe Critical Value"}; mattrib S_table colname=namz;
  do i = 1 to numdf;
    s_table(|i,1|) = i;
    s_table(|i,2|) = numdf/i * critval;
  end;
  reset noname; /* Makes output look nicer in this case */
  print "Initial test has" numdf " and " dendf "degrees of freedom."
        "Using significance level alpha = " alpha;
  print s_table;

```

Rotten potatoes
Two-factor ANOVA several different ways

1

The MEANS Procedure

Analysis Variable : rot

bact	temp	N Obs	N	Mean	Std Dev	Minimum
1	Cool	9	9	3.5555556	4.2752518	0
	Warm	9	9	7.0000000	3.5355339	0
2	Cool	9	9	4.7777778	3.1135903	0
	Warm	9	9	13.5555556	6.3267510	3.0000000
3	Cool	9	9	8.0000000	4.5552168	2.0000000
	Warm	9	9	19.5555556	5.5251948	8.0000000

Analysis Variable : rot

bact	temp	N Obs	Maximum
1	Cool	9	9.0000000
	Warm	9	10.0000000
2	Cool	9	10.0000000
	Warm	9	23.0000000
3	Cool	9	15.0000000
	Warm	9	26.0000000

Rotten potatoes
Two-factor ANOVA several different ways

2

	bact			All
	1	2	3	
	Mean	Mean	Mean	
	rot	rot	rot	
temp				
Cool	3.56	4.78	8.00	5.44
Warm	7.00	13.56	19.56	13.37
All	5.28	9.17	13.78	9.41

Rotten potatoes
Two-factor ANOVA several different ways
Standard 2-way ANOVA with proc glm

3

The GLM Procedure

Class Level Information

Class	Levels	Values
bact	3	1 2 3
temp	2	Cool Warm
Number of Observations Read		54
Number of Observations Used		54

Rotten potatoes
Two-factor ANOVA several different ways
Standard 2-way ANOVA with proc glm

4

The GLM Procedure

Dependent Variable: rot

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1652.814815	330.562963	15.05	<.0001
Error	48	1054.222222	21.962963		
Corrected Total	53	2707.037037			

R-Square	Coeff Var	Root MSE	rot Mean
0.610562	49.81676	4.686466	9.407407

Source	DF	Type I SS	Mean Square	F Value	Pr > F
temp	1	848.0740741	848.0740741	38.61	<.0001
bact	2	651.8148148	325.9074074	14.84	<.0001
bact*temp	2	152.9259259	76.4629630	3.48	0.0387

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	1	848.0740741	848.0740741	38.61	<.0001
bact	2	651.8148148	325.9074074	14.84	<.0001
bact*temp	2	152.9259259	76.4629630	3.48	0.0387

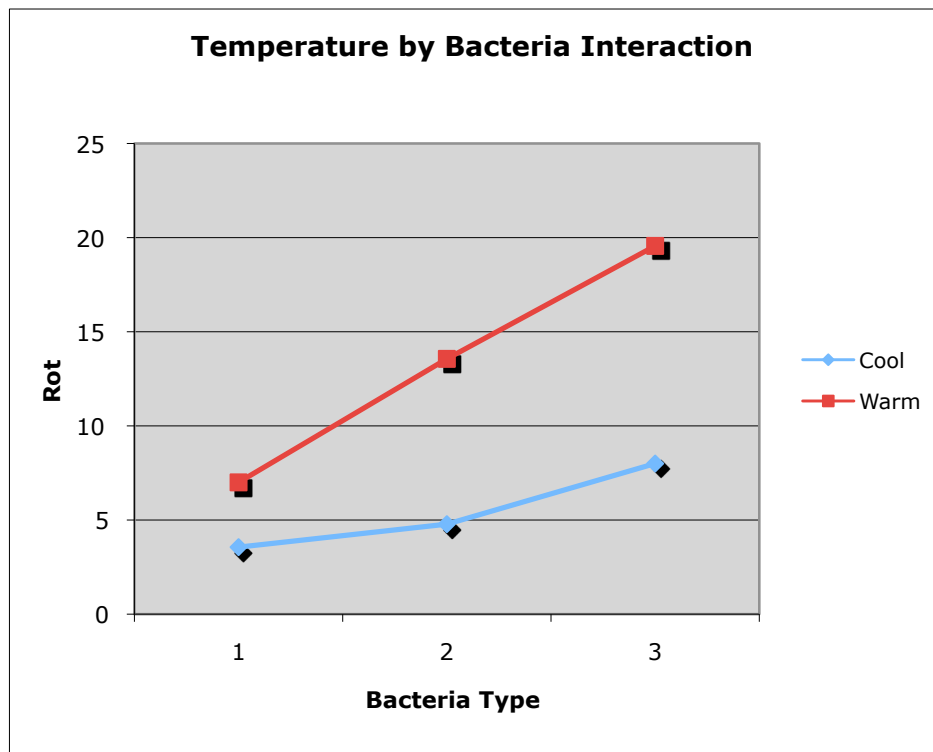
Rotten potatoes
Two-factor ANOVA several different ways
Standard 2-way ANOVA with proc glm

The GLM Procedure

Level of temp	N	-----rot----- Mean	Std Dev
Cool	27	5.4444444	4.31752541
Warm	27	13.3703704	7.27031979

Level of bact	N	-----rot----- Mean	Std Dev
1	18	5.2777778	4.19811660
2	18	9.1666667	6.61771242
3	18	13.7777778	7.71214135

Level of bact	Level of temp	N	-----rot----- Mean	Std Dev
1	Cool	9	3.5555556	4.27525178
1	Warm	9	7.0000000	3.53553391
2	Cool	9	4.7777778	3.11359028
2	Warm	9	13.5555556	6.32675097
3	Cool	9	8.0000000	4.55521679
3	Warm	9	19.5555556	5.52519482



Rotten potatoes 6
 Two-factor ANOVA several different ways
 Using the proc reg test statement and cell means coding

The REG Procedure
 Model: MODEL1
 Dependent Variable: rot

Number of Observations Read 54
 Number of Observations Used 54

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	6431.77778	1071.96296	48.81	<.0001
Error	48	1054.22222	21.96296		
Uncorrected Total	54	7486.00000			

Root MSE	4.68647	R-Square	0.8592
Dependent Mean	9.40741	Adj R-Sq	0.8416
Coeff Var	49.81676		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
mu11	1	3.55556	1.56216	2.28	0.0273
mu12	1	4.77778	1.56216	3.06	0.0036
mu13	1	8.00000	1.56216	5.12	<.0001
mu21	1	7.00000	1.56216	4.48	<.0001
mu22	1	13.55556	1.56216	8.68	<.0001
mu23	1	19.55556	1.56216	12.52	<.0001

7

Rotten potatoes
 Two-factor ANOVA several different ways
 Using the proc reg test statement and cell means coding

The REG Procedure
 Model: MODEL1

Test Overall Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	330.56296	15.05	<.0001
Denominator	48	21.96296		

Test Temperature Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	848.07407	38.61	<.0001
Denominator	48	21.96296		

Test Bacteria Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	325.90741	14.84	<.0001
Denominator	48	21.96296		

Test Bact_by_Temp1 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	76.46296	3.48	0.0387
Denominator	48	21.96296		

Test Bact_by_Temp2 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	76.46296	3.48	0.0387
Denominator	48	21.96296		

Rotten potatoes
 Two-factor ANOVA several different ways
 Proc glm: Using contrasts on the combination variable

12

The GLM Procedure

Class Level Information

Class	Levels	Values
combo	6	11 12 13 21 22 23
Number of Observations Read		54
Number of Observations Used		54

Dependent Variable: rot

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1652.814815	330.562963	15.05	<.0001
Error	48	1054.222222	21.962963		
Corrected Total	53	2707.037037			

R-Square	Coeff Var	Root MSE	rot Mean
0.610562	49.81676	4.686466	9.407407

Source	DF	Type I SS	Mean Square	F Value	Pr > F
combo	5	1652.814815	330.562963	15.05	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
combo	5	1652.814815	330.562963	15.05	<.0001

Contrast	DF	Contrast SS	Mean Square
Main Effect for Temperature	1	848.0740741	848.0740741
Main Effect for Bacteria	2	651.8148148	325.9074074
Temperature by Bacteria Interaction	2	152.9259259	76.4629630

Contrast	F Value	Pr > F
Main Effect for Temperature	38.61	<.0001
Main Effect for Bacteria	14.84	<.0001
Temperature by Bacteria Interaction	3.48	0.0387

Rotten potatoes
Two-factor ANOVA several different ways
Effect coding

14

The REG Procedure
Dependent Variable: rot

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1652.81481	330.56296	15.05	<.0001
Error	48	1054.22222	21.96296		
Corrected Total	53	2707.03704			

Root MSE	4.68647	R-Square	0.6106
Dependent Mean	9.40741	Adj R-Sq	0.5700

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	9.40741	0.63775	14.75	<.0001
b1	1	-4.12963	0.90191	-4.58	<.0001
b2	1	-0.24074	0.90191	-0.27	0.7907
t	1	-3.96296	0.63775	-6.21	<.0001
tb1	1	2.24074	0.90191	2.48	0.0165
tb2	1	-0.42593	0.90191	-0.47	0.6389

Test Temperature Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	848.07407	38.61	<.0001
Denominator	48	21.96296		

Test Bacteria Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	325.90741	14.84	<.0001
Denominator	48	21.96296		

Test Bact_by_Temp Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	76.46296	3.48	0.0387
Denominator	48	21.96296		

Rotten potatoes
Two-factor ANOVA several different ways
Further exploration using cell means coding

18

Showing only the output from the test statements ...

Test Bact1vs2 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	136.11111	6.20	0.0163
Denominator	48	21.96296		

Test Bact1vs3 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	650.25000	29.61	<.0001
Denominator	48	21.96296		

Test Bact2vs3 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	191.36111	8.71	0.0049
Denominator	48	21.96296		

Test Temp_for_Bac1 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	53.38889	2.43	0.1255
Denominator	48	21.96296		

Test Temp_for_Bac2 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	346.72222	15.79	0.0002
Denominator	48	21.96296		

Test Temp_for_Bac3 Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	600.88889	27.36	<.0001
Denominator	48	21.96296		

Test Bact_for_CoolTemp Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	47.44444	2.16	0.1264
Denominator	48	21.96296		

Test Bact_for_WarmTemp Results for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	354.92593	16.16	<.0001
Denominator	48	21.96296		

Test Bact1vs2_for_WarmTemp Results
for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	193.38889	8.81	0.0047
Denominator	48	21.96296		

Test Bact1vs3_for_WarmTemp Results
for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	709.38889	32.30	<.0001
Denominator	48	21.96296		

Test Bact2vs3_for_WarmTemp Results
for Dependent Variable rot

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	162.00000	7.38	0.0092
Denominator	48	21.96296		

Rotten potatoes
Two-factor ANOVA several different ways
Table of critical values for all possible Scheffe tests

30

Initial test has 5 and 48 degrees of freedom.
Using significance level alpha = 0.05

Number of Contrasts in followup test	Scheffe Critical Value
1	12.042571
2	6.0212853
3	4.0141902
4	3.0106426
5	2.4085141

SENIC data are unbalanced.

```

/***** senic2way.sas *****/
%include 'senicread.sas'; /* senicread.sas reads data, etc. */
title2 'Two-way ANCOVA on SENIC Data';

proc glm;
  class region medschl;
  model infrisk = nbeds census nurses region|medschl;
  lsmeans region|medschl;

/* Check relationships among explanatory variables */
proc freq;
  tables medschl*region / nocol nopercnt chisq;
proc corr nosimple;
  var nbeds census nurses;

```

Study of the Effectiveness of Nosocomial Infection Control
Two-way ANCOVA on SENIC Data

1

The GLM Procedure

Class Level Information

Class	Levels	Values
region	4	North Central Northeast South West
medschl	2	No Yes

Number of Observations Read 113
 Number of Observations Used 113

Study of the Effectiveness of Nosocomial Infection Control 2
 Two-way ANCOVA on SENIC Data

The GLM Procedure

Dependent Variable: infrisk Prob of acquiring infection in hospital

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	55.4449776	5.5444978	3.88	0.0002
Error	102	145.9348454	1.4307338		
Corrected Total	112	201.3798230			

R-Square 0.275325
 Coeff Var 27.46657
 Root MSE 1.196133
 infrisk Mean 4.354867

Source	DF	Type I SS	Mean Square	F Value	Pr > F
nbeds	1	26.06404496	26.06404496	18.22	<.0001
census	1	4.34156448	4.34156448	3.03	0.0845
nurses	1	4.13729357	4.13729357	2.89	0.0921
region	3	11.95336893	3.98445631	2.78	0.0446
medschl	1	0.72613430	0.72613430	0.51	0.4778
region*medschl	3	8.22257135	2.74085712	1.92	0.1317

Source	DF	Type III SS	Mean Square	F Value	Pr > F
nbeds	1	0.30778506	0.30778506	0.22	0.6438
census	1	1.23214806	1.23214806	0.86	0.3556
nurses	1	1.96235291	1.96235291	1.37	0.2443
region	3	4.99986749	1.66662250	1.16	0.3269
medschl	1	0.58857891	0.58857891	0.41	0.5227
region*medschl	3	8.22257135	2.74085712	1.92	0.1317

The GLM Procedure
Least Squares Means

region	infrisk LSMEAN
North Central	4.09619975
Northeast	4.77314443
South	4.45672739
West	4.03829266

medschl	infrisk LSMEAN
No	4.48428350
Yes	4.19789861

region	medschl	infrisk LSMEAN
North Central	No	4.52790273
North Central	Yes	3.66449677
Northeast	No	4.76991489
Northeast	Yes	4.77637397
South	No	3.86163956
South	Yes	5.05181523
West	No	4.77767684
West	Yes	3.29890848

The FREQ Procedure

Table of medschl by region

medschl(Medical school affiliation)
region(Region of country (usa))

Frequency					Total
Row Pct	Northeas t	North Ce ntral	South	West	
Yes	6 35.29	7 41.18	2 11.76	2 11.76	17
No	23 23.96	25 26.04	34 35.42	14 14.58	96
Total	29	32	36	16	113

Statistics for Table of medschl by region

Statistic	DF	Value	Prob
Chi-Square	3	4.5084	0.2115
Likelihood Ratio Chi-Square	3	5.0108	0.1710
Mantel-Haenszel Chi-Square	1	2.3105	0.1285
Phi Coefficient		0.1997	
Contingency Coefficient		0.1959	
Cramer's V		0.1997	

WARNING: 38% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 113

Study of the Effectiveness of Nosocomial Infection Control 5
 Two-way ANCOVA on SENIC Data

The CORR Procedure

3 Variables: nbeds census nurses

Pearson Correlation Coefficients, N = 113
 Prob > |r| under H0: Rho=0

	nbeds	census	nurses
nbeds	1.00000	0.98100	0.91550
Average # beds during study period		<.0001	<.0001
census	0.98100	1.00000	0.90790
Aver # patients in hospital per day	<.0001		<.0001
nurses	0.91550	0.90790	1.00000
Aver # nurses during study period	<.0001	<.0001	