

Factorial ANOVA with R: Rotten Potato Data*

```
> rm(list=ls()); options(scipen=999)
> potato =
read.table("https://www.utstat.toronto.edu/~brunner/data/illegal/potato2.data.txt")
> # Type of bacteria: 1, 2, 3
> # Storage temperature: 1=10 degrees C, 2=16 degrees C
> head(potato)

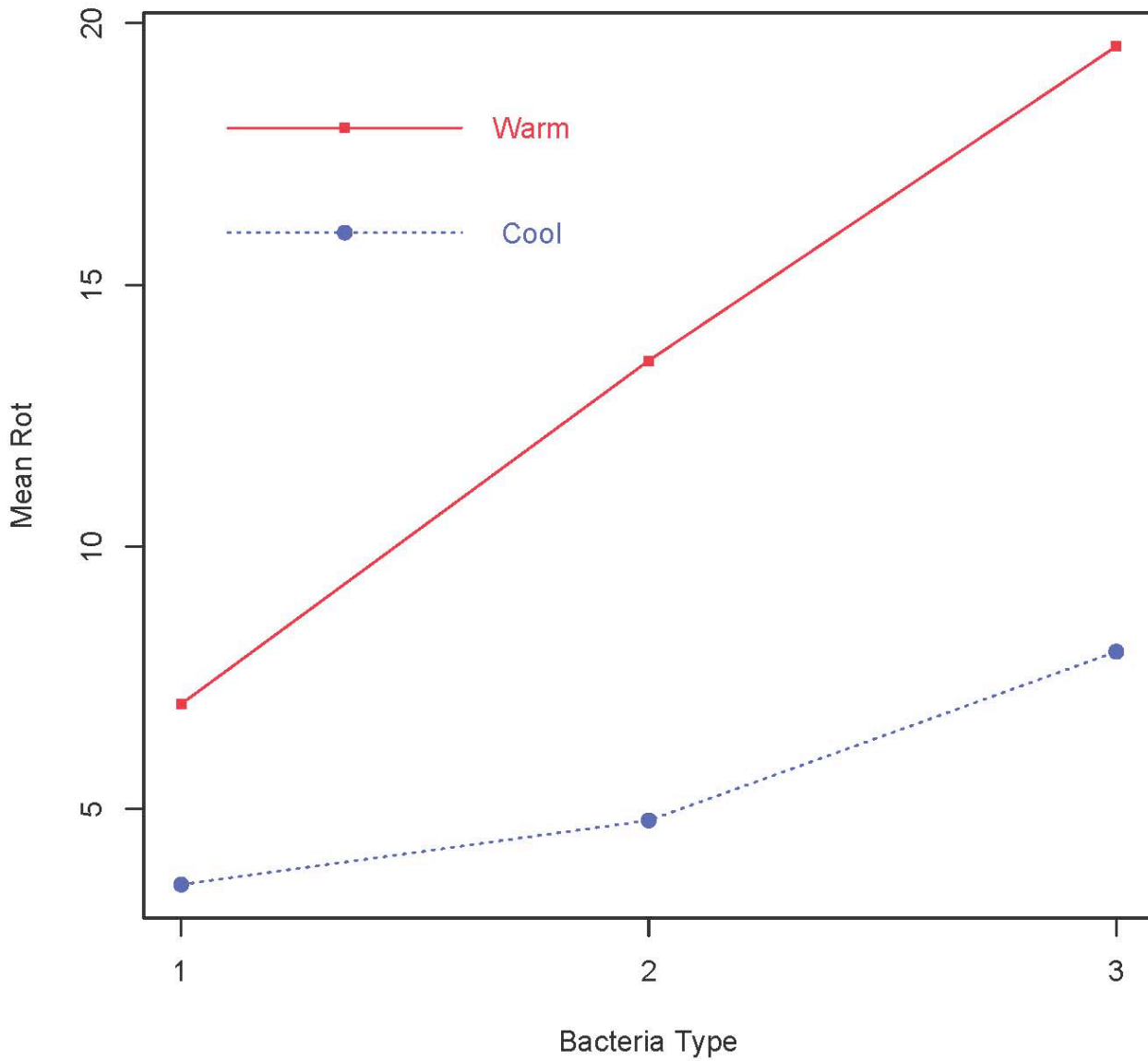
  Bact Temp Rot
1    1    1    7
2    1    1    7
3    1    1    9
4    1    1    0
5    1    1    0
6    1    1    0
>
> spud = within(potato,{
+ Bact = factor(Bact)
+ Temp = factor(Temp, labels = c('10 Degrees','16 Degrees'))
+ }) # End within potato
>
> # Check sample sizes
> with(spud,table(Temp,Bact, useNA = 'ifany'))
      Bact
Temp   1 2 3
10 Degrees 9 9 9
16 Degrees 9 9 9
> # Look at means
> with(spud,{
+ meanz = tapply(Rot,INDEX=list(Temp,Bact),FUN=mean)
+ print(meanz)
+ meanzz = addmargins(meanz,FUN=mean) # Add marginal means
+ print(round(meanzz,2))
+ # Plot the means
+ cool = meanz[1,]; warm = meanz[2,]
+ Bacteria = c(1:3,1:3); MeanRot = c(cool,warm)
+ # Invisible points at first, x axis points at 1,2,3; see help(plot)
+ plot(Bacteria,MeanRot,pch=" ",xaxp=c(1,3,2), xlab="Bacteria Type",ylab="Mean
Rot")
+ title("Mean Rot as a Function of Temperature and Bacteria Type")
+ points(1:3,warm,col='red',pch=15) # Red squares
+ points(1:3,cool,col='blue',pch=19) # Blue circles
+ lines(1:3,warm,lty=1,col='red'); lines(1:3,cool,lty=3,col='blue')
+ # Annotate the plot
+ x1 = c(1.1,1.6); y1 = c(18,18); lines(x1,y1,lty=1,col='red')
+ points(1.35,18,col='red',pch=15)
+ text(1.75,18,'Warm',col='red')
+ x2 = c(1.1,1.6); y2 = c(16,16); lines(x2,y2,lty=3,col='blue')
+ points(1.35,16,col='blue',pch=19)
+ text(1.75,16,'Cool',col='blue')
+ }) # End displaying the means with spud
```

* See last page for copyright information

	1	2	3
10 Degrees	3.555556	4.777778	8.000000
16 Degrees	7.000000	13.555556	19.555556

	1	2	3	mean
10 Degrees	3.56	4.78	8.00	5.44
16 Degrees	7.00	13.56	19.56	13.37
mean	5.28	9.17	13.78	9.41

Mean Rot as a Function of Temperature and Bacteria Type



```

>
> # Two-factor ANOVA, several different ways
> # First with Dummy variable regression, cell means coding and contrasts.
> spud = within(spud,{
+ # Make 6 indicator dummy variables.
+ m11 = m12 = m13      =
+ m21 = m22 = m23      = numeric(length(Rot))
+ TB = paste(Temp,Bact,sep="") # All combinations
+ m11[TB=='10Degrees1'] = 1; m12[TB=='10Degrees2'] = 1; m13[TB=='10Degrees3'] = 1
+ m21[TB=='16Degrees1'] = 1; m22[TB=='16Degrees2'] = 1; m23[TB=='16Degrees3'] = 1
+ })
> # data.frame(Temp,Bact,TB,p11,p12,p13,p21,p22,p23) # To check
>
> cellmeans_model = lm(Rot ~ 0 + m11+m12+m13 + m21+m22+m23, data=spud)
> summary(cellmeans_model)

```

Call:

```
lm(formula = Rot ~ 0 + m11 + m12 + m13 + m21 + m22 + m23, data = spud)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-11.5556	-3.5556	0.2222	3.4444	9.4444

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
m11	3.556	1.562	2.276	0.02734 *
m12	4.778	1.562	3.058	0.00363 **
m13	8.000	1.562	5.121	0.0000053299166 ***
m21	7.000	1.562	4.481	0.0000460076838 ***
m22	13.556	1.562	8.677	0.00000000000213 ***
m23	19.556	1.562	12.518	< 0.0000000000000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.686 on 48 degrees of freedom

Multiple R-squared: 0.8592, Adjusted R-squared: 0.8416

F-statistic: 48.81 on 6 and 48 DF, p-value: < 0.00000000000000022

```

>
> # Define fttest = function(model,L,h=0)
> source("http://www.utstat.utoronto.ca/~brunner/Rfunctions/fttest.txt")
> # For any factorial design, THE BASICS are an overall test for equality of
> # treatment means, and also tests for all the main effects and interactions.
> # For the potato data, first the overall test for equality of the 6
> # treatment means
> L1 = rbind(c(1,-1, 0, 0, 0, 0),
+           c(0, 1,-1, 0, 0, 0),
+           c(0, 0, 1,-1, 0, 0),
+           c(0, 0, 0, 1,-1, 0),
+           c(0, 0, 0, 0, 1,-1))
> round(fttest(cellmeans_model,L1),5)
      F      df1      df2 p-value
15.05093 5.00000 48.00000 0.00000
> # Test for differences between marginal means (Main effects)
> # Better LOOK at the marginal means again
> with(spud,{
+ meanz = tapply(Rot,INDEX=list(Temp,Bact),FUN=mean)
+ meanzz = addmargins(meanz,FUN=mean, quiet=TRUE) # Add marginal means
+ print(round(meanzz,2))
+ }) # End with
      1      2      3 mean
10Degrees 3.56  4.78  8.00  5.44
16Degrees 7.00 13.56 19.56 13.37
mean      5.28  9.17 13.78  9.41

```

```

> # Averaging across Bacteria Types, does Storage Temperature affect
> # the amount of rot?
> L2 = rbind(c(1,1,1,-1,-1,-1))
> round(ftest(cellmeans_model,L2),5)
      F      df1      df2  p-value
38.61383  1.00000 48.00000  0.00000
> # Averaging across Storage Temperatures, does Bacteria Type affect
> # the amount of rot?
> L3 = rbind(c(1,-1, 0, 1,-1, 0),
+           c(0, 1,-1, 0, 1,-1))
> round(ftest(cellmeans_model,L3),5)
      F      df1      df2  p-value
14.83895  2.00000 48.00000  0.00001
> # Now the Interaction: Does the effect of Temperature depend on type of bacteria?
> # H0: mu21-mu11 = mu22-mu12 = mu23-mu13
> # <=> (-1)*mu11 + (1)*mu12 + (0)*mu13 + (1)*mu21 + (-1)*mu22 + (0)*mu23 = 0 and
> #      (0)*mu11 + (-1)*mu12 + (1)*mu13 + (0)*mu21 + (1)*mu22 + (-1)*mu23 = 0
> L4 = rbind(c(-1, 1, 0, 1,-1, 0),
+           c( 0,-1, 1, 0, 1,-1) )
> round(ftest(cellmeans_model,L4),5)
      F      df1      df2  p-value
 3.48145  2.00000 48.00000  0.03874
> # There are easier ways
>
> cm2 = lm(Rot ~ 0 + TB, data=spud) # TB is the combination variable
> summary(cm2)

```

Call:

```
lm(formula = Rot ~ 0 + TB, data = spud)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.5556	-3.5556	0.2222	3.4444	9.4444

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
TB10Degrees1	3.556	1.562	2.276	0.02734 *
TB10Degrees2	4.778	1.562	3.058	0.00363 **
TB10Degrees3	8.000	1.562	5.121	0.0000053299166 ***
TB16Degrees1	7.000	1.562	4.481	0.0000460076838 ***
TB16Degrees2	13.556	1.562	8.677	0.00000000000213 ***
TB16Degrees3	19.556	1.562	12.518	< 0.0000000000000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.686 on 48 degrees of freedom

Multiple R-squared: 0.8592, Adjusted R-squared: 0.8416

F-statistic: 48.81 on 6 and 48 DF, p-value: < 0.00000000000000022

```
> summary(cellmeans_model) # For comparison
```

Call:

```
lm(formula = Rot ~ 0 + m11 + m12 + m13 + m21 + m22 + m23, data = spud)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.5556	-3.5556	0.2222	3.4444	9.4444

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
m11	3.556	1.562	2.276	0.02734 *
m12	4.778	1.562	3.058	0.00363 **
m13	8.000	1.562	5.121	0.0000053299166 ***
m21	7.000	1.562	4.481	0.0000460076838 ***
m22	13.556	1.562	8.677	0.00000000000213 ***

```

m23    19.556      1.562  12.518 < 0.00000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.686 on 48 degrees of freedom
Multiple R-squared:  0.8592, Adjusted R-squared:  0.8416
F-statistic: 48.81 on 6 and 48 DF,  p-value: < 0.000000000000000022

```

```

> # With cm2, you don't have to make the dummy variables.
>
> ez1 = lm(Rot ~ Temp + Bact + Temp:Bact, data=spud)
> # Syntax Rot ~ Temp*Bact would work too.
> summary(ez1)

```

```

Call:
lm(formula = Rot ~ Temp + Bact + Temp:Bact, data = spud)

```

```

Residuals:
      Min       1Q   Median       3Q      Max
-11.5556  -3.5556   0.2222   3.4444   9.4444

```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.556      1.562   2.276  0.0273 *
Temp16Degrees    3.444      2.209   1.559  0.1255
Bact2             1.222      2.209   0.553  0.5827
Bact3             4.444      2.209   2.012  0.0499 *
Temp16Degrees:Bact2  5.333      3.124   1.707  0.0943 .
Temp16Degrees:Bact3  8.111      3.124   2.596  0.0125 *

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 4.686 on 48 degrees of freedom
Multiple R-squared:  0.6106, Adjusted R-squared:  0.57
F-statistic: 15.05 on 5 and 48 DF,  p-value: 0.000000007003

```

```

>
> # There's a problem. Compare F = 38.61383, p = 0.000 for main effect of
temperature

```

```

> # ECspud is spud with effect coding
> ECspud = within(spud,{
+ contrasts(Temp) = contr.sum
+ contrasts(Bact) = contr.sum
+ }) # End making ECspud
> with(ECspud,{
+ print(contrasts(Temp)); cat('\n')
+ print(contrasts(Bact)); cat('\n')
+ }) # End with

```

```

      [,1]
10Degrees    1
16Degrees   -1

```

```

      [,1] [,2]
1         1    0
2         0    1
3        -1   -1

```

```

>
> ez2 = lm(Rot ~ Temp + Bact + Temp:Bact, data=ECspud)
> summary(ez2)

Call:
lm(formula = Rot ~ Temp + Bact + Temp:Bact, data = ECspud)

Residuals:
    Min       1Q   Median       3Q      Max
-11.5556  -3.5556   0.2222   3.4444   9.4444

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)   9.4074     0.6377  14.751 < 0.0000000000000002 ***
Temp1        -3.9630     0.6377  -6.214   0.0000000118 ***
Bact1        -4.1296     0.9019  -4.579   0.000033285 ***
Bact2        -0.2407     0.9019  -0.267    0.7907
Temp1:Bact1   2.2407     0.9019   2.484   0.0165 *
Temp1:Bact2  -0.4259     0.9019  -0.472   0.6389
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.686 on 48 degrees of freedom
Multiple R-squared:  0.6106, Adjusted R-squared:  0.57
F-statistic: 15.05 on 5 and 48 DF, p-value: 0.000000007003

>
> 6.244^2 # Compare F = 38.61383 for main effect of temperature.
[1] 38.98754
> # Difference is just rounding error.
>
> # Main effect for Bacteria type. Compare F = 14.83895
> coefficients(ez2)
(Intercept)      Temp1      Bact1      Bact2 Temp1:Bact1 Temp1:Bact2
 9.4074074 -3.9629630 -4.1296296 -0.2407407  2.2407407 -0.4259259
> L5 = rbind(c(0,0,1,0,0,0),
+           c(0,0,0,1,0,0))
> round(ftest(ez2,L5),5)
      F      df1      df2 p-value
14.83895  2.00000 48.00000 0.00001
>
> # Interaction of temperature by bacteria type. Compare F = 3.48145, p = 0.03874
> L6 = rbind(c(0,0,0,0,1,0),
+           c(0,0,0,0,0,1))
> round(ftest(ez2,L6),5)
      F      df1      df2 p-value
 3.48145  2.00000 48.00000 0.03874
>
> # Full-restricted test for interaction: Compare F = 3.48145, p = 0.03874
> ez2nointer = lm(Rot ~ Temp + Bact, data=ECspud)
> anova(ez2nointer,ez2)
Analysis of Variance Table

Model 1: Rot ~ Temp + Bact
Model 2: Rot ~ Temp + Bact + Temp:Bact
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     50 1207.2
2     48 1054.2  2    152.93 3.4815 0.03874 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

>
> # To test non-standard hypotheses, it is easiest to use the cell means model
> # Just a few examples out of MANY possible
>
> # Is there an effect of bacteria type at cool temperatures?
> # H0: mu11=mu12=mu13
> L7 = rbind(c(1,-1, 0, 0, 0, 0),
+           c(0, 1,-1, 0, 0, 0) )
> round(ftest(cellmeans_model,L7),5)
      F      df1      df2  p-value
2.16020  2.00000 48.00000  0.12638
>
> # Is there an effect of bacteria type at warm temperatures?
> # H0: mu21=mu22=mu23
> L8 = rbind(c(0, 0, 0, 1,-1, 0),
+           c(0, 0, 0, 0, 1,-1) )
> round(ftest(cellmeans_model,L8),5)
      F      df1      df2  p-value
16.1602  2.0000 48.0000  0.0000
>
> # At warm temperatures, which bacteria types are different
> # from which others? Three pairwise comparisons:
>
> L9 = rbind(c(0, 0, 0, 1,-1, 0)) # One versus two
> round(ftest(cellmeans_model,L9),5)
      F      df1      df2  p-value
8.80523  1.00000 48.00000  0.00467
>
> L10 = rbind(c(0, 0, 0, 1, 0,-1)) # One versus three
> round(ftest(cellmeans_model,L10),5)
      F      df1      df2  p-value
32.29933  1.00000 48.00000  0.00000
>
> L11 = rbind(c(0, 0, 0, 0, 1,-1)) # Two versus three
> round(ftest(cellmeans_model,L11),5)
      F      df1      df2  p-value
7.37605  1.00000 48.00000  0.00916
>
>
> # All three are significant at alpha = 0.05. For DIRECTIONAL conclusions, look at
the picture.

```

This document was prepared by [Jerry Brunner](#), University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License:

http://creativecommons.org/licenses/by-sa/3.0/deed.en_US. Use any part of it as you like and share the result freely. It is available in OpenOffice.org format from the course website: <http://www.utstat.toronto.edu/brunner/oldclass/312f22>