

Ecol 145 Assignment 7
Dahl Winters
3/24/06

Question 1

Fit a single regression model (or at most two models) and test these data for a treatment effect. Use whatever methodology you think appropriate.

Inputting the data:

```
num.borers1<-c(0:26)
trt1<-c(19,12,18,18,11,12,7,8,4,4,1,0,1,1,0,1,0,1,0,1,0,0,0,0,0,0,1)
counts1<-rep(num.borers1,trt1)
```

```
num.borers2<-c(0:12)
trt2<-c(24,16,16,18,15,9,6,5,3,4,3,0,1)
counts2<-rep(num.borers2,trt2)
```

```
num.borers3<-c(0:8)
trt3<-c(43,35,17,11,5,4,1,2,2)
counts3<-rep(num.borers3,trt3)
```

```
num.borers4<-c(0:11)
trt4<-c(47,23,27,9,7,3,1,1,0,0,1,1)
counts4<-rep(num.borers4,trt4)
```

```
counts<-c(counts1,counts2,counts3,counts4)
length(counts)
[1] 480
```

```
trts<-rep(1:4,rep(length(counts1),4))
table(trts) #to check if I used the rep function correctly
trts
  1  2  3  4
120 120 120 120
```

Converting the numeric values of the treatments to factors

```
trt.f<-factor(trts)
class(trt.f)
[1] "factor"
contrasts(trt.f)
  2 3 4
1 0 0 0
2 1 0 0
3 0 1 0
4 0 0 1
```

Model 1: Fitting a model without treatments as predictors

```
library(MASS)
modell1<-glm.nb(counts~1)
summary(modell1)
Call:
glm.nb(formula = counts ~ 1, init.theta = 1.08235181293464, link = log)
```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6185 -1.6185 -0.2003  0.4337  3.4552
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.93528    0.05237   17.86  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial(1.0824) family taken to be 1)
    Null deviance: 533.43  on 479  degrees of freedom
Residual deviance: 533.43  on 479  degrees of freedom
AIC: 2029.0
Number of Fisher Scoring iterations: 1
            Theta:  1.082
            Std. Err.:  0.112
 2 x log-likelihood: -2024.952

```

Model 2: Fitting a model with treatments as predictors

```

model12<-glm.nb(counts~trt.f)
summary(model12)
Call:
glm.nb(formula = counts ~ trt.f, init.theta = 1.47144974025270, link = log)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9705 -1.4324 -0.3071  0.2768  2.9291
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.39459    0.08792   15.863  < 2e-16 ***
trt.f2       -0.24191    0.12659   -1.911    0.056  .
trt.f3       -1.00030    0.13788   -7.255 4.02e-13 ***
trt.f4       -0.98359    0.13754   -7.151 8.60e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial(1.4714) family taken to be 1)
    Null deviance: 623.56  on 479  degrees of freedom
Residual deviance: 538.74  on 476  degrees of freedom
AIC: 1958.5
Number of Fisher Scoring iterations: 1
Correlation of Coefficients:
      (Intercept) trt.f2 trt.f3
trt.f2 -0.69
trt.f3 -0.64      0.44
trt.f4 -0.64      0.44  0.41
            Theta:  1.471
            Std. Err.:  0.174
 2 x log-likelihood: -1948.453

```

Is there a treatment effect?

The AIC is considerably lower for model 2 (with predictors) than for the model without predictors, model 1 (1958.5 compared to 2029.0). This suggests that model 2 is the better one. Also, when doing an ANOVA comparing model 1 with model 2 (below), the p-value was very small, which also suggests that including predictors gives a model with a greater goodness of fit to the data.

```

anova(model11,model12)
Likelihood ratio tests of Negative Binomial Models

```

```

Response: counts
      Model      theta Resid. df    2 x log-lik.   Test      df LR stat.      Pr(Chi)
1          1 1.082352     479      -2024.952
2 trt.f + I(trt.dummy) 1.471450     476      -1948.453 1 vs 2      3 76.49927 2.220446e-16

```

Question 2

I would now like you to fit a single regression model that not only tests for an overall treatment effect, but also tests the following three hypotheses simultaneously. What conclusions do you draw for each of these tests?

Coming up with a new coding scheme to test the three hypotheses simultaneously

```

trt.fb<-factor(trts)
contrasts(trt.fb)<-cbind(c(0,1,1,-2),c(0,1,-1,0),c(3,-1,-1,-1))
contrasts(trt.fb)
  [,1] [,2] [,3]
1     0     0     3
2     1     1    -1
3     1    -1    -1
4    -2     0    -1

```

Trt1: $\log \mu_1 = \beta_0 + 3\beta_3$
 Trt2: $\log \mu_2 = \beta_0 + \beta_1 + \beta_2 - \beta_3$
 Trt3: $\log \mu_3 = \beta_0 + \beta_1 - \beta_2 - \beta_3$
 Trt4: $\log \mu_4 = \beta_0 - 2\beta_1 - \beta_3$

Fitting another negative binomial glm using the new coding scheme

```

model3<-glm.nb(counts~trt.fb)
summary(model3)
Call:
glm.nb(formula = counts ~ trt.fb, init.theta = 1.47144974025270,
       link = log)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9705  -1.4324  -0.3071   0.2768   2.9291
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.83814    0.04905  17.088 < 2e-16 ***
trt.fb1      0.12083    0.04227   2.858  0.00426 **
trt.fb2      0.37919    0.06996   5.420 5.95e-08 ***
trt.fb3      0.18548    0.02640   7.027 2.11e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial(1.4714) family taken to be 1)
Null deviance: 623.56  on 479  degrees of freedom
Residual deviance: 538.74  on 476  degrees of freedom
AIC: 1958.5
Number of Fisher Scoring iterations: 1
Correlation of Coefficients:
      (Intercept) trt.fb1 trt.fb2
trt.fb1 -0.06
trt.fb2 -0.11      -0.08
trt.fb3 -0.12      0.03   0.07
      Theta: 1.471

```

Std. Err.: 0.174
2 x log-likelihood: -1948.453

Conclusions drawn for each of the 3 tests

1. Treatment 4 involves a double application of fungal spores. Test whether a double application yields a different result than a single application (ignoring the timing of the single application). This should be a contrast of treatment 4 with treatments 2 and 3 combined.

This is a test of whether μ_4 is different from the average of μ_2 and μ_3 . When the appropriate substitutions are made, this reduces to a test of whether $\beta_1 = 0$. From the summary output, we see that the p-value for this test is 0.00426, which is under 0.05. Thus we can say that this test doesn't give us a reason to reject the hypothesis that a double application of spores yields a different result from a single application.

2. Treatments 2 and 3 are both single applications of fungal spores, but their timing is different. Test whether the timing of the application makes a difference. This should be a contrast of treatment 2 with treatment 3.

This is a test of whether μ_2 is different from μ_3 . When the appropriate substitutions are made, this reduces to a test of whether $\beta_2 = 0$. The p-value for this test was 5.95e-08, which is very small. Again, we can say there isn't a cause for rejecting the hypothesis that the timing of the application makes a difference.

3. Test whether any application of fungal spores at all yields a result different from no application. This should be a contrast of treatment 1 with treatments 2, 3, and 4 combined.

This is a test of whether μ_1 is different from the average of μ_2 , μ_3 , and μ_4 . When the appropriate substitutions are made, this reduces to a test of whether $\beta_3 = 0$. The p-value for this test was 2.11e-12, the smallest yet (though I know I'm not supposed to read much into the magnitude of these numbers). Because the p-value is much smaller than 0.05, this test does not provide a reason to reject the hypothesis that applying spores yields a different result than not applying any.

Question 3

Are the contrasts specified in Question 2 orthogonal contrasts? Why or why not?

```
contrasts(trt.fb)
  [,1] [,2] [,3]
1     0     0     3
2     1     1    -1
3     1    -1    -1
4    -2     0    -1
```

They are orthogonal contrasts because the sum of all columns = 0, the sum of all pairwise multiplications $(0+0+1-1-1+1+0+0) = 0$, and the number of observations in each treatment group are the same (treatments 1-4 each have 120 observations), which are the three required conditions for contrasts to be orthogonal.

Question 4

Create an error bar plot in which you display 95% Wald confidence intervals for the mean much like you did in Question 3 of Assignment 4. This time though use only the results from the regression model you fit in **Question 1** of this assignment.

Getting the formulas for the treatment means

The form of the model is $\log \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$.

```
contrasts(trt.f)
  2 3 4
1 0 0 0
2 1 0 0
3 0 1 0
4 0 0 1
```

Trt1: $\log \mu_1 = \beta_0$
 Trt2: $\log \mu_2 = \beta_0 + \beta_1$
 Trt3: $\log \mu_3 = \beta_0 + \beta_2$
 Trt4: $\log \mu_4 = \beta_0 + \beta_3$

```
coef(model2)
(Intercept)      trt.f2      trt.f3      trt.f4
  1.3945932   -0.2419137   -1.0003014   -0.9835879
```

```
#the following are actually log means
mean1<-coef(model2)[1]
mean2<-coef(model2)[1]+coef(model2)[2]
mean3<-coef(model2)[1]+coef(model2)[3]
mean4<-coef(model2)[1]+coef(model2)[4]
```

To get the standard errors of the individual means

```
out<-summary(model2)
out$cov.unscaled #gives the covariance matrix
(Intercept)      trt.f2      trt.f3      trt.f4
(Intercept)  0.007729465 -0.007729465 -0.007729465 -0.007729465
trt.f2      -0.007729465  0.016024393  0.007729465  0.007729465
trt.f3      -0.007729465  0.007729465  0.019010791  0.007729465
trt.f4      -0.007729465  0.007729465  0.007729465  0.018917675
```

```
err1<-sqrt( c(1,0,0,0) %*% out$cov.unscaled %*% c(1,0,0,0) )
err1
      [,1]
[1,] 0.08791737
err2<-sqrt( c(0,1,0,0) %*% out$cov.unscaled %*% c(0,1,0,0) )
err2
      [,1]
[1,] 0.1265875
err3<-sqrt( c(0,0,1,0) %*% out$cov.unscaled %*% c(0,0,1,0) )
err3
      [,1]
[1,] 0.1378796
err4<-sqrt( c(0,0,0,1) %*% out$cov.unscaled %*% c(0,0,0,1) )
err4
      [,1]
[1,] 0.1375415
```

Back-transforming the log means as a check to see if we will get the same means as in Assignment 4

```
mean.count<-c(exp(mean1), exp(mean2), exp(mean3), exp(mean4))
mean.count
(Intercept) (Intercept) (Intercept) (Intercept)
```

```
4.033333 3.166667 1.483333 1.508333
#yes, these are the same means, which is good.
```

Set up plot area

```
plot(1:4, mean.count, axes=FALSE, xlab="Treatment", ylab="Mean Count",
xlim=c(0.5,4.5), ylim=c(1,5), cex=1.5)
axis(1,cex.axis=.9)
axis(2,cex.axis=.9)
box()
mtext('95% Wald confidence interval for negative binomial means', side=3,
line=.5)
```

Getting lower and upper bounds for all 4 treatments – calculated on the scale of the log link

Lower bounds

```
mean1-qnorm(.975)*err1
      [,1]
[1,] 1.222278
mean2-qnorm(.975)*err2
      [,1]
[1,] 0.9045726
mean3-qnorm(.975)*err3
      [,1]
[1,] 0.1240527
mean4-qnorm(.975)*err4
      [,1]
[1,] 0.1414288
```

Upper bounds

```
mean1+qnorm(.975)*err1
      [,1]
[1,] 1.566908
mean2+qnorm(.975)*err2
      [,1]
[1,] 1.400786
mean3+qnorm(.975)*err3
      [,1]
[1,] 0.6645309
mean4+qnorm(.975)*err4
      [,1]
[1,] 0.6805818
```

Drawing the arrows

```
#draws arrows from y0=meanval to y1=meanval+/-stderrval, x0=x1=1,2,3,or 4.
```

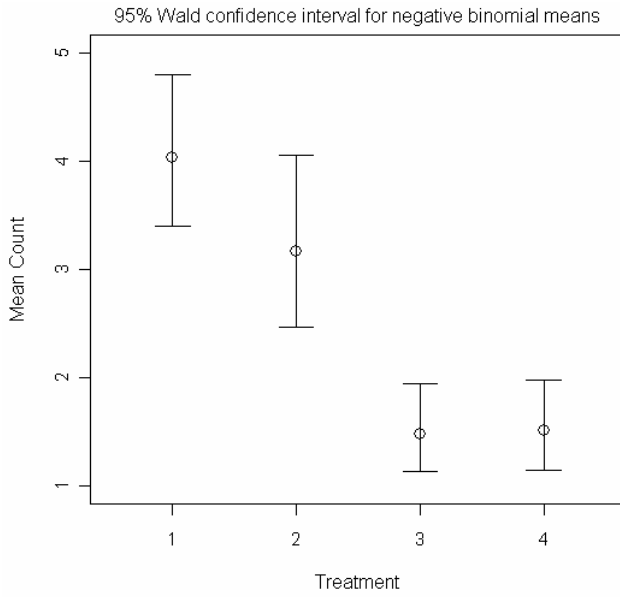
```
arrow.draw<-function(xpos,meanval,stderrval) arrows(xpos,
exp(meanval+stderrval), xpos, exp(meanval-stderrval), length=0.15, angle=90,
code=3)
```

```
arrow.draw(1,mean1,qnorm(.975)*err1)
arrow.draw(2,mean2,qnorm(.975)*err2)
arrow.draw(3,mean3,qnorm(.975)*err3)
arrow.draw(4,mean4,qnorm(.975)*err4)
```

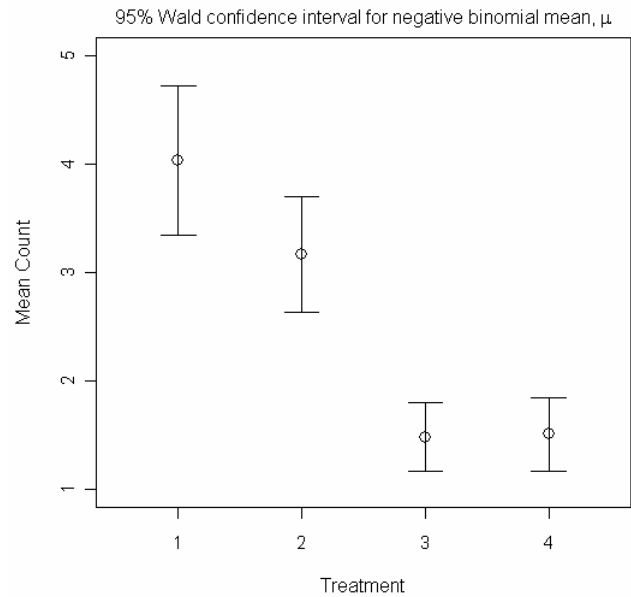
Question 5

How does the picture you drew in **Question 4** compare to the comparable picture you constructed in Assignment 4?

New error bar plot, displaying 95% Wald confidence intervals for the mean



Old error bar plot from Assignment 4



The means of both plots are the same, and the error bars are about the same length as before. These differences in error bar length could be due to differences in the functions used to generate values for the standard errors. In Assignment 4 we used the `nlm` function and the `out$hessian` from the model output to compute the standard errors. This time, we used a negative binomial GLM and the covariance matrix from its summary output to compute the standard errors.