

ST516 Homework 3

Sam Wozniak

March 23, 2018

```
knitr::opts_chunk$set(echo = TRUE, error = TRUE)
library(Sleuth3)
library(ggplot2)
library(trtools)
library(COUNT)
library(npregfast)
library(DAAG)
library(tidyr)
options(digits = 4)
```

Aflatoxicol and Liver Tumors in Trout

1. Estimate a logistic regression model using dose as the explanatory variable. Report parameter estimates, standard errors, and CI's. Write a brief summary that summarizes the effect of aflatoxicol using an odds ratio. Plot the model. Plot studentized residuals against predicted values. Comment on if you think the model is a good fit for the data

Summary: For a 0.1 ppm increase in Dose, the odds of a trout getting liver tumors increases by a factor of 4.19.

Comment: The model is not a good fit for the data. A majority of the studentized residuals are more extreme than 2 or -2.

```
#logistic regression model
m <- glm(cbind(Tumor, Total - Tumor) ~ Dose, family = binomial, data = ex2116)
cbind(summary(m)$coefficients, confint(m))
```

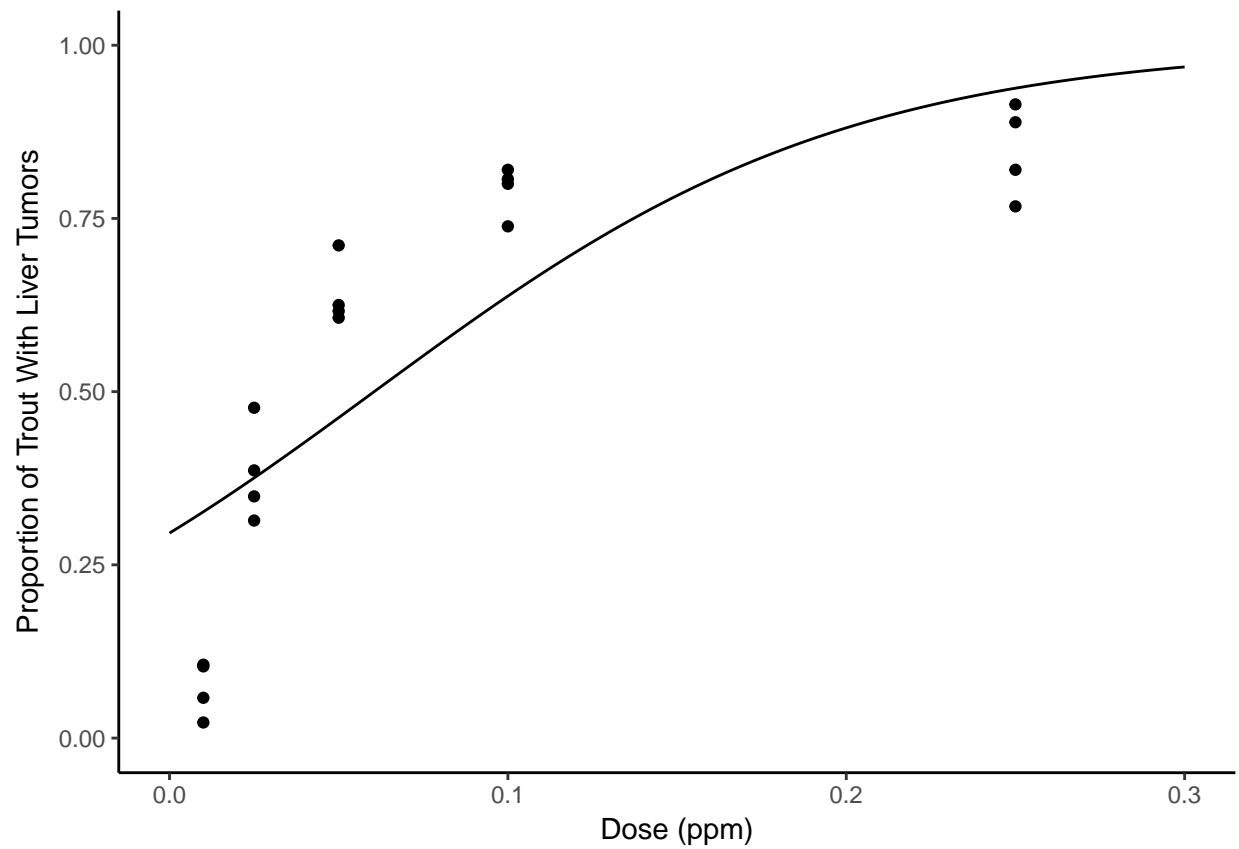
```
##           Estimate Std. Error z value Pr(>|z|)  2.5 %  97.5 %
## (Intercept)  -0.867    0.07673   -11.3 1.321e-29 -1.019 -0.7179
## Dose         14.334    0.93695    15.3 7.838e-53 12.558 16.2346
```

```
#odds ratio
contrast(m, a = list(Dose = .2), b = list(Dose = .1), tf = exp)
```

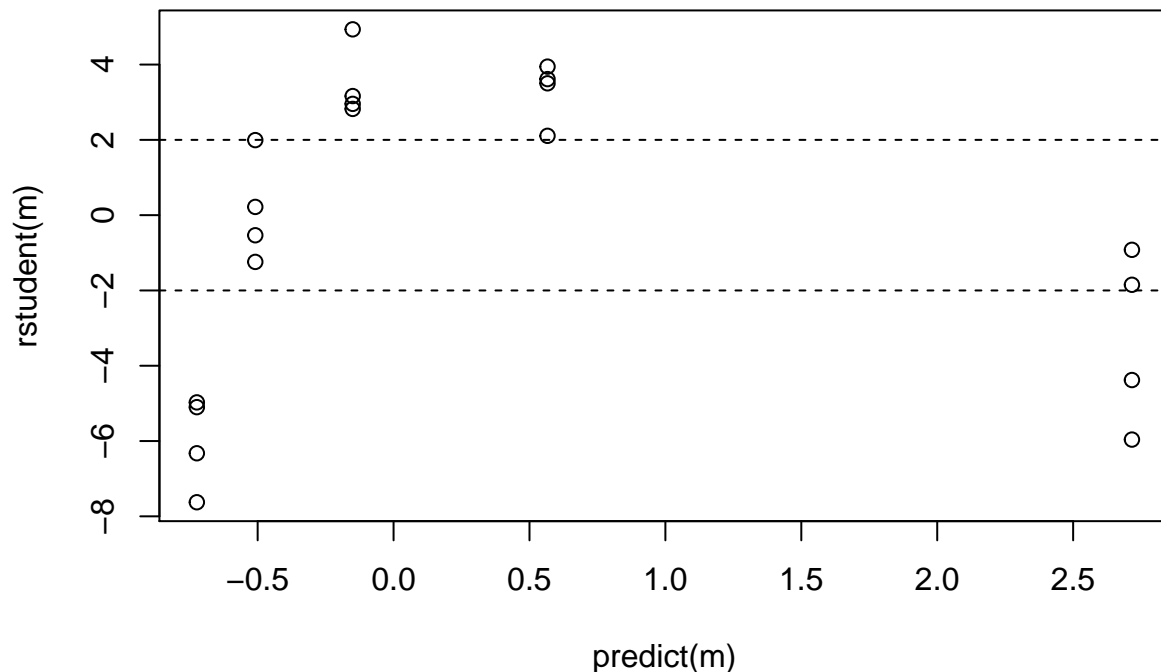
```
## estimate      se lower upper tvalue df      pvalue
##      4.193 0.09369 3.489 5.038   15.3 Inf 7.838e-53
```

```
#plot model
d <- expand.grid(Dose = seq(0, .3, length = 100))
d$yhat <- predict(m, newdata = d, type = "response")

p <- ggplot(ex2116, aes(x = Dose, y = Tumor/Total))
p <- p + geom_point() + theme_classic() + ylim(0,1)
p <- p + geom_line(aes(y = yhat), data = d)
p <- p + labs(x = "Dose (ppm)", y = "Proportion of Trout With Liver Tumors")
plot(p)
```



```
#plot studentized residuals vs predicted  
plot(predict(m), rstudent(m))  
abline(-2, 0, lty = 2)  
abline(2, 0, lty = 2)
```



2. Repeat the previous problem except using the base-2 log of dose as the explanatory variable. Be sure your interpretation of the odds ratio accounts for the log transformation of dose.

Summary: For a doubling of dose, the odds of a trout getting liver tumors increases by a factor of 2.46.

Comment: This model is a better fit than the previous model, but still has a fair number of studentized residuals fall outside of (-2,2).

```
#logistic regression model with log2 transformation of dose
m <- glm(cbind(Tumor, Total - Tumor) ~ log2(Dose), family = binomial, data = ex2116)
cbind(summary(m)$coefficients, confint(m))
```

```
##           Estimate Std. Error z value Pr(>|z|)  2.5 % 97.5 %
## (Intercept)   4.1634    0.2085   19.97 9.564e-89 3.7631  4.581
## log2(Dose)    0.8997    0.0446   20.17 1.628e-90 0.8141  0.989
```

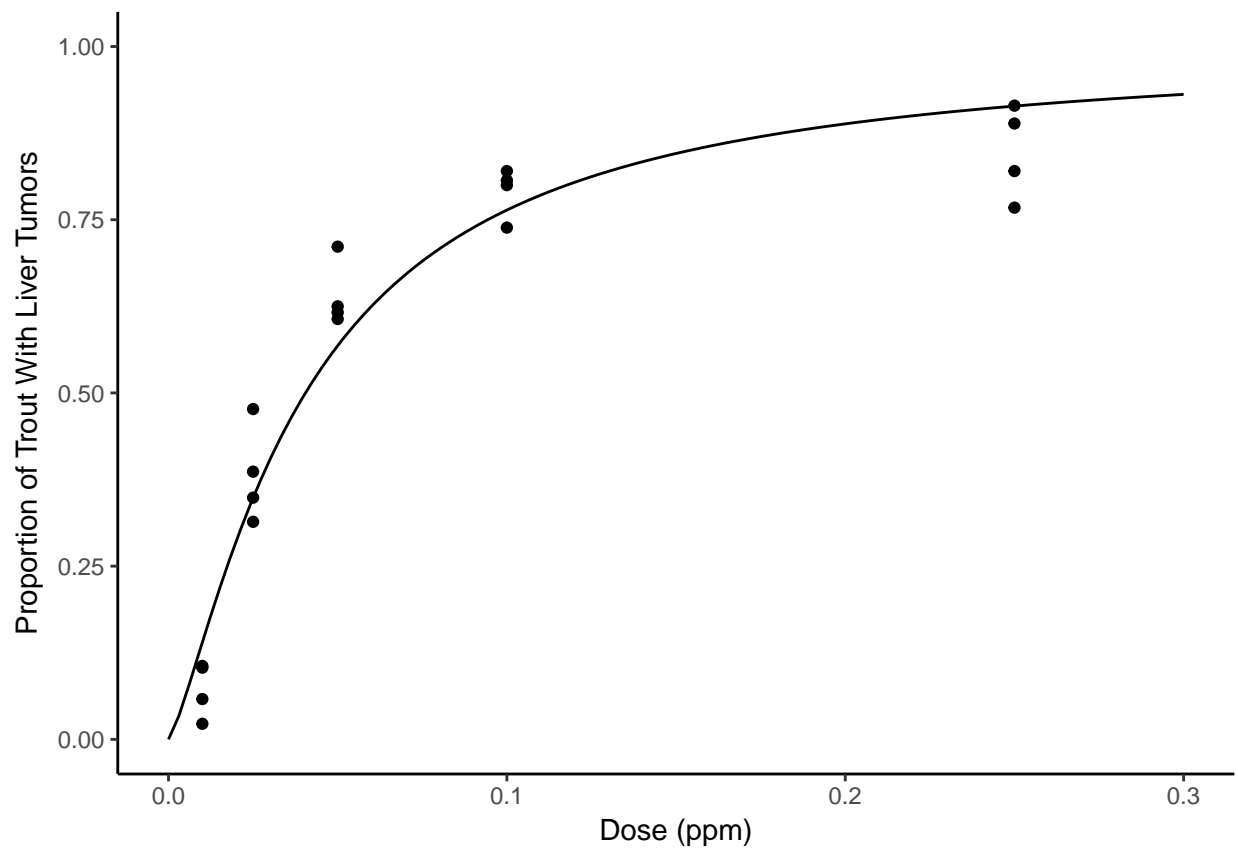
```
#odds ratio
contrast(m, a = list(Dose = .2), b = list(Dose = .1), tf = exp)
```

```
## estimate      se lower upper tvalue df    pvalue
##      2.459 0.0446 2.253 2.684  20.17 Inf 1.628e-90
```

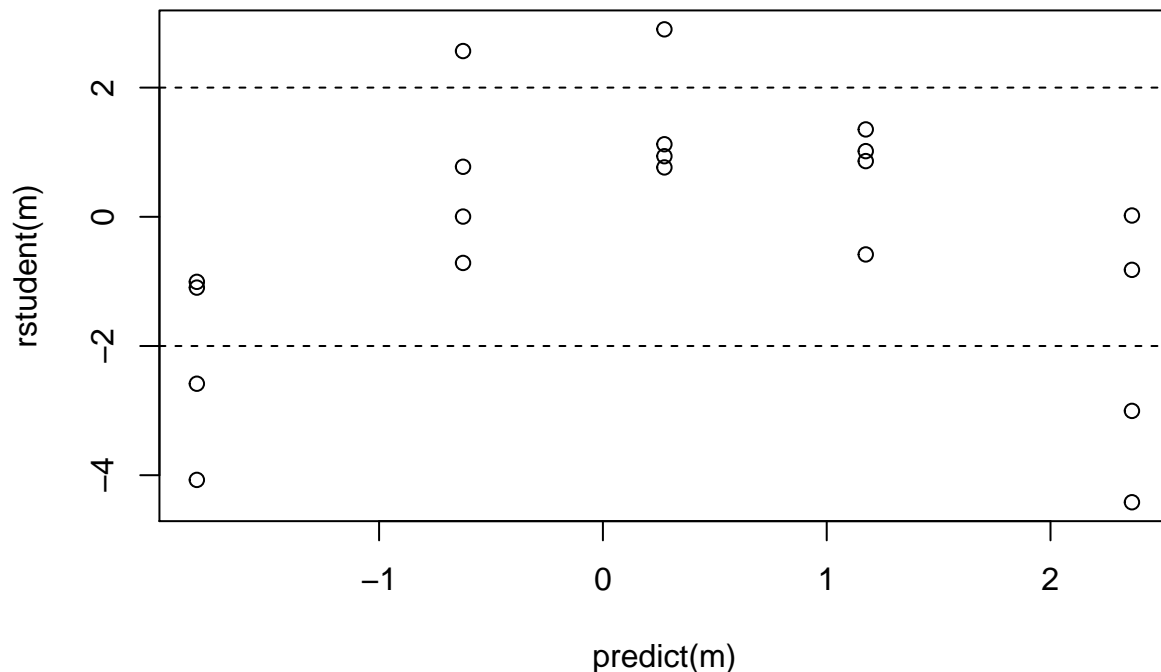
```
#plot model
d <- expand.grid(Dose = seq(0, .3, length = 100))
d$yhat <- predict(m, newdata = d, type = "response")
```

```
p <- ggplot(ex2116, aes(x = Dose, y = Tumor/Total))
p <- p + geom_point() + theme_classic() + ylim(0,1)
```

```
p <- p + geom_line(aes(y = yhat), data = d)
p <- p + labs(x = "Dose (ppm)", y = "Proportion of Trout With Liver Tumors")
plot(p)
```



```
#plot studentized residuals vs predicted
plot(predict(m), rstudent(m))
abline(-2, 0, lty = 2)
abline(2, 0, lty = 2)
```



3. Repeat the previous two problems, but specify dose as a factor.

Summary:

For an increase in Dose from 0.01 to 0.025 ppm, the odds of a trout getting liver tumors increases by a factor of 7.95.

For an increase in Dose from 0.01 to 0.05 ppm, the odds of a trout getting liver tumors increases by a factor of 22.29.

For an increase in Dose from 0.01 to 0.1 ppm, the odds of a trout getting liver tumors increases by a factor of 48.91.

For an increase in Dose from 0.01 to 0.25 ppm, the odds of a trout getting liver tumors increases by a factor of 70.84.

Comment: The studentized residuals for this model look much better than either of the previous two models—there are few studentized residuals that fall outside $(-2, 2)$, and those that do, are not very far from -2 or 2 .

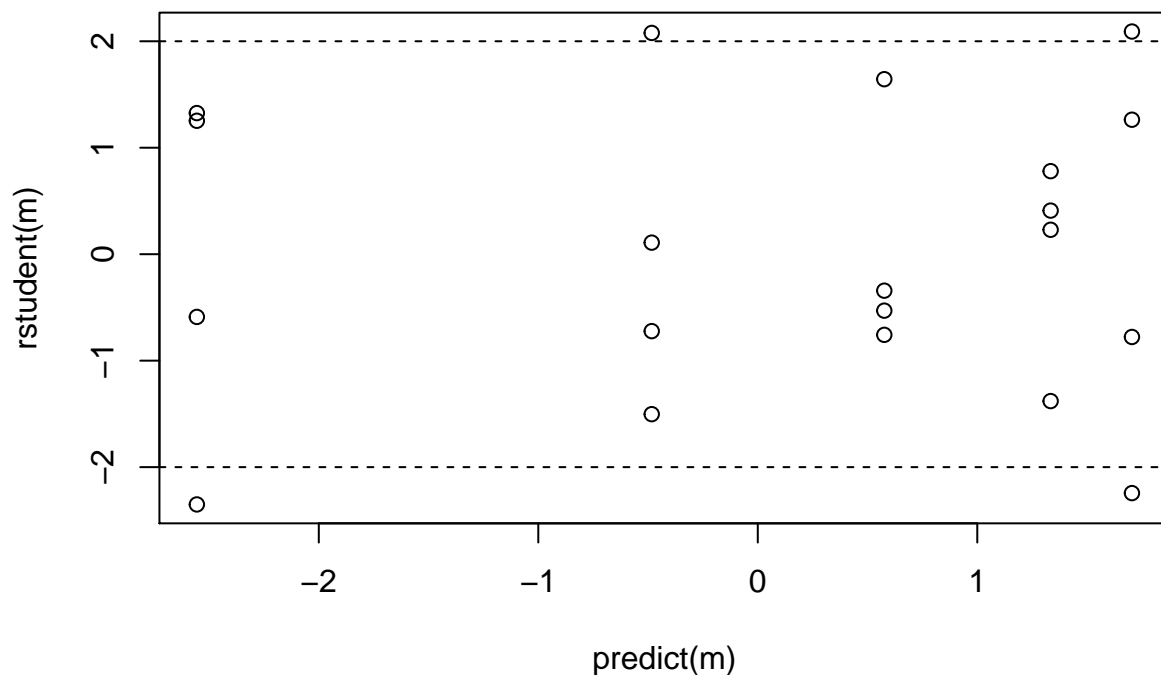
```
#logistic regression model
m <- glm(cbind(Tumor, Total - Tumor) ~ factor(Dose), family = binomial, data = ex2116)
cbind(summary(m)$coefficients, confint(m))
```

##	Estimate	Std. Error	z value	Pr(> z)	2.5 %	97.5 %
## (Intercept)	-2.556	0.2076	-12.310	8.049e-35	-2.988	-2.171
## factor(Dose)0.025	2.073	0.2353	8.809	1.264e-18	1.628	2.553
## factor(Dose)0.05	3.132	0.2354	13.306	2.130e-40	2.688	3.614
## factor(Dose)0.1	3.890	0.2453	15.857	1.252e-56	3.427	4.391

```
## factor(Dose)0.25      4.260      0.2566  16.605 6.436e-62  3.775  4.784
#odds ratio
contrast(m, a = list(Dose = c('0.025', '0.05', '0.1', '0.25')),
          b = list(Dose = '0.01'),
          cnames = c("0.025 vs 0.01", "0.05 vs 0.01", "0.1 vs 0.01", "0.25 vs 0.01"), tf = exp)

##           estimate      se lower  upper tvalue  df    pvalue
## 0.025 vs 0.01    7.945 0.2353  5.01  12.60  8.809 Inf 1.264e-18
## 0.05 vs 0.01   22.920 0.2354 14.45  36.36 13.306 Inf 2.130e-40
## 0.1 vs 0.01    48.909 0.2453 30.24  79.10 15.857 Inf 1.252e-56
## 0.25 vs 0.01   70.840 0.2566 42.84 117.13 16.605 Inf 6.436e-62

#plot studentized residuals vs predicted
plot(predict(m), rstudent(m))
abline(-2, 0, lty = 2)
abline(2, 0, lty = 2)
```



Effects of Commercial Fishing on Abundance

1. Estimate a Poisson regression model where expected number of fish depends on trawl depth and period. Include an interaction. Account for the response variable being a rate (# of fish per unit area) and that trawls vary in terms of their area. Report parameter estimates, se's, and CI's for model parameters.

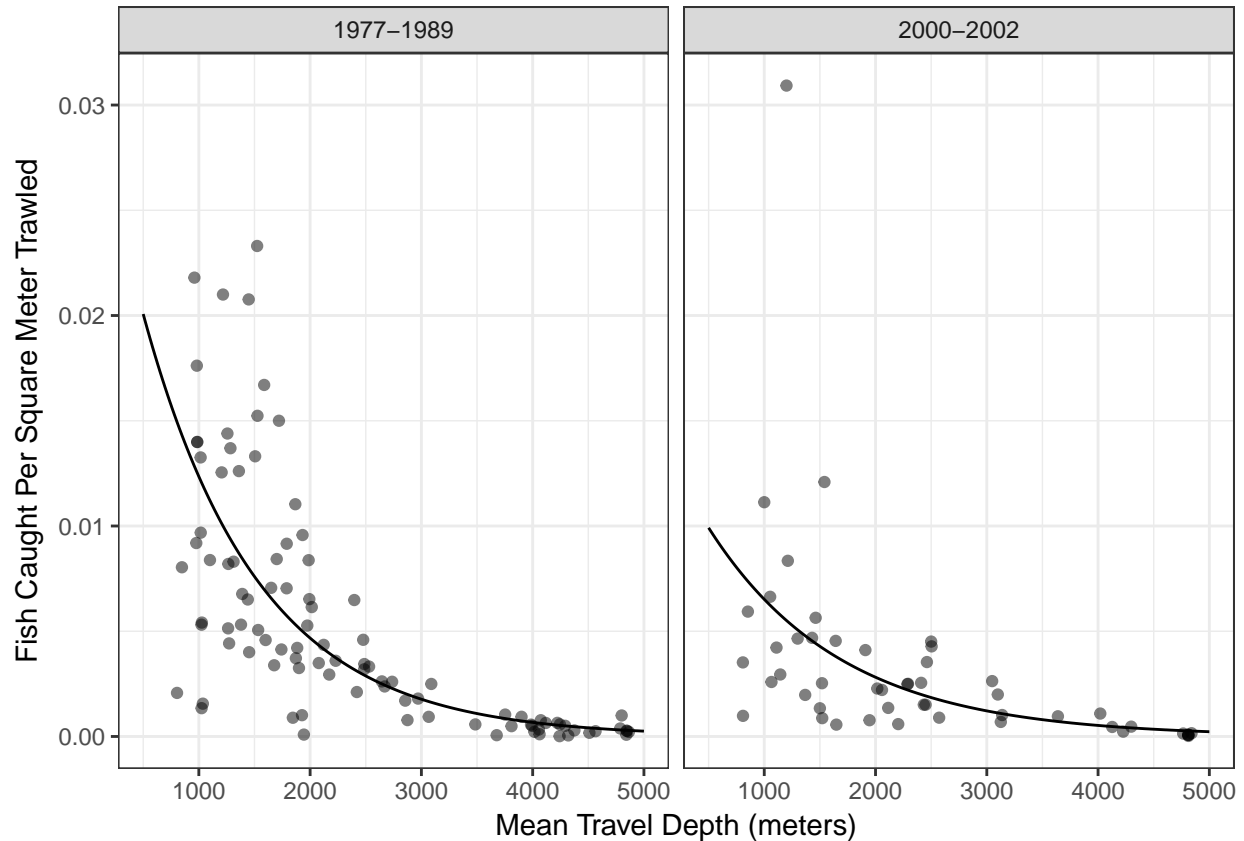
```
data(fishing)
m <- glm(totabund ~ offset(log(sweptarea)) + meandepth * period,
        family = poisson, data = fishing)
cbind(summary(m)$coefficients, confint(m))
```

##		Estimate	Std. Error	z value	Pr(> z)
##	(Intercept)	-3.4228194	1.490e-02	-229.672	0.000e+00
##	meandepth	-0.0009713	7.965e-06	-121.945	0.000e+00
##	period2000-2002	-0.7711169	2.973e-02	-25.937	2.547e-148
##	meandepth:period2000-2002	0.0001318	1.524e-05	8.651	5.090e-18
##		2.5 %	97.5 %		
##	(Intercept)	-3.4520234	-3.3936038		
##	meandepth	-0.0009869	-0.0009557		
##	period2000-2002	-0.8293945	-0.7128522		
##	meandepth:period2000-2002	0.0001019	0.0001616		

2. Plot.

```
d <- expand.grid(period = levels(fishing$period),
                meandepth = seq(500, 5000, length = 1000), sweptarea = 1)
d$yhat <- predict(m, newdata = d, type = "response")

p <- ggplot(fishing, aes(x = meandepth, y = totabund/sweptarea))
p <- p + geom_point(alpha = 0.5) + facet_wrap(~period) + theme_bw()
p <- p + geom_line(aes(y = yhat, group = period), data = d)
p <- p + labs(x = "Mean Travel Depth (meters)",
              y = "Fish Caught Per Square Meter Trawled")
plot(p)
```



3. Write out regression model implied by results of summary

$$E(Y_i) = \exp(b_0^* + b_1 d_i + b_2 p_i + b_3 d_i p_i)$$

where

$$b_0^* = \log(S) + b_0 = \log(\text{sweptarea}) + b_0$$

$$d_i = \text{meandepth}$$

$$p_i =$$

- 1 if period = 2000-2002
- 0 if period = 1977-1989

Casewise:

$$E(Y_i) = \exp(b_0 + b_1 d_i), \text{ for the period 1977-1989}$$

$$E(Y_i) = \exp(b_0 + b_2 + (b_1 + b_3) * d_i), \text{ for the period 2000-2002}$$

4. Using contrast briefly explain (a) how the estimated fish per area changes with mean depth for each period, and (b) how the estimated fish per area changed between the two periods at mean depths of 1000, 2000, 3000, 4000, 5000 meters.

Interpretation:

- a) In the period 1977-1989, for a 1000 m increase in mean depth, the estimated fish per area changes by a factor of 0.37.

In the period 2000-2002, for a 1000 m increase in mean depth, the estimated fish per area changes by a factor of 0.43.

- b) From the period 1977-1989 to the period 2000-2002, the estimated fish per area changes by a factor of 0.52 at a depth of 1000 m.

From the period 1977-1989 to the period 2000-2002, the estimated fish per area changes by a factor of 0.60 at a depth of 2000 m.

From the period 1977-1989 to the period 2000-2002, the estimated fish per area changes by a factor of 0.69 at a depth of 3000 m.

From the period 1977-1989 to the period 2000-2002, the estimated fish per area changes by a factor of 0.78 at a depth of 4000 m.

From the period 1977-1989 to the period 2000-2002, the estimated fish per area changes by a factor of 0.89 at a depth of 5000 m.

#a)

```
contrast(m, a = list(period = "1977-1989", meandepth = 2000, sweptarea = 1),
          b = list(period = "1977-1989", meandepth = 1000, sweptarea = 1),
          cname = "1977-1989", tf = exp)
```

```
##           estimate      se lower upper tvalue df pvalue
## 1977-1989   0.3786 0.007965 0.3727 0.3846 -121.9 Inf      0
```

```
contrast(m, a = list(period = "2000-2002", meandepth = 2000, sweptarea = 1),
          b = list(period = "2000-2002", meandepth = 1000, sweptarea = 1),
          cname = "2000-2002", tf = exp)
```

```
##           estimate      se lower upper tvalue df pvalue
## 2000-2002   0.432 0.01299 0.4211 0.4431 -64.61 Inf      0
```

#b)

```
contrast(m, a = list(period = "2000-2002", meandepth = c(1000, 2000, 3000, 4000, 5000), sweptarea = 1),
          b = list(period = "1977-1989", meandepth = c(1000, 2000, 3000, 4000, 5000), sweptarea = 1),
          cname = c('1000 m', '2000 m', '3000 m', '4000 m', '5000 m'), tf = exp)
```

```
##           estimate      se lower upper tvalue df pvalue
## 1000 m   0.5277 0.01740 0.5100 0.5460 -36.740 Inf 1.700e-295
## 2000 m   0.6020 0.01364 0.5861 0.6183 -37.199 Inf 7.157e-303
## 3000 m   0.6869 0.02311 0.6565 0.7187 -16.256 Inf 2.011e-59
## 4000 m   0.7837 0.03669 0.7293 0.8421 -6.644 Inf 3.045e-11
## 5000 m   0.8941 0.05121 0.8087 0.9885 -2.186 Inf 2.883e-02
```

Efficacy of Two Topical Cream Preparations

1. Estimate a logistic regression model, which includes clinic but the odds ratio for the effect of treatment should not vary between clinics. Treat clinic as a factor.

#treat clinic as factor

```
topical$clinic <- as.factor(topical$clinic)
```

#logistic regression

```
m <- glm(cbind(favorable, unfavorable) ~ clinic + treatment, family = binomial, data = topical)
cbind(summary(m)$coefficients, confint(m))
```

```
##           Estimate Std. Error z value Pr(>|z|) 2.5 % 97.5 %
```

```
## (Intercept)      -1.3220      0.3165 -4.1775 2.947e-05 -1.9709 -0.7246
## clinic2          2.0554      0.4201  4.8929 9.938e-07  1.2568  2.9099
## clinic3          1.1529      0.4246  2.7155 6.618e-03  0.3298  2.0012
## clinic4         -1.4185      0.6636 -2.1376 3.255e-02 -2.9264 -0.2404
## clinic5         -0.5199      0.5338 -0.9740 3.301e-01 -1.6375  0.4840
## clinic6         -2.1469      1.0614 -2.0228 4.310e-02 -5.0699 -0.4672
## clinic7         -0.7977      0.8149 -0.9789 3.276e-01 -2.7230  0.6318
## clinic8          2.2079      0.7195  3.0687 2.150e-03  0.8926  3.7967
## treatmentDrug    0.7769      0.3067  2.5333 1.130e-02  0.1851  1.3916
```

2. Estimate the odds ratio for the effect of treatment two ways: 1) by using the appropriate output from either summary or coef, and 2) by using contrast. Summarize the effect of the treatment using an odds ratio.

Summary: The odds of successful treatment of infection under the 'Drug' treatment is 2.17 times that of the 'Control' treatment.

```
exp(cbind(coef(m), confint(m)))
```

```
##              2.5 %  97.5 %
## (Intercept)  0.2666 0.139334 0.4845
## clinic2      7.8102 3.514126 18.3559
## clinic3      3.1674 1.390724  7.3980
## clinic4      0.2421 0.053589  0.7863
## clinic5      0.5946 0.194463  1.6225
## clinic6      0.1168 0.006283  0.6267
## clinic7      0.4504 0.065675  1.8810
## clinic8      9.0967 2.441547 44.5524
## treatmentDrug 2.1748 1.203376  4.0212
```

```
contrast(m,
  a = list(treatment = "Drug", clinic = levels(topical$clinic)),
  b = list(treatment = "Control", clinic = levels(topical$clinic)),
  cnames = paste("clinic", c(levels(topical$clinic))), tf = exp)
```

```
##      estimate      se lower upper tvalue  df pvalue
## clinic 1      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 2      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 3      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 4      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 5      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 6      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 7      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
## clinic 8      2.175 0.3067 1.192 3.967  2.533 Inf 0.0113
```

3. Produce point estimates and CI's from the model for the probability of a favorable application of the topical cream preparation for each clinic using contrast or glmint.

```
#favorable application with 'DRUG' version of topical cream
```

```
contrast(m,
  a = list(clinic = levels(topical$clinic), treatment = 'Drug'),
  cnames = paste("clinic", c(levels(topical$clinic))), tf = plogis)
```

```
##      estimate      se lower upper tvalue  df pvalue
## clinic 1  0.36700 0.2929 0.24616 0.5072 -1.861 Inf 6.275e-02
```

```
## clinic 2  0.81911 0.3830 0.68127 0.9056  3.943 Inf 8.042e-05
## clinic 3  0.64744 0.3688 0.47129 0.7909  1.648 Inf 9.930e-02
## clinic 4  0.12308 0.6190 0.04005 0.3208 -3.172 Inf 1.514e-03
## clinic 5  0.25636 0.4726 0.12012 0.4654 -2.253 Inf 2.424e-02
## clinic 6  0.06345 1.0322 0.00888 0.3387 -2.608 Inf 9.104e-03
## clinic 7  0.20705 0.7868 0.05290 0.5497 -1.707 Inf 8.790e-02
## clinic 8  0.84062 0.6951 0.57456 0.9537  2.392 Inf 1.674e-02

#favorable application with 'CONTROL' version of topical cream
contrast(m,
  a = list(clinic = levels(topical$clinic), treatment = 'Drug'),
  cnames = paste("clinic", c(levels(topical$clinic))), tf = plogis)

##          estimate      se   lower  upper tvalue  df    pvalue
## clinic 1  0.36700 0.2929 0.24616 0.5072 -1.861 Inf 6.275e-02
## clinic 2  0.81911 0.3830 0.68127 0.9056  3.943 Inf 8.042e-05
## clinic 3  0.64744 0.3688 0.47129 0.7909  1.648 Inf 9.930e-02
## clinic 4  0.12308 0.6190 0.04005 0.3208 -3.172 Inf 1.514e-03
## clinic 5  0.25636 0.4726 0.12012 0.4654 -2.253 Inf 2.424e-02
## clinic 6  0.06345 1.0322 0.00888 0.3387 -2.608 Inf 9.104e-03
## clinic 7  0.20705 0.7868 0.05290 0.5497 -1.707 Inf 8.790e-02
## clinic 8  0.84062 0.6951 0.57456 0.9537  2.392 Inf 1.674e-02
```

An Allometric Model for Barnacles Revisited

1. Estimate the model three ways: as a gamma GLM with the distribution/family specified as a gamma distribution, using a quasi-likelihood approach to specify the variance structure (using either `quais` or `tweedie` for the family argument), and by using iteratively weight least squares with nonlinear regression—similar to the last homework. Use the base-2 logarithm of rostro-carinal length rather than natural log. Report the parameter estimates, se's, and CI's.

```
#subset barnacle data to remove values of 0 and an outlier
barnacle.subset <- barnacle[-c(2, 991, 1451),]
head(barnacle.subset)

##      DW   RC   F
## 1 0.14  9.5 barca
## 3 0.42 13.1 barca
## 4 0.01  3.7 barca
## 5 0.03  5.6 barca
## 6 1.56 18.6 barca
## 7 0.04  5.7 barca

#gamma generalized linear model
m <- glm(DW ~ F * log2(RC), family = Gamma(link = log), data = barnacle.subset)
cbind(summary(m)$coefficients, confint(m))

##          Estimate Std. Error  t value  Pr(>|t|)    2.5 %   97.5 %
## (Intercept) -8.07452    0.030794 -262.210 0.000e+00 -8.13389 -8.01487
## Flens      -0.14365    0.045212  -3.177 1.509e-03 -0.23098 -0.05631
## log2(RC)     1.97218    0.008679  227.225 0.000e+00  1.95539  1.98892
## Flens:log2(RC) 0.05263    0.012677   4.152 3.438e-05  0.02815  0.07711
```

```

#quasi-likelihood model
m <- glm(DW ~ F * log2(RC), family = quasi(link = log, variance = "mu^2"), data = barnacle.subset)
cbind(summary(m)$coefficients, confint(m))

##           Estimate Std. Error  t value Pr(>|t|)    2.5 %   97.5 %
## (Intercept)  -8.07452    0.030794 -262.210 0.000e+00 -8.13389 -8.01487
## Flens        -0.14365    0.045212  -3.177 1.509e-03 -0.23098 -0.05631
## log2(RC)      1.97218    0.008679  227.225 0.000e+00  1.95539  1.98892
## Flens:log2(RC) 0.05263    0.012677   4.152 3.438e-05  0.02815  0.07711

#iteratively weight least squares model
m2 <- nls(DW ~ exp(t0 + t1 * (F == "lens") +
  (t2 + t3 * (F == "lens"))) * log2(RC)), data = barnacle.subset,
  start = c(t0 = -8.07, t1 = 0.14, t2 = 1.97, t3 = 0.05))

for (i in 1:10) {
  barnacle.subset$w <- 1/predict(m2)^2
  m2 <- nls(DW ~ exp(t0 + t1 * (F == "lens") + (t2 + t3 * (F == "lens"))) * log2(RC)),
    data = barnacle.subset, start = coef(m2), weights = w)
}

cbind(summary(m2)$coefficients, confint(m2))

##      Estimate Std. Error  t value Pr(>|t|)    2.5%   97.5%
## t0 -8.07452    0.030794 -262.210 0.000e+00 -8.13621 -8.01365
## t1 -0.14366    0.045212  -3.177 1.509e-03 -0.23370 -0.05368
## t2  1.97218    0.008679  227.225 0.000e+00  1.95497  1.98953
## t3  0.05263    0.012677   4.152 3.437e-05  0.02739  0.07789

```

2. Write out the regression model implied by the results of summary from glm in its mathematical form to show how the $E(Y_i)$ is related to the explanatory variables. Be sure to define each explanatory variable to show that you know how it is computed based on the raw data. Also write the model case-wise for each location.

$$E(Y_i) = \exp(B_0 + B_1 X_{i1} + B_2 X_{i2} + B_3 X_{i3})$$

$$x_{i1} = 1 \text{ if 'F' is 'lens', and 0 if 'F' is 'barca'}$$

$$x_{i2} = \log_2(RC)$$

$$x_{i3} = x_{i1} * x_{i2}$$

Casewise:

$$\text{if 'F' == 'barca': } E(Y_i) = \exp(B_0 + B_2 * \log_2(RC))$$

$$\text{if 'F' == 'lens': } E(Y_i) = \exp(B_0 + B_1 + (B_2 + B_3) * \log_2(RC))$$

3. Use contrast to estimate the multiplicative effect of rostro-carinal length for each location. Use base-2 log in model. Summarize the effect.

Summary: For each doubling of RC length, the dry weight of barnacles: at Punta de Barca increases by a factor of 7.18 and at Punta de Lens by a factor of 7.58.

```

contrast(m,
  a = list(RC = 20, F = c('barca', 'lens')),
  b = list(RC = 10, F = c('barca', 'lens')),
  cnames = c('barca', 'lens'), tf = exp)

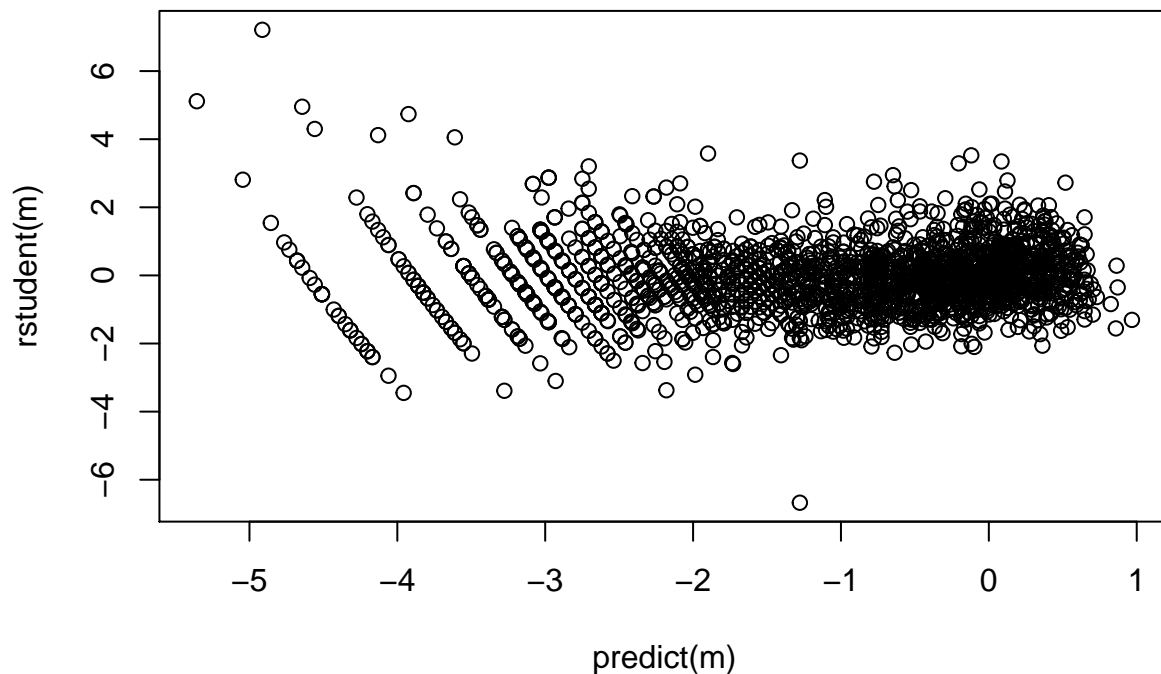
```

```
##      estimate      se lower upper tvalue  df pvalue
## barca    7.186 0.008679 7.065 7.310  227.2 1993     0
## lens     7.575 0.009240 7.439 7.713  219.1 1993     0
```

4. Produce a studentized residual plot to check to see if the specified variance structure was appropriate. Comment on whether you believe the variance structure was modeled correctly and why.

Comment: The variance structure was not modeled correctly because the studentized residuals show decreasing variance.

```
plot(predict(m), rstudent(m))
```



5. Conduct a significance test to determine if the difference between the multiplicative effect in the different locations is statistically significant.

HOW?

```
m <- glm(DW ~ F * log2(RC), family = Gamma(link = log), data = barnacle.subset)
m.null <- glm(DW ~ F + log2(RC), family = Gamma(link = log), data = barnacle.subset)
anova(m.null, m, test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: DW ~ F + log2(RC)
## Model 2: DW ~ F * log2(RC)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1      1994      58.4
## 2      1993      57.9 1      0.552 2.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Moth Habitat Study

1. Estimate the model and report output from summary. Try obtaining profile likelihood ratio CI's (you will get an error but still report). Obtain Wald CI's for the model parameters using `coefest` or `lincon`. Change the habitat reference level to “Disturbed” using `relevel`. Report output from summary, try obtaining both types of CI's again. Estimate a model without an interaction and report the results from summary and try to obtain both types of CI's. Discuss what you observed when comparing results from all three models/parameterizations. What did you notice that was unusual?

Discussion: When the reference level for habitat was changed to “Disturbed”, the parameter estimates for the “Bank” level were much greater than the other habitat levels, and the standard errors for the “Bank” level were many orders of magnitude greater than the other habitat levels. Something is going wrong with the model in relation to the Bank habitat level. When the interaction between habitat and moth type is removed from the model, there is nothing odd about the ‘Bank’ level parameter estimates and standard errors.

```
#put data into long form
moths.long <- gather(moths, A, P, key = 'type', value = 'count')

#model rate of count/m with effect of habitat, type, and interaction
m <- glm(count ~ offset(log(meters)) + type * habitat,
         family = poisson, data = moths.long)
summary(m)$coefficients
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -17.35      773.8 -0.02242  0.9821
## typeP           15.69      773.8  0.02028  0.9838
## habitatDisturbed 15.53      773.8  0.02008  0.9840
## habitatLowerside 15.81      773.8  0.02043  0.9837
## habitatNEsoak    14.45      773.8  0.01867  0.9851
## habitatNWsoak    17.44      773.8  0.02253  0.9820
## habitatSEsoak    15.70      773.8  0.02028  0.9838
## habitatSWsoak    15.59      773.8  0.02015  0.9839
## habitatUpperside 13.82      773.8  0.01786  0.9857
## typeP:habitatDisturbed -14.27  773.8 -0.01844  0.9853
## typeP:habitatLowerside -16.57  773.8 -0.02141  0.9829
## typeP:habitatNEsoak    -15.69  773.8 -0.02028  0.9838
## typeP:habitatNWsoak    -17.01  773.8 -0.02198  0.9825
## typeP:habitatSEsoak    -17.51  773.8 -0.02263  0.9819
## typeP:habitatSWsoak    -14.81  773.8 -0.01914  0.9847
## typeP:habitatUpperside -16.94  773.8 -0.02189  0.9825
```

```
#try obtaining profile likelihood ratio CI
confint(m)
```

```
## Warning: glm.fit: fitted rates numerically 0 occurred
```

```
## Warning: glm.fit: fitted rates numerically 0 occurred
```

```
## Warning: glm.fit: fitted rates numerically 0 occurred
```

```
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Error in glm.fit(x = Xi, y = Y, weights = W, etastart = LP, offset = o, : NA/NaN/Inf in 'x'
#Obtain Wald CI for model parameters with coeftest or lincon
lincon(m)

##               estimate      se lower upper    tvalue df pvalue
## (Intercept)      -17.35 773.8 -1534  1499 -0.02242 Inf  0.9821
## typeP             15.69 773.8 -1501  1532  0.02028 Inf  0.9838
## habitatDisturbed  15.53 773.8 -1501  1532  0.02008 Inf  0.9840
## habitatLowerside  15.81 773.8 -1501  1532  0.02043 Inf  0.9837
## habitatNEsoak     14.45 773.8 -1502  1531  0.01867 Inf  0.9851
## habitatNWsoak     17.44 773.8 -1499  1534  0.02253 Inf  0.9820
## habitatSEsoak     15.70 773.8 -1501  1532  0.02028 Inf  0.9838
## habitatSWsoak     15.59 773.8 -1501  1532  0.02015 Inf  0.9839
## habitatUpperside  13.82 773.8 -1503  1530  0.01786 Inf  0.9857
## typeP:habitatDisturbed -14.27 773.8 -1531  1502 -0.01844 Inf  0.9853
## typeP:habitatLowerside -16.57 773.8 -1533  1500 -0.02141 Inf  0.9829
## typeP:habitatNEsoak  -15.69 773.8 -1532  1501 -0.02028 Inf  0.9838
## typeP:habitatNWsoak  -17.01 773.8 -1534  1500 -0.02198 Inf  0.9825
## typeP:habitatSEsoak  -17.51 773.8 -1534  1499 -0.02263 Inf  0.9819
## typeP:habitatSWsoak  -14.81 773.8 -1531  1502 -0.01914 Inf  0.9847
## typeP:habitatUpperside -16.94 773.8 -1534  1500 -0.02189 Inf  0.9825

#change reference level to "Disturbed" using relevel. Report output from summary, try obtaining both ty
moths.long$habitat <- relevel(moths.long$habitat, ref = "Disturbed")
m <- glm(count ~ offset(log(meters)) + type * habitat,
         family = poisson, data = moths.long)
summary(m)$coefficients

##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.81238    0.3536 -5.12618 2.957e-07
## typeP             1.41707    0.3941  3.59584 3.233e-04
## habitatBank      -15.53473   773.7839 -0.02008 9.840e-01
## habitatLowerside   0.27368    0.3865  0.70807 4.789e-01
## habitatNEsoak     -1.08590    0.4432 -2.45012 1.428e-02
## habitatNWsoak      1.90067    0.3729  5.09645 3.461e-07
## habitatSEsoak      0.16061    0.3899  0.41191 6.804e-01
## habitatSWsoak      0.05452    0.4183  0.13033 8.963e-01
## habitatUpperside  -1.71398    0.4009 -4.27542 1.908e-05
## typeP:habitatBank  14.27181   773.7841  0.01844 9.853e-01
## typeP:habitatLowerside -2.29742    0.4884 -4.70416 2.549e-06
## typeP:habitatNEsoak -1.41707    0.5460 -2.59517 9.455e-03
## typeP:habitatNWsoak -2.73531    0.4712 -5.80512 6.432e-09
```

```
## typeP:habitatSEsoak      -3.23622      0.5908 -5.47807 4.300e-08
## typeP:habitatSWsoak      -0.54160      0.4755 -1.13891 2.547e-01
## typeP:habitatUpperside   -2.66983      0.5622 -4.74928 2.041e-06
```

lincon(m)

```
##              estimate      se      lower      upper      tvalue
## (Intercept)    -1.81238    0.3536    -2.5053    -1.1194   -5.12618
## typeP           1.41707    0.3941     0.6447     2.1895    3.59584
## habitatBank    -15.53473  773.7839  -1532.1233  1501.0539  -0.02008
## habitatLowerside  0.27368    0.3865    -0.4839     1.0312    0.70807
## habitatNEsoak   -1.08590    0.4432    -1.9546    -0.2172   -2.45012
## habitatNWsoak    1.90067    0.3729     1.1697     2.6316    5.09645
## habitatSEsoak    0.16061    0.3899    -0.6036     0.9248    0.41191
## habitatSWsoak    0.05452    0.4183    -0.7654     0.8744    0.13033
## habitatUpperside -1.71398    0.4009    -2.4997    -0.9282   -4.27542
## typeP:habitatBank 14.27181  773.7841  -1502.3171  1530.8608    0.01844
## typeP:habitatLowerside -2.29742    0.4884    -3.2546    -1.3402   -4.70416
## typeP:habitatNEsoak -1.41707    0.5460    -2.4873    -0.3468   -2.59517
## typeP:habitatNWsoak -2.73531    0.4712    -3.6588    -1.8118   -5.80512
## typeP:habitatSEsoak -3.23622    0.5908    -4.3941    -2.0784   -5.47807
## typeP:habitatSWsoak -0.54160    0.4755    -1.4736     0.3904   -1.13891
## typeP:habitatUpperside -2.66983    0.5622    -3.7716    -1.5680   -4.74928
##              df      pvalue
## (Intercept)    Inf 2.957e-07
## typeP           Inf 3.233e-04
## habitatBank     Inf 9.840e-01
## habitatLowerside Inf 4.789e-01
## habitatNEsoak   Inf 1.428e-02
## habitatNWsoak   Inf 3.461e-07
## habitatSEsoak   Inf 6.804e-01
## habitatSWsoak   Inf 8.963e-01
## habitatUpperside Inf 1.908e-05
## typeP:habitatBank Inf 9.853e-01
## typeP:habitatLowerside Inf 2.549e-06
## typeP:habitatNEsoak Inf 9.455e-03
## typeP:habitatNWsoak Inf 6.432e-09
## typeP:habitatSEsoak Inf 4.300e-08
## typeP:habitatSWsoak Inf 2.547e-01
## typeP:habitatUpperside Inf 2.041e-06
```

confint(m)

```
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
## Warning: glm.fit: fitted rates numerically 0 occurred
```



```
## Warning: glm.fit: fitted rates numerically 0 occurred

## Warning: glm.fit: fitted rates numerically 0 occurred

## Error in glm.fit(x = Xi, y = Y, weights = W, etastart = LP, offset = o, : NA/NaN/Inf in 'x'
#Estimate model without an interaction
m <- glm(count ~ offset(log(meters)) + type + habitat,
          family = poisson, data = moths.long)
summary(m)$coefficients

##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.6973     0.1620  -4.304 1.674e-05
## typeP         -0.3851     0.1062  -3.627 2.872e-04
## habitatBank   -1.4800     0.5238  -2.825 4.723e-03
## habitatLowerside -1.0136     0.2040  -4.968 6.778e-07
## habitatNEsoak  -2.0269     0.2452  -8.268 1.368e-16
## habitatNWsoak   0.5037     0.1884   2.673 7.514e-03
## habitatSEsoak  -1.3232     0.2183  -6.062 1.343e-09
## habitatSWsoak  -0.3558     0.1977  -1.800 7.192e-02
## habitatUpperside -3.0968     0.2284 -13.558 7.059e-42

lincon(m)

##              estimate      se  lower  upper  tvalue  df    pvalue
## (Intercept)   -0.6973 0.1620 -1.0147 -0.3798  -4.304  Inf  1.674e-05
## typeP         -0.3851 0.1062 -0.5933 -0.1770  -3.627  Inf  2.872e-04
## habitatBank   -1.4800 0.5238 -2.5067 -0.4533  -2.825  Inf  4.723e-03
## habitatLowerside -1.0136 0.2040 -1.4135 -0.6137  -4.968  Inf  6.778e-07
## habitatNEsoak  -2.0269 0.2452 -2.5074 -1.5464  -8.268  Inf  1.368e-16
## habitatNWsoak   0.5037 0.1884  0.1344  0.8730   2.673  Inf  7.514e-03
## habitatSEsoak  -1.3232 0.2183 -1.7511 -0.8954  -6.062  Inf  1.343e-09
## habitatSWsoak  -0.3558 0.1977 -0.7434  0.0317  -1.800  Inf  7.192e-02
## habitatUpperside -3.0968 0.2284 -3.5445 -2.6491 -13.558  Inf  7.059e-42

confint(m)

##              2.5 %   97.5 %
## (Intercept)   -1.0304 -0.39389
## typeP         -0.5948 -0.17817
## habitatBank   -2.6826 -0.57349
## habitatLowerside -1.4102 -0.60772
## habitatNEsoak  -2.5170 -1.55132
## habitatNWsoak   0.1418  0.88261
## habitatSEsoak  -1.7520 -0.89305
## habitatSWsoak  -0.7384  0.03918
## habitatUpperside -3.5485 -2.64944
```

2. Estimate the model with an interaction, but omitting data from transects from the “Bank” habitat type. Report parameter estimates, se’s, and profile likelihood and Wald CI’s.

```
#omit data from transects from the Bank habitat type
moths.long.subset <- subset(moths.long, habitat != "Bank")

#estimate model
m <- glm(count ~ offset(log(meters)) + type * habitat,
```

```
family = poisson, data = moths.long.subset)
summary(m)$coefficients
```

##	Estimate	Std. Error	z value	Pr(> z)
## (Intercept)	-1.81238	0.3536	-5.1262	2.957e-07
## typeP	1.41707	0.3941	3.5958	3.233e-04
## habitatLowerside	0.27368	0.3865	0.7081	4.789e-01
## habitatNEsoak	-1.08590	0.4432	-2.4501	1.428e-02
## habitatNWsoak	1.90067	0.3729	5.0964	3.461e-07
## habitatSEsoak	0.16061	0.3899	0.4119	6.804e-01
## habitatSWsoak	0.05452	0.4183	0.1303	8.963e-01
## habitatUpperside	-1.71398	0.4009	-4.2754	1.908e-05
## typeP:habitatLowerside	-2.29742	0.4884	-4.7042	2.549e-06
## typeP:habitatNEsoak	-1.41707	0.5460	-2.5952	9.455e-03
## typeP:habitatNWsoak	-2.73531	0.4712	-5.8051	6.432e-09
## typeP:habitatSEsoak	-3.23622	0.5908	-5.4781	4.300e-08
## typeP:habitatSWsoak	-0.54160	0.4755	-1.1389	2.547e-01
## typeP:habitatUpperside	-2.66983	0.5622	-4.7493	2.041e-06

```
lincon(m)
```

##	estimate	se	lower	upper	tvalue	df
## (Intercept)	-1.81238	0.3536	-2.5053	-1.1194	-5.1262	Inf
## typeP	1.41707	0.3941	0.6447	2.1895	3.5958	Inf
## habitatLowerside	0.27368	0.3865	-0.4839	1.0312	0.7081	Inf
## habitatNEsoak	-1.08590	0.4432	-1.9546	-0.2172	-2.4501	Inf
## habitatNWsoak	1.90067	0.3729	1.1697	2.6316	5.0964	Inf
## habitatSEsoak	0.16061	0.3899	-0.6036	0.9248	0.4119	Inf
## habitatSWsoak	0.05452	0.4183	-0.7654	0.8744	0.1303	Inf
## habitatUpperside	-1.71398	0.4009	-2.4997	-0.9282	-4.2754	Inf
## typeP:habitatLowerside	-2.29742	0.4884	-3.2546	-1.3402	-4.7042	Inf
## typeP:habitatNEsoak	-1.41707	0.5460	-2.4873	-0.3468	-2.5952	Inf
## typeP:habitatNWsoak	-2.73531	0.4712	-3.6588	-1.8118	-5.8051	Inf
## typeP:habitatSEsoak	-3.23622	0.5908	-4.3941	-2.0784	-5.4781	Inf
## typeP:habitatSWsoak	-0.54160	0.4755	-1.4736	0.3904	-1.1389	Inf
## typeP:habitatUpperside	-2.66983	0.5622	-3.7716	-1.5680	-4.7493	Inf
##	pvalue					
## (Intercept)	2.957e-07					
## typeP	3.233e-04					
## habitatLowerside	4.789e-01					
## habitatNEsoak	1.428e-02					
## habitatNWsoak	3.461e-07					
## habitatSEsoak	6.804e-01					
## habitatSWsoak	8.963e-01					
## habitatUpperside	1.908e-05					
## typeP:habitatLowerside	2.549e-06					
## typeP:habitatNEsoak	9.455e-03					
## typeP:habitatNWsoak	6.432e-09					
## typeP:habitatSEsoak	4.300e-08					
## typeP:habitatSWsoak	2.547e-01					
## typeP:habitatUpperside	2.041e-06					

```
confint(m)
```

##	2.5 %	97.5 %
----	-------	--------

```
## (Intercept)          -2.5955 -1.1911
## typeP                0.6954  2.2619
## habitatLowerside     -0.4290  1.1071
## habitatNEsoak        -1.9337 -0.1682
## habitatNWsoak        1.2318  2.7135
## habitatSEsoak        -0.5506  0.9991
## habitatSWsoak        -0.7288  0.9356
## habitatUpperside     -2.4529 -0.8589
## typeP:habitatLowerside -3.3099 -1.3808
## typeP:habitatNEsoak   -2.5280 -0.3702
## typeP:habitatNWsoak   -3.7170 -1.8549
## typeP:habitatSEsoak   -4.4818 -2.1457
## typeP:habitatSWsoak   -1.5201  0.3633
## typeP:habitatUpperside -3.8388 -1.6195
```

3. Use contrast or glmint to report point estimates and CI's for the expected number of moths per meter for each of the 14 combinations of habitat and moth type.

```
#moth A point estimates
contrast(m,
  a = list(type = 'A',
    habitat = c('Disturbed', 'Lowerside', 'NEsoak', 'NWsoak', 'SEsoak', 'SWsoak', 'Upperside'),
    meters = 1),
  tf = exp,
  cnames = paste("Moth A at", c('Disturbed', 'Lowerside', 'NEsoak', 'NWsoak', 'SEsoak', 'SWsoak', 'Upperside')))
```

	estimate	se	lower	upper	tvalue	df	pvalue
## Moth A at Disturbed	0.16327	0.3536	0.08165	0.32647	-5.126	Inf	2.957e-07
## Moth A at Lowerside	0.21466	0.1562	0.15806	0.29153	-9.853	Inf	6.686e-23
## Moth A at NEsoak	0.05512	0.2673	0.03264	0.09307	-10.844	Inf	2.121e-27
## Moth A at NWsoak	1.09231	0.1187	0.86562	1.37836	0.744	Inf	4.569e-01
## Moth A at SEsoak	0.19171	0.1644	0.13890	0.26459	-10.047	Inf	9.438e-24
## Moth A at SWsoak	0.17241	0.2236	0.11123	0.26724	-7.861	Inf	3.799e-15
## Moth A at Upperside	0.02941	0.1890	0.02031	0.04260	-18.660	Inf	1.052e-77

```
#moth b point estimates
contrast(m,
  a = list(type = 'P',
    habitat = c('Disturbed', 'Lowerside', 'NEsoak', 'NWsoak', 'SEsoak', 'SWsoak', 'Upperside'),
    meters = 1),
  tf = exp,
  cnames = paste("Moth P at", c('Disturbed', 'Lowerside', 'NEsoak', 'NWsoak', 'SEsoak', 'SWsoak', 'Upperside')))
```

	estimate	se	lower	upper	tvalue	df	pvalue
## Moth P at Disturbed	0.673469	0.1741	0.478787	0.94731	-2.271	Inf	2.315e-02
## Moth P at Lowerside	0.089005	0.2425	0.055331	0.14317	-9.974	Inf	1.980e-23
## Moth P at NEsoak	0.055118	0.2673	0.032644	0.09307	-10.844	Inf	2.121e-27
## Moth P at NWsoak	0.292308	0.2294	0.186449	0.45827	-5.361	Inf	8.266e-08
## Moth P at SEsoak	0.031088	0.4082	0.013967	0.06920	-8.502	Inf	1.863e-17
## Moth P at SWsoak	0.413793	0.1443	0.311834	0.54909	-6.113	Inf	9.755e-10
## Moth P at Upperside	0.008403	0.3536	0.004203	0.01680	-13.517	Inf	1.235e-41

4. Use contrast to estimate the effect of moth type as a multiplicative effect for the Disturbed, Lowerside, and Upperside habitats. Summarize the effect.

Summary: The expected number of P moths per meter in the Disturbed habitat is 4.1 times greater than the expected number of A moths per meter.

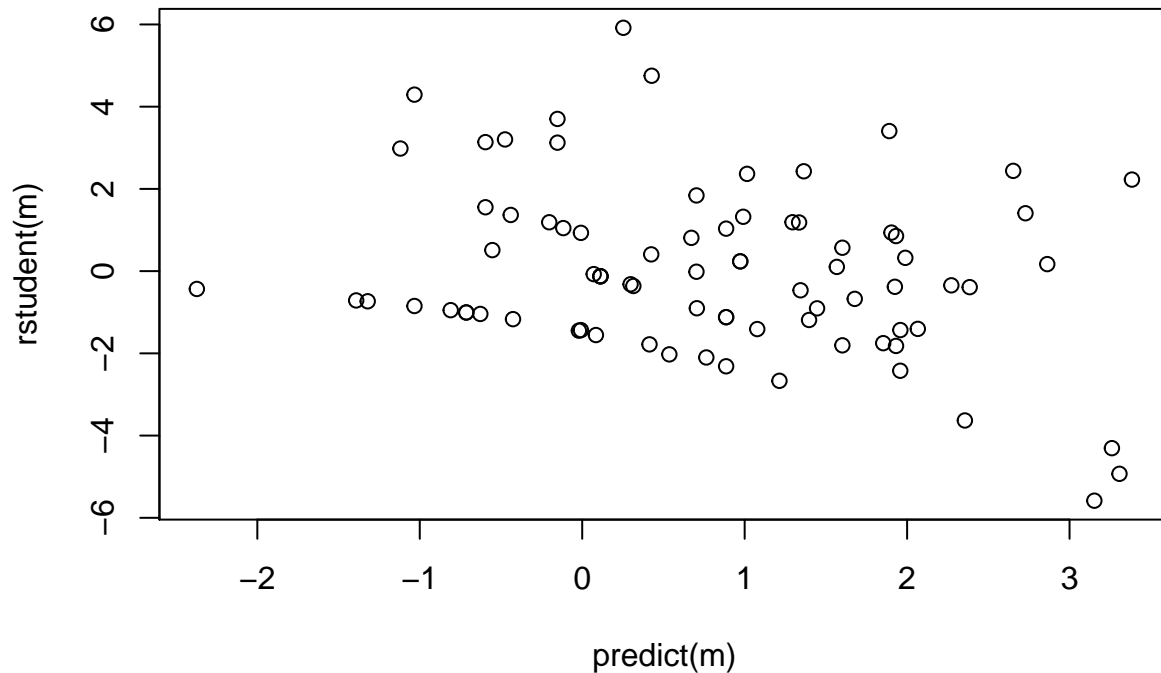
```
contrast(m,
  a = list(type = 'P',
    habitat = c('Disturbed', 'Lowerside', 'Upperside'),
    meters = 1),
  b = list(type = 'A',
    habitat = c('Disturbed', 'Lowerside', 'Upperside'),
    meters = 1),
  tf = exp,
  cnames = paste("P vs A at", c('Disturbed', 'Lowerside', 'Upperside')))
```

##		estimate	se	lower	upper	tvalue	df	pvalue
##	P vs A at Disturbed	4.1250	0.3941	1.9054	8.9304	3.596	Inf	0.0003233
##	P vs A at Lowerside	0.4146	0.2885	0.2356	0.7298	-3.052	Inf	0.0022744
##	P vs A at Upperside	0.2857	0.4009	0.1302	0.6269	-3.125	Inf	0.0017784

5. Check the model for overdispersion. Comment and justify if and why you think overdispersion is or is not present.

Comment: There is overdispersion because variance is increasing; not quite a classic megaphone shape, but similar.

```
plot(predict(m), rstudent(m))
```



6. Using data from the df moths, estimate a logistic regression model for the proportion of moths in each transect that are of type P, but using only habitat type as your explanatory variables. Be sure to drop data from the “Bank” habitat type. Report parameter estimates, se’s, and CI’s.

```
#remove 'Bank' habitat data
```

```
moths.subset <- subset(moths, habitat != "Bank")
```

```
head(moths.subset)
```

```
##   meters A P   habitat
## 1     25 9 8   NWsoak
## 2     37 3 20  SWsoak
## 3    109 7 9  Lowerside
## 4     10 0 2  Lowerside
## 5    133 9 1  Upperside
## 6     26 3 18 Disturbed
```

```
#logistic regression model
```

```
m <- glm(cbind(P, A) ~ habitat, family = binomial, data = moths.subset)
```

```
cbind(summary(m)$coefficients, confint(m))
```

```
##               Estimate Std. Error z value Pr(>|z|)    2.5 % 97.5 %
## (Intercept)      1.4171     0.3941   3.596 3.233e-04  0.6954 2.2619
## habitatLowerside -2.2974     0.4884  -4.704 2.549e-06 -3.3099 -1.3808
## habitatNEsoak    -1.4171     0.5460  -2.595 9.455e-03 -2.5280 -0.3702
## habitatNWsoak    -2.7353     0.4712  -5.805 6.432e-09 -3.7170 -1.8549
## habitatSEsoak    -3.2362     0.5907  -5.478 4.292e-08 -4.4818 -2.1457
## habitatSWsoak    -0.5416     0.4755  -1.139 2.547e-01 -1.5201  0.3633
## habitatUpperside -2.6698     0.5622  -4.749 2.041e-06 -3.8388 -1.6195
```

7. For the Disturbed, Lowerside, and Upperside habitat types, use contrast to estimate: (a) the odds that a moth found a given habitat is P vs A, and (b) the probability that a moth found in a given habitat will be of type P, noting that the log of the odds produced by contrast can be converted to odds and probabilities by transforming them with the R functions ‘exp’ and ‘plogis’ respectively.

```
#a) odds
```

```
contrast(m,
```

```
  a = list(habitat = c('Disturbed', 'Lowerside', 'Upperside')),
```

```
  tf = exp, cnames = paste('P vs. A at', c('Disturbed', 'Lowerside', 'Upperside'))))
```

```
##               estimate      se lower upper tvalue df    pvalue
## P vs. A at Disturbed    4.1250 0.3941 1.9054 8.9304  3.596 Inf 0.0003233
## P vs. A at Lowerside    0.4146 0.2885 0.2356 0.7298 -3.052 Inf 0.0022741
## P vs. A at Upperside    0.2857 0.4009 0.1302 0.6269 -3.125 Inf 0.0017784
```

```
#b) probability
```

```
contrast(m,
```

```
  a = list(habitat = c('Disturbed', 'Lowerside', 'Upperside')),
```

```
  tf = plogis, cnames = paste('P vs. A at', c('Disturbed', 'Lowerside', 'Upperside'))))
```

```
##               estimate      se lower upper tvalue df    pvalue
## P vs. A at Disturbed    0.8049 0.3941 0.6558 0.8993  3.596 Inf 0.0003233
## P vs. A at Lowerside    0.2931 0.2885 0.1907 0.4219 -3.052 Inf 0.0022741
## P vs. A at Upperside    0.2222 0.4009 0.1152 0.3853 -3.125 Inf 0.0017784
```