STA303/1002 - Week 8 R Markdown

March 4-8, 2019

Case Study IV: The Data

Get the data (from R library):

```
#load Sleuth3 R data library; see case2101
library(Sleuth3); krunnit = case2101
str(krunnit)
```

18 islands

```
## 'data.frame':    18 obs. of  4 variables:
##  $ Island : Factor w/ 18 levels "Hietakraasukka",..: 16 6 11 2 1 3 4 7 15 12
##  $ Area   : num  185.8 105.8 30.7 8.5 4.8 ...
##  $ AtRisk : int  75 67 66 51 28 20 43 31 28 32 ...
##  $ Extinct: int  5 3 10 6 3 4 8 3 5 6 ...
```

x_i

m_i

y_i

Case Study IV: New variables

Get the data (from R library):

```
attach(krunnit); head(krunnit)
```

x_i m_i y_i y_i/m_i

```
##              Island  Area AtRisk Extinct
## 1       Ulkokrunni 185.8     75       5
## 2        Maakrunni 105.8     67       3
## 3        Ristikari  30.7     66      10
## 4 Isonkivenletto   8.5     51       6
## 5 Hietakraasukka   4.8     28       3
## 6        Kraasukka   4.5     20       4
```

$\log\left(\frac{\pi_i}{1-\pi_i}\right)$

$\log\left(\frac{y_i}{m_i-y_i}\right)$

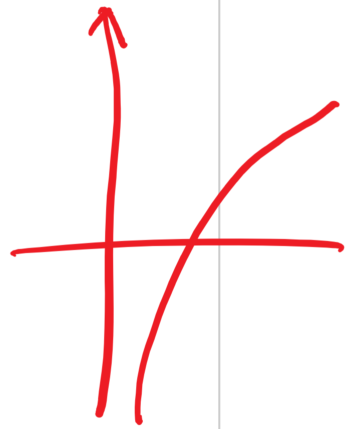
```
logitpi<-log(Extinct/AtRisk/(1-(Extinct/AtRisk))) #observed logits
logarea<-log(Area) # log transformed Area
NExtinct<-AtRisk-Extinct
pis<-Extinct/AtRisk
```

$= m_i - y_i$

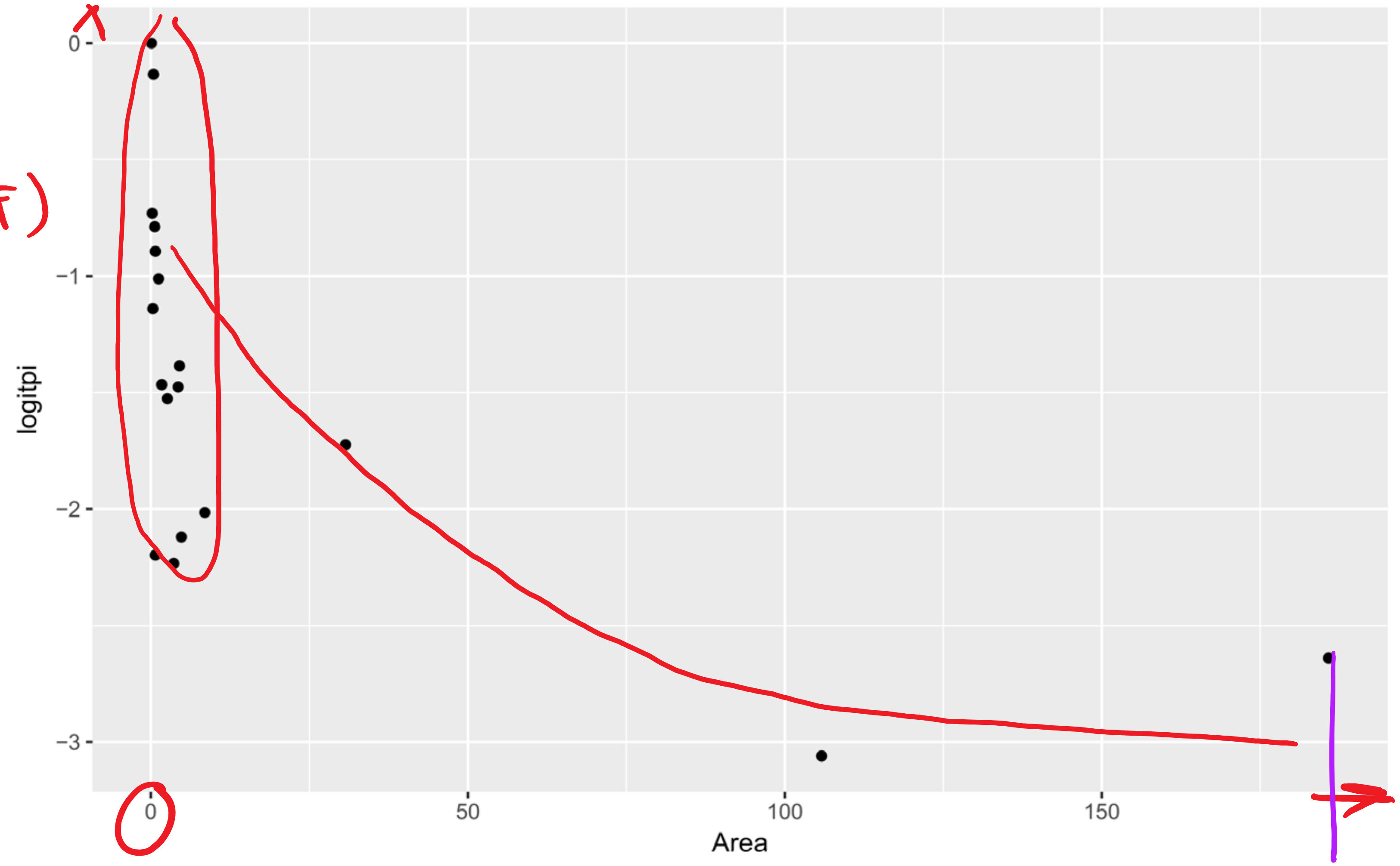
$= y_i/m_i$

Case Study IV: Visualizing the data

```
library(ggplot2)
ggplot(krunnit, aes(x=Area, y=logitpi))+geom_point()
```



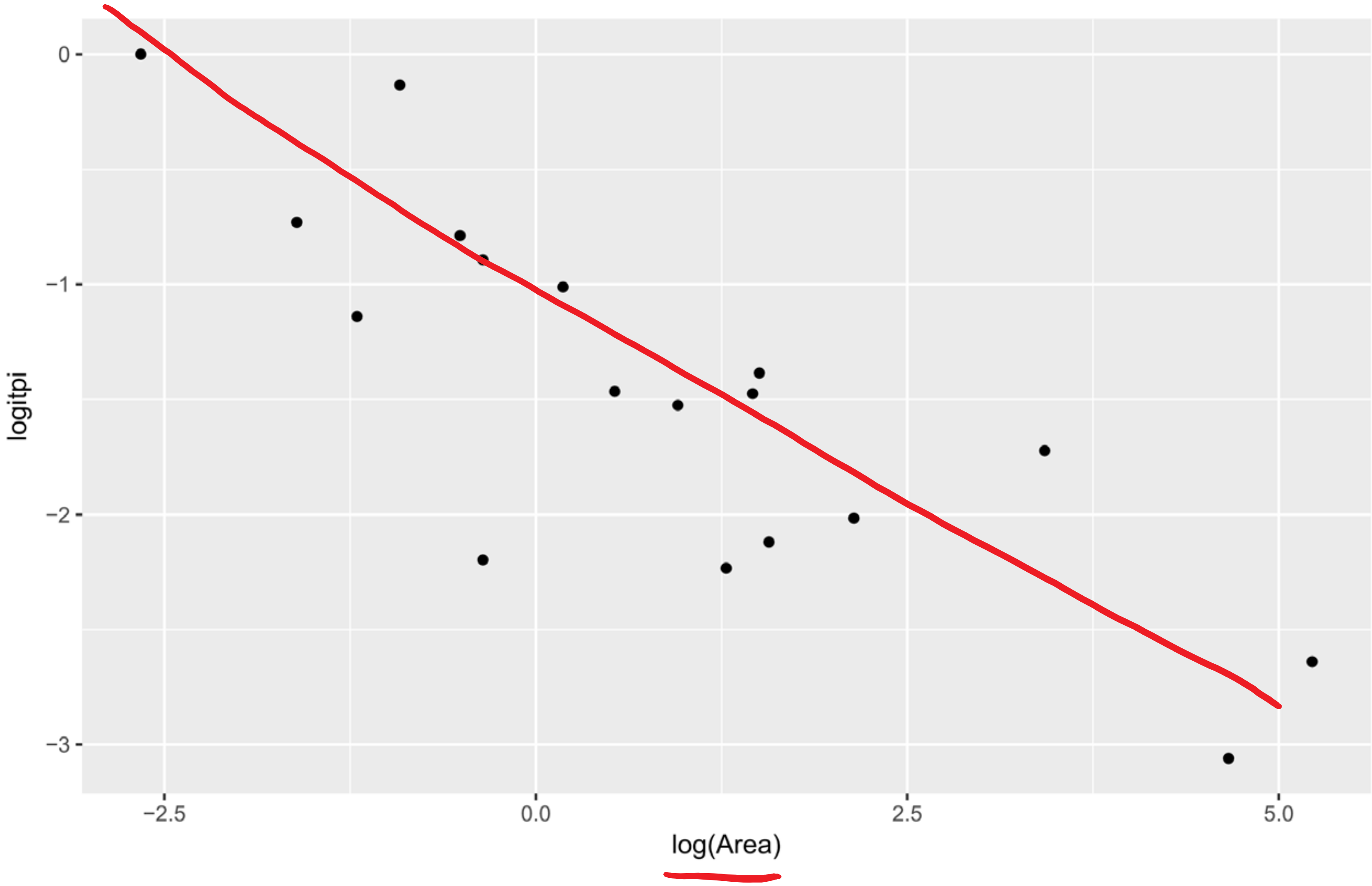
$\text{logit}(\bar{\pi})$



x_i

Case Study IV: Visualizing the data

```
ggplot(krunnit, aes(x=log(Area), y=logitpi))+geom_point()
```



response predictor

Case Study IV: Logisitc Model with logged explanatory variable

```r
fitbl<-glm(cbind(Extinct,NExtinct)~log(Area), family=binomial, data=krunnit)
summary(fitbl)
```

y_i $m_i - y_i$ logistic regression.

```
##
## Call:
## glm(formula = cbind(Extinct, NExtinct) ~ log(Area), family = binomial,
##     data = krunnit)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.71726  -0.67722   0.09726   0.48365   1.49545
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.19620    0.11845 -10.099  < 2e-16 ***
## log(Area)   -0.29710    0.05485  -5.416 6.08e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45.338  on 17  degrees of freedom
## Residual deviance: 12.062  on 16  degrees of freedom
## AIC: 75.394
##
## Number of Fisher Scoring iterations: 4
```

$\hat{\beta}$ $se(\hat{\beta})$ $\frac{\hat{\beta}}{se(\hat{\beta})}$ Wald

$\Rightarrow p \approx 0$ or $p < 0.0001$

Deviance (Fitted vs Sat.).

Case IV: Deviance test and Estimated Var-Cov of β

```r
anova(fitbl, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(Extinct, NExtinct)
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         17     45.338
## log(Area)  1   33.277        16     12.062 7.994e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

LRT

```r
print(vcov(fitbl))
```

```
##              (Intercept)    log(Area)
## (Intercept)  0.014029452 -0.002602237
## log(Area)   -0.002602237  0.003008830
```

$0.003 = (0.05485)^2$

$Var(\hat{\beta}_1) = (se(\hat{\beta}_1))^2$

Case IV: Wald tests in R

```
library(aod) # Analysis of Overdispersed Data
wald.test(Sigma=vcov(fitbl), b=coef(fitbl), Terms=2)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 29.3, df = 1, P(> X2) = 6.1e-08
```

$= (-5.416)^2 = 29.3$

Case IV: Confidence Intervals for β's

```
CL=cbind(bhat=coef(fitbl), confint.default(fitbl))  # 95% CI for betas
CL
```

```
##                   bhat      2.5 %     97.5 %
## (Intercept) -1.1961955 -1.4283454 -0.9640456
## log(Area)   -0.2971037 -0.4046132 -0.1895942
```

$\hat{\beta}_1 \pm 1.96(se(\hat{\beta}_1))$

$-0.297 \pm 1.96(0.05485)$

```
2^(CL) # doubling Area
```

```
##                  bhat     2.5 %    97.5 %
## (Intercept) 0.4364247 0.3715568 0.5126174
## log(Area)   0.8138847 0.7554388 0.8768524
```

```
.5^(CL) # halving Area
```

```
##                 bhat    2.5 %   97.5 %
## (Intercept) 2.291346 2.691379 1.950773
## log(Area)   1.228675 1.323734 1.140443
```

Case IV: Estimated probabilities of extinction per island

```
phats<-predict.glm(fitbl, type="response") # estimated probability of extinctio
options(digits=4)
rbind(Extinct, NExtinct, pis,phats)
```

5/75

```
##    i =          1        2        3       4       5      6       7
## Extinct   5.00000  3.00000 10.00000  6.0000  3.0000  4.000  8.0000
## NExtinct 70.00000 64.00000 56.00000 45.0000 25.0000 16.000 35.0000
## pis       0.06667  0.04478  0.15152  0.1176  0.1071  0.200  0.1860
## phats     0.06017  0.07036  0.09854  0.1380  0.1595  0.162  0.1639
##                 8       9      10      11      12      13      14     15
## Extinct   3.00000  5.0000  6.0000  8.0000  2.0000  9.0000  5.0000 7.0000
## NExtinct 28.00000 23.0000 26.0000 22.0000 18.0000 22.0000 11.0000 8.0000
## pis       0.09677  0.1786  0.1875  0.2667  0.1000  0.2903  0.3125 0.4667
## phats     0.17125  0.1854  0.2052  0.2226  0.2516  0.2516  0.2603 0.2842
##                16      17     18
## Extinct    8.0000 13.0000 3.0000
## NExtinct  25.0000 27.0000 3.0000
## pis        0.2424  0.3250 0.5000
## phats      0.3019  0.3278 0.3998
```

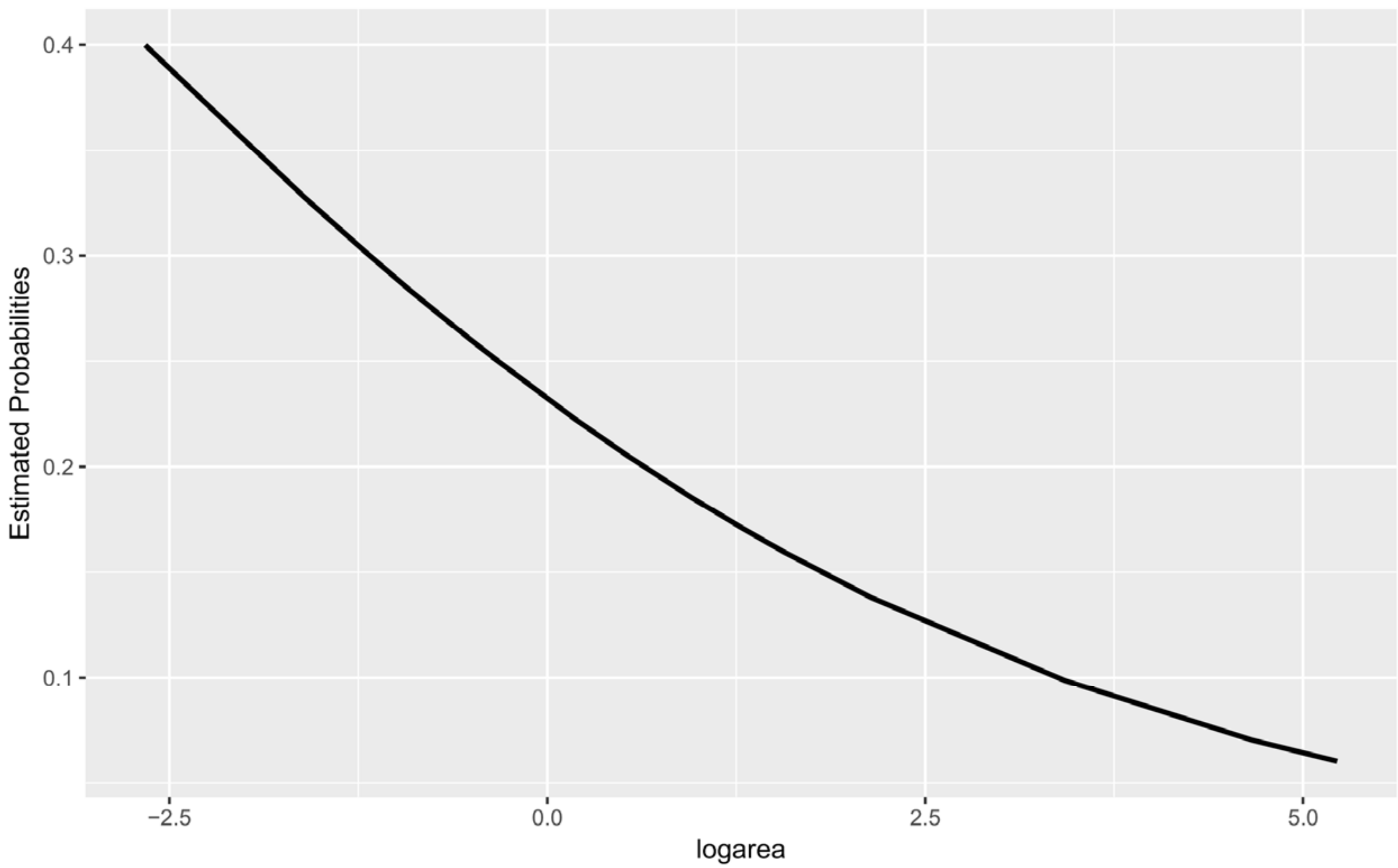
$\bar{\pi}_{s,i}$

$\hat{\pi}_{M,i}$

Case IV Effect Plot

```
ggplot(krunnit,aes(x=logarea, y=phats))+ylab("Estimated Probabilities")+
  geom_line(size=1)
```

$\hat{\pi}_{M,i}$



As logarea $\uparrow$ $\hat{\pi}_{M,i}$ $\downarrow$

$\Downarrow$

Larger preserves associated with lower extinction.

Case IV: Deviance test and Estimated Var-Cov of β

```r
anova(fitbl, test="Chisq")
```

$\text{logit}(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 \log(\text{Area})$
was adequate

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(Extinct, NExtinct)
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                         17       45.3
## log(Area)  1     33.3        16       12.1    8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
1-pchisq(12.062, 16)
```

```
## [1] 0.7397
```

0.74

Case IV: Estimated probabilities of extinction per island

```
phats<-predict.glm(fitbl, type="response") # estimated probability of extinctio
options(digits=4)
rbind(Extinct, NExtinct, pis,phats)
```

π_i $\hat{\pi}_i$

```
##                  1        2        3       4       5      6       7
## Extinct    5.00000  3.00000 10.00000  6.0000  3.0000  4.000  8.0000
## NExtinct  70.00000 64.00000 56.00000 45.0000 25.0000 16.000 35.0000
## pis        0.06667  0.04478  0.15152  0.1176  0.1071  0.200  0.1860
## phats      0.06017  0.07036  0.09854  0.1380  0.1595  0.162  0.1639
##                  8       9      10      11      12      13      14     15
## Extinct    3.00000  5.0000  6.0000  8.0000  2.0000  9.0000  5.0000 7.0000
## NExtinct  28.00000 23.0000 26.0000 22.0000 18.0000 22.0000 11.0000 8.0000
## pis        0.09677  0.1786  0.1875  0.2667  0.1000  0.2903  0.3125 0.4667
## phats      0.17125  0.1854  0.2052  0.2226  0.2516  0.2516  0.2603 0.2842
##                16      17     18
## Extinct    8.0000 13.0000 3.0000
## NExtinct  25.0000 27.0000 3.0000
## pis        0.2424  0.3250 0.5000
## phats      0.3019  0.3278 0.3998
```

Case IV Fit Statistics

```
AIC(fitbl)
```

```
## [1] 75.39
```

```
BIC(fitbl)
```

```
## [1] 77.17
```

Case IV Residuals

fitted Model type of residuals.

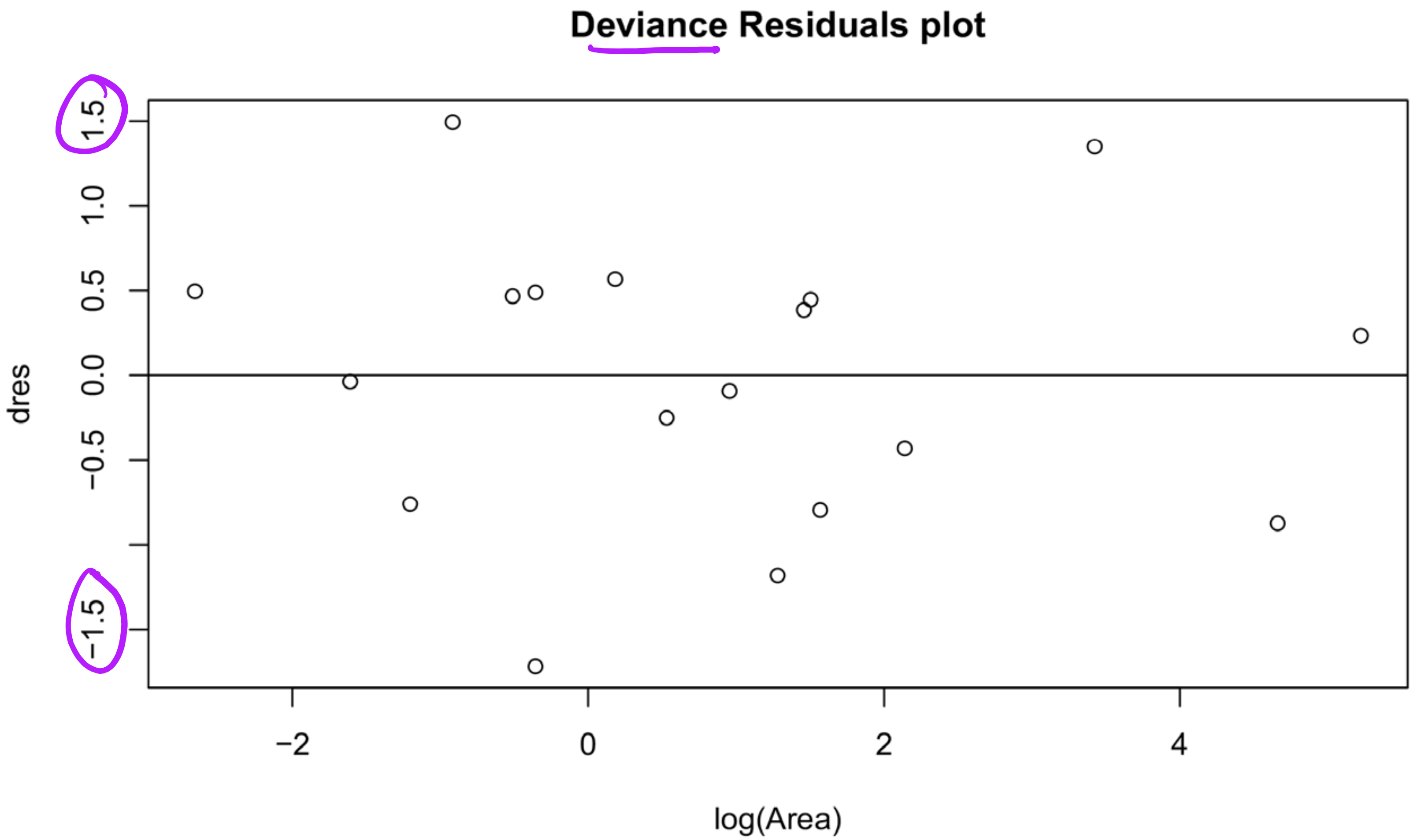
```
rres<-residuals(fitbl, type=c("response"))
pres<-residuals(fitbl, type=c("pearson"))
dres<-residuals(fitbl, type=c("deviance"))
rbind(pis,phats,rres, pres,dres)
```

$\pi_i - \hat{\pi}_i$

```
##              1        2       3        4        5       6       7        8
## pis   0.066667  0.04478 0.15152  0.11765  0.10714 0.20000 0.18605  0.09677
## phats 0.060173  0.07036 0.09854  0.13800  0.15946 0.16205 0.16389  0.17125
## rres  0.006493 -0.02558 0.05298 -0.02035 -0.05232 0.03795 0.02216 -0.07448
## pres  0.236464 -0.81883 1.44400 -0.42139 -0.75619 0.46058 0.39247 -1.10075
## dres  0.232656 -0.87369 1.34958 -0.43071 -0.79584 0.44746 0.38577 -1.18097
##               9       10      11      12      13     14     15       16
## pis    0.178571  0.18750 0.26667  0.1000 0.29032 0.3125 0.4667  0.24242
## phats  0.185415  0.20524 0.22264  0.2516 0.25158 0.2603 0.2842  0.30185
## rres  -0.006844 -0.01774 0.04403 -0.1516 0.03875 0.0522 0.1825 -0.05943
## pres  -0.093181 -0.24850 0.57969 -1.5622 0.49717 0.4759 1.5673 -0.74367
## dres  -0.093632 -0.25127 0.56727 -1.7173 0.48934 0.4666 1.4954 -0.75939
##              17     18
## pis    0.325000 0.5000
## phats  0.327828 0.3998
## rres  -0.002828 0.1002
## pres  -0.038101 0.5008
## dres  -0.038129 0.4957
```

Case IV Residuals Plot

```
plot(log(Area), dres, main="Deviance Residuals plot")
abline(h=0)
```



Case IV Residuals Plot

```r
plot(log(Area), pres, main="Pearson Residuals plot")
abline(h=0)
```