Oscar Huang
May 12, 2017

Statistics 516

Homework 05

Random Effects and the Delta Method

Student: Oscar Huang

Date: 2017/05/12

Oscar Huang
May 12, 2017

Estimation of the Crossing Point of Regression Curve

a. Code and output:

deltaMethod(m,"-b2/b3",parameterNames = c("b0","b1","b2","b3"))

	Estimate	SE	2.5 %	97.5 %	
-b2/b3	14.79925		6.562727	1.936545	27.66196

b. Discussion:

The estimated d_c as a function of β_j could be shown as below:

$$\beta_0+\beta_1 d_c=\beta_0+\beta_2+(\beta_1+\beta_3)d_c$$

$$d_c=-\beta_2/\beta_3$$

When using delta method to estimated d_c, the estimated value was 14.799, standard error: 6.563, and confident interval: 1.937 to 27.662. The estimated value (14.799) was agreed with Figure 1.

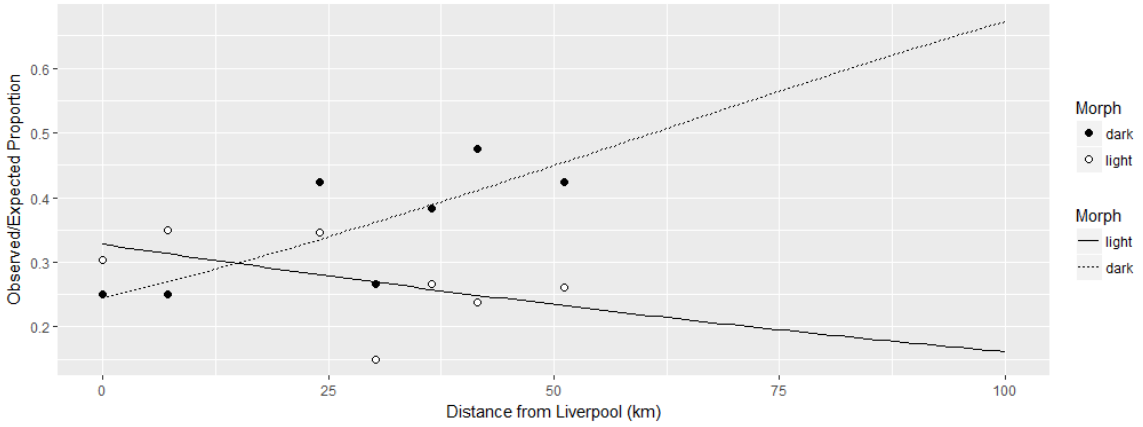

Figure 1 The expected proportion of the dark and light moths being removed with different distance from Liverpool.

Oscar Huang
May 12, 2017

Quadratic Regression and the Quadratic Formula

a. Code and output:

```
library(hcc)
data("tensile")
m<-nls(Y ~ b0 + b1 * x + b2 * x^2, data = tensile, start=c(b0=-10,b1=2,b2=-0.1))

deltaMethod(m,"(-b1-sqrt(b1^2-4*b2*(b0-30)))/(2*b2)",parameterNames =
c("b0","b1","b2"))
```

	Estimate	SE	2.5 %	97.5 %
(-b1 - sqrt(b1^2 - 4 * b2 * (b0 - 30)))/(2 * b2)	14.57329	0.3916506	13.80567	15.34091

```
deltaMethod(m,"(-b1+sqrt(b1^2-4*b2*(b0-30)))/(2*b2)",parameterNames =
c("b0","b1","b2"))
```

	Estimate	SE	2.5 %	97.5 %
(-b1 + sqrt(b1^2 - 4 * b2 * (b0 - 30)))/(2 * b2)	3.965862	0.1933071	3.586987	4.344736

```
p<-ggplot(tensile, aes(x=x, y=Y))+ geom_point()
p<- p + geom_line(aes(y=yhat), data = newdata)
p<- p + ylab("Tensile Strength")+ xlab("Percent Hardwood")
p<- p + geom_line(aes(y=30), linetype = 2)
plot(p)
```

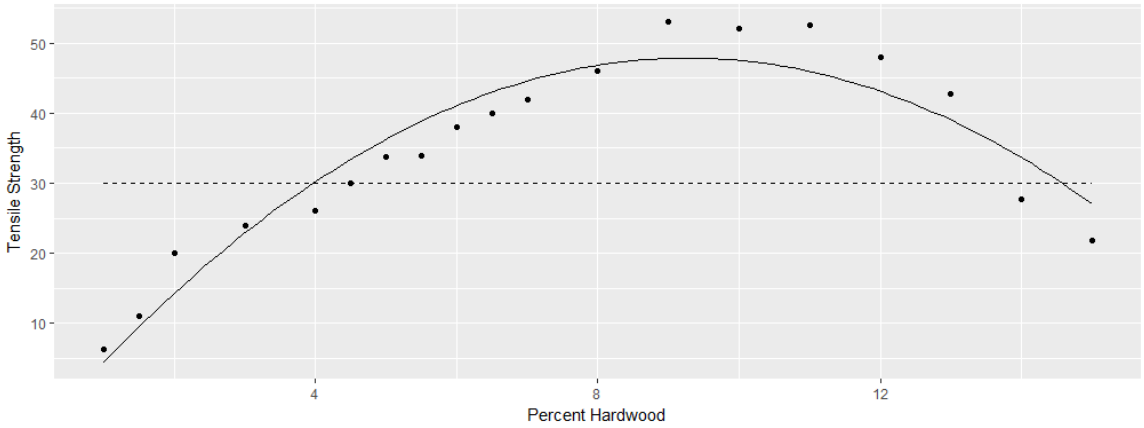

Figure 2 The expected tensile strength with different percent hardwood.

b. Discussion:

The estimated values percent hardwood that has tensile strength at 30 were:
$x_l = 3.966$, standard error: 0.1933, confident interval: 3.587 to 4.345;
$x_u = 14.573$, standard error: 0.3917, confident interval: 13.806 to 15.341;
Both estimated values agreed with Figure 2.

Oscar Huang
May 12, 2017

A Mixed Effects Model for the Snow Geese Data

a. Code and output:

```
m<-glmer(count~ observer*photo + (1|flock), data=snowgeese, family=
poisson(link=identity))
summary(m)$coefficients
```

```
                  Estimate Std. Error    z value     Pr(>|z|)
(Intercept)       -2.2509537 7.25766117 -0.3101486 7.564479e-01
observerobs2      -3.8027546 1.80286018 -2.1092898 3.491957e-02
photo              0.8198270 0.06116979 13.4024804 5.847564e-41
observerobs2:photo  0.3078387 0.02954585 10.4190159 2.030432e-25
```

```
summary(m)$varcor
```

```
 Groups Name        Std.Dev.
 flock  (Intercept) 33.164
```

b. Discussion:

The estimated parameters for the fixed effects: β_0 = -2.2510, β_1 = -3.8028, β_2 = 0.8198, and β_3 = 0.3078. The estimated variance of the γ_i: 33.164^2 = 1100

Oscar Huang
May 12, 2017

Orthodontic Measurements on Children Over Time

1. Estimated parameters:

a. Code and output:

```
m<-lmer(distance~Sex*age+(1+age|Subject), data = dental)
summary(m)$coefficients
summary(m)$varcor
```

```
             Estimate Std. Error    t value
(Intercept)   16.3406250 1.01852881 16.0433606
SexFemale      1.0321023 1.59572797  0.6467909
age            0.7843750 0.08599931  9.1207122
SexFemale:age -0.3048295 0.13473503 -2.2624372
```

```
summary(m)$varcor
```

```
 Groups   Name        Std.Dev. Corr
 Subject  (Intercept) 2.40547
          age         0.18034  -0.668
 Residual             1.31004
```

```
#plot
dental$yhat<-predict(m, re.form = NA)
p <- ggplot(dental, aes(x=age, y=distance))
p <- p + geom_line(aes(group= Subject, y=yhat, linetype=Sex))
p <- p + labs(x="Age(years)",y="Distance (mm)")
p <- p + theme_bw()
plot(p)
```

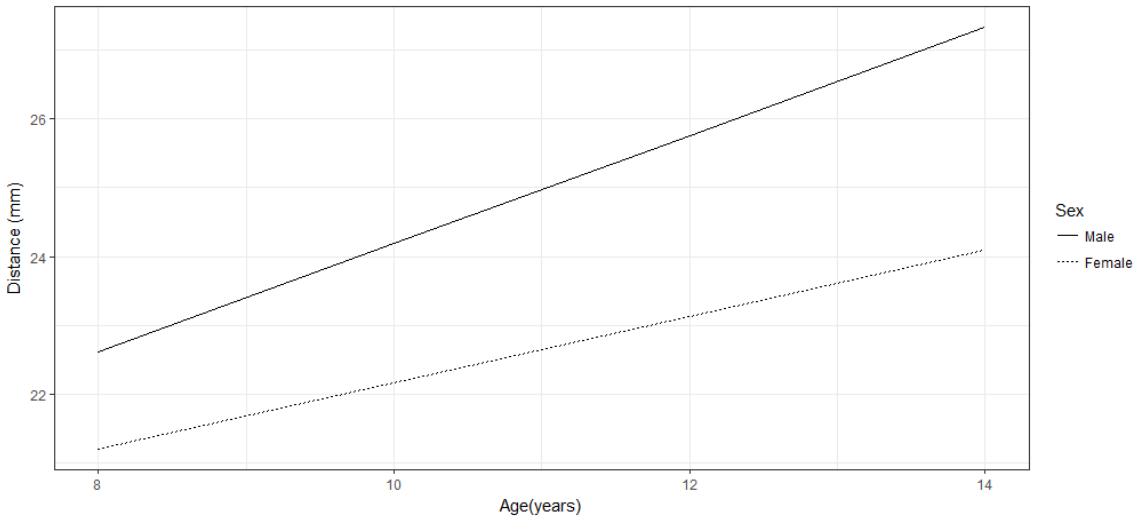

Figure 3 The expected distance changed with different ages and sexual.

b. Discussion:

The estimated parameters for the fixed effects are: β_0 =16.3406, β_1 = 1.0321, β_2 = 0.7844, and β_3 = -0.3483. The standard deviation of the random parameters are: γ_i: 2.4055 ,δ_i: 0.1803, and ε_{ij}: 1.3100

Oscar Huang
May 12, 2017

2. Slope of the Estimated parameters:

a. Code and output:

```
contrast(m,
        a=list(Sex=c("Male","Female"),age=11),
        b=list(Sex=c("Male","Female"),age=10),
        cnames = c("Male","Female"))
```

```
       estimate        se     lower     upper   tvalue  df       pvalue
Male   0.7843750 0.08599931 0.6158194 0.9529306 9.120712 Inf 7.463224e-20
Female 0.4795455 0.10371908 0.2762598 0.6828311 4.623503 Inf 3.773134e-06
```

b. Discussion:

The estimated slope of the line for an average male was 0.7844, and 0.4795 for female. The slope agreed with the Figure 3.

3. Expected distance difference between male and female with same age

a. Code and output:

```
contrast(m,
        a=list(Sex="Male",age=c(8,10,12,14)),
        b=list(Sex="Female",age=c(8,10,12,14)),
        cnames = c("8","10","12","14"))
```

```
   estimate       se      lower    upper   tvalue  df       pvalue
8  1.406534 0.8248721 -0.2101855 3.023254 1.705154 Inf 0.0881656685
10 2.016193 0.7596059  0.5273930 3.504993 2.654262 Inf 0.0079482025
12 2.625852 0.7866489  1.0840488 4.167656 3.338023 Inf 0.0008437665
14 3.235511 0.8976971  1.4760575 4.994965 3.604235 Inf 0.0003130734
```

b. Discussion:

The estimated distance difference between male and female at same age (male - female) were: 1.4065 for 8 years old, 2.0162 for 10 years old, 2.6259 for 12 years old, and 3.2355 for 14 years old child. The estimated differences agreed with the Figure 3.

Oscar Huang
May 12, 2017

Efficacy of Two Topical Cream Preparations

1. Logistic regression model

a. Code and output:

```
m.bio<-glm(cbind(favorable,unfavorable)~treatment, data= topical, family = binomial)
summary(m.bio)
```

```
Call:
glm(formula = cbind(favorable, unfavorable) ~ treatment, family = binomial,
    data = topical)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-3.093  -2.491  -0.913   1.594   4.145

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.7142     0.1780  -4.012 6.03e-05 ***
treatmentDrug   0.4040     0.2514   1.607    0.108
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 93.555  on 15  degrees of freedom
Residual deviance: 90.960  on 14  degrees of freedom
AIC: 133.35

Number of Fisher Scoring iterations: 4
```

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

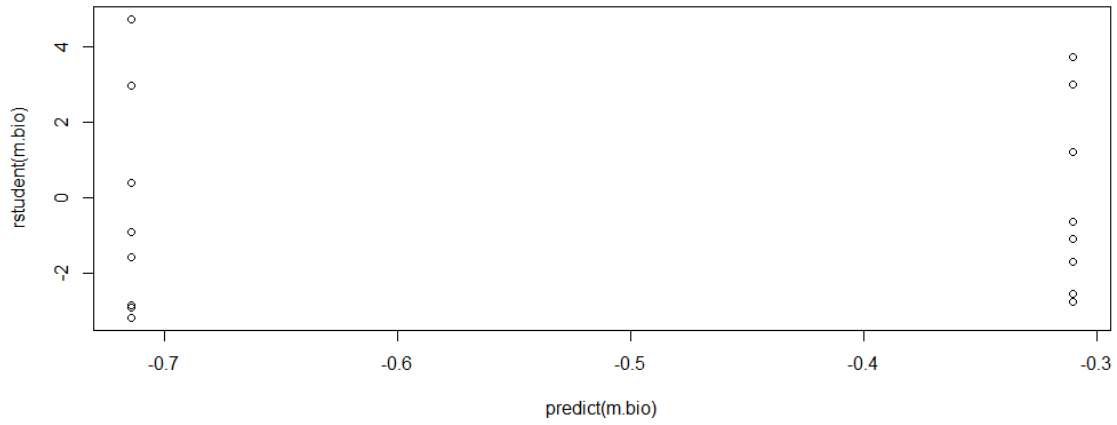

Figure 4 The residuals with different prediction value of the logistic model (without clinic variable)

```
m.quabio<-glm(cbind(favorable,unfavorable)~treatment, data= topical, family =
quasibinomial)
```

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

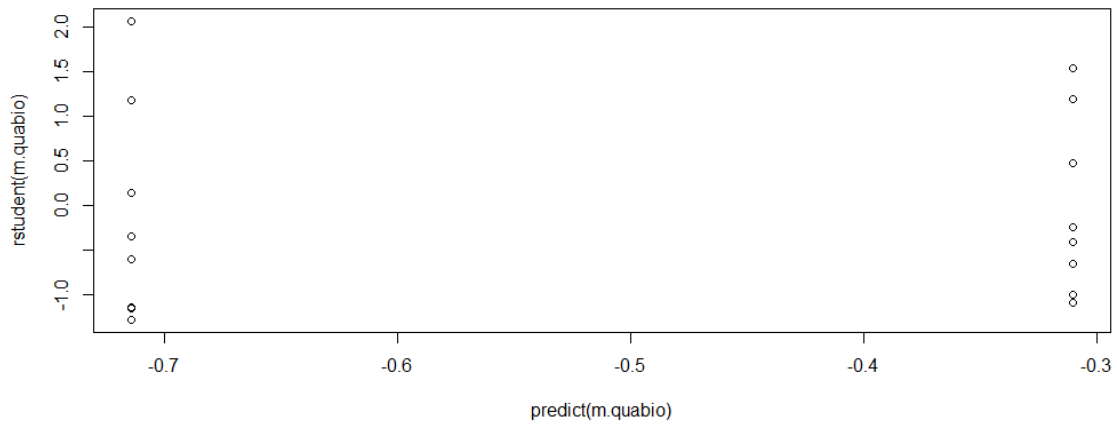

Figure 5 The residuals with different prediction value of the quasilogistic model (without clinic variable)

```
contrast(m.quabio,
      a=list(treatment="Drug"),
      b=list(treatment="Control"),tf =exp)
```

```
estimate       se     lower   upper    tvalue df    pvalue
1.497872 0.6066467 0.4077577 5.50234 0.6660312 14 0.5162163
```

b. Discussion:

When using binomial model without considering the effect of clinic variable, the residual variance (90.960) was much higher than the degree of freedom (14), and when plotting the Figure 4, there are some residuals higher than 2 or lower than -2. Both tests show the evidences of over dispersion.

When applying the quasi-binomial model, and plotting the Figure 5, all the residuals are within -2 and 2, which means the model had no overdispersion.

The odds ratio of “Drug” treatment group and “Control” group was 1.4979. The confident interval was 0.4078 to 5.5023, which include 1.00 (when the two treatment has no different), and the p-value was 0.5162 > 0.05. Both tests fail to reject the null hypothesis that there was no difference between the “Drug” treatment group and “Control” group. There was no clear difference between two groups

2. Including clinic variable as a fixed effect

a. Code and output:

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

```
Call:
```

Oscar Huang
May 12, 2017

```
glm(formula = cbind(favorable, unfavorable) ~ treatment + clinic,
    family = binomial, data = topical)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.87919  -0.77729  -0.00401   0.47293   0.92934

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.3220     0.3165  -4.177 2.95e-05 ***
treatmentDrug   0.7769     0.3067   2.533  0.01130 *
clinic2         2.0554     0.4201   4.893 9.94e-07 ***
clinic3         1.1529     0.4246   2.715  0.00662 **
clinic4        -1.4185     0.6636  -2.138  0.03255 *
clinic5        -0.5199     0.5338  -0.974  0.33007
clinic6        -2.1469     1.0614  -2.023  0.04310 *
clinic7        -0.7977     0.8149  -0.979  0.32764
clinic8         2.2079     0.7195   3.069  0.00215 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 93.5545  on 15  degrees of freedom
Residual deviance:  9.7463  on  7  degrees of freedom
AIC: 66.136

Number of Fisher Scoring iterations: 4
```

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

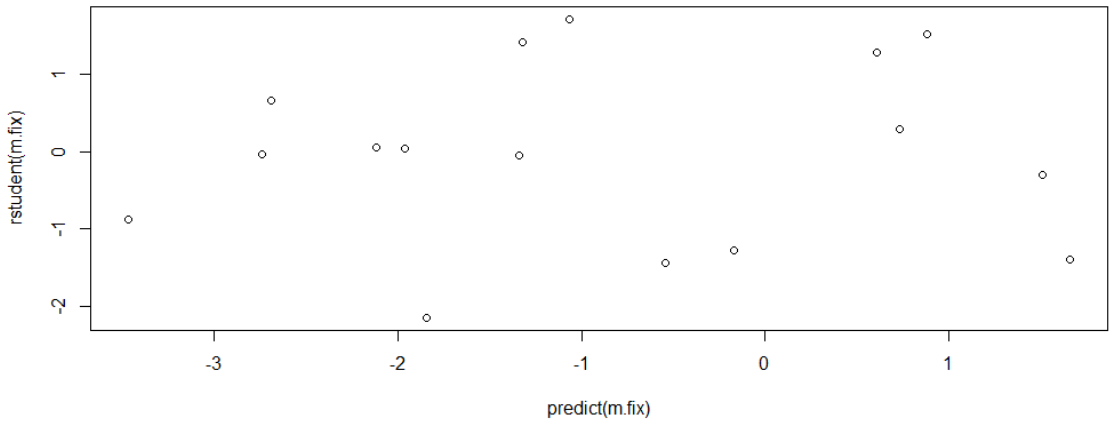

Figure 6 The residuals with different prediction value of the logistic model (include clinic variable as fixed effect)

```
contrast(m.fix,
      a=list(treatment="Drug", clinic = "1"),
      b=list(treatment="Control", clinic = "1"), tf=exp)
```

```
estimate      se    lower    upper   tvalue  df     pvalue
```

```
2.174764 0.306687 1.192229 3.967022 2.533268 Inf 0.01130045
```

b. Discussion:

When considering the clinic variances as a fixed effect in the binomial model and plotting the Figure 6, all the residuals are within -2 and 2, which means the model had no overdispersion.

The odds ratio of “Drug” treatment group and “Control” group was 2.1748. The confident interval was 1.1922 to 3.9670, which did not include 1.00 (when the two treatment has no different), and the p-value was 0.0113 < 0.05. Both tests rejected the null hypothesis that there was no difference between the “Drug” treatment group and “Control” group. There was significant difference between two groups

3. Including the clinic variable as a random effect

a. Code and output:

```
m.rand<-glmer(cbind(favorable,unfavorable)~treatment + (1|clinic), data= topical,
family = binomial)
contrast(m.fix,
      a=list(treatment="Drug", clinic= "1"),
      b=list(treatment="Control", clinic= "1"), tf=exp)
```

```
estimate      se    lower    upper   tvalue  df     pvalue
 2.174764 0.306687 1.192229 3.967022 2.533268 Inf 0.01130045
```

b. Discussion:

The odds ratio of “Drug” treatment group and “Control” group was 2.1748. The confident interval was 1.1922 to 3.9670, which did not include 1.00 (when the two treatment has no different), and the p-value was 0.0113 < 0.05. Both tests rejected the null hypothesis that there was no difference between the “Drug” treatment group and “Control” group. There was significant difference between two groups