Alyssa Semerdjian

Stat 510

26 Feb 2018

Homework 6

1.

**a)**

by hand:

|  |  |  |
| --- | --- | --- |
|  | Mean | variance |
| r=1, p=0.5 | = = = 1 | = = = 2 |
| r= 3, p=0.25 | = = = 9 | = = = 36 |

In R:

> x=rnbinom(1000,size=1,prob=.5)

> mean(x);var(x)

[1] 1.014

[1] 2.198002

> x=rnbinom(1000,size=3,prob=.25)

> mean(x);var(x)

[1] 8.977

[1] 36.72319

The hand calculated mean and variance are similar to the values calculated in R. The variance is greater than the mean in both cases.

**b)** = = 1/p >1

This is true when 0 < p < 1. Variance and mean are the same when p=1.

**c)**  r=3, p=0.5

p(x=1)= = = = = 0.1875

p(x≤2) = = = = = 0.1875

> dnbinom(1,size = 3,prob = .5)

[1] 0.1875

> dnbinom(2,size = 3,prob = .5)

[1] 0.1875

The answer is the same for both x values whether it’s calculated in R or by hand.

**2)**

**a)** The p-values in the model summaries are pretty similar (see below). This means that the overdispersion doesn’t affect the models much.

**b)** model 4 has the largest p-value for nonwhite, indicating that it is the most conservative.

> summary(model3)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.943029 0.107234 8.794 < 2e-16 \*\*\*

female 0.058407 0.037151 1.572 0.116

nonwhite 0.186667 0.043599 4.281 1.86e-05 \*\*\*

educate -0.063834 0.007015 -9.100 < 2e-16 \*\*\*

income 0.033232 0.006779 4.902 9.47e-07 \*\*\*

---

> summary(model4)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.943029 0.128110 7.361 2.69e-13 \*\*\*

female 0.058407 0.044384 1.316 0.188341

nonwhite 0.186667 0.052087 3.584 0.000347 \*\*\*

educate -0.063834 0.008380 -7.617 4.03e-14 \*\*\*

income 0.033232 0.008098 4.104 4.24e-05 \*\*\*

---

> summary(model5)

Call:

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.925852 0.130067 7.118 1.09e-12 \*\*\*

female 0.056289 0.044982 1.251 0.210794

nonwhite 0.185085 0.053889 3.435 0.000593 \*\*\*

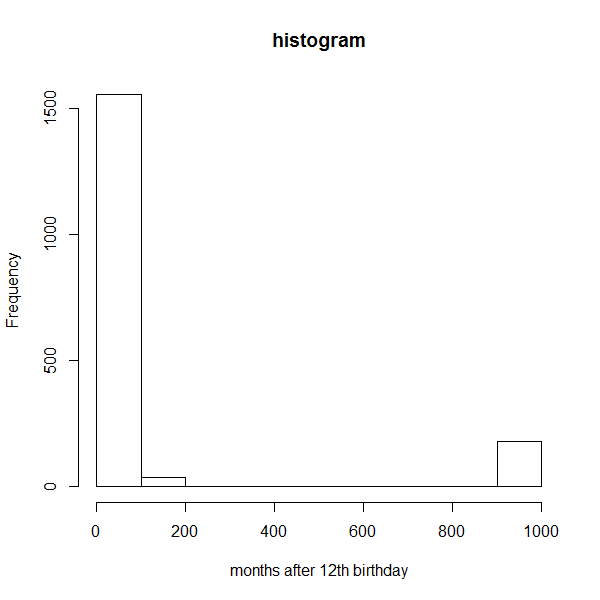
educate -0.062635 0.008530 -7.343 2.09e-13 \*\*\*

income 0.033448 0.008083 4.138 3.50e-05 \*\*\*

---

**3)**

**a)** The firstsex column has a lot of 999 values. The code 999 is often used to indicate that an event under scrutiny had not occurred by the end of the study. There are 177 cases that are right-censored.



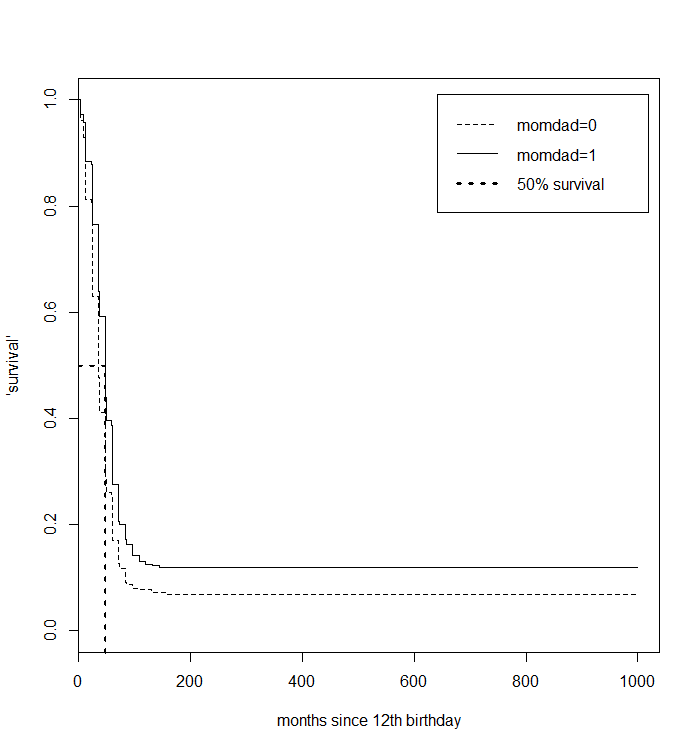
**b)** The mean for firstsex is about 138 while the median is 48 (meaning most of the subjects had sex after 48 months). These numbers are very different because of the high number of right censored cases.

> mean(first$firstsex);median(first$firstsex)

[1] 138.5456

[1] 48

**c)** The median “survival” time is 48 months with a survival value of 0.384.



**d)** The median survival time for males that do not live with their parents is 36, which is smaller than the median survival time for males that do live with their parents, which is 48. The survdiff output tells us that this difference is significant (p=1.82e-14).

> survdiff(firstsex1~momdad,data=1,rho=1)

Call:

survdiff(formula = firstsex1 ~ momdad, data = 1, rho = 1)

N Observed Expected (O-E)^2/E (O-E)^2/V

momdad=0 659 393 306 25.1 58.7

momdad=1 1108 539 627 12.3 58.7

Chisq= 58.7 on 1 degrees of freedom, p= 1.82e-14

**e)** Model 3 has the lowest AIC score and is therefore the ‘best’ model.

The coefficient for the momdad indicates that those who live with both mom and dad are expected to ‘survive’ about 1.36 times longer than those who don’t.

Each additional member of family size is associated with about 0.04% decrease (1-0.959) in the time until first sexual experience, adjusting for the effects of other variables in the model.

Parent’s education is not strongly associated with the time until first sexual experience (p = 0.645).

> AIC=AIC(model3,model4,model5,model6)

> AIC$dAIC = AIC$AIC - min(AIC$AIC)

> AIC=AIC[order(AIC$dAIC),]

> AIC

df AIC dAIC

model3 8 17218.77 0.0000

model5 8 18090.04 871.2688

model4 7 19050.40 1831.6261

model6 6 21331.60 4112.8316

> exp(coef(model3))

(Intercept) famsize pareduc lowincom relschol momdad white

39.8093409 0.9599486 0.9942644 1.0136514 1.1144056 1.3584854 1.4338169

> summary(model3)

Call:

survreg(formula = firstsex1 ~ famsize + pareduc + lowincom +

relschol + momdad + white, data = data1, dist = "lognormal")

Value Std. Error z p

(Intercept) 3.68410 0.1850 19.917 2.88e-88

famsize -0.04088 0.0164 -2.493 1.27e-02

pareduc -0.00575 0.0125 -0.461 6.45e-01

lowincom 0.01356 0.0727 0.186 8.52e-01

relschol 0.10832 0.0891 1.216 2.24e-01

momdad 0.30637 0.0686 4.467 7.95e-06

white 0.36034 0.0671 5.372 7.78e-08

Log(scale) 0.28907 0.0184 15.701 1.49e-55