

Logistic Regression Analysis for Breastfeeding of Nepalese Children

Scientific Question: determine whether and how bf behavior varies with age and/or sex of child.

Data: Nepal Data, $Y_{ij}=I(\text{breastfeeding}_{ij})$

Use the equally-spaced times (same method as in the Midterm);

```
. egen newid=group(id)
. summ newid
```

Variable	Obs	Mean	Std. Dev.	Min	Max
newid	1000	100.5	57.76319	1	200

```
. drop id
. ren newid id
. sort id age
. by id: gen obsno=_n
```

```
. tab obsno
```

obsno	Freq.	Percent	Cum.
1	200	20.00	20.00
2	200	20.00	40.00
3	200	20.00	60.00
4	200	20.00	80.00
5	200	20.00	100.00
Total	1000	100.00	

```
. tsset id obsno
```

```
. tsset id obsno
panel variable: id, 1 to 200
time variable: obsno, 1 to 5
```

. Generate binary breastfeeding variable (ever vs. never)

```
. xttab bf
```

bf	Overall		Between		Within
	Freq.	Percent	Freq.	Percent	Percent
0	564	59.56	142	71.36	82.46
1	151	15.95	77	38.69	40.81
2	232	24.50	86	43.22	56.45
Total	947	100.00	305	153.27	64.61 (n = 199)

```
. label list
```

```

bf:
    0    none
    1    <10 times/day
    2    >=10 times/day
sex:
    1    male
    2    female
    3

```

```

. gen bfbinsm=1*(bf==1|bf==2)
. gen bfbinsf=1*(bf==1|bf==2)

. tab bf bfbinsm

```

Current breastfeeding level	bfbinsm		Total
	0	1	
none	564	0	564
<10 times/day	0	151	151
>=10 times/day	0	232	232
Total	564	383	

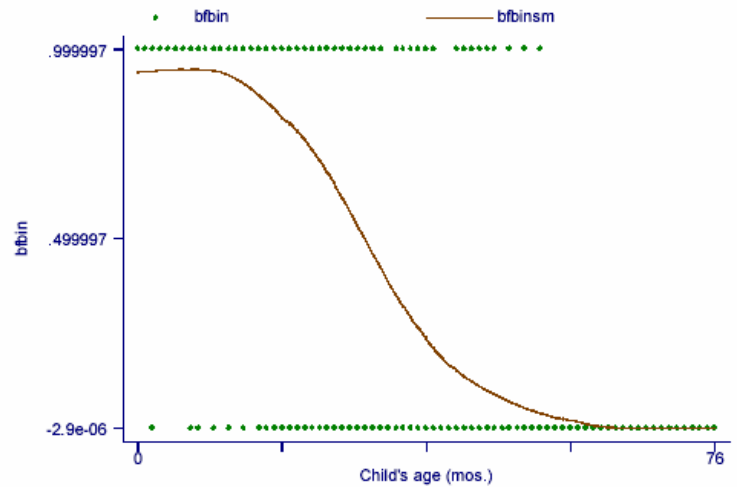
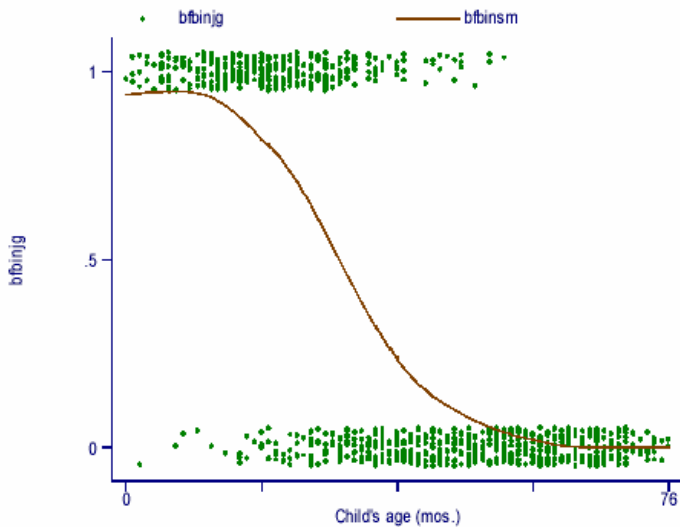
Explore marginal mean model wrt age

```

ksm bfbinsm age, lowess bw(.4) xlab ylab(0(.2)1) nograph gen(bfbinsm)
sort age
gr bfbinsm bfbinsm age, c(.L) s(oi) ylab

gen bfbinsf=bfbinsm+.1*(uniform()-0.5)
sort age
gr bfbinsf bfbinsm age, c(.L) s(oi) ylab

```



It appears that the logistic function will be appropriate for modeling the effects of child's age on prevalence of bf.

Review of Logistic regression in STATA without taking into account the correlation

```
. gen agesex=age*sex
```

```
. logit bfbn age sex agesex
```

```
Iteration 0: log likelihood = -665.51369
Iteration 1: log likelihood = -372.84807
Iteration 2: log likelihood = -327.36492
Iteration 3: log likelihood = -318.37149
Iteration 4: log likelihood = -317.73989
Iteration 5: log likelihood = -317.73456
Iteration 6: log likelihood = -317.73456
```

```
Logit estimates                               Number of obs =      1000
LR chi2(3) =      695.56
Prob > chi2 =      0.0000
Pseudo R2 =      0.5226
Log likelihood = -317.73456
```

bfbn	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.1313039	.0301624	-4.35	0.000	-.1904212	-.0721865
sex	.3728255	.6397586	0.58	0.560	-.8810782	1.626729
agesex	-.0196382	.0204261	-0.96	0.336	-.0596727	.0203963
_cons	4.215441	.9741133	4.33	0.000	2.306214	6.124668

Report Odds Ratio instead of log Odds Ratio as coefficients:

```
. logistic bfbn age sex agesex
```

```
Logit estimates                               Number of obs =      1000
LR chi2(3) =      695.56
Prob > chi2 =      0.0000
Pseudo R2 =      0.5226
Log likelihood = -317.73456
```

bfbn	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.8769513	.026451	-4.35	0.000	.8266109	.9303573
sex	1.451831	.9288213	0.58	0.560	.4143359	5.087209
agesex	.9805533	.0200289	-0.96	0.336	.9420728	1.020606

```
. predict prob
(option p assumed; Pr(bfbn))
```

```
. *** sensitivity and specificity from 2 by 2 table ***
```

```
. gen c = .5
. gen bfhat = 1 if prob > c
(608 missing values generated)
. replace bfhat = 0 if bfhat == .
(608 real changes made)
. tab bfbn bfhat
```

bfbn	bfhat		Total
	0	1	
0	544	73	617
1	64	319	383
Total	608	392	1000

. *** test for statistical significance ***

. test agesex

(1) agesex = 0.0

chi2(1) = 0.92
Prob > chi2 = 0.3363

. test sex agesex

(1) sex = 0.0

(2) agesex = 0.0

chi2(2) = 1.98
Prob > chi2 = 0.3710

Fit GEE model w/ AR(1)

. xtgee bfbfn sex age agesex, f(bin) link(logit) corr(ar1)

Iteration 1: tolerance = .1830486
Iteration 2: tolerance = .01876749
Iteration 3: tolerance = .00125261
Iteration 4: tolerance = .00004341
Iteration 5: tolerance = 2.666e-06
Iteration 6: tolerance = 1.167e-07

GEE population-averaged model
Group and time vars: id obsno Number of obs = 1000
Link: logit Number of groups = 200
Family: binomial Obs per group: min = 5
Correlation: AR(1) avg = 5.0
max = 5
Wald chi2(3) = 112.36
Scale parameter: 1 Prob > chi2 = 0.0000

bfbfn	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sex	.1411955	.9366279	0.15	0.880	-1.694562	1.976953
age	-.1409907	.0449762	-3.13	0.002	-.2291423	-.052839
agesex	-.0103829	.0297624	-0.35	0.727	-.0687161	.0479504
_cons	4.334149	1.447124	3.00	0.003	1.497837	7.17046

. test sex agesex

(1) sex = 0.0

(2) agesex = 0.0

chi2(2) = 0.39
Prob > chi2 = 0.8218

Remove interaction:

. xtgee bfbfn age, f(bin) link(logit) corr(ar1)

Iteration 1: tolerance = .03838573
Iteration 2: tolerance = .0042243
Iteration 3: tolerance = .0001424
Iteration 4: tolerance = 6.738e-06
Iteration 5: tolerance = 2.780e-07

GEE population-averaged model
Group and time vars: id obsno Number of obs = 1000
Link: logit Number of groups = 200
Family: binomial Obs per group: min = 5
avg = 5.0

```

Correlation:                AR(1)                max =          5
                               Wald chi2(1)      =    112.34
Scale parameter:            1          Prob > chi2    =    0.0000

```

bfbin	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.1554856	.0146696	-10.60	0.000	-.1842375 -.1267336
_cons	4.525001	.4653168	9.72	0.000	3.612997 5.437005

```
. xtcorr, compact
```

```

Error structure: AR(1)
Estimated within-id autocorrelation: 0.6518

```

```
. xtcorr
```

```
Estimated within-id correlation matrix R:
```

```

      c1      c2      c3      c4      c5
r1  1.0000
r2  0.6518  1.0000
r3  0.4248  0.6518  1.0000
r4  0.2769  0.4248  0.6518  1.0000
r5  0.1805  0.2769  0.4248  0.6518  1.0000

```

Interpretation and conclusion: Sex of child is not significant predictor of bf behavior, but age is. The estimated odds ratio for bf for two children who differ by one year in age is 0.86, 95%CI: [0.83, 0.88]

Assess the model fit

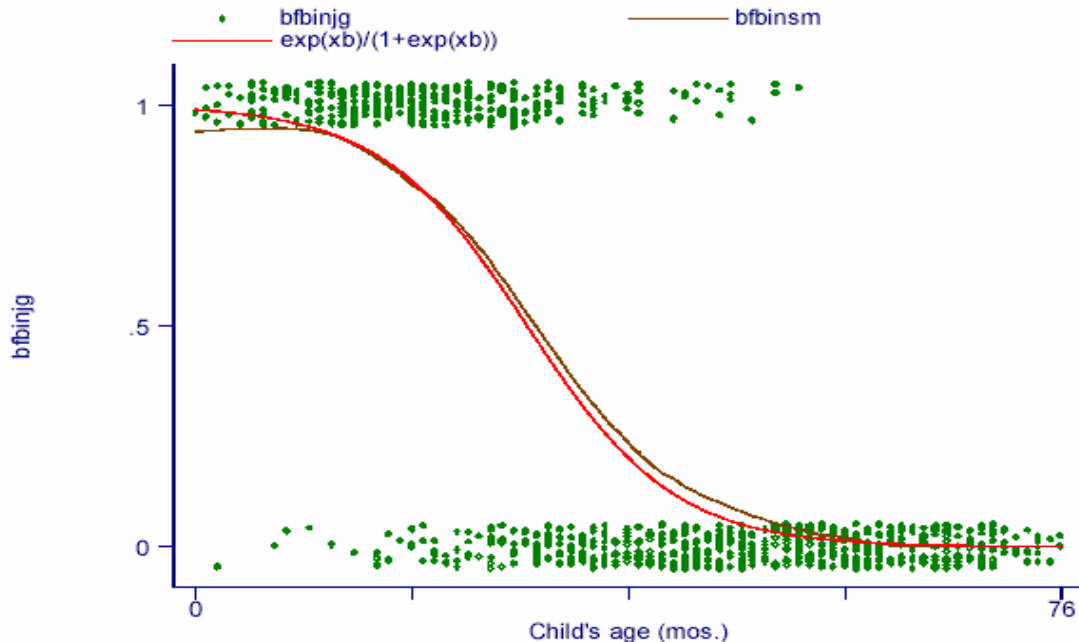
```

. predict bffitar1
(option mu assumed; exp(xb)/(1+exp(xb)))

. sort age

. gr bfbinjg bfbinsm bffitar1 age,c(.11) s(oi) ylab

```



Fit GEE with uniform corr model

```
. xtgee bfbin sex age agesex, f(bin) link(logit) corr(exc) nolog
```

```
GEE population-averaged model
Group variable:          id      Number of obs      =      1000
Link:                    logit   Number of groups =      200
Family:                  binomial Obs per group: min =      5
Correlation:             exchangeable avg =      5.0
Scale parameter:        1      Wald chi2(3)      =      152.54
                          Prob > chi2      =      0.0000
```

bfbin	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sex	-.7438961	.7987586	-0.93	0.352	-2.309434	.8216421
age	-.1928599	.0405235	-4.76	0.000	-.2722845	-.1134352
agesex	.0248351	.0252077	0.99	0.325	-.0245711	.0742413
_cons	5.468235	1.284332	4.26	0.000	2.95099	7.985481

```
. xtgee bfbin age , f(bin) link(logit) corr(exc) nolog
```

```
GEE population-averaged model
Group variable:          id      Number of obs      =      1000
Link:                    logit   Number of groups =      200
Family:                  binomial Obs per group: min =      5
Correlation:             exchangeable avg =      5.0
Scale parameter:        1      Wald chi2(1)      =      150.49
                          Prob > chi2      =      0.0000
```

bfbin	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-.155618	.0126853	-12.27	0.000	-.1804806	-.1307553
_cons	4.367159	.4021635	10.86	0.000	3.578933	5.155385

**** display exponential of coefficient ****

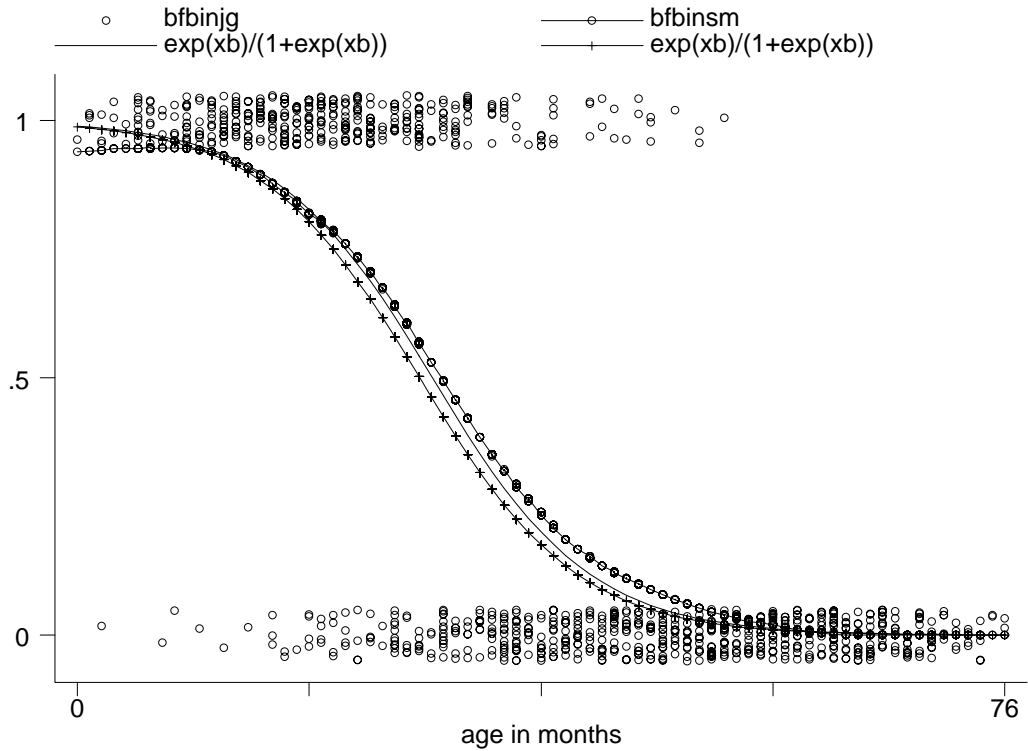
```
. xtgee bfbin age , f(bin) link(logit) corr(exc) nolog eform
```

```
GEE population-averaged model
Group variable:          id      Number of obs      =      1000
Link:                    logit   Number of groups =      200
Family:                  binomial Obs per group: min =      5
Correlation:             exchangeable avg =      5.0
Scale parameter:        1      Wald chi2(1)      =      150.49
                          Prob > chi2      =      0.0000
```

bfbin	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.8558861	.0108571	-12.27	0.000	.8348689	.8774324

```
. predict bffitexc
(option mu assumed; exp(xb)/(1+exp(xb)))
```

```
. gr bfbinjg bfbinsm bffitar1 bffitexc age,c(.111) s(oop) ylab
```



. Use the unstructured option to choose a correlation structure

```
. xtgee bfbins age , f(bin) link(logit) corr(unst) nolog
```

```
GEE population-averaged model
Group and time vars:      id obsno      Number of obs      =      1000
Link:                     logit        Number of groups   =      200
Family:                   binomial     Obs per group: min =      5
Correlation:              unstructured avg =              5.0
Scale parameter:         1              Wald chi2(1)       =      122.86
                          Prob > chi2   =      0.0000
```

bfbins	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.1552873	.01401	-11.08	0.000	-.1827465 - .1278282
_cons	4.434984	.443211	10.01	0.000	3.566307 5.303662

```
. xtcorr
```

Estimated within-id correlation matrix R:

```

      c1    c2    c3    c4    c5
r1  1.0000
r2  0.4902  1.0000
r3  0.4747  0.7397  1.0000
r4  0.4191  0.6107  0.6662  1.0000
r5  0.3440  0.5264  0.5516  0.7136  1.0000
```

The unstructured std. err. estimate and the correlation matrix also suggests an ar(1) model for the correlation structure.