

Applied Regression II Final - Part One

Nick Williams

Data import and preparation

```
proc import out = depression
  datafile = "~\final\regression_II_final\data.csv"
  dbms = csv replace;
  getnames = yes;
run;

data depression;
  set depression;
  rename PARDEP = parent_dep
         DSMDEPHR = child_dep
         PTSEX = child_sex
         PTAGE = child_age
         BEDEPON = age_child_dep
         DSMSUBHR = sub_abuse_child
         BESUBON = age_sub_child
         SESCLASS = ses_parent
         MSPARENT = mar_stat_parent;
run;
```

Section One

I defined the start time as the time of birth for all children. The end time was defined as either the age of onset of depression in children that were classified as ever having depression or as the age of a child at the interview for children that classified as never having depression. To hold this survival time, I created a variable called `follow_time`:

```
data depression;
  set depression;

  if child_dep = 1 then
    follow_time = age_child_dep;
  else if child_dep = 0
    then follow_time = child_age;
run;
```

Section Two

```
proc means data = depression median clm maxdec = 2;  
    var age_child_dep;  
    class parent_dep;  
    where child_dep = 1;  
run;
```

Among children that developed depression but did not have a parent with depression, the median age of onset of depression was 17. Among children that developed depression but did have a parent with depression, the median age of onset of depression was 13. We are 95% confident that the average age of onset of depression among children without parental depression is between 15.47 and 18.08. In addition, we are 95% confident that the average age of onset of depression among children with parental depression is between 11.51 and 13.90.

Section Three

Descriptive statistics table:

Table 1: Offspring characteristics stratified by parental depression status

Covariate	Parent depression status	
	Never depressed (N = 95)	Ever depressed (N = 125)
Child sex - n(%)		
Male	42 (44.21)	63 (50.40)
Female	53 (55.79)	62 (49.60)
Child depression status - n(%)		
Never depressed	73 (76.84)	78 (62.40)
Ever depressed	22 (23.16)	47 (37.60)
Child substance abuse - n(%)		
No substance abuse	88 (92.63)	104 (83.20)
Substance abuse	7 (7.370)	21 (16.80)
Parent SES class - n(%)		
1	1 (1.050)	16 (13.01)
2	19 (20.00)	19 (15.45)
3	16 (16.84)	31 (25.20)
4	47 (49.47)	49 (39.84)
5	12 (12.63)	8 (6.500)
Parent marital status - n(%)		
Married w/Spouse	84 (88.42)	90 (72.00)
Separated/Divorced	11 (11.58)	35 (28.00)
Age covariates - mean(sd)		
Average age at interview	16.78 (4.110)	16.68 (5.180)
Average age of depression onset	16.77 (2.940)	12.7 (4.070)
Average age of substance abuse	8.43 (9.000)	8.9 (8.180)

Section Four

```
proc lifetest data = depression method = km conftype = loglog stderr plots = survival(cl);  
    strata parent_dep;  
    time follow_time * child_dep(0);  
run;
```

- (i) $H_0 : S_1(t) = S_2(t)$ for all $t \leq \tau$
 $H_A : S_1(t) \neq S_2(t)$ for some $t \leq \tau$
- (ii) $Q = 7.6876 \sim \chi_1^2$
 $p = P(\chi_1^2 \geq 7.6876) = 0.0056$
- (iii) $0.0056 < 0.05 \rightarrow$ reject the null hypothesis
- (iv) At the 5% significance level, there is sufficient evidence to claim that age of onset of depression in children differs between children with parental history of depression and children without parental history of depression.