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

Using a Kaplan-Meier estimator, we estimate that 50% children who do not have a depressed parent or a parent with a history of depression will develop depression by the age of 23. We are 95% confident that this estimate could be as low as 19. In addition, we estimate that 50% of children with a depressed parent or a parent with a history of depression will develop depression by the age of 20. We are 95% confident that this estimate could be as low as 17.

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	.	LOGLOG	23.0000	.
50	23.0000	LOGLOG	19.0000	.
25	18.0000	LOGLOG	17.0000	19.0000

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	.	LOGLOG	.	.
50	20.0000	LOGLOG	17.0000	.
25	13.0000	LOGLOG	12.0000	15.0000

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) $Q = 7.6876 \sim \chi^2_1$
$p = P(\chi^2_1 \geq 7.6876) = 0.0056$

(iii) $0.0056 < 0.05 \rightarrow$ reject the null hypothesis

(iv) At the 5% signficance 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.

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	7.6876	1	0.0056
Wilcoxon	15.1936	1	<.0001
-2Log(LR)	5.7178	1	0.0168