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;
```

1. 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;
```

2.

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

3.

4.

```
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% 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.