Statistics 641, Fall 2009
Homework #3
Answers

1. Suppose all subjects in a trial are followed for 1 year, and at the end of that time they either survive disease free (DFS), survive but experience a recurrence of disease, or die. Subjects are randomly assigned either treatment 1 (control) or treatment 2 (experimental). We observe the following:

	DFS	Recurrence	Dead
Treatment 1	33	17	41
Treatment 2	44	18	25

   (a) Compute (by hand) the test statistic for the Wilcoxon rank-sum test assuming that responses are ordered as shown.

   The ranks for the DFS subjects range from 1 to 77, for Recurrence from 78 to 112, and the Dead subjects from 113 to 178. Thus the mean ranks in the three groups are (1+77)/2=39, (78+112)/2=95, and (113+178)/2=145.5. The rank sum for Treatment 2 is $T_2 = 44 \times 39 + 18 \times 95 + 25 \times 145.5 = 7063.5$. The expected value of T_2 is $87 \times (178 + 1)/2 = 7785.5$, so $U = 7063.5 - 7785.5 = -723$. The variance is

$$\frac{87 \times 91}{178 - 1}\left(\frac{1}{178}(77 \times 39^2 + 35 * 975^2 + 66 \times 145.5^2 - \frac{(1 + 178)^2}{4}\right) = 101620.8$$

   The test statistic is $723^2/101620.8 = 5.14$.

   (b) Compute (by hand) the test statistic for the Mann-Whitney U test.

   Compute the U-statistic $U = \sum_j \sum_j I(Y_{i1} > Y_{j2}) + I(Y_{i1} = Y_{j2})/2 - 1/2$. using the table:

		DFS	Recurrence	Death
	$n_{j2} \backslash n_{i1}$:	33	17	41
DFS	44	.5	1	1
Recurrence	18	0	.5	1
Death	25	0	0	.5

   $U = 17 \times 44 + 41 \times 44 + 41 \times 18 + .5 * (33 \times 44 + 17 \times 18 + 41 * 25) - 87 \times 91/2 = 723$. The variance is as above, and the test statistic is identical.

   (c) What do these results suggest regarding the effect of treatment.

   Lower average rank in group 2 suggests that treatment group 2 has better outcomes than treatment group 1.

2. The data file "data3a.csv" (in csv format, comma delimited) contains columns

   - `x0`: baseline value of response variable
   - `x1`: value of response variable at first follow-up time
   - `x2`: value of response variable at second follow-up time
   - `z`: treatment variable (0,1)

   Assume that the responses are normally distributed.

   (a) For the first follow-up response (`x1`) test the null hypothesis that there is no difference by treatment by

      i. ignoring baseline

Fit model for `x1`with just `z`:

```
> summary(lm(x1~z, data=D))
[snip]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   60.914      1.809  33.677   <2e-16 ***
z              3.886      2.515   1.545    0.128
[snip]
```

The mean difference is 3.886 with SE 2.515, and t-statistic 1.545.

      ii. using change from baseline (`x1-x0`)

```
> summary(lm(x1-x0~z, data=D))
[snip]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.2679     0.9919  -0.270   0.7881
z             3.5745     1.3792   2.592   0.0122 *
[snip]
```

The mean difference is 3.5745 with SE 1.3892 and t-statistic 2.592.

      iii. fitting regression model $x_1 = \alpha_0 + \alpha_1 x_0 + \beta z + \epsilon$.

```
> summary(lm(x1~z+x0, data=D))
[snip]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.7465     3.6149   3.803 0.000361 ***
z             3.6458     1.2252   2.976 0.004337 **
x0            0.7709     0.0573  13.454  < 2e-16 ***
[snip]
```

The mean difference is 3.6458 with SE 1.2252 and t-statistic 2.976.

Why do the conclusions differ from these three analysis?

In (c) the coefficient for `x0` is .771, suggesting (assuming equal variances for `x0` and `x1`) that the correlation is greater than 1, so the change from baseline should have smaller variance than the follow-up value alone. This is borne out in the differences between (a) and (b). Since this coefficient is not too close to one, we expect that the regression model in (c) should have smaller variance than either (a) or (b), and again this is borne out in the results.

(b) Repeat each of (i), (ii), and (iii) above for the response at the second follow-up time (`x2`).

```
> summary(lm(x2~z, data=D))
[snip]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   61.618      2.063  29.868   <2e-16 ***
z              6.872      2.868   2.396   0.0200 *
[snip]

> summary(lm(x2-x0~z, data=D))
[snip]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4357     2.6931   0.162   0.8721
z             6.5610     3.7446   1.752   0.0852 .
[snip]
> summary(lm(x2~z+x0, data=D))
[snip]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  52.7430     8.4511   6.241 6.55e-08 ***
z             6.8270     2.8644   2.383   0.0206 *
x0            0.1451     0.1340   1.083   0.2836
[snip]
```

In the third analysis, the coefficient for `x0` is small, suggesting that there is much less correlation between `x2` and `x0` than between `x1` and `x0`. Therefore, we expect that the change from baseline will be much less efficient than ignoring baseline altogether. Again this is borne out in the results. In this case the third analysis gives essentially the same result as the first.

(c) Comment on the differences between (a) and (b).

Change from baseline beats observed follow-up value alone when correlation between baseline and follow-up is high, and loses when correlation is low. In either case, the regression model is at least as good as the others and should always be preferred.

3. Suppose that a hypothetical trial is conducted whose primary outcome is time to a non-fatal event (there are no deaths during the study, so all subjects can continue to be followed beyond the occurrence of the event). The treatment is intended to be given during the entire study, however, some subjects discontinue their assigned treatment prior to the end of follow-up. Suppose further that we are able to follow all subjects until their planned administrative censoring time so that we can observe events occurring after treatment discontinuation. We also observe two baseline variables that are associated with outcomes.

   The data file "data3b.csv" (in csv format, comma delimited) contains the following variables:

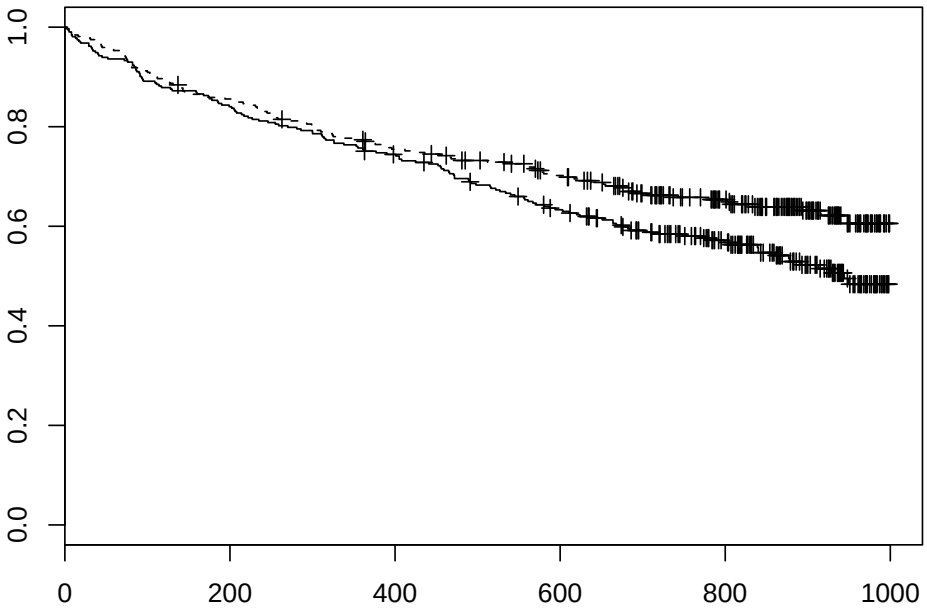
   - `z`: Treatment variable (0=control, 1=experimental)
   - `var1`, `var2`: Baseline covariates associated with risk
   - `time`: Time to the event of interest
   - `event`: Indicator of the event of interest (0=censored, 1=event)
   - `time.ot`: Time to the event of interest or end of treatment
   - `event.ot`: Indicator that the event of interest happened on treatment (0=censored, 1=event)
   - `eot.time`: Time of treatment termination or end of follow-up
   - `off.trt`: Indicator of treatment termination (0=censored, 1=treatment terminated)

   Note that the variable `time.ot` is the earliest of `time` and `eot.time` and that `event.ot` = `event` unless `time`>`eot.time`.

   (a) Conduct the *Intention-to-treat* analysis of the primary outcome:

      i. Plot Kaplan-Meier survival curves,

```
> km <- survfit(Surv(time, event)~z, data=surv)
> plot(km, lty=1:2)  ## experimental treatment is dashed line
```



ii. Conduct log-rank test,

```
>  survdiff(Surv(time, event)~z, data=surv)
Call:
survdiff(formula = Surv(time, event) ~ z, data = surv)

      N Observed Expected (O-E)^2/E (O-E)^2/V
z=0 313      144      126      2.64      5.18
z=1 319      113      131      2.53      5.18

 Chisq= 5.2  on 1 degrees of freedom, p= 0.0228
```

iii. Compute hazard ratio for the experimental versus the control.

```
>  coxph(Surv(time, event)~z, data=surv)
Call:
coxph(formula = Surv(time, event) ~ z, data = DD)


    coef exp(coef) se(coef)     z     p
z -0.285     0.752    0.126 -2.27 0.023

Likelihood ratio test=5.19  on 1 df, p=0.0227  n= 632 Call:
```

iv. Compute hazard ratio for the experimental versus the control adjusted for the baseline variable `var1`.

```
>  coxph(Surv(time, event)~z + var1, data=surv)
Call:
coxph(formula = Surv(time, event) ~ z + var1, data = surv)


       coef exp(coef) se(coef)     z       p
z    -0.349     0.705    0.126 -2.77 5.6e-03
var1  0.367     1.444    0.046  7.97 1.6e-15

Likelihood ratio test=74.2  on 2 df, p=1.11e-16  n= 632
```

What do you conclude regarding the effect of

i. treatment

The experimental treatment reduces the risk of the event by approximately 30% (the adjusted and unadjusted analyses agree closely).

ii. `var1`

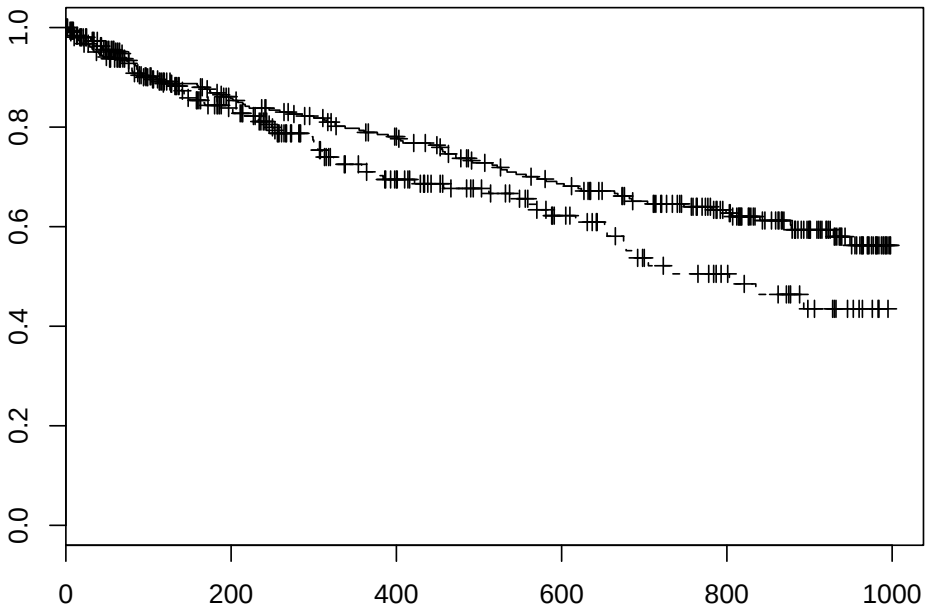
> Each unit increase in `var1` appears to increase risk of the event by about 44% (after accounting for treatment).

on the risk of a primary outcome event?

(b) Conduct the "on-treatment" analysis of the primary outcome (i.e., censor failure times at the time subjects discontinue treatment).

i. Plot Kaplan-Meier survival curves,

```
> km.ot <- survfit(Surv(time.ot, event.ot)~z, data=surv)
> plot(km.ot, lty=1:2)  ## experimental treatment is dashed line
```



ii. Conduct log-rank test,

```
> survdiff(Surv(time.ot, event.ot)~z, data=surv)
Call:
survdiff(formula = Surv(time.ot, event.ot) ~ z, data = surv)

      N Observed Expected (O-E)^2/E (O-E)^2/V
z=0 313       98    110.5      1.42      3.91
z=1 319       80     67.5      2.33      3.91

 Chisq= 3.9  on 1 degrees of freedom, p= 0.0481
```

iii. Compute hazard ratio for the experimental versus the control.

```
> coxph(Surv(time.ot, event.ot)~z, data=surv)
Call:
coxph(formula = Surv(time.ot, event.ot) ~ z, data = surv)


   coef exp(coef) se(coef)    z     p
z 0.303      1.35    0.154 1.97 0.049

Likelihood ratio test=3.83  on 1 df, p=0.0503  n= 632
```

iv. Compute hazard ratio for the experimental versus the control adjusted for the baseline variable `var1`.

```
> coxph(Surv(time.ot, event.ot)~z+var1, data=surv)
Call:
coxph(formula = Surv(time.ot, event.ot) ~ z + var1, data = surv)


       coef exp(coef) se(coef)     z       p
z    0.0175      1.02   0.1572 0.111 9.1e-01
var1 0.3545      1.43   0.0573 6.191 6.0e-10

Likelihood ratio test=44.9  on 2 df, p=1.82e-10  n= 632
```

v. Compute hazard ratio for the experimental versus the control adjusted for the baseline variable `var2`.

```
> coxph(Surv(time.ot, event.ot)~z+var2, data=surv)
Call:
coxph(formula = Surv(time.ot, event.ot) ~ z + var2, data = surv)


       coef exp(coef) se(coef)     z    p
z    -0.231     0.793   0.1563 -1.48 0.14
var2  0.991     2.693   0.0977 10.14 0.00

Likelihood ratio test=135  on 2 df, p=0  n= 632
```

Recall that covariate adjustment helps in two ways: reducing variability in the estimates of the parameter of interest when the covariate explains much of the variability between subjects (note change in the standard error of $\hat{\beta}$ in problem 2(a)—`x0` explains much of the variation in `x1`), and to account for potential confounding between the treatment and the response. How do the results of (b)iv and (b)v help explain the differences you see in the models in part (a) and part (b)iii.

The ITT analysis in part (a) indicates that assignment to the experimental treatment reduces the risk of the event. On the other hand, the "on-treatment" analysis seems to

suggest that treatment increases risk. While the covariate adjustment in (a) does not alter in any meaningful way the apparent effect of treatment, The covariate adjustment in (b) has no effect on the standard error of the coefficient, it significantly changes the estimates, actually reversing the sign of the log-HR from positive to negative. This suggests that there is confounding between the treatment, outcome and covariate. (Because treatment is randomly assigned, there cannot be confounding with the assignment itself, however, since we're only counting events prior to treatment discontinuation, we're implicitly considering "treatment" to be both assignment and *duration* and the latter is *not* randomly assigned and therefore subject to confounding.) Hence, the unadjusted OT analysis may be driven as much by confounding as by an effect of treatment.
In addition, adjustment for `var1` induces a smaller change in the coefficient for treatment, suggesting that it is less strongly associated with treatment and/or outcome.

(c) Consider the association between treatment discontinuation and outcomes by creating 2×2 tables of outcome (`event`=0/1) versus treatment discontinuation (`off.trt`=0/1) for each treatment group separately. What does this suggest regarding the assumptions underlying the models in part (b)?

```
> tab0 <- table(surv$event[surv$z==0], surv$off.trt[surv$z==0])
> tab0

     0   1
  0 125  44
  1  56  88
> chisq.test(tab0, correct=F)

Pearson's Chi-squared test

data:  tab0
X-squared = 39.224, df = 1, p-value = 3.779e-10

> tab1 <- table(surv$event[surv$z==1], surv$off.trt[surv$z==1])
> tab1

     0   1
  0  33 173
  1  56  57
> chisq.test(tab1, correct=F)

Pearson's Chi-squared test

data:  tab1
X-squared = 40.8033, df = 1, p-value = 1.683e-10
```

In both treatment groups treatment discontinuation is strongly associated with events, however, the relationship is reversed between groups. For control (`z`=0), the OR is $125 \times 88/44 \times 56 = 4.46$ whereas for experimental (`z`=1), the OR is $33 \times 57/56 \times 173 = 0.194$.

The key assumption for any time-to-event analysis is independence between censoring and the event. These tables clearly show an association between censoring (going off treatment) and events, violating the key assumption.

(d) Suppose that we had discontinued follow-up at the time of treatment discontinuation (so that we observed only `time.ot` and `event.ot` and *not* `time` and `event`) and that we measured baseline variable `var1` but not `var2`. What would we have concluded regarding the effect of treatment and why?

The unadjusted OT analysis suggests that the experimental treatment has an adverse effect on events, while the adjusted (for `var1`) analysis suggests no difference between treatments. The variable `var2` accounts for most of the confounding observed induced by the dependent censoring, however, we are now assuming that we do not observe `var2` so it will be impossible to adjust for `var2`. It would be impossible to reach the correct conclusion that the experimental treatment is superior. (A "valid" on-treatment analysis requires an assumption of "no unmeasured confounders" which could be clearly violated in this case.)