Statistics 641, Spring 2018
Homework #4
Solutions

1. Suppose the table below represents the potential outcomes (Y_A and Y_B) for 10 subjects enrolled in a randomized trial. In this trial A is placebo and B is the active experimental treatment. Imagine further that non-adherence to placebo (A) yields the same response as adherence to placebo (no "placebo effect"), and that non-adherence to the active treatment (B) is equivalent to placebo. Subjects are randomly assigned treatment according to the third column. The fourth column indicates which subjects are adherent to assigned treatment (Y=Yes, N=No). The last column is the response that would be observed given assigned treatment and adherence.

Subject	Y_A	Y_B	Assigned Treatment	Adhere?	Y
1	13	13	A	Y	
2	14	14	A	N	
3	4	14	B	Y	
4	2	13	A	Y	
5	7	20	B	Y	
6	11	11	B	N	
7	4	4	B	N	
8	10	10	A	N	
9	4	4	B	Y	
10	1	7	A	Y	

   (a) What is the true average treatment effect of treatment B relative to treatment A in this population?

   The mean of the Y_As is 7, and the mean of the Y_Bs is 11, so the average causal effect of B relative to A is 11-7=4,

   (b) Fill in the last column of observed responses given treatment assignments and adherence. (You can just provide this as a list with values and subject numbers rather than reproducing the entire table.)

   Subjects 1,2,4,8,10 are assigned A, but subjects 2 and 8 are non-adherent. However, given that non-adherence to A has no effect on the response, the responses for subjects 1,2,4,8,10 are 13, 14, 2, 10, and 1 respectively. Subjects 3,5,6,7,9 are assigned B and, while subjects 6 and 7 are non-adherent, their A and B responses are identical. Hence, we observe 14, 20, 11, 4, and 4 for subjects 3,5,6,7,9 respectively.

   (c) Find the ITT estimate of the causal effect of treatment.

   The mean of observed responses among subjects assigned A is 8 and the mean of observed responses among subjects assigned B is 10.6. The ITT estimate of the causal effect is 2.6.

(d) (At least for placebo controlled trials) the ITT analysis is generally considered "conservative" in that it underestimates the full adherence treatment effect. Is that true in this case? Why or why not?

In the case, subjects who are non-adherent to assigned treatment have identical responses to both A and B, i.e., they are "non-responders". At least for this population of 10 subjects, the causal effect of receipt is identical to the causal effect of assignment and therefore the ITT estimate is exactly equal to the full-adherence estimate had we been able to observe it. The "conventional wisdom" that the expected value of the ITT estimate is smaller than the full adherence causal effect relies on unknowable assumptions, e.g., the response under partial adherence is strictly smaller than the response under partial adherence.

(e) Find the "Per-protocol" estimate of the treatment effect (exclude non-adherers). Does this estimate reflect the true causal effect of treatment? Why or why not?

The means among adherent subjects assigned A is 5.33 and the mean among adherent subjects assigned B is 12.67 and the difference is 7.33. Because the population average causal effect is an average over the whole population, including non-responders, and the "per-protocol" estimate excludes only non-responders, (for this population) the "per-protocol" estimate *over-estimates* the causal effect of receipt of treatment.

(f) Find the "As-treated" estimate of the treatment effect. Does this estimate reflect the true causal effect of treatment? Why or why not?

The "As-treated" estimate considers the two subjects assigned B who are non-adherent to be equivalent to to placebo, and would reassign them to the A group. Non-adherers in the A group would be considered to be receiving the equivalent of placebo and would remain in the A group. The mean of the three B-adherers is 12.67, and the mean of the remaining responses is 7.86. The "As-treated" estimate of the treatment effect is 4.8. Because the "As-treated" estimate is no based on the randomization, it cannot be an estimate of a causal effect. The direction of the bias is, in general, unpredictable.

2. A randomized, two-arm trial is conducted comparing a control treatment (A) to a experimental treatment (B). The primary outcome is all-cause mortality.

(a) After completion of the trial (between 1.5 and 3 years follow-up), we observe the following table:

	Subjects	Deaths	Person-years follow-up
A	500	241	835
B	500	219	876

Assume exponential survival and compute the hazard ratio, the Wald test statistic for the log-hazard ratio and a 95% confidence interval for the hazard ratio. (I skipped this in class, but you can use Section 5.8.1 of the notes. The definitions of δ_i and T_i are on

the previous page of the notes.).

The hazard ratio is

$$\frac{\widehat{\lambda}_B}{\widehat{\lambda}_A} = \frac{219/876}{241/835} = 0.866$$

From Section 5.8.1, the variance of $\widehat{\beta} = \log \widehat{\lambda}_B/\widehat{\lambda}_A$ is

$$\text{Var}\left[\widehat{\beta}\right] = \frac{1}{241} + \frac{1}{219} = 0.008715$$

and the Wald statistic is

$$\frac{\widehat{\beta}^2}{\text{Var}\left[\widehat{\beta}\right]} = \frac{(\log 0.866)^2}{0.00875} = 2.37 \sim \chi^2_1 \text{ under } H_0$$

or

$$\frac{\widehat{\beta}}{\sqrt{\text{Var}\left[\widehat{\beta}\right]}} = -1.54 \sim N(0,1) \text{ under } H_0$$

(This corresponds to a two-sided p-value of 0.12, so it does not reach conventional levels of statistical significance.)

You could also calculate a Wald test on the hazard ratio scale, rather than the log-hazard ratio scale. By the delta-method

$$\text{Var}\left[e^{\widehat{\beta}}\right] = e^{2\widehat{\beta}}\text{Var}\left[\widehat{\beta}\right] = 0.866^2 \times 0.00875 = 0.00656$$

and the test is

$$\frac{(0.866-1)^2}{0.00656} = 2.73$$

or

$$\frac{0.866-1}{\sqrt{0.00656}} = -1.65$$

The 95% CI for β is $\log 0.866 \pm 1.961\sqrt{0.008715} = (-0.327, 0.039)$, and the 95% CI for $\lambda_B/\lambda_A = (0.721, 1.040)$

(b) It is noted that many subjects do not adhere to their assigned treatment and when subjects are classified by their level of adherence, we observe the following table:

	Adherence	Subjects	Deaths	Person-years follow-up
A	$> 80\%$	110	52	180
	50%–80%	240	153	352
	$\leq 50\%$	150	36	303
B	$> 80\%$	145	51	262
	50%–80%	280	148	463
	$\leq 50\%$	75	20	151

Compute hazard ratios for subjects within each stratum based on adherence ($> 80\%$, 50%–80%, $\leq 50\%$), and note the differences with the overall comparison in part (a). Which result is more credible as an assessment of the causal effect of treatment and why?

The within-stratum hazard ratios are

$$\begin{aligned} > 80\%: &\quad \frac{51/262}{52/180} = 0.674 \\ 50\%\text{-}80\%: &\quad \frac{148/463}{153/352} = 0.735 \\ \leq 50\%: &\quad \frac{20/151}{36/303} = 1.115 \end{aligned}$$

The hazard ratios for the "better compliers" ($> 80\%$ and 50%-80% strata) are much smaller than the overall comparison (the *intention-to-treat (ITT) analysis*). It might be tempting to conclude that this analysis provides evidence of treatment benefit that is not evident in the intention-to-treat analysis, however, the within-stratum analysis is not credible for several reasons.

- The randomization ensures that the treatment assignments are independent of outcomes, and therefore, the ITT analysis is a valid test of the null hypothesis that there is no net causal effect of treatment assignment on outcomes.
- Adherence to assigned treatment is a post-randomization characteristic, and analysis conditional on post-randomization characteristics *do not* in general have valid causal interpretations.

Thus, the apparent benefit of B relative to A in the stratified analysis does not represent a valid causal effect of treatment.

The table below shows subjects (hypothetically) assigned to one of four unobservable strata as a function of what their adherence *would have been* had they been assigned either A or B. For simplicity, the 50%-80% and $\leq 50\%$ categories have been collapsed to a single $\leq 80\%$ category. The four strata are

I. adherence $> 80\%$ on both A and B

II. adherence $> 80\%$ were they to be assigned to A but $\leq 80\%$ were they to be assigned to B

III. adherence $\leq 80\%$ were they to be assigned to A but $> 80\%$ were they to be assigned to B

IV. adherence $\leq 80\%$ on both A and B.

Note that these are inherent subject characteristics that are present at baseline, but we cannot fully observe them. We cannot observe which of the B-adherence categories subjects assigned A are in because we cannot observe their B-adherence status, and vice versa for subjects assigned to B.

The first column of this table corresponds to the good "A-adherers" and we observe the overall result at the bottom of the column, but we cannot know how these are divided between good "B-adherers" and poor "B-adherers". Similarly, the first row corresponds to the good "B-adherers" and we observe the overall result on the right, but we cannot know how these are divided between good "A-adherers" and poor "A-adherers". Each entry is the color-coded sample size and observed hazard ratio for

subjects Assigned A, Assigned B. We can see that the good B-adherers, composed of strata I and III which have low risk (about .2), have overall low risk, but the good A-adherers, composed of strata I (low risk) and II (high risk) has higher risk.

Assigned B	Assigned A $> 80\%$	Assigned A $\leq 80\%$	All
$> 80\%$	I 83, $\widehat{\lambda} = 26/127 = 0.205$ 81, $\widehat{\lambda} = 27/134 = 0.201$	III 63, $\widehat{\lambda} = 26/123 = 0.211$ 64, $\widehat{\lambda} = 24/128 = 0.188$	145, $\widehat{\lambda} = 51/262 = 0.195$
$\leq 80\%$	II 27, $\widehat{\lambda} = 26/53 = 0.491$ 39, $\widehat{\lambda} = 22/55 = 0.400$	IV 327, $\widehat{\lambda} = 163/532 = 0.306$ 316, $\widehat{\lambda} = 146/559 = 0.261$	355, $\widehat{\lambda} = 168/614 = 0.274$
All	110, $\widehat{\lambda} = 52/180 = 0.289$	390, $\widehat{\lambda} = 189/555 = 0.289$	

Note that within each cell, the observed event rates are similar by treatment. This is guaranteed by randomization if treatment has no effect in any stratum. The differences are simply due to sampling variability.

From a potential outcomes point of view, the potential outcomes given assignment to A for subjects in stratum II are included in the *per-protocol* analysis, but would *not* be had they been assigned to treatment B. Conversely, the potential outcomes given assignment to B for subjects in stratum III are included in the *per-protocol* analysis, but would *not* be had they been assigned to treatment A. Hence the difference in rates for the adherers does not represent a causal effect.

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 "data4.csv" (in csv format, comma delimited) contains the following variables:

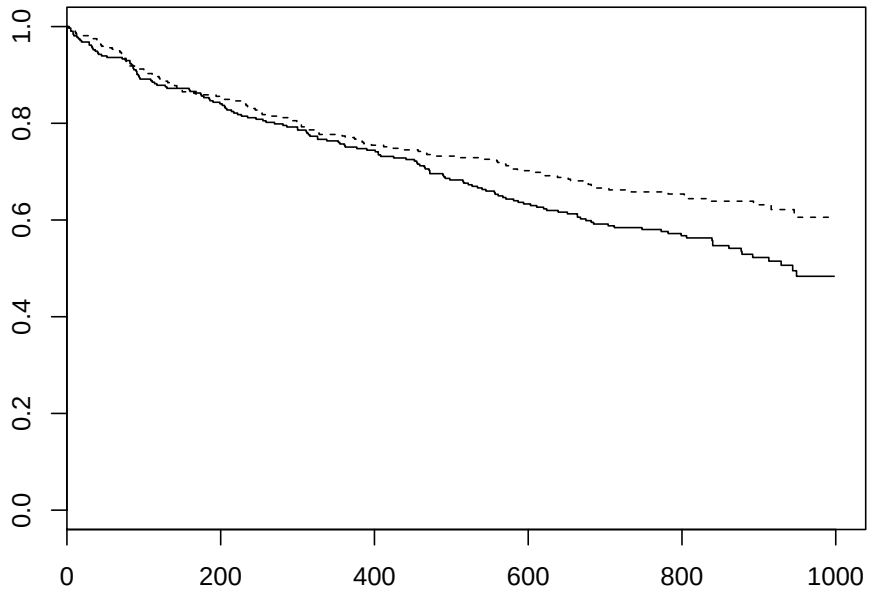
   - `z`: Treatment variable (0=control, 1=experimental)
   - `risk1`, `risk2`: Baseline covariates associated with risk
   - `time`: Time to the event of interest
   - `event`: Indicator of the event of interest (0=censored, 1=event)
   - `time.ot`: Time to the event of interest or end of treatment
   - `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)
...
      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

>  coxph(Surv(time, event)~z, data=surv)
...
    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, number of events= 257
```

iii. Compute hazard ratio for the experimental versus the control adjusted for the baseline variable `risk1`.

```
>  coxph(Surv(time, event)~z + risk1, data=surv)
...
        coef exp(coef) se(coef)     z       p
z     -0.349     0.705    0.126 -2.77 5.6e-03
risk1  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 roughly agree).

ii. `risk1`

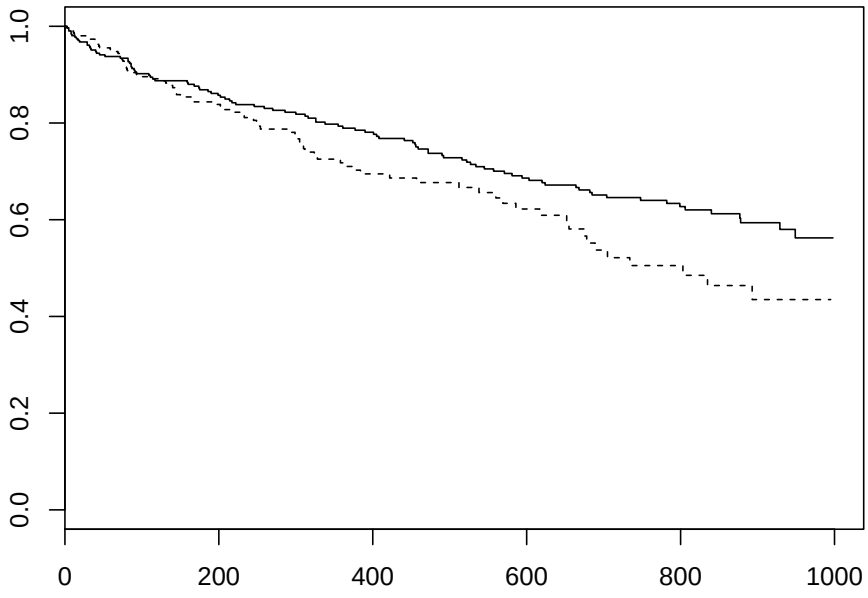
Each unit increase in `risk1` appears to increase risk of the event by about 44% (after accounting for treatment).

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)
...
      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)
...
   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, number of events= 178
```

iv. Compute hazard ratio for the experimental versus the control adjusted for the baseline variable `risk1`.

```
> coxph(Surv(time.ot, event.ot)~z+risk1, data=surv)
...
        coef exp(coef) se(coef)     z       p
z     0.0175      1.02   0.1572 0.111 9.1e-01
risk1 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, number of events= 178
```

v. Compute hazard ratio for the experimental versus the control adjusted for the baseline variable `risk2`.

```
> coxph(Surv(time.ot, event.ot)~z+risk2, data=surv)
...
        coef exp(coef) se(coef)     z    p
z     -0.231     0.793   0.1563 -1.48 0.14
risk2  0.991     2.693   0.0977 10.14 0.00
Likelihood ratio test=135  on 2 df, p=0
n= 632, number of events= 178
```

Recall that in general 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, 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, but significantly changes the estimates, actually reversing the sign of the log-HR from positive to negative. This suggests that there may be 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 `risk1` 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 `risk1` but not `risk2`. 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 `risk1`) analysis suggests no difference between treatments. The variable `risk2` accounts for most of the confounding observed induced by the dependent censoring, however, we are now assuming that we do not observe `risk2` so it will be impossible to adjust for `risk2`. 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.)