STA 580 — Fall 2008 — Dr. Charnigo

Lecture 14

Time-to-Event Response Variables

Motivation #1: better power. Lecture 13 addressed scenarios with a dichotomous response variable and a dichotomous explanatory variable. These scenarios often entail a prospective study and a test of $H_0 : p_1 = p_2$,[1] where p_1 and p_2 are the risks in the two groups defined by the dichotomous explanatory variable. For instance, we could test whether presently-healthy smokers and nonsmokers have equal risks of developing lung cancer within 15 years. Yet, the time frame seems arbitrary. Why not ask about the risks of developing lung cancer within 10 years? Or 12 years?

Indeed, what we have called p_1 and p_2 depend on the time frame. Explicitly, let $p_1(t)$ and $p_2(t)$ denote the risks of developing lung cancer within t years. Note that $p_1(t)$ and $p_2(t)$ are increasing functions of t.[2] Rejecting $H_{0;15} : p_1(15) = p_2(15)$ should persuade us that smoking is relevant to the development of lung cancer. But rejecting $H_{0;10} : p_1(10) = p_2(10)$ or $H_{0;12} : p_1(12) = p_2(12)$ should also be persuasive. In fact, rejecting $H_{0;all}$: for all t, $p_1(t) = p_2(t)$ should be persuasive.

Notice, the only way that $H_{0;all}$ can be true is if $H_{0;15}$, $H_{0;10}$, $H_{0;12}$, and all other "constituent" null hypotheses are true. Hence, our best chance at demonstrating the relevance of smoking (more formally, the best power at a fixed significance level and with fixed sample sizes) seems to lie in testing $H_{0;all}$ rather than in testing one of the constituent null hypotheses.

[1] Such a test is conducted implicitly whenever we construct a confidence interval for the relative risk or the odds ratio, as exclusion of 1 from either confidence interval is tantamount to rejection of H_0.

[2] Someone who has developed lung cancer at an earlier date cannot "undevelop" lung cancer at a later date; he or she may be successfully treated for lung cancer at a later date, but the person is still regarded as having developed lung cancer. On the other hand, someone who has not developed lung cancer at an earlier date may go on to develop lung cancer at a later date.

Motivation #2: accommodating uneven follow-up. Randomized clinical trials (Definition 6.6) are characterized by experimental assignment of study participants to treatment groups rather than circumstantial or self assignment to exposure groups. Moreover, eligibility for a clinical trial often depends on the occurrence of some precipitating event. For instance, a myocardial infarction would be the precipitating event qualifying a person to participate in a clinical trial to assess how well a new medication prevented repeat myocardial infarctions. As such, participants in clinical trials are usually not enrolled simultaneously. Thus, when the study ends, some participants will have had longer follow-up than others.

For concreteness, suppose that we are conducting a clinical trial to assess how well a new medication prevents repeat myocardial infarctions. We can view the response variable as dichotomous (is there a repeat myocardial infarction?), and we can view the explanatory variable as dichotomous (does the participant receive the new medication?). Also, suppose that 1000 participants are to be enrolled in 2010, that 1000 participants are to be enrolled in 2011, and that the study is to conclude in 2013.

Using obvious notation, we could test $H_{0;2} : p_1(2) = p_2(2)$, but such a test would disregard third-year data from the 1000 participants enrolled in 2010. Testing $H_{0;3} : p_1(3) = p_2(3)$ would also be problematic since none of the 1000 participants enrolled in 2011 would be observed for three years. Testing $H_{0;all}$: for all t, $p_1(t) = p_2(t)$ is the easiest way out of this dilemma.

Time-to-event response variables. Let T denote the (nonnegative) time at which a person reaches an "endpoint". The endpoint may be the development of disease, death, some other adverse event, or even a "composite" that includes some combination of the preceding. Time zero may correspond to a person's birth or to a precipitating event that makes the person eligible for the study. Even if the endpoint is not death, we still refer to T as a "survival" time or as the time at which the person "expires".

Although T may be regarded as a continuous random variable, we cannot employ a t-test or a rank-sum test since the value of T will not be observed for all participants (more on this below). Hence, to distinguish T from continuous response variables that are amenable to a t-test or a rank-sum test, we refer to T as a "time-to-event" response variable.

Example (time-to-event response variables). We will repeatedly use the following "recidivism" example. Let T denote the time in days at which a person who has attempted to quit smoking relapses. Time zero is when the person initiates the attempt to quit. We do not observe the value of T for participants who remain abstinent past the end of the study.[3]

Right-censoring. A participant's survival time is said to be right-censored when we know only a lower bound for the survival time (Definition 14.5); sometimes we speak loosely and say that the participant is right-censored. Right-censoring occurs when the participant's survival time exceeds the duration of follow-up, either because the participant drops out of the study or because the study ends before the participant expires. There does exist a phenomenon called left-censoring,[4] but any instance of "censoring" in the rest of this lecture will refer to right-censoring.

[3] As defined here, T is technically not a continuous random variable since presumably some participants will *never* relapse. If desired, we can alleviate this minor technical issue by redefining the event of interest to be the composite of relapse and death.

[4] Left-censoring occurs when we know only an upper bound for a participant's survival time. For instance, suppose that a medical condition manifests so gradually that the exact time at which a person first had the condition — call it T — is not necessarily observed. Hypertension is one example of such a condition: if the person is formally diagnosed at age 45, we know that $T \leq 45$ but may not feel comfortable asserting that $T = 45$.

Estimation of Survival Functions

Survival functions. Let $S_1(t) := 1 - p_1(t) = P_1(T > t)$ denote the probability that a person in exposure/treatment group 1 does not expire by time t, and let $S_2(t) := 1 - p_2(t) = P_2(T > t)$ denote the probability that a person in exposure/treatment group 2 does not expire by time t. Above, the subscripts 1 and 2 on the probability symbols distinguish probabilities for people in group 1 from those for people in group 2.

We refer to $S_1(t)$ and $S_2(t)$ as survival functions or survival curves (Definition 14.3). Since $p_1(t)$ and $p_2(t)$ are increasing functions of t, $S_1(t)$ and $S_2(t)$ are decreasing functions of t.

Kaplan-Meier estimation. The most common approach for estimating survival functions is the "Kaplan-Meier" or "product-limit" method (Equations 14.38 and 14.39). Let $t_1 < \cdots < t_k$ be nonnegative numbers chosen according to the precision with which T will be measured. For instance, suppose that T is the number of days to smoking relapse and that we plan to monitor participants for no more than one year. If we will obtain only the number of full days that a participant avoids relapse (i.e., we will not record hours or half-days), then $t_1 = 0, t_2 = 1, t_3 = 2, \ldots, t_{366} = 365$.

First I will describe estimation of $S_1(t)$. Let S_0 be the number of participants in group 1 with which we start. Let d_1 be the number who expire by time t_1, let l_1 be the number who do not expire by time t_1 but who will not remain in our study afterward (i.e., who are censored at time t_1), and let S_1 be the number who do not expire by time t_1 and who will remain in our study. A natural estimate for $S_1(t_1) = P_1(T > t_1)$ is

$$\widehat{S_1(t_1)} := \frac{l_1 + S_1}{S_0} = 1 - \frac{d_1}{S_0}.$$

Of the S_1 participants in group 1 who have not expired by time t_1 and who have remained in our study, d_2 will expire by time t_2, l_2 will not expire by time t_2 but will not remain in our study afterward, and S_2 will not expire by time t_2 and will remain in our study. Since

$$S_1(t_2) = P_1(T > t_2) = P_1(T > t_1 \cap T > t_2) = P_1(T > t_1)\, P_1(T > t_2 \mid T > t_1),$$

a natural estimate for $S_1(t_2)$ is

$$\widehat{S_1(t_2)} := \left(1 - \frac{d_1}{S_0}\right)\left(1 - \frac{d_2}{S_1}\right).$$

We may continue in similar fashion. With obvious notation, we put

$$\widehat{S_1(t_3)} := \left(1 - \frac{d_1}{S_0}\right)\left(1 - \frac{d_2}{S_1}\right)\left(1 - \frac{d_3}{S_2}\right),$$

$$\widehat{S_1(t_4)} := \left(1 - \frac{d_1}{S_0}\right)\left(1 - \frac{d_2}{S_1}\right)\left(1 - \frac{d_3}{S_2}\right)\left(1 - \frac{d_4}{S_3}\right),$$

and so forth. The estimated survival function $\widehat{S_1(t)}$ is determined uniquely up to time t_k once we impose the requirement that it be a right-continuous step function.

Estimation of $S_2(t)$ is carried out in the same manner.

Remarks. The Kaplan-Meier approach is a favorite of data analysts because there is no need to assume that $S_1(t)$ and $S_2(t)$ have simple mathematical forms characterized by a limited number of parameters.[5] As such, statisticians classify the Kaplan-Meier approach as "nonparametric".

[5] An example of such an assumption would be that $S_1(t) = \exp[-\lambda_1 t]$ and $S_2(t) = \exp[-\lambda_2 t]$ for some positive parameters λ_1 and λ_2. Such an assumption may be undesirable because it may impose unrealistic "hidden constraints" on the survival functions. With $S_1(t)$ as defined above, for instance, a person in group 1 who has been abstinent during the past six months must have the same probability of being abstinent during the next six months as another person in group 1 who just attempted to quit smoking today.

If no censoring has taken place prior to time t_i, then the estimates of $S_1(t_i)$ and $S_2(t_i)$ simplify to

$$\frac{\text{number of participants in group 1 not expired by time } t_i}{\text{total number of participants in group 1}}$$

and

$$\frac{\text{number of participants in group 2 not expired by time } t_i}{\text{total number of participants in group 2}},$$

respectively.

Example (Kaplan-Meier estimation). Refer to {Survival.pdf}. Survival functions have been estimated separately for each gender (male = 1, female = 2). Graphical representations of the estimated survival functions are on page 10. The box "Product-Limit Survival Estimates" (page 1, page 5) contains the following information:

• In column "Survival" are the estimates $\widehat{S_1(t_i)}$ or $\widehat{S_2(t_i)}$ for various t_i. We have, for example,

$$\widehat{S_1(0)} = 0.9818 = \left(1 - \frac{2}{110}\right) = \frac{108}{110}$$

and

$$\widehat{S_1(1)} = 0.9091 = \left(1 - \frac{2}{110}\right)\left(1 - \frac{8}{108}\right) = \frac{100}{110}.$$

These are the estimated probabilities of a male avoiding smoking relapse for at least one full day and for at least two full days, respectively.

• In column "Survival Standard Error" are standard errors that help define $100(1-\alpha)\%$ confidence intervals for $S_1(t_i)$ and $S_2(t_i)$ via the formulas[6]

$$\widehat{S_1(t_i)} \pm z_{1-\alpha/2} \text{ stderr of } (\widehat{S_1(t_i)})$$

[6]These differ from the formulas in Equations 14.40 and 14.41, which entail exponentiating the endpoints of confidence intervals for $\log S_1(t_i)$ and $\log S_2(t_i)$.

and

$$\widehat{S_2(t_i)} \pm z_{1-\alpha/2} \text{ stderr of } (\widehat{S_2(t_i)}).$$

• In columns "Number Failed" and "Number Left" are the numbers of participants in group 1 or group 2 who have expired and who remain in the study not having expired.

Comparison of Survival Functions

Introduction. The Kaplan-Meier estimates of $S_1(t)$ and $S_2(t)$ almost always differ. The question is whether the estimates of $S_1(t)$ and $S_2(t)$ differ enough to convince us that $S_1(t)$ and $S_2(t)$ differ. This is essentially the same question encountered in testing whether two population means μ_1 and μ_2 are equal: the sample means $\bar{x}$ and $\bar{y}$ almost always differ, but sometimes they do not differ enough to convince us that the population means differ.

Formally, our null hypothesis is that $S_1(t) = S_2(t)$ for all t. This is logically equivalent to having $p_1(t) = p_2(t)$ for all t, which I denoted by $H_{0;all}$ in the first two pages of this lecture. For simplicity, I hereafter omit the subscript *all*. Testing H_0 : for all t, $S_1(t) = S_2(t)$ requires some preliminary computations involving contingency tables; these are described next.

Contingency tables. We construct a contingency table for each t_i following the pattern in Table 14.18. Entry a_i is the number of participants in the first group who had not expired by time t_{i-1} but who have expired by time t_i. Entry b_i is the number of participants in the first group who have not expired by time t_i. Entries c_i and d_i are defined analogously for participants in the second group. In what follows we let $n_i := a_i + b_i + c_i + d_i$.

Example (contingency tables). Here are the first two contingency tables for our recidivism example.

	Expire	Survive	Row Total
Male	2	108	110
Female	11	113	124
Column Total	13	221	234

	Expire	Survive	Row Total
Male	8	100	108
Female	7	106	113
Column Total	15	206	221

The log rank test. The "log rank" test statistic is (Equation 14.43)

$$\chi^2 := \left(\sum_{i=1}^{k} a_i - \sum_{i=1}^{k} \frac{(a_i + b_i)(a_i + c_i)}{n_i} \right)^2 \Big/ \sum_{i=1}^{k} \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2 (n_i - 1)}.$$

We reject H_0 : for all t, $S_1(t) = S_2(t)$ at level α if $\chi^2 > \chi^2_{1,1-\alpha}$.

Notice, the log rank test statistic has the form $(O - E)^2/V$. Thus, the log rank test loosely resembles the chi-square test for association described in Lecture 8. For us to apply the log rank test, the sample sizes should be large enough so that $V \geq 5$.

Example (the log rank test). Refer to {Survival.pdf}. In "Test of Equality over Strata" (page 11), we find that $\chi^2 = 0.1733$ and that the p-value is 0.6772. In the context of smoking recidivism, we may not conclude that the survival functions for males and females differ.