Chapter 7 Hypothesis Testing

PubH 7450

©Wei Pan

Email: weip@biostat.umn.edu

Http: www.biostat.umn.edu/∼weip

§7.2 One-sample tests

- H_0: $S(t) = S_0(t)$ for $t \leq \tau$
  or, H_0: $h(t) = h_0(t)$ for $t \leq \tau$
  typically, $\tau = T_{(n)}$
- Given: a censored sample from $S(t)$; $S_0(t)$ or $h_0(t)$.
  From the data $\Longrightarrow$ t_i, d_i and y_i, $i = 1, ..., D$.
- Idea: compare $\hat{h}(t)$ or $\hat{H}(t)$ with ...
  $Z(\tau) = \sum_{i=1}^{D} \frac{d_i}{y_i} - \int_0^\tau h_0(s)ds = \hat{H}(\tau) - H_0(\tau) = O(\tau) - E(\tau)$.
- More generally, use a weight function $W(t)$:
  $Z(\tau) = \sum_{i=1}^{D} W(t_i)\frac{d_i}{y_i} - \int_0^\tau W(s)h_0(s)ds$.
- Assuming that the terms are independent and d_i has a Poisson distribution, we have
  $Var(Z(\tau)) \approx \sum_{i=1}^{D} W(t_i)^2 \frac{d_i}{y_i^2} = V_1$.
- Under H_0, replace d_i/y_i by its estimand $h_0(t_i)$, so

$Var(Z(\tau)) \approx \int_0^\tau W(s)^2 \frac{h_0(s)}{y(s)} ds = V_2$.
$y(s)$=# of subjects at risk at s^-; $y(t_i) = y_i$; a step function.

- Test statistic

$$\frac{Z(\tau)}{\sqrt{Var(Z(\tau))}} \overset{a.}{\sim} N(0,1) \text{ under } H_0.$$

- Choice of V_1 vs V_2: where did we have a similar issue?
  1) if $h(t) = h_0(t)$, then ...
  2) if $h(t) > h_0(t)$, then ...
  3) if $h(t) < h_0(t)$, then ...

- Most popular: $W(t) = y(t)$; one-sample log-rank test.
  $Z(\tau) = \sum_{i=1}^D d_i - \int_0^\tau y(s)h_0(s)ds$ = #(obs'ed events) - #(exp'ed events) $= O - E$.
  $V_1 = ...$

- For left-truncated data (L_i, T_i, δ_i), $i = 1, ..., n$,
$\int_0^\tau y(s)h_0(s)ds = \sum_{i=1}^n \int y_i(s)h_0(s)ds = \sum_{i=1}^n \int_{L_i}^{T_i} h_0(s)ds = \sum_{i=1}^n [H_0(T_i) - H_0(L_i)]$.

- Table 7.1.

$$\chi^2 = (O - E)^2/E = \frac{(15 - 4.4740)^2}{4.4740} = 24.8 \sim \chi^2_d$$

with $d =?$ under $H_0 \Longrightarrow p = ...$

- Fleming-Harrington family:
$W(t) = S_0(t)^p[1 - S_0(t)]^q$, $p \geq 0$, $q \geq 0$.
early departure: $p > q$;
late departure: $p < q$;
middle departure: $p = q > 0$.

TABLE 7.1

Computation of One-Sample, Log-Rank Test

Subject j	*Sex*	*Status* d_i	*Age at Entry* L_i	*Age at Exit* T_j	$H_0(L_j)$	$H_0(T_j)$	$H_0(T_j) - H_0(L_j)$
1	f	1	51	52	0.0752	0.0797	0.0045
2	f	1	58	59	0.1131	0.1204	0.0073
3	f	1	55	57	0.0949	0.1066	0.0117
4	f	1	28	50	0.0325	0.0711	0.0386
5	m	0	21	51	0.0417	0.1324	0.0907
6	m	1	19	47	0.0383	0.1035	0.0652
7	f	1	25	57	0.0305	0.1066	0.0761
8	f	1	48	59	0.0637	0.1204	0.0567
9	f	1	47	61	0.0606	0.1376	0.0770
10	f	1	25	61	0.0305	0.1376	0.1071
11	f	0	31	62	0.0347	0.1478	0.1131
12	m	0	24	57	0.0473	0.1996	0.1523
13	m	0	25	58	0.0490	0.2150	0.1660
14	f	0	30	67	0.0339	0.2172	0.1833
15	f	0	33	68	0.0365	0.2357	0.1992
16	m	1	36	61	0.0656	0.2704	0.2048
17	m	0	30	61	0.0561	0.2704	0.2143
18	m	1	41	63	0.0776	0.3162	0.2386
19	f	1	43	69	0.0503	0.2561	0.2058
20	f	1	45	69	0.0548	0.2561	0.2013
21	f	0	35	65	0.0384	0.1854	0.1470
22	m	0	29	63	0.0548	0.3162	0.2614
23	m	0	35	65	0.0638	0.3700	0.3062
24	m	1	32	67	0.0590	0.4329	0.3739
25	f	1	36	76	0.0395	0.4790	0.4395
26	m	0	32	71	0.0590	0.5913	0.5323
Total		15					4.4740

§7.3 K-sample tests

- Given data: $(T_{i1}, \delta_{i1}), ..., (T_{iK}, \delta_{iK})$, $i = 1, ..., n_K$.
- Goal: Test H_0: $h_1(t) = ... = h_K(t)$ for all $t \le \tau$
  vs H_1: at least one equality does not hold for some $t \le \tau$.
  Assumption: nonparametric; X_{ij} and C_{ij} are independent.
  $\tau = \min_j T_{(n),j}$.
  Distributions of $C_{i1}, ..., C_{iK}$ are identical?
- A key: you are comparing ...
- General idea: assuming H_0 true, then compare $\hat{h}_j(t)$ to ...
- Notation: pool the samples together
  1. Define $t_1 < t_2 < ... < t_D$ as *distinct* event times;
  2. $d_{ij} = \#$ events at t_i from sample j;
     $d_i = \sum_j d_{ij}$;
  3. $y_{ij} = \#$ subjects at risk at t_i^- from sample j;

$y_i = \sum_j y_{ij}$.

- Components of test statistics:

$$Z_j(\tau) = \sum_{i=1}^{D} W_j(t_i) \left(\frac{d_{ij}}{y_{ij}} - \frac{d_i}{y_i} \right)$$

for $j = 1, ..., K$.

- Under H_0, let $W_j(t_i) = W(t_i) y_{ij}$,

$$Var(Z_j(\tau)) = \sum_{i=1}^{D} W(t_i)^2 \frac{y_{ij}}{y_i} \left(1 - \frac{y_{ij}}{y_i}\right) \left(\frac{y_i - d_i}{y_i - 1}\right) d_i,$$

$$Cov(Z_j(\tau), Z_g(\tau)) = -\sum_{i=1}^{D} W(t_i)^2 \frac{y_{ij}}{y_i} \frac{y_{ig}}{y_i} \left(\frac{y_i - d_i}{y_i - 1}\right) d_i$$

for $j \neq g$.

Idea of the derivation: 1) $(y_i - d_i)/(y_i - 1)$ is a correction for ties; 2) other terms related to

$$(d_{i1}, ..., d_{iK})' \sim Multinomial(d_i, (p_{i1}, ..., p_{iK})')$$

with $\hat{p}_{ij} = y_{ij}/y_i$, $j = 1, ..., K$.
$Var(d_{ij}) = d_i p_{ij}(1 - p_{ij})$.
$Cov(d_{ij}, d_{ig}) = -d_i p_{ij} p_{ig}$.

- Use just any $(K - 1)$ Z_j's because $\sum_{j=1}^{K} Z_j(\tau) = ...$,
- Test statistic:
  $Z(\tau) = (Z_1(\tau), ..., Z_{K-1}(\tau))'$,
  $\Sigma = Cov(Z(\tau))$,
  $\chi^2 = Z(\tau)'\Sigma^{-1}Z(\tau) \overset{a.}{\sim} \chi^2_{K-1}$ under H_0.
- $K = 2$: $W_1(t_i) = W(t_i)y_{i1}$,

$$\chi = \frac{\sum_{i=1}^{D} W(t_i)\left(d_{i1} - d_i \frac{y_{i1}}{y_i}\right)}{\sqrt{\sum_{i=1}^{D} W(t_i)^2 \frac{y_{i1}}{y_i}\frac{y_{i2}}{y_i}\frac{y_i - d_i}{y_i - 1} d_i}} \overset{a.}{\sim} N(0, 1) \text{ under } H_0.$$

- Various choices of weight leads to various tests:
  1. $W(t) = 1$: log-rank test;
  2. $W(t_i) = y_i$: Gehan's generalization of Mann-Whitney-Wilcoxon test;
  3. $W(t_i) = \sqrt{y_i}$: Tarone-Ware class;
  4. Peto-Peto and Kalbfleisch-Prentice's generalization of M-W-W test:
     $\tilde{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{y_i + 1}\right)$,
     $W(t_i) = \tilde{S}(t_i)$.
  5. Fleming-Harrington class:
     $\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{y_i}\right)$,
     $W(t_i) = \hat{S}(t_{i-1})^p [1 - \hat{S}(t_{i-1})]^q$, $p \ge 0$, $q \ge 0$.
     Technicality: $W(t_i)$ is known prior to t_i.
     $p = q = 0$: log-rank test;
     $p = 1$, $q = 0$: a version of M-W-M test;

$q = 0$, $p > 0$: give more weights to ...

$p = 0$, $1 > 0$: give more weights to ...

Choice of (p, q) reflects one's emphasis on ...

- Example 7.2: SAS

  Fig 7.2: relative weight $W(t_i)/\sum_{i=1}^{D} W(t_i)$.

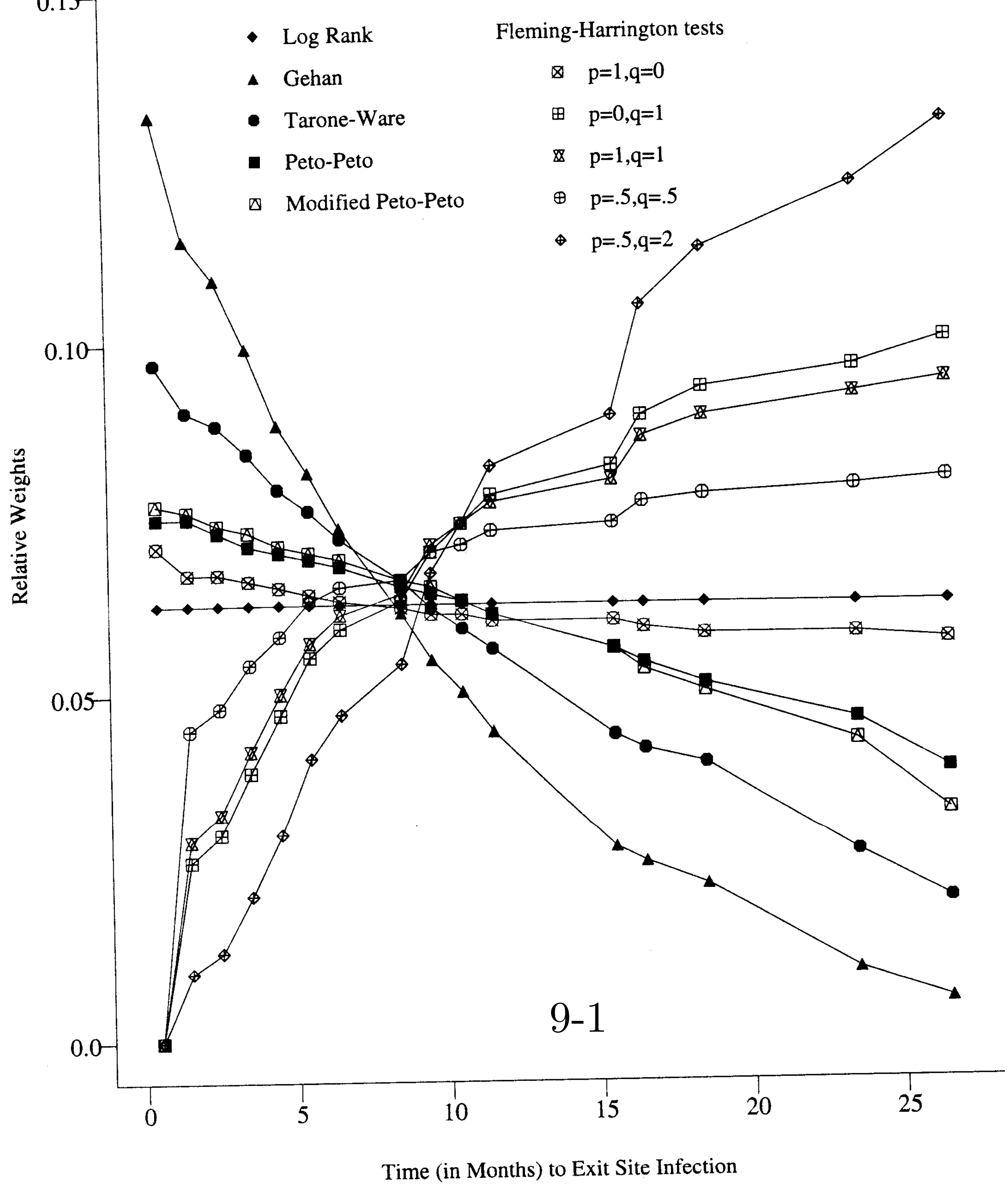


Figure 7.2 *Relative weights for comparison of observed and expected numbers of deaths for kidney dialysis patients.*

- Log-rank test: optimal for proportional hazards

$$h_j(t) = h_0(t)\exp(\theta_j), \qquad j = 1, 2.$$

also called Lehmann alternative.
optimal: asymptotically most powerful; related to a score test.

- Log-rank test: another derivation for $K = 2$.
At t_i:

Group/Event	Dead	Alive	total
Grp 1	d_{i1}		y_{i1}
Grp 2	d_{i2}		$y_{i2} = y_i - y_{i1}$
	d_i		y_i

$H_{0,i}$: No association b/w group and event at t_i
$\Leftrightarrow$ Pr(Death at t_i|given alive at t_i^-, Grp 1)
=Pr(Death at t_i|given alive at t_i^-, Grp 2)
$\Leftrightarrow h_1(t_i) = h_2(t_i)$.

$$MH = \frac{\sum_{i=1}^{D}(d_{i1} - E(d_{i1}|H_{0,i}))}{\sqrt{\sum_{i=1}^{D} \frac{y_{i1}y_{i2}d_i(y_i-d_i)}{y_i^2(y_i-1)}}} \overset{a.}{\sim} N(0,1).$$

Note: $E(d_{i1}|H_{0,i}) = d_i y_{i1}/y_i \Longrightarrow MH = \chi$ with $W(t) = 1$.

- Gehan's test: generalization of M-W-W test with $K = 2$.
- Review of M-W-W test: no censoring

  Wilcoxon test:

  $X_{11}, ..., X_{m1} \sim F_1$

  $X_{12}, ..., X_{n2} \sim F_2$

  H_0: $F_1 = F_2$ vs H_1: $F_1 \neq F_2$.

  Define R_{i1} = rank of X_{i1} in the pooled sample.

  Test stat: $R_1 = \sum_{i=1}^{m} R_{i1}$

  Decision rule: reject H_0 if R_1 is too small or too large.

M-W form:

$$U(X_{i1}, X_{j2}) = \begin{cases} 1 & \text{if } X_{i1} > X_{j2}; \\ 0 & \text{if } X_{i1} = X_{j2}; \\ -1 & \text{if } X_{i1} < X_{j2}. \end{cases}$$

Test stat: $U = \sum_{i=1}^{m} \sum_{j=1}^{n} U(X_{i1}, X_{j2})$

Decision rule: reject H_0 if $|U|$ is too large.

The two tests are equivalent because

$$R_1 = \frac{m(m+n+1)}{2} + \frac{U}{2}.$$

- Now, with right-censored data

  Sample 1: (T_{i1}, δ_{i1}), $i = 1, ..., n_1$.

  Sample 2: (T_{i2}, δ_{i2}), $i = 1, ..., n_2$.

$$\phi[(T_{i1}, \delta_{i1}), (T_{h2}, \delta_{h2})]$$

$$= \begin{cases} 1 & \text{if } (T_{i1} \le T_{h2},\ \delta_{i1} = 1,\ \delta_{h2} = 0) \\ & \text{or } (T_{i1} < T_{h2},\ \delta_{i1} = 1,\ \delta_{h2} = 1)\ ; \\ -1 & \text{if } (T_{i1} \ge T_{h2},\ \delta_{i1} = 0,\ \delta_{h2} = 1) \\ & \text{or } (T_{i1} > T_{h2},\ \delta_{i1} = 1,\ \delta_{h2} = 1)\ ; \\ 0 & \text{otherwise.} \end{cases}$$

$$Z_1(\tau) = \sum_{i=1}^{n_1} \sum_{h=1}^{n_2} \phi[(T_{i1}, \delta_{i1}), (T_{h2}, \delta_{h2})].$$

§7.4 Test for trend

- H_0: $h_1(t) = h_2(t) = ... = h_K(t)$ for all $t \leq \tau$
  vs H_1: $h_1(t) \leq h_2(t) \leq ... \leq h_K(t)$ for all $t \leq \tau$ and at least one strict inequality holds.

- Use previous tests: fine?
  Yes, ...
  No,

- Choose score $a_1 < a_2 < ... < a_K$, often $a_j = j$.

$$Z = \frac{\sum_{j=1}^{K} a_j Z_j(\tau)}{\sqrt{\sum_{j=1}^{K} \sum_{g=1}^{K} a_j a_g \sigma_{jg}^2}} \overset{a.}{\sim} N(0,1) \text{ under } H_0,$$

  where $\sigma_{jg}^2 = Cov(Z_j(\tau), Z_g(\tau))$ as given before.

- Note: $\sum_j (a_j - \bar{a}) Z_j = \sum_j a_j Z_j$,

- Corr coef for (x_i, y_i), $i = 1, .., n$

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

- $y_i = b_0 + b_1 x_i + \epsilon_i$,
  $\hat{b}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$.
  So, Z is testing H_0: $b_1 = 0$ with (x_i, y_i) replaced by ....

- Example 7.6: SAS.

§7.5 Stratified tests

- Why to stratify?
  Short answer:
  assoc(X,Y) $\neq$ assoc(X,Y|Z) $\Longrightarrow$ Z is ...
  Example: StarTribune, Oct 16, 2005

	Birth type	
Race	Out of wedlock	Others
Black	69%	
others	35%	

  $\Longrightarrow$ assoc b/w ...
  Q: Is it really due to race or ...?
  Other examples: earning difference b/w genders; mortality rates of newborns in USA and Swede; mortality rates of the general populations in USA and Japan;...
- How to handle?

for a binary outcome and a binary risk factor, use ... then ...

- A deep understanding of confounding:
  Z is a confounder iff 1) Z is associated with the outcome and 2) Z is associated with the risk factor.

- Back to the current context: M strata and K groups.
  H_0: $h_{1s}(t) = h_{2s}(t) = ... = h_{Ks}(t)$, $s = 1, ..., M$.
  H_1: not H_0.

- Approach:
  1) form M strata.
  2) for each stratum s, get Z_{js} and Σ_s as before, $j = 1, ..., K$ and $s = 1, ..., M$.
  3) $Z_{j.} = \sum_{s=1}^{M} Z_{js}$, $Z_{.} = (Z_{1.}, ..., Z_{K-1,.})'$, $\Sigma_{.} = \sum_{s=1}^{M} \Sigma_s$
  4) Test stat

$$\chi^2 = Z_{.}'\Sigma_{.}^{-1} Z_{.} \overset{a.}{\sim} \chi^2_{K-1} \text{ under } H_0.$$

- Example 7.4: SAS and R.
- Application to matched data: each matched set is a ...
  read Example 7.8.

§7.6 Renyi type tests

- Consider the logrank test for $K = 2$: $h_1(t)$ going down while $h_2(t)$ going down in t; they cross.
  $Z_1 = \sum_i (O_i - E_i)$
  In early times, $O_i > E_i$; in later times, $O_i < E_i$
  $\Longrightarrow$ early $(O_i - E_i) > 0$ terms cancel out with late $(O_i - E_i) < 0$ terms
  $\Longrightarrow$ small $|Z_1| \Longrightarrow$ ...
- The strategy used so far:
- An alternative: an analog of the Kolmogorov-Smirnov test

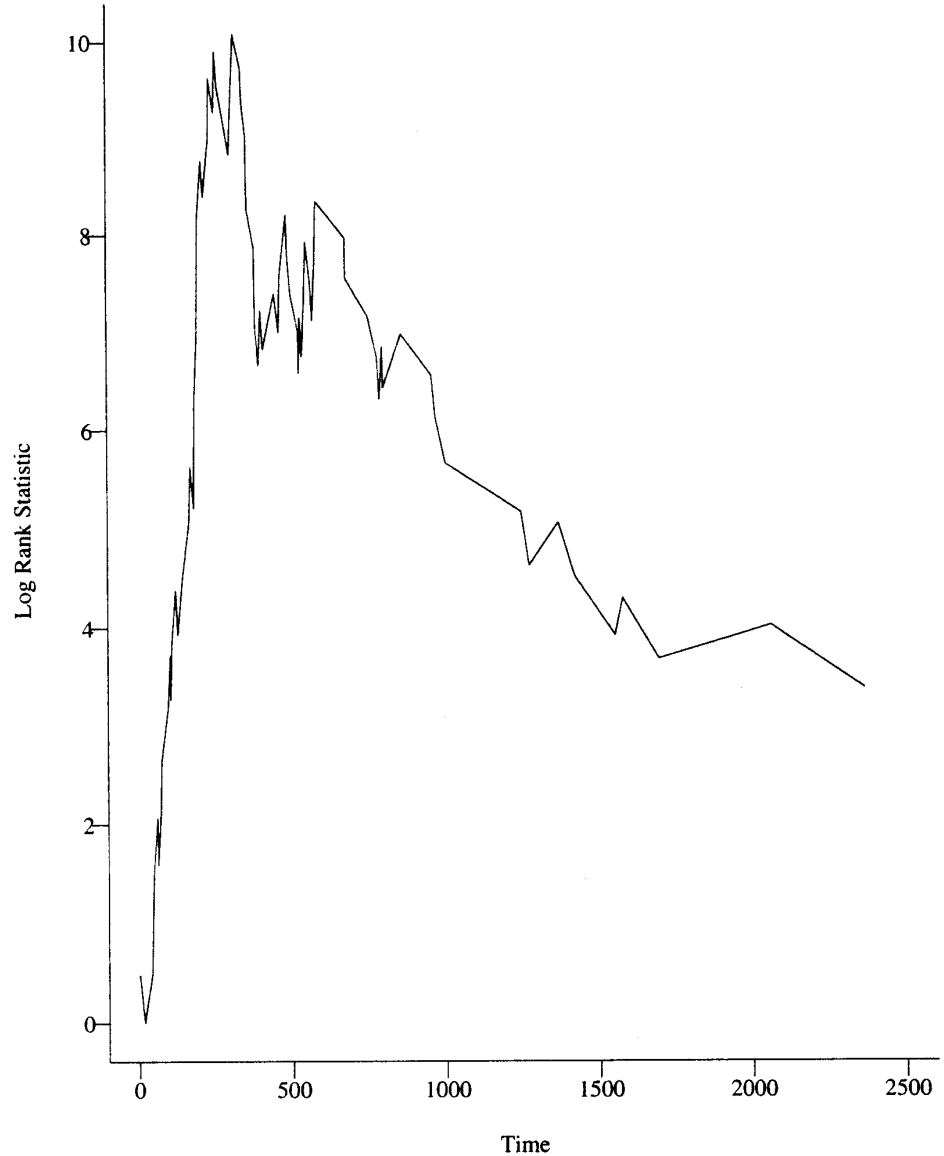
$$KS = \max_t |\hat{F}_1(t) - \hat{F}_2(t)|$$

- Generalized linear rank tests:
  $Z(t) = \sum_{t_i < t} W(t_i) \left(d_{i1} - y_{i1} \frac{d_i}{y_i} \right)$;
  It is a function of t;

Fig 7.4: $W(t) = 1$.

- $Var(Z(\tau)) = \sigma^2(\tau)$.
  $Q = sup_{t \le \tau}\{|Z(t)|\}/\sigma(\tau) \overset{a.}{\sim}$ some distribution (see Table C.5 in Appendix C) under H_0.

- Example 7.9
  Fig 7.5: Log-rank test: $Z(\tau) = -2.15$, $\sigma(\tau) = 4.46$, $p = 0.6295$; $Q = 2.20$, $p = 0.053$.



20-1

Figure 7.4 *Values of* $|Z(t_i)|$ *for the gastrointestinal tumor study*

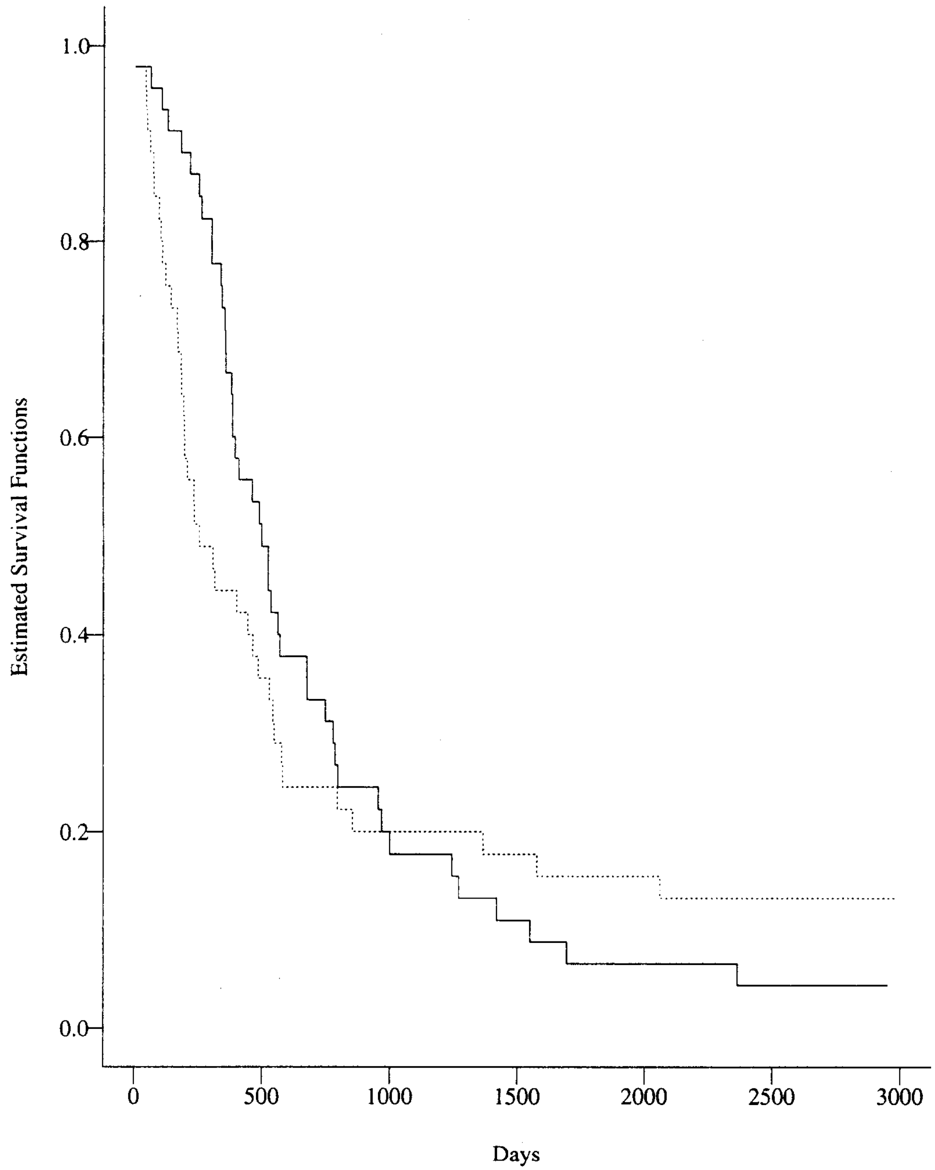


Figure 7.5 *Estimated survival functions for the gastrointestinal tumor study. Chemotherapy only (———) Chemotherapy plus radiation (------)*

§7.7 Other tests

- Cramer-von Mises test: no censoring,
  $\int [\hat{F}_1(t) - \hat{F}_2(t)]^2 d\hat{F}_p(t)$,
  where $\hat{F}_j$ is eCDF for sample j, and $\hat{F}_p$ is eCDF from the pooled sample.

- For censored data,
  $\int_0^\tau [\hat{S}_1(t) - \hat{S}_2(t)]^2 d[-\hat{S}_p(t)]$.
  but its asymptotic distribution is hard to derive

- Use
  $\tilde{H}_j(t) = \sum_{t_i \le t} \frac{d_{ij}}{y_{ij}}$, $j = 1, 2$.
  $\sigma_j^2(t) = \sum_{t_i \le t} \frac{d_{ij}}{y_{ij}^2}$ or ...
  $\sigma^2(t) = \sigma_1^2(t) + \sigma_2^2(t)$.
  $Q_1 = \left(\frac{}{\sigma^2(\tau)}\right)^2 \int_0^\tau [\tilde{H}_1(t) - \tilde{H}_2(t)]^2 d\sigma^2(t) = ... \overset{a.}{\sim}$ some distribution given in Table C.6 in Appendix C.

- Weighted K-M test:

$$W_{KM} = \sqrt{\frac{n_1 n_2}{n}} \int_0^{\tau} W(t)[\hat{S}_1(t) - \hat{S}_2(t)]dt = ...$$

$W_{KM}/\sqrt{Var} \overset{a.}{\sim} N(0,1)$; see (7.7.8) on p.230 for the formula for Var.

A special case: $W(t) = 1 \Longrightarrow W_{KM} = c[\hat{\mu}_1(\tau) - \hat{\mu}_2(\tau)]$.

§7.8 Test survival difference at a given t_0

- H_0: $S_1(t_0) = ... = S_K(t_0)$ vs H_1: at least one equality does not hold.
- Given data $\Longrightarrow \hat{S}_j = \hat{S}_j(t_0)$, $V_j = Var(\hat{S}_j)$.
- $K = 2$,
  $Z = \frac{\hat{S}_1 - \hat{S}_2}{\sqrt{V_1 + V_2}} \overset{a.}{\sim} N(0, 1)$.
- $K > 2$, $H_0 \Leftrightarrow H_0'$: $LS = b$, $S = (S_1, ..., S_K)'$,
  $X^2 = (L\hat{S} - b)'(LVL')^{-1}(L\hat{S} - b) \sim \chi^2_k$ under H_0,
  $k = rank(L)$, $V = Cov(\hat{S}) = ...$
- Examples:
  1. H_0: $S_1(t_0) = S_2(t_0) = S_3(t_0)$; $L = ...$, $b = ...$, $k = ...$
  2. H_0: $S_1(t_0) = S_2(t_0) = S_3(t_0) = 0.5$; $L = ...$, $b = ...$, $k = ...$
  3. H_0: $S_1(t_0) = S_2(t_0) = 2S_3(t_0) + 0.1$; $L = ...$, $b = ...$, $k = ...$