Statistics for Biology and Health
Chapter 6 Hypothesis Testing

Qi Guo

July 21,2019

THE UNIVERSITY OF TEXAS AT DALLAS
School of Natural Sciences and Mathematics

Contents

Introduction

Introduction

- In this chapter, we focus on hypothesis tests that are based on comparing the NA estimator, obtained directly from the data, to an expected estimator of the cumulative hazard rate, based on the assumed model under the null hypothesis(H_0).
- Rather than a direct comparison of these two rates, we examine tests at weighted differences between the observed and expected hazard rates. The weights will allow us to put more emphasis on certain parts of the curves.

One-Sample Tests

Introduction

- Suppose that we have a censored sample of size n from some population, test the hypothesis that the population hazard rate is $h_0(t)$ for all $t \leq \tau$, against the alternative that not $h_0(t)$.
- An estimate of the cumulative hazard rate function $H(t)$ is the NA estimator, given by $\sum_{t_i \leq t} \frac{d_i}{Y(t_i)}$, where d_i is the number of events at the observed event times, $t_1, \cdots, t_D$ and $Y(t_i)$ is the number of individuals under study just prior to the observed event time t_i, and hence get the crude estimate of the $h_0(t_i)$.
- Compare the sum of weighted differences between the observed and expected hazard rates to test the H_0

Test statistic

- Let $W(t)$ be a weight function with the property that $W(t)$ is zero whenever $Y(t)$ is zero. The test statistic is:

$$Z(\tau) = O(\tau) - E(\tau) = \sum_{i=1}^{D} W(t_i)\frac{d_i}{Y(t_i)} - \int_0^{\tau} W(s)h_0(s)ds$$

- If H_0 is true, the variance is:

$$V[Z(\tau)] = \int_0^{\tau} W^2(s)\frac{h_0(s)}{Y(s)}ds$$

- For large samples, the statistic $Z(\tau)^2/V[Z(\tau)]$ has a central chi-squared distribution when the H_0 is true.

Choose weight function

- The most popular choice of a weight function is the weight $W(t) = Y(t)$ which yields the one-sample log-rank test.
- For left truncation, let T_j be the time on study and L_j be the delayed entry time for the jth patient. When τ is equal to the largest time on study:

$$O(\tau) = \textit{observed number of events at or prior to time } \tau$$

And

$$E(\tau) = V[Z(\tau)] = \sum_{j=1}^{n} [H_0(T_j) - H_0(L_j)]$$

Harrington and Fleming function

- Other weight functions like Harrington and Fleming function $W_{HF} = Y(t)S_0(t)^p[1 - S_0(t)]^q, p \geq 0, q \geq 0$.
- By choice of p and q, one can put more weight on early departures from the H_0(p much larger than q).
- The log-rank weight is a special case of this model with $p = q = 0$.

Tests for Two or More Samples

- Now we can extend these methods to the problem of comparing hazard rates of $K(K \geq 2)$ populations, that is, test the following set of hypothesis: $H_0 : h_1(t) = h_2(t) = \cdots = h_K(t),\ \textit{for all } t \leq \tau$ vs $H_A\ :\ \textit{at least one of them is different}$.Here τ is the largest time at which all of the groups have at least one subject at risk.
- Let $t_1 < t_2 < \cdots < t_D$ be the distinct death times in the pooled sample, at time t_i we observed d_{ij} events in the jth sample out of Y_{ij} individuals at risk, $j = 1, 2, \cdots, k$, $i = 1, 2, \cdots, D$. Let $d_i = \sum_{j=1}^{K} d_{ij}$ and $Y_i = \sum_{j=1}^{K} Y_{ij}$ be the number of deaths and at risk in the combined sample at time t_i.

Test statistic

- The test of H_0 based on weighted comparisons of the estimated hazard rate of the jth population under H_0 and H_A, based on the NA estimator.
- Let $W_j(t)$ be a positive weight function with the property that $W_j(t_i) = 0$ whenever $Y_{ij} = 0$.
- For H_0, the test statistics:

$$Z_j(\tau) = \sum_{i=1}^{D} W_j(t)\{\frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i}\}. \; j = 1, \cdots, K$$

- If all the $Z_j(\tau)$'s are close to zero then there is little evidence to believe that H_0 is false, otherwise...

Test statistic

- All of commonly used tests have a weight function $W_j(t_i) = Y_{ij}W(t_i)$, $W(t_i)$ is a common weight shared by each group, and Y_{ij} is the number at risk in the jth group at time t_i, so now:

$$Z_j(\tau) = \sum_{i=1}^{D} W(t_i)[d_{ij} - Y_{ij}(\frac{d_i}{Y_i})] \; j = 1, \cdots, K$$

- This class of weights the test statistic is the sum of the weighted difference between the observed number of deaths and the expected number of death under H_0 in the jth sample.
- The variance of $Z_j(\tau)$ is given by:

$$\hat{\delta}_{ij} = \sum_{i=1}^{D} W(t_i)^2 \frac{Y_{ij}}{Y_i}(\frac{Y_i - d_i}{Y_i - 1})d_i. \; j = 1, \cdots, K$$

Test statistic

- and the covariance of $Z_j(\tau)$ and $Z_g(\tau)$ is:

$$\hat{\delta}_{jg} = -\sum_{i=1}^{D} W(t_i)^2 \frac{Y_{ij}}{Y_i} \frac{Y_{jg}}{Y_i} (\frac{Y_i - d_i}{Y_i - 1}) d_i.\ g \neq j$$

- The test statistic is given by the quadratic form:

$$\chi^2 = (Z_1(\tau), \cdots, Z_{K-1}(\tau)) \sum^{-1} (Z_1(\tau), \cdots, Z_{K-1}(\tau))'$$

- When H_0 is true, it has the χ^2_{K-1}, when $K = 2$, the test statistic can be written as a $N(0, 1)$.

Choose weight function

- Common choice of weight function $W(t) = 1$ for all t, that's so called log-rank test
- A second choice of weights is $W(t_i) = Y_i$, it yields Gehan's generlization of the two-sample Mann-Whitney- Wilcoxon test and Breslow's generalization of the Kruskal-Wallis test.
- An alternate censored-data version of the Mann-Whitney-Wilcoxon test was proposed by Peto and Peto, and Kalbfleisch and Prentice, which define an estimate of the common survival function by

$$\tilde{S}(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{Y_i + 1})$$

which is close to the pooled KM estimator, suggest using $W(t_i) = \tilde{S}(t_i)$.

Tests for Trend

Determine the bandwidth

- Now we test the trend for K population hazard rates, the H_0 is still the same as before, but the H_A is expressed:

$$H_A: \ h_1(t) \leq h_2(t) \leq \cdots h_K(t) \textit{ for } t \leq \tau, \textit{with at least one strict inequalit}$$

- The test will be based on the statistics $Z_j(\tau)$ given by last section, and the weight functions discussed before can be used here.

Test statistic

- To construct the test, a sequence of scores $a_1 < a_2 < \cdots < a_K$ is selected, any increasing set of scores can be used, and the test is invariant under linear transformations of the scores.
- In most cases, the scores $a_j = j$ are used, and the test statistic is:

$$Z = \frac{\sum_{j=1}^{K} a_j Z_j(\tau)}{\sqrt{\sum_{j=1}^{K} \sum_{g=1}^{K} a_j a_g \hat{\sigma}_{jg}}}$$

- When the H_0 is true and the sample sizes are sufficiently large, then, this statistic has a $N(0, 1)$ distribution.

Stratified Tests and Renyi Type Tests

Stratified Tests

- When some other covariates that affect the event rates in the K populations, one approach is to imbed the problem of comparing the K populations into a regression function, the other one is to base the test on a stratified version of one of the tests in two or more samples.
- Assume that over test is to be stratified on M levels of a set of covariates.

$$H_0 : h_{1s}(t) = h_{2s}(t) = \cdots = h_{Ks}(t) \ for \ s = 1, \cdots, M, \ t < \tau$$

- Based only on the data from the sth strata, let $Z_{js}(\tau)$ be the statistic as before, and let $Z_{j.}(\tau) = \sum_{s=1}^{M} Z_{js}(\tau)$ and $\hat{\sigma}_{jg.} = \sum_{s=1}^{M} \sigma_{jgs.}$, the test statistic is:

$$\frac{\sum_{s=1}^{M} Z_{1s}(\tau)}{\sqrt{\sum_{s=1}^{M} \hat{\sigma}_{11s}}} \sim N(0,1)$$

Renyi Type tests

- When these tests are applied to samples from populations where the hazard rates cross, these tests have little power because early differences in favor of one group are canceled out by late differences in favor of the other treatment.
- The test statistics to be used are called Renyi statistics.

Renyi Type tests

- When the hazard rates cross, the absolute value of these sequential evaluations of the test statistic will have a maximum value at some time point prior to the largest death time.
- When this value is too large, then, the $H_0 : h_1(t) = h_2(t), t < \tau$ is rejected in favor of $H_A : h_1(t) < h_2(t)$, for some t, and we have to construct multiple test statistics on the same set of data, a correction is made to the critical value of the test.
- Suppose that we have two independent samples of size n_1 and n_2,respectively. Let $n = n_1 + n_2$, and $t_1 < t_2 < \cdots < t_D$ be the distinct death times, d_{ij} and Y_{ij} be the number of deaths and at risk in t_i and sample j, where $i = 1, \cdots, D$, $j = 1, 2$.Let $Y_i = Y_{i1} + Y_{i2}$, and $d_i = d_{i1} + d_{i2}$ be the total.

Renyi Type tests

- Let $W(t)$ be the weight function, for example, for log-rank function, let $W(t) = 1$ as before, and for the Gehan-Wilcoxon version, let $W(t_i) = Y_{i1} + Y_{i2}$, for each value of t_i,we compute,$Z(t_i)$ as:

$$Z(t_i) = \sum_{t_k \leq t_i} W(t_k)[d_{k1} - Y_{k1}\frac{d_k}{Y_k}],\ i = 1, \cdots, D$$

- Let $\sigma(\tau)$ be the standard error of $Z(\tau)$ which from tests for two or more samples that section

$$\sigma^2(\tau) = \sum_{t_k \leq \tau} W(t_k)^2 \left(\frac{Y_{k1}}{Y_k}\right)\left(\frac{Y_{k2}}{Y_k}\right)\left(\frac{Y_k - d_k}{Y_k - 1}\right)(d_k)$$

where τ is the largest t_k with $Y_{k1}, Y_{k2} > 0$

Renyi Type tests

- The test statistic for a two-sided alternative is given by:

$$Q = sup\{|Z(t)|,\ t \leq \tau\}/\sigma(\tau)$$

- When the H_0 is true, then, the distribution of Q can be approximated by the distribution of the $sup(|B(x)|),\ 0 \leq x \leq 1$, where B is a standard Brownian motion process.

Other Two-Sample Tests

Introduction

- In this section, we present three other two-sample tests which constructed to have greater power than the tests to detect crossing hazard rates
- And here just introduce the first one, a censored-data version of the Cramer-von Mises test.
- For right-censored data, it is more appropriate to base a test on the integrated, squared difference between the two estimated hazard rates, and we present two versions of the test.

Cramer-von Mises test

- Recall that the NA estimator of the cumulative hazard function in the jth sample is given by:

$$\tilde{H}_j(t) = \sum_{t_i \le t} \frac{d_{ij}}{Y_{ij}}, \; j = 1, 2.$$

and the variance is given by:

$$\sigma^2{}_j(t) = \sum_{t_j \le t} \frac{d_{ij}}{Y_{ij}(Y_{ij} - 1)}, \; j = 1, 2.$$

- This test is based on the difference between $\tilde{H}_1(t)$ and $\tilde{H}_2(t)$, so that we need to compute $\sigma^2(t) = \sigma^2{}_1(t) + \sigma^2{}_2(t)$, which is estimated variance of $\tilde{H}_1(t) - \tilde{H}_2(t)$.

Cramer-von Mises test

- Let $A(t) = n\sigma^2(t)/[1 + n\sigma^2(t)]$, the first version of the Cramer-von Mises statistic is given by

$$Q_1 = \left(\frac{1}{\sigma^2(\tau)}\right) \int_0^\tau [\tilde{H}_1(t_i) - \tilde{H}_2(t)]^2 [\sigma^2(t_i) - \sigma^2(t_{i-1})]$$

  where $t_0 = 0$, and the sum is over the distinct death times less than τ.

- When the H_0 is true, the large sample distribution of Q_1 is the same as that of $R_1 = \int_0^1 [B(x)]^2 dx$, where $B(x)$ is a standard Brownian motion process.

Cramer-von Mises test

- An alternate version of the Cramer-von Mises statistic is given by:

$$Q_2 = n \int_0^{\tau} \frac{[\tilde{H}_1(t) - \tilde{H}_2(t)]^2}{1 + n\sigma^2(t)} dA(t)$$

  which is computed as

$$Q_2 = n \sum_{t_i \leq \tau} \frac{[\tilde{H}_1(t_i) - \tilde{H}_2(t_i)]^2}{1 + n\sigma^2(t_i)} [A(t_i) - A(t_{i-1})]$$

- When the H_0 is true, the large sample distribution of Q_1 is the same as that of $R_2 = \int_0^{A(\tau)} [B^0(x)]^2 dx$, where $B^0(x)$ is a standard Brownian motion process.