Estimation of Location Shift with Stratification

Jiarui Chi, Adam Boyd

Pfizer Inc.

October 2020

Overview

Motivation

- Patients with dilated cardiomyopathy are randomized (2:1) to receive experimental treatment:placebo.
- Primary endpoint: Change from baseline (CFB) in 6-minute walk distance (6MWD) at Week 12.
- Model:

$$X_{ik} = \Delta + t_k + \epsilon_{ik},$$
$$Y_{jk} = t_k + \eta_{jk},$$

  where $X_{ik}(Y_{jk})$ is the CFB value for the ith (jth) patient in the kth stratum from the experimental (placebo) arm, $i = 1, \ldots, m_k(n_k)$, $k = 1, \ldots, d$, Δ is the treatment effect, t_k is the (fixed) stratum effect and $\epsilon_{ik}, \eta_{jk} \sim$ i.i.d. continuous distribution F.
- Goal: Estimate Δ and construct confidence interval without strict assumption on F.

Unstratified point estimator

- Suppose we ignore the strata in our model.
- Note that under the model assumptions, $X_1 - \Delta, \ldots, X_m - \Delta$ and $Y_1, \ldots, Y_n$ should have the same distribution.
- Thus, a natural estimator is the amount $\hat{\Delta}$ to be subtracted from X to make the observations from the 2 arms align as closely as possible.
- We charaterize this by the pairwise differences $((X_i - \Delta) - Y_j)$ that should be symmetrically distributed around zero.
- Choose $\hat{\Delta}$ such that half of the differences $((X_i - \hat{\Delta}) - Y_j)$ are greater than zero and half are less than zero. A viable estimator $\hat{\Delta}_{HL}$ is the point $\hat{\Delta}$ such that the median of pairwise differences $((X_i - \hat{\Delta}) - Y_j)$ is zero.
- This estimator $\hat{\Delta}_{HL}$ is known as the "Hodges-Lehmann" (HL) estimator [1]. One of the most important properties is the asymptotic normality $\sim \mathcal{N}(\Delta, \sigma^2_{HL})$.

[1] Hodges and Lehmann, 1963

Unstratified interval estimator

- We can also use a non-parameteric testing procedure for constructing point and interval estimators, simply through the p-value function.
- Wilcoxon rank-sum test (WRST):
  1. Rank all observations in the combined sample $\{X, Y\}$, and let S_i be the rank of the ith observation from the experimental group;
  2. Test statistic: $W(X,Y) = \sum_{i=1}^{m} S_i$;
  3. Asymptotic Normality: $W(X,Y) \sim \mathcal{N}(\mu_W, \sigma_W^2)$.
- $W(X - \Delta, Y)$ should be close to μ_W in large sample approximation.
- We would expect the p-value using $W(X - \hat{\Delta}, Y)$, i.e. $p(W(X - \hat{\Delta}, Y))$, to be close to 0.5 in a one-sided test under the null. Thus, the inverse of $p(W(X - h, Y)) \approx 0.5$ as an estimator is similar to the HL estimator (e.g. same asymptotic efficiency).

Unstratified interval estimator

- Lehmann [2] proposed a $(1 - \alpha)$ confidence interval composed of the quantiles of the ordered sequence of pairwise differences, i.e. the $\alpha/2$ quantile and the $1 - \alpha/2$ quantile.
- A $(1 - \alpha)$ confidence interval (Δ^L, Δ^U) follows by finding the points such that $p(W(X - h, Y)) \approx \alpha/2$ and $p(W(X - h, Y)) \approx 1 - \alpha/2$, respectively.

[2] Lehmann, 1963

Stratified case

- A stratified version of the WRST was developed by van Elteren [3], which simply combines the within-stratum Wilcoxon rank-sum statistics, $W_{vE} = \sum_{k=1}^{d} c_k W_k$ for some weights c_k.
- Weights $c_k = 1/(m_k + n_k + 1)$ are chosen by van Elteren such that the vE test has the largest power in our model.
- Interestingly, to our knowledge, a stratified estimator has not been evaluated in the literature.
- We have investigated several candidate stratified estimators utilizing the following approaches:
  1. Inverting the van Elteren test through the p-value function;
  2. Using a linear combination of the within-stratum HL estimators;
  3. Using the unstratified HL estimator after first removing the within-stratum effects (so-called "ranking after alignment").
- For each of the above estimators, we provide large sample properties and demonstrate performance via simulations.

[3] van Elteren, 1960

Inverting the van Elteren test

- We can use the same approach used when inverting the WRST to find point and interval estimators for Δ by inverting the van Elteren test.
- That is to find the points such that $p(W_{vE}(X_k - h), Y_k) \approx 0.5, \alpha/2, 1-\alpha/2$.
- We have established that
  1. This point estimator $p_{0.5}$ is consistent and asymptotic normal $\sim \mathcal{N}(\Delta, \sigma_d^2)$.
  2. The interval estimator $(p_{\alpha/2}, p_{1-\alpha/2})$ has asymptotic confidence level $(1-\alpha)$.
  3. A conservative confidence interval has also been established to achieve at least $(1-\alpha)$.
  4. When $d = 1$, i.e. no stratification, $\sigma^2_{d=1}$ of this point estimator becomes σ^2_{HL} of the HL estimator. This means $p_{0.5}$ and the HL estimator have the same asymptotic efficiency.

Linear combination of the within-stratum Hodges Lehmann estimators

- A natural estimator arises by considering independently estimating Δ within each stratum and then combine these d estimates into an overall estimator. That is $\hat{\Delta}_{LCHL} = \sum_{k=1}^{d} w_k \hat{\Delta}_k$, where $\hat{\Delta}_k$ is the HL estimator in kth stratum and $\sum_{k=1}^{d} w_k = 1, w_k > 0$.
- We have established that
  1. This LCHL is asymtotically median unbiased and normally distributed $\sim \mathcal{N}(\Delta, \sigma_s^2)$.
  2. The asymptotic $(1-\alpha)$ confidence interval can be constructed by the asymptotic normal distribution. σ_s^2 can be replaced with the consistent estimators proposed by Lehmann [4], which estimated by a combination of the length of within-stratum confidence intervals through Delta method.

[4] Lehmann, 1963

Ranking after alignment

- A third option arises by considering an approach proposed by Hodges and Lehmann [5] in the testing situation.
- They proposed first substracting off an estimate of the within-stratum effect across each stratum from the observed data and then testing the pseudo unstratified data using an unstratified approach (e.g. WRST).
- The within-stratum estimate can be obvious choices such as the within-stratum mean or Hodges Lehmann estimate [6].
- Similarly, we investigated using the HL estimator on these pseudo data. We call such estimator an aligned HL (AHL) estimator.

[5] Hodges and Lehmann, 1962
[6] Mehrotra, 2010

Simulation results

Consider the following model

$$x_{ik} = \Delta + t_k + \epsilon_{ik},$$
$$y_{jk} = t_k + \eta_{jk},$$

where $\{\epsilon_{ik}, \eta_{jk}\}$ are independent identically distributed random errors, t_k is the fixed stratum effect and Δ is the fixed treatment effect (location shift). We present the point estimate and coverage probability $(\alpha = 0.05)$ of

- $p_{0.5}$, with $(p_{\alpha/2}, p_{1-\alpha/2})$ as the confidence interval and a conservative confidence interval denoted by (C) in the following tables.
- LCHL with $w_k = (m_k + n_k)/(\sum_{k=1}^{d} m_k + \sum_{k=1}^{d} n_k)$.
- AHL by mean subtraction.
- Hodges Lehmann (HL) estimator.

Simulation results

		$t \neq 0$	
Δ	d	Point Estimate	Coverage Probability%
		$p_{0.5}$/LCHL/AHL/HL	$p_{0.5}$/$p_{0.5}$(C)/LCHL/AHL/HL
0	2	0.00/0.00/0.00/0.00	95.24/95.29(C)/95.24/95.26/97.58
	5	0.00/0.00/0.00/0.00	94.80/94.86(C)/95.08/94.77/96.23
	30	0.00/0.00/0.00/0.00	94.76/95.05(C)/96.59/94.23/95.75
	60	-0.01/0.00/0.00/0.00	94.89/95.57(C)/99.07/93.65/95.74
1	2	1.00/1.00/1.00/1.00	95.19/95.28(C)/95.24/95.26/97.58
	5	1.00/1.00/1.00/1.00	94.74/94.85(C)/95.07/94.77/96.23
	30	1.00/1.00/1.00/1.00	94.37/95.16(C)/96.59/94.23/95.75
	60	1.00/1.00/1.00/1.00	93.62/95.55(C)/99.07/93.65/95.74

Table: Overall Results for $N = 600, \mathcal{N}(0, 2^2)$ with equal allocation ($1:2$) and stratification after 10000 simulation runs. The conservative confidence intervals are denoted by (C).

Simulation results

		$t \neq 0$	
Δ	d	Point Estimate	Coverage Probability%
		$p_{0.5}$/LCHL/AHL/HL	$p_{0.5}/p_{0.5}$(C)/LCHL/AHL/HL
0	2	0.00/0.00/0.00/0.00	95.08/95.08(C)/98.68/96.74/97.02
	5	0.00/0.00/0.00/0.00	95.01/95.02(C)/98.93/98.02/96.43
	30	0.00/-0.01/-0.01/0.00	95.22/95.39(C)/99.96/98.39/96.08
	60	0.00/-0.01/-0.01/0.00	94.94/95.39(C)/100.0/98.71/96.13
1	2	1.00/1.01/1.01/1.00	95.63/95.66(C)/98.80/96.93/97.26
	5	1.00/1.02/1.00/1.00	95.16/95.19(C)/99.01/98.04/96.36
	30	1.00/1.17/1.00/1.00	95.17/95.47(C)/99.85/98.26/95.96
	60	1.00/1.88/1.00/1.00	94.87/95.45(C)/99.92/97.79/96.13

Table: Overall Results for $N = 600, \mathcal{LN}(0, 2^2)$ with equal allocation $(1:2)$ and stratification after 10000 simulation runs. The conservative confidence intervals are denoted by (C).

Conclusions

- $p_{0.5}$ is overall robust but needs a good optimization algorithm or a fine grid search.
- LCHL is robust to the allocation but sensitive to heavy-tailed distributions. A rule of thumb of sample size in each cell (per treatment group per stratum) ≥ 50 for not heavy-tailed distributions.
- AHL is robust to distributions (not heavy-tailed) but sensitive to the allocation.
- For other interesting discussion, remember to attend *GSC* in October!

References

van Elteren (1960) On the combination of independent two sample tests of wilcoxon. *Bull Inst Intern Staist*, 37:351-361.

Hodges Jr, J. L. and Lehmann, E. L. (1963) Estimates of location based on rank tests. *Annals of Mathematical Statistics*, pages 598-611.

Hodges Jr, J. L. and Lehmann, E. L. (1962) Rank methods for combination of independent experiments in analysis of variance. *The Annals of Mathematical Statistics*, 33(2):482-497.

Lehmann, E. (1963) Nonparametric confidence intervals for a shift parameter. *The Annals of Mathematical Statistics*, 34(4):1507-1512.

Mehrotra et al. (2010) Rank-based analyses of stratified experiments: alternatives to the van elteren test. *The American Statistician*, 64(2):121-130.


Thank You!