Lecture 18: ANOVA

Pratheepa Jeganathan

05/13/2019

Recall

- One sample sign test, Wilcoxon signed rank test, large-sample approximation, median, Hodges-Lehman estimator, distribution-free confidence interval.
- Jackknife for bias and standard error of an estimator.
- Bootstrap samples, bootstrap replicates.
- Bootstrap standard error of an estimator.
- Bootstrap percentile confidence interval.
- Hypothesis testing with the bootstrap (one-sample problem.)
- Assessing the error in bootstrap estimates.
- Example: inference on ratio of heart attack rates in the aspirin-intake group to the placebo group.
- The exhaustive bootstrap distribution.

- Discrete data problems (one-sample, two-sample proportion tests, test of homogeneity, test of independence).
- Two-sample problems (location problem - equal variance, unequal variance, exact test or Monte Carlo, large-sample approximation, H-L estimator, dispersion problem, general distribution).
- Permutation tests (permutation test for continuous data, different test statistic, accuracy of permutation tests).
- Permutation tests (discrete data problems, exchangeability.)
- Rank-based correlation analysis (Kendall and Spearman correlation coefficients.)
- Rank-based regression (straight line, multiple linear regression, statistical inference about the unknown parameters, nonparametric procedures - does not depend on the distribution of error term.)
- Smoothing (density estimation, bias-variance trade-off, curse of dimensionality)
- Nonparametric regression (Local averaging, local regression, kernel smoothing, local polynomial, penalized regression)

- Cross-validation, Variance Estimation, Confidence Bands, Bootstrap Confidence Bands.
- Wavelets (wavelet representation of a function, coefficient estimation using Discrete wavelet transformation, thresholding - VishuShrink and SureShrink).

ANOVA (Analysis of variance)

Overview

- Decomposing total variance in the observed data into variability explained by treatment, block, error (considering the experimental design).

The One-Way Layout

The One-Way Layout

- Primary interest is about the relative locations (medians) of three or more populations (k populations).
- Data: random samples from each of k population.

Treatments			
1	2	$\cdots$	k
X_{11}	X_{12}	$\cdots$	X_{1k}
X_{21}	X_{22}	$\cdots$	X_{2k}
$\vdots$	$\vdots$		$\vdots$
$X_{n_1 1}$	$X_{n_2 2}$	$\cdots$	$X_{n_k k}$

- $N = n_1 + n_2 + \cdots + n_k$.

The One-Way Layout

- Factor A has k levels/treatments.
- Complete randomized design: N subjects are randomly assigned to k different treatments.
- Let n_j subjects randomly assigned to j-th treatment $j = 1, 2, \cdots, k$.
- Let X_{ij} be the i-th response in j-th treatment.
- Assumptions:
  - N random variables $\{X_{1j}, X_{2j}, \cdots, X_{n_j j}\}, j = 1, 2, \cdots, k$ are mutually independent.
  - $\{X_{1j}, X_{2j}, \cdots, X_{n_j j}\} \sim F_j$, F_j is continuous.
  - $F_j(t) = F(t - \tau_j)$ for $j = 1, 2, \cdots, k$, where F is a distribution function for a continuous distribution with unknown median θ and τ_j is the unknown treatment effect for the j-th population.

The One-Way Layout

- The one-way layout model

$$X_{ij} = \theta + \tau_j + e_{ij}, i = 1, 2 \cdots, n_j \quad \text{and} \quad j = 1, 2, \cdots, k,$$

  - θ - overall median,
  - τ_j is the treatment effect,
  - N $\boldsymbol{e}$'s from a random sample from a continuous distribution with median 0.

General Alternatives

General alternatives (KRUSKAL–WALLIS)

- $H_0 : [\tau_1 = \tau_2 = \cdots = \tau_k]$ versus
  $H_A : [\tau_1, \tau_2, \cdots, \tau_k$ not all equal].
- Test Statistic

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(R_{.j} - \frac{N+1}{2} \right)^2,$$

  where $R_{.j} = \frac{R_j}{n}$ and $R_j = \sum_{i=1}^{n} r_{ij}$, sum of the ranks in treatment j.
  - Motivation:
    - Under H_0, $F_1 = F_2 = \cdots = F_k = F$.
    - Combine k samples and rank.
    - Let r_{ij} denote the rank of X_{ij} in this joint ranking.
    - $\mathbb{E}(R_{.j}) = \frac{N+1}{2}$.
    - If H_A is true $R_{.j}$ would be larger than $\frac{N+1}{2}$ and H is expected to be large.
- Distribution of H under H_0 can be computed using permutation method.

Large-sample approximation

- As $\min\{n_1, n_2, \cdots, n_k\}$ goes to ∞,

$$H \sim \chi^2_{k-1}.$$

Ties

- Modification needed.
  - Small-sample procedure is only approximately of the significance level α in the presence of tied X observations.

Example 6.1 (Half-Time of Mucociliary Clearance)

- Assess mucociliary efficiency from the rate of removal of dust in the three groups:
  - normal subjects,
  - subjects with obstructive airway disease, and
  - subjects with asbestosis.
- Responses - half-times mucociliary clearance of the subjects.

Example 6.1 (Half-Time of Mucociliary Clearance)

```
normal = c(2.9, 3.0, 2.5, 2.6, 3.2)
obstruct = c(3.8, 2.7, 4.0, 2.4)
asbestosis = c(2.8, 3.4, 3.7, 2.2, 2.0)
x = c(normal,obstruct,asbestosis)
treatment = c(rep(1,length(normal)),
  rep(2,length(obstruct)),
  rep(3,length(asbestosis)))

KW = kruskal.test(x, treatment)
```

Example 6.1 (Half-Time of Mucociliary Clearance)

```
KW$statistic
```

```
## Kruskal-Wallis chi-squared
##                   0.7714286
```

```
KW$p.value
```

```
## [1] 0.6799648
```

Other alternatives

Ordered alternatives

- Interest in testing increasing treatment effect (according to natural labeling of the treatments).
  - treatments corresponding to degrees of knowledge of performance.
  - treatments corresponding to degrees of drug dosage levels, etc.
- $H_A : [\tau_1 \leq \tau_2 \cdots \leq \tau_k$ with at least one strict inequality].
- Use JONCKHEERE–TERPSTRA test statistic.
  - Use $k(k-1)/2$ Mann–Whitney counts.

Umbrella alternatives

- Interest in testing umbrella alternative treatment effect (said to have a peak at treatment population p).
  - reaction to increasing drug dosage levels where a downturn in effect may occur after the optimal dosage is exceeded.
  - effect of age on responses to certain stimuli.
- H_A :
  $[\tau_1 \leq \tau_2 \cdots \leq \tau_{p-1} \leq \tau_p \geq \tau_{p+1} \cdots \geq \tau_k$ with at least one strict ine
- Use MACK–WOLFE test statistic.
  - Use $p(p-1)/2$ Mann–Whitney counts for every pair of treatments with labels less than or equal to the hypothesized peak.
- Test is available for umbrella alternatives with unknown peak.

Test for treatments versus a control

- One of the treatments belongs to baseline set of conditions or control.
- Interest in testing which of the treatments are better than control.
- $H_0 : [\tau_i = \tau_1, i = 2, \cdots, k]$ versus H_A :
  $[\tau_i \geq \tau_1, i = 2, \cdots, k,$ with at least one strict inequality].
- Use FLIGNER–WOLFE test statistic (FW).
  - Combine sample and rank.
  - FW is the sum of ranks in noncontrol treatments.

Notes

- With all the above alternative hypotheses, we can consider k-Sample Behrens–Fisher Problem.
  - permit the possibility of differences in scale parameters as well as medians within the common F.
- Rationale for multiple comparison procedure
  - Upon rejection of H_0, test about specific pairs of treatment effects $\tau_1, \cdots, \tau_k$.

Multiple comparison procedure

- Consider a collection of hypothesis test $H_0^{(1)}, H_0^{(2)}, H_0^{(3)}, \cdots, H_0^{(s)}$.
- Individual type I error rate α_I.
  - $\alpha_I =$ Type I error rate for each test conducted as if H were alone.
- Experiment-wise type I error rate α_E.
  - $\alpha_E = P\,(\text{Type I error on at least one test})$.

Multiple comparison procedure

- If the tests in the multiple comparison procedure are independent

$$\begin{aligned} \alpha_E =& P\left(\text{Type I error on at least one test}\right) \\ &= 1 - P\left(\text{No Type I errors}\right) \\ &= 1 - (1 - \alpha_I)^s \end{aligned} \tag{1}$$

- Thus,

$$\alpha_I = 1 - (1 - \alpha_E)^{1/s}.$$

Multiple comparison procedure

- Bonferroni correction/adjustment
  - $\alpha_s = \frac{\alpha_E}{s}$.
  - Apply regardless whether test are independent or not.
  - Conservative.
- Holm procedure (modification to Bonferroni)
  - Calculate p-values for each hypothesis test.
  - Order the p-values from smallest to largest.
    - Compare smallest p-value to $\frac{\alpha_E}{s}$.
    - Compare next smallest p-value to $\frac{\alpha_E}{s-1}$.
    - Compare largest p-value to $\frac{\alpha_E}{1}$
  - Do the above procedure until we fail to reject one of $H_0^{(i)}$.

Contrast

- A contrast is $C = a_1\tau_1 + a_2\tau_2 + \cdots + a_k\tau_k$ such that $\sum_{j=1}^{k} a_j = 0$.
- If we consider k treatment effects, $\tau_1, \tau_2, \cdots, \tau_k$:
  - $H_0^1 : \tau_1 = \tau_2$, $H_0^2 : \tau_2 = \tau_3$, $\cdots$, $H_0^{k(k-1)/2} : \tau_{k-1} = \tau_k$.

Testing k(k-1)/2 (contrast) pairs of treatment effects

- To make decisions about individual differences between pairs of treatment effects (τ_i, τ_j) for $i < j$.
- Do multiple hypotheses procedure, after rejection of H_0 with the Kruskal–Wallis procedure.
- $H_0 : \tau_j = \tau_l$ versus $H_A : \tau_j \neq \tau_l$, $j \neq l = 1, 2, \cdots, k$.
  - Test statistic: For each pair of treatments (i, j),

$$W_{ij} = \sum_{j=1}^{n_j} R_{ij}, 1 \leq i < j \leq k.$$

  - Wilcoxon rank sum test.
  - We will use p-value approach to make decision.

Multiple comparison procedure (Example)

- Length of YOY Gizzard Shad from Kokosing Lake, Ohio, Sampled in Summer, 1984 (mm).
- Let $\alpha = .01$.

```
num.of.contrasts = 4*(4-1)/2;num.of.contrasts
```

```
## [1] 6
```

```
library(NSM3)
data(gizzards)
grp = factor(c(rep("I", length(gizzards[[1]])),
  rep("II", length(gizzards[[2]])),
  rep("III", length(gizzards[[3]])),
  rep("IV", length(gizzards[[4]]))))
leng = as.numeric(unlist(gizzards))
```

Multiple comparison procedure (Example)

```
kw.test  = kruskal.test(leng, grp)
kw.test$p.value
```

```
## [1] 4.334659e-05
```

- We reject the null hypothesis at .01 significance level and conclude that the length of YOY Gizzard Shad is different in at least two sites of the river.

Multiple comparison procedure (Example)

- Experimentwise error rate $\alpha = .01$.

```
p.value12 = wilcox.test(gizzards[[1]],
  gizzards[[2]])$p.value
p.value13 = wilcox.test(gizzards[[1]],
  gizzards[[3]])$p.value
p.value14 = wilcox.test(gizzards[[1]],
  gizzards[[4]])$p.value
p.value23 = wilcox.test(gizzards[[2]],
  gizzards[[3]])$p.value
p.value24 = wilcox.test(gizzards[[2]],
  gizzards[[4]])$p.value
p.value34 = wilcox.test(gizzards[[3]],
  gizzards[[4]])$p.value
```

Multiple comparison procedure (Example)

```
round(c(p.value12, p.value13,
  p.value14, p.value23,
  p.value24, p.value34), digits = 3)
```

```
## [1] 0.255 0.001 0.001 0.001 0.001 1.000
```

- Bonferroni correction to p-values (multiply each p-value by number of contrasts and set the p-value more than one to one.)

```
round(p.adjust(c(p.value12, p.value13,
  p.value14, p.value23,
  p.value24, p.value34),
  method = "bonferroni"),digits = 3)
```

```
## [1] 1.000 0.007 0.004 0.008 0.006 1.000
```

At an experimentwise error rate of .01, the six multiple comparison decisions can be summarized by the statement
$(\tau_1 = \tau_2) \neq (\tau_3 = \tau_4)$)

Multiple comparison procedure (Example)

- Holm procedure.

```
round(p.adjust(c(p.value12, p.value13,
  p.value14, p.value23,
  p.value24, p.value34),
  method = "holm"), digits = 3)
```

```
## [1] 0.509 0.005 0.004 0.005 0.005 1.000
```

We reach the same conclusion using Holm multiple comparison adjustment procedure.

References for this lecture

HWC Chapter 6