STAT 135 Lecture 16

Henry Liev

30 July 2025

Remark 0.1

T_Y = Sum of the ranks of the "treatment group" (Y's)
$U_Y = \sum_{i=1}^{n} \sum_{j=1}^{m} 1\{X_i < Y_j\}$
T_Y and U_Y differ by a constant, will result in the same p-value
$\sum_{i=1}^{n} \sum_{j=1}^{m} 1\{X_i < Y_j\}, V_{ij} = 1\{X_{(i)} < Y_{(j)}\}$
$= \sum_{i=1}^{n} \sum_{j=1}^{m} V_{ij}$
$= \sum_{j=1}^{m} (X's < Y_{(j)})$
$= \sum_{j=1}^{m} (\text{Rank}(Y_{(j)}) - j) = \sum_{j=1}^{m} \text{Rank}(Y_{(j)}) - \frac{m(m+1)}{2}$

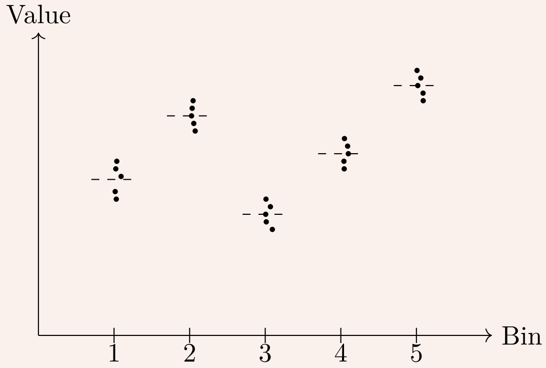
Remark 0.2 (Statistics is a Patchwork)

Problem Framing: NP Paradigm, Estimation Framework
Useful Constructs: Likelihood function, Suffienct statistics
Math Techniques: Delta Method, $\mathbb{E}((X-Y)^2) = E((X-a+a-Y)^2)$
Statistical Techniques: Method of moments, Chi-squared tests
Deep Math: Limit Theorems, CLT, Wilk's Theorem, MLE Asymptotics
Optimality Results: Cramer-Rao Lower Bound, MLE Efficiency, Likelihood Ratio Optimality

Remark 0.3 (Analysis of Variance)

Two sample problem
$H_0 : \mu_X = \mu_Y$
$H_1 : \mu_X \neq \mu_Y$
I-Sample problem
I groups with J observations in each
$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$
H_1 : At least one μ_i is different

Example 0.4 ("One-way" Analysis of Variance)



Method 0.5 ("One Way" Analysis of Variance)

J iid observation in each of I groups. with common variance σ^2
Model: $Y_{ij} = \mu + \alpha_i + \epsilon$
$\epsilon \sim \mathcal{N}(0, \sigma^2)$
$\sum \alpha_i = 0$
$\mu + \alpha_i$ constant
α_i incremental effect of treatment i
$H_0 : \alpha_i = 0 \forall i$ (Groups are all the same)
$H_1 : \alpha_i$
Assumptions:

- Normal Data
- Common Variance
- Everything Independent

Compare variation between groups to variation within groups

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i.})^2 + J\sum_{i=1}^{I}(\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$\bar{Y}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}, \bar{Y}_{i.} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij}$ Overall average vs group average
$SS_{TOT} = SS_W + SS_B$
Idea: Compare SS_B to SS_W

Theorem 0.6 (12.2A)

$E(SS_B) = J\sum_{i=1}^{I}\alpha_i^2 + (I-1)\sigma^2$
$E(SS_W) = I(J-1)\sigma^2$
Under H_0
$E(SS_B) = (I-1)\sigma^2$
$E(SS_W) = I(J-1)\sigma^2$
$\frac{SS_B}{I-1} \approx \frac{SS_W}{I(J-1)}$

Remark 0.7

Distributional Fact
$\frac{SS_B}{\sigma^2} \sim \chi^2_{I-1}$
$\frac{SS_W}{\sigma^2} \sim \chi^2_{I(J-1)}$
SS_B and SS_W are independent
$A \sim \chi^2_a, B \sim \chi^2_b$ independent
$\frac{A/a}{B/b} \sim F_{a,b}$
$T = \frac{SS_B/(I-1)}{SS_W/(I(J-1))} = \frac{(SS_B/\sigma^2)/(I-1)}{(SS_W/\sigma^2)/(I(J-1))} \sim F_{I-1,I(J-1)}$
Reject for large values (larger than 1)

Example 0.8

Data from section 12.2 in textbook 7 labs, 10 measurements/lab
I=7, J=10

Solution.

$$T = \frac{SS_B/(I-1)}{SS_W/(I(J-1))} = \frac{SS_B/6}{SS_W/63}$$

■

Remark 0.9 (Multiple Comparisons)

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$
So we've rejected H_0. Now what?
Could we do all possible t-test comparing μ_i vs μ_j?
Would get a lot of false positives
Test 10 independent hypotheses at the 5% level
What is the probability of at least one type 1 error
$1 - (0.95)^{10} \approx 0.4$

Method 0.10 (Bonferroni Correction)

If you do k tests, do them at level $\frac{\alpha}{k}$
Why? $P(A_1 \cup A_2 \cup \cdots \cup A_k) \leq P(A_1) + P(A_2) + \cdots + P(A_k) = \alpha$
Assume all H_0 are true, A_k be a type 1 error for k^{th} test
Bonferroni is conservative, works pretty well

Proof. Suppose $1\{A_1 \cup A_2 \cup \cdots \cup A_n\} > 1\{A_1\} + 1\{A_2\} + \cdots + 1\{A_n\}$
If one value on the left is non zero, then that implies that one of the values on the right is non zero
$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n)$ ■

Method 0.11 (Tukey's Method)

Less conservative, makes more assumptions, is understood by no one

Method 0.12 (Kruskal-Wallis Test)

- Non-parametric versino of ANOVA
- Follow usual routine, Y_{ij} I groups, J trials/group, $Y_{ij} \sim F_1, F_2, \ldots, F_I$, $H_0 : F_1 = F_2 = \cdots F_I$
  $Y_{ij} \to R_{ij}$ (Rank among all IJ observations)

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(R_{ij} - \bar{R}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(R_i j - \bar{R}_i)^2 + J\sum_{i=1}^{I}(\bar{R}_i - \bar{R})^2 \to SS_{TOT} = SS_W + SS_B$$

$$\bar{R}_{..} = \frac{1 + IJ}{2}$$

We can look at the between $SS_B = J\sum_{i=1}^{I}(\bar{R}_i - \bar{R})^2$, under H_0, assuming numbers 1,. . ., IJ put in at random

$$\frac{12}{IJ(IJ+1)}SS_B \approx \chi^2_{I-1}$$

Get approximate p-value