Univariate One-way Analysis of Variance (ANOVA)

Study Notes March 19, 2011 Kenneth Chu

1 Motivating Example

GOAL: To evaluate the effect of smoking on heart rate.

- Suppose the collection $\mathcal{U}$ of all human beings has been divided into four overlapping groups: $\mathcal{U}_1$ non-smokers, $\mathcal{U}_2$ light smokers, $\mathcal{U}_3$ moderate smokers, and $\mathcal{U}_4$ heavy smokers, i.e.

$$\mathcal{U} \;=\; \bigsqcup_{j=1}^{4} \mathcal{U}_j$$

- In each group, we take a random sample of 6 individuals, and measure their heart rates after three minutes of resting.
- Let μ_j be the population mean heart rate of $\mathcal{U}_j$, $j = 1, \ldots, 4$, respectively. We would like to use the collected data to test:

$$\begin{array}{c} H_0: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ \text{versus} \\ H_1: \quad \text{not all of } \mu_1,\, \mu_2,\, \mu_3, \text{ and } \mu_4 \text{ are equal.} \end{array}$$

Suppose the collected data are as follows:

non-smokers $j = 1$	light smokers $j = 2$	moderate smokers $j = 3$	heavy smokers $j = 4$
69	55	66	91
52	60	81	72
71	78	70	81
58	58	77	67
59	62	57	95
65	66	79	84

2 Sums of Squares and Their Relevant Properties

$$\begin{aligned} SS_{\text{inter-level}} &:= \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)^2 \;=\; \sum_{j=1}^{k} n_j\left(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}\right)^2 \\ SS_{\text{intra-level}} &:= \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot j}\right)^2 \\ SS_{\text{total}} &:= \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 \end{aligned}$$

Proposition 2.1

$$SS_{\text{total}} \;=\; SS_{\text{inter-level}} \;+\; SS_{\text{intra-level}}$$

Proposition 2.2

Suppose:

1. Y_{ij}*, for* $j = 1, \ldots, k$ *and* $i = 1, \ldots, n_j$*, are independent random variables.*

2. *there exist* $\sigma^2 > 0$ *and* $\mu_1, \ldots, \mu_k \in \mathbb{R}$ *such that* $Y_{ij} \sim N(\mu_j, \sigma^2)$, *for each* $j = 1, \ldots, k$, *and each* $i = 1, \ldots, n_j$.

Then, regardless of whether H_0 $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$ *is true, we have:*

- $\dfrac{SS_{\text{intra-level}}}{\sigma^2}$ *has a* χ^2 *distribution with* $(n-k)$ *degrees of freedom. In particular,* $E\left[\dfrac{SS_{\text{intra-level}}}{\sigma^2}\right] = n - k$.
- $SS_{\text{intra-level}}$ *and* $SS_{\text{inter-level}}$ *are independent random variables.*

Proposition 2.3

Suppose:

1. Y_{ij}, *for* $j = 1, \ldots, k$ *and* $i = 1, \ldots, n_j$, *are independent random variables.*
2. *there exist* $\sigma^2 > 0$ *and* $\mu_1, \ldots, \mu_k \in \mathbb{R}$ *such that* $Y_{ij} \sim N(\mu_j, \sigma^2)$, *for each* $j = 1, \ldots, k$, *and each* $i = 1, \ldots, n_j$.

Then,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \Longrightarrow \quad \frac{SS_{\text{inter-level}}}{\sigma^2} \quad \sim \quad \chi^2_{k-1}$$

Corollary 2.4

Suppose:

1. Y_{ij}, *for* $j = 1, \ldots, k$ *and* $i = 1, \ldots, n_j$, *are independent random variables.*
2. *there exist* $\sigma^2 > 0$ *and* $\mu_1, \ldots, \mu_k \in \mathbb{R}$ *such that* $Y_{ij} \sim N(\mu_j, \sigma^2)$, *for each* $j = 1, \ldots, k$, *and each* $i = 1, \ldots, n_j$.

Then,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \Longrightarrow \quad F \;:=\; \frac{SS_{\text{inter-level}}/(k-1)}{SS_{\text{intra-level}}/(n-k)} \quad \sim \quad \mathcal{F}^{k-1}_{n-k}$$

Lemma 2.5

Suppose:

1. Y_{ij}, *for* $j = 1, \ldots, k$ *and* $i = 1, \ldots, n_j$, *are independent random variables.*
2. *there exist* $\sigma^2 > 0$ *and* $\mu_1, \ldots, \mu_k \in \mathbb{R}$ *such that* $Y_{ij} \sim N(\mu_j, \sigma^2)$, *for each* $j = 1, \ldots, k$, *and each* $i = 1, \ldots, n_j$.

Then,

$$E\left[\, SS_{\text{inter-level}} \,\right] \;=\; (k-1)\,\sigma^2 \;+\; \sum_{j=1}^{k} n_j(\mu_j - \mu)^2\,,$$

where $\mu := \dfrac{1}{n}\sum_{j=1}^{k} \mu_j$, *and* $n = n_1 + n_2 + \cdots + n_k$. *In particular,*

$$\frac{E\left[\, SS_{\text{inter-level}}/(k-1)\,\right]}{E\left[\, SS_{\text{intra-level}}/(n-k)\,\right]} \;=\; \frac{\sigma^2 + \dfrac{1}{k-1}\displaystyle\sum_{j=1}^{k} n_j(\mu_j - \mu)^2}{\sigma^2} \;\geq\; 1$$

3 Experimental Design of Univariate One-way ANOVA

Let $n, N \in \mathbb{N}$, with $n \leq N$. Let $\mathcal{U} = \{\, 1, 2, \ldots, N \,\}$, which represents the finite population, or universe, of N elements.

A Technical Results

Theorem A.1

Let $Y_1, Y_2, \ldots, Y_n$ be independent identically distributed random variables with common distribution $N(\mu, \sigma^2)$. Then,

- *the following two random variables are independent:*

$$\overline{Y} \quad := \quad \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad S^2 \quad := \quad \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$

- *the random variable*

$$\frac{1}{\sigma^2}\cdot(n-1)\,S^2 \quad = \quad \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

  has a χ^2 distribution with $(n-1)$ degrees of freedom.