Query2

Amane Chibana and Harry Wang

2024-04-04

Q1: SSB

$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_i - \overline{X} \right)^2 = \sum_{i=1}^{k} n_i \left(X_i - \overline{X} \right)^2$

$SSB \ = \ \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_i - \overline{X} \right)^2$

$SSB \ = \ \sum_{i=1}^{k} n_i \left(X_i - \overline{X} \right)^2$

Since j is not used in the equation, the summation can be simplified to n. Which would verify the formula of SSB.

Q2: SSE

$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_i \right)^2 = \sum_{i=1}^{k} (n_i - 1) S_i^2$

$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_i \right)^2$

$SSE = \sum_{i=1}^{k} (n_i - 1) \frac{1}{(n_i-1)} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_i \right)^2$

$SSE = \sum_{i=1}^{k} (n_i - 1) S_i^2$

Since the group sample variance is $S_i^2 \ = \frac{1}{(n_i-1)} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_i \right)^2$, when the $\frac{n_i-1}{n_i-1}$ is taken out it is possible to see the sample group variance equation inside SSE. Where then you can substitute S_i^2 verifying SSE.

Q3

For constants $a_1 \ldots a_k$ and sample contrast C_ψ, the corresponding contrast on the population r.v's are

$\psi = a_1 \mu_1 + a_2 \mu_2 + \ldots + a_k \mu_k$

Q4

$E\left[C_\psi\right]$

$E\left[\sum_{i=1}^{k}\ a_i\overline{X}_i\right]$

$\sum_{i=1}^{k} E\left[a_i\overline{X}_i\right]$

$\sum_{i=1}^{k} a_i E\left[\overline{X}_i\right]$

$\sum_{i=1}^{k} a_i\mu_i \ = \ \psi$

Q5

$Var\left(C_\psi\right) = Var\left(a_1\overline{X}_1, +\ldots a_k\overline{X}_k\right)$

$Var\left(a_1\overline{X}_1, +\ldots a_k\overline{X}_k\right) = a_1^2 Var\left(\overline{X}_1), +\ldots (a_k^2\overline{X}_k\right)$

$Var\left(\overline{X}_i\right) = \frac{\sigma^2}{n_i}$

$Var\left(C_\psi\right) = a_1^2\frac{\sigma_1^2}{n_1} + a_2^2\frac{\sigma_2^2}{n_2} +\ldots a_k^2\frac{\sigma_k^2}{n_k}$

Q6

$C_\psi = a_1\overline{X}_1 + a_2\overline{X}_2 +\ldots a_k\overline{X}_k$

$\overline{X}_i. \sim N(\mu, \frac{\sigma_i^2}{n_i})$

$E[C\psi] = a_1\mu_1 + a_2\mu_2 +\ldots a_k\mu_k$

$Var\left(C_\psi\right) = a_1^2\frac{\sigma_1^2}{n_1} + a_2^2\frac{\sigma_2^2}{n_2} +\ldots a_k^2\frac{\sigma_k^2}{n_k}$

$C_\psi \sim N(E[C\psi], Var(C_\psi))$

Q7

Since Z is a standard normal variable, $P(C\psi \geq 0)$ is equal to $P(Z \geq 0)$. The probability that a standard normal variable is greater than or equal to 0 is 0.5, because the standard normal distribution is symmetric around zero.

Under the null hypothesis $H_0 : \psi = 0$, the probability that the sample contrast $C_\psi$is greater than or equal to 0 is 0.5. This is because under H_0, C_ψ is normally distributed with a mean of 0, and the standard normal distribution is symmetric, with exactly half of the distribution lying above zero.