

Teste de Tukey para Comparações de Médias no Delineamento Inteiramente Casualizado (DIC)

Martha Liliana Huertas Madrid

18/11/2025

1. Introdução

Na pesquisa experimental, um dos desafios fundamentais consiste em determinar se existem diferenças significativas entre as médias populacionais de diversos tratamentos avaliados. Para isso, empregam-se desenhos experimentais que permitem realizar comparações válidas e confiáveis. Entre estes, o Delineamento Inteiramente Casualizado (DIC) destaca-se como um dos mais utilizados na prática estatística, caracterizando-se pela atribuição aleatória dos tratamentos a unidades experimentais homogêneas.

A análise inicial do DIC é realizada por meio de uma ANOVA de um único fator, cujo propósito é contrastar a hipótese global:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{vs} \quad H_1 : \text{pelo menos duas médias } \mu_i \text{ diferem.}$$

Embora a ANOVA nos indique que existem diferenças entre os grupos, ela não especifica exatamente quais são os que realmente diferem entre si. Para realizar essas comparações específicas entre pares de tratamentos, precisamos utilizar métodos que mantenham sob controle o erro acumulado. Sem um procedimento adequado de comparações múltiplas, o erro Tipo I aumenta o risco de concluir que existem diferenças quando, na realidade, não as há. Esse erro acumulado é conhecido como taxa de erro familiar (FWER). Segundo (?), o Teste de Tukey é um dos mais utilizados e confiáveis para esse propósito, pois se baseia na distribuição do alcance studentizado, o qual controla explicitamente o erro familiar mantendo o nível de significância α global.

2. Modelo Estatístico do DIC

Considera-se um Delineamento Inteiramente Casualizado (DIC) com k tratamentos, onde cada tratamento é observado r vezes, conformando assim um delineamento balanceado. O modelo linear pode ser expresso da seguinte forma:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, r,$$

com as seguintes condições clássicas:

1. $\sum_{i=1}^k \tau_i = 0$
2. $\varepsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.

O modelo é válido tanto em delineamentos balanceados quanto não balanceados, mas a estrutura equilibrada ocorre apenas quando todos os tratamentos possuem o mesmo número de repetições. A partir deste modelo, podemos derivar a decomposição da variância total, que constitui a base para a análise de variância e, posteriormente, para o teste de Tukey.

3. Análise de Variância

O procedimento de Análise de Variância (ANOVA) parte da ideia fundamental de decompor a variabilidade total observada nos dados em duas componentes: uma associada às diferenças entre tratamentos e outra associada ao erro aleatório. Esta decomposição é essencial para o teste de Tukey, pois fornece a estimativa do erro padrão comum necessário para as comparações múltiplas. Como projeções ortogonais de vetores normais são independentes (?), segue-se que SSTr e SSE são independentes e conduzem à estatística F exata.

A soma de quadrados total é definida como:

$$\text{SST} = \sum_{i=1}^k \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2$$

se decompõe em

$$\text{SST} = \text{SSTr} + \text{SSE},$$

onde

$$\text{SSTr} = r \sum_{i=1}^k (\bar{Y}_{i\cdot} - \bar{Y}_{..})^2, \quad \text{SSE} = \sum_{i=1}^k \sum_{j=1}^r (Y_{ij} - \bar{Y}_{i\cdot})^2.$$

Sob $H_0 : \tau_1 = \dots = \tau_k = 0$,

$$\frac{\text{SSTr}/(k-1)}{\text{SSE}/(k(r-1))} \sim F_{k-1, k(r-1)}.$$

A estimativa não viesada de σ^2 é

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{k(r-1)}.$$

4. Construção do teste de Tukey

Quando a ANOVA indica diferenças significativas entre tratamentos, surge naturalmente a necessidade de identificar quais pares de médias diferem entre si. O problema das comparações múltiplas requer métodos que controlem a taxa de erro familiar (FWER), evitando que o acúmulo de testes aumente indevidamente a probabilidade de erro Tipo I.

Tukey propõe construir intervalos simultâneos para todas as diferenças $\mu_i - \mu_j$ de modo que a probabilidade de que todas as diferenças verdadeiras estejam contidas seja $1 - \alpha$. A estatística base é a amplitude studentizada (studentized range), denotada por q (?).

4.1. Estatística do Alcance Studentizado e Intervalos Simultâneos

Define-se a estatística do alcance studentizado como:

$$Q = \frac{\bar{Y}_{\max} - \bar{Y}_{\min}}{\sqrt{\text{MSE}/r}},$$

onde \bar{Y}_{\max} e \bar{Y}_{\min} são, respectivamente, a maior e a menor média amostral.

A distribuição de Q sob normalidade depende de k (número de médias) e de ν (graus de liberdade do erro). O quantil crítico $q_{\alpha;k,\nu}$ satisfaz:

$$P(Q \leq q_{\alpha;k,\nu}) = 1 - \alpha,$$

garantindo controle simultâneo do erro tipo I. Denota-se por $q_{\alpha;k,\nu}$ o quantil $(1 - \alpha)$ da distribuição do alcance studentizado com parâmetros k e ν .

O procedimento de Tukey testa simultaneamente todas as hipóteses $H_0 : \mu_i = \mu_j$ versus $H_1 : \mu_i \neq \mu_j$ para $i \neq j$. A diferença mínima significativa (DMS) de Tukey é:

$$\text{DMS} = q_{\alpha;k,\nu} \cdot \sqrt{\text{MSE}/r} \quad (1)$$

O procedimento de Tukey rejeita $H_0 : \mu_i = \mu_j$ se e somente se:

$$|\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| > \text{DMS} \quad (2)$$

Este procedimento controla a FWER ao nível α , ou seja, $P(\text{pelo menos um erro tipo I}) \leq \alpha$.

Os intervalos de confiança simultâneos para todas as diferenças $\mu_i - \mu_j$ são:

$$\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} \pm \text{DMS}, \quad \forall i \neq j \quad (3)$$

com nível de confiança simultâneo $1 - \alpha$, significando que a probabilidade de todos os intervalos conterem as respectivas diferenças verdadeiras é pelo menos $1 - \alpha$.

4.2. Extensão para Delineamentos Não Balanceados: Tukey–Kramer

Quando os tamanhos por tratamento são desiguais ($n_i \neq n_j$ para alguns pares), a fórmula anterior requer ajuste. A extensão proposta por Kramer, conhecida como *Tukey–Kramer*, adapta o procedimento para o caso não balanceado:

$$\text{DMS}_{ij}^{\text{TK}} = q_{\alpha;k,\nu} \sqrt{\frac{1}{2} \text{MSE} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad \nu = \text{gl do erro.}$$

A condição de significância torna-se:

$$|\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| > \text{DMS}_{ij}^{\text{TK}}.$$

Esta extensão mantém o controle da FWER mesmo quando os tamanhos amostrais diferem entre tratamentos.

5. Conclusão

O Teste de Tukey constitui um método teoricamente sólido que permite realizar inferências simultâneas sobre todas as comparações entre médias em um Delineamento Inteiramente Casualizado. Este teste não apenas indica se existem diferenças significativas, mas também fornece informações valiosas sobre a magnitude e a direção dessas diferenças, por meio de intervalos de confiança simultâneos.

A principal vantagem do método reside no controle explícito da taxa de erro familiar (FWER) ao nível α , garantindo que a probabilidade de cometer pelo menos um erro Tipo I não exceda o nível de significância especificado. Quando trabalhamos com grupos de tamanhos diferentes, a extensão de Tukey–Kramer mantém essas propriedades, adaptando o procedimento para delineamentos não balanceados.

É importante destacar que a validade desses resultados depende fundamentalmente do cumprimento dos pressupostos do modelo (normalidade, independência e homocedasticidade) e de que estes sejam devidamente verificados antes da aplicação do teste. O Teste de Tukey representa, portanto, uma ferramenta essencial na análise de experimentos com múltiplos tratamentos, oferecendo uma abordagem rigorosa e confiável para comparações múltiplas.