مقدرات تحليل التباين في النموذج العشوائي الاحادي لمركبات التباين عندما يكون حجم العينة غير متوازن

د.شونم عبد القادر محي الدين *   د.عبد الرحيم خلف راهي الحارثي **

المستخلص:

في هذا البحث تم دراسة مقدرات تحليل التباين لمركبات التباين للنموذج العشوائي الاحادي عندما تكون التاثيرات موزعة توزيعاً طبيعياً فضلا عن دراسة تاثير حجم العينة غير المتوازن على تباين المقدر لمركبات التباين بين المجموعات ، كما تم تحديد الحد الادنى لتباين المقدر فيما يتعلق بمجموع المربعات لحجوم العينة.

المقدمة :

يهدف تحليل التباين الاحادي الى معرفة اثر متغير توضيحي ذا مستويات مختلفة على متغير تابع واحد وان النموذج المعروف عندما تكون التاثيرات عشوائية للنموذج الاحادي.

$$y_{ij} = \mu + t_i + e_{ij} \qquad , i = 1,2,\ldots,k$$
$$, j = 1,2,\ldots\ldots,n_i$$

حيث ان t_i يتوزع توزيع طبيعي بمتوسط صفر، وتباين σ_t^2 وان e_{ij} ايضا لـه توزيـع طبيعي بمتوسط صفر وتباين σ_t^2 فضلا عن ان t_i ، e_{ij} يفترض ان يكونا مستقلان، والشيء المهم هنا هو كيف يمكن تقدير المعلومات m ، σ_t^2 ، σ_e^2 حيث ان الطريقة الشائعة للتقدير تعتمد على متوسط مجموع مربعات تحليل التباين.

---

* استاذ مساعد/ كلية الادارة والاقتصاد جامعة السليمانية

** استاذ مساعد/ كلية الادارة والاقتصاد جامعة السليمانية

$$Mse = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 /(N - K)$$

$$Mst = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x}..)/(k-1)$$

حيث ان :

$$\bar{x}_i = 1/n_i \sum x_{ij}. \qquad N = n_1 + ... + n_k$$

$$\bar{x}.. = 1/N \sum_i \sum_j x_{ij}$$

ومن المعروف ان :

$$E(Mse) = \sigma_e^2$$

حيث ان :

$$E(Mst) = \sigma_e^2 + W\sigma_t^2$$

$$W = \frac{1}{k-1}\left(N - \frac{1}{N}\sum n_i^2\right)$$

والمقدرات غير المتحيزة تعطى كما يلي :

$$\hat{\sigma}_e^2 = Mse$$

$$\hat{\sigma}_t^2 = (Mst - Mse)/W$$

ولايجاد تباينات العينة لهذه المقدرات يجب ان يكون σ_t^2 ، σ_e^2 كلاهما يتبع التوزيع الطبيعي ويكون :

$$Var(\hat{\sigma}_e^2) = 2\sigma_e^4 /(N - K)$$

وكذلك

$$Var(\hat{\sigma}_t^2) = A\sigma_t^4 + B\sigma_t^2\sigma_e^2 + C\sigma_e^4 \qquad ...(1)$$

حيث ان :

$$A = \frac{2\left\{N^2\sum n_i^2 + \left(\sum n_i^2\right) - 2N\sum n_i^2\right\}}{\left\{N^2 - \sum n_i^2\right\}^2} \qquad ...(2)$$

$$B = \frac{4N}{N^2 - \sum n_i^2} , C = \frac{2(K-1)N^2(N-1)}{(N-K)\left\{N^2 - \sum n_i^2\right\}^2} \qquad ...(3)$$

A ، B وكذلك C عبارة عن معاملات يمكن ايجادهما كما يمكن ايجاد تباين Mse بسهولة لان $(N-K)Mse/\sigma_e^2$ يتبع توزيع مربعا كاي، وعندما يكون

$n_1 = n_2 = ... = n_k = n = N/K$ (Balance Case) فان تباين $\hat{\sigma}_t^2$ يمكن ايجاده بسهولة ايضا لكون $(K-1)Mst/(\sigma_e^2 + n\sigma_t^2)$ يتبع توزيع مربع كاي ولكن في هذا البحث كيف يمكن تحديد اثر (Imbalance) بين $n_1,....,n_k$ على تباين $\hat{\sigma}_t^2$ حيث بين Hammersley (1949) عندما $N = n_1 + n_2 + ... + n_k$ ان تباين $\hat{\sigma}_t^2$ يمكن تقليله عندما تكون كل n_i متساوية ما في Anderson Grump (1967) بين ان تباين $\hat{\sigma}_t^2$ يكون عند حده الادنى عندما $|n_i - n_j| \le 1$ لكل الازواج n_i ، n_j.

وبما ان تباين $\hat{\sigma}_t^2$ يعتمد على المعلومات σ_t^2 ، σ_e^2 حسب المعادلة (1) فان A,B وكذلك C يمكن ان يكونا عند حدهم الادنى عندما $|n_i - n_j| \le 1$ وحيث ان B,C يعتمدان على n_i ، من خلال ايجاد $\sum_{i=1}^{k} n_i^2$ اذن يمكن قياس تاثير عدم التوزان (Imbalance) من خلال مجموع المربعات غير المصحح $\sum_{i=1}^{k}(n_i - N/K)$ الا ان هذا لايصح لايجاد قيمة المعامل A لانه يتطلب ايجاد المقدار $\sum_{1}^{k} n_i^2$.

لو افترضنا ان N=am+r حيث ان m=(N/K) (عدد صحيح) وان $0 \le r \le K-1$ وجميع العينات تكون بحجم $(n_i \ge 1)$ اي ان K تكون معلومة وكما معروف ان A,B وكذلك C تكون مختلفة طبقا لترتيب $n_1,....,n_k$ ونفترض ايضا ان $n_1 \ge n_2 \ge .... \ge n_k$ وحيث ان n_i^2 دالة مقعرة (Convex) لقيم n_i فان مجموع المربعات $q = \sum_{1}^{k} n_i^2$ سيصل الى حده الاعلى اي ان:

$q_{\max} = (N-K+1)^2 + K - 1$ عندما $n_1 = N-(K-1)$ ، $n_2 = ... = n_k = 1$

بينما نحصل على اقل قيمة للمقدار q عندما :

$n_{r+1} = ... = n_k = m$ و $n_1 = ... = n_r = m+1$ ...(4)

فيكون

$q_{\min} = r(m+1)^2 + (K-r)m^2 = am^2 + (2m+1)r$

بما ان $q_{\max} < N^2$ فان المقدار الموجود في المقام لقيم A,B وكذلك سيبقى دائما موجب . لذلك فان C,B تكون عبارة عن دوال متزايدة للمقدار $\sum n_i^2$ وحيث ان مجموع

بما ان $q_{max} < N^2$ فان المقدار الموجود في المقام لقيم A,B وكذلك سيبقى دائما موجب . لذلك فان C,B تكون عبارة عن دوال متزايدة للمقدار $\sum n_i^2$ وحيث ان مجموع المربعات هذا يكون عند حده الادنى $n_1 ...... n_k$ كما في المعادلة (4) فان المعاملين C,B تكونان ايضا عند قيمهما الدنيا وكما موضحة في كل من **Anderson** و **Crump** عام (1967) اما بالنسبة الى المعامل A فيمكن كتابته:

$$A = \frac{2\left\{N^4 - 2N^2\sum_1^k n_i^2 + \left(\sum_1^k n_i^2\right)^2 + 3N^2\sum_1^k n_i^2 - 2N\sum_1^k n_i^3 - N^4\right\}}{\left\{N^2 - \sum_1^k n_i^2\right\}^2}$$

$$= 2\left[1 + \frac{3Nq - 2\sum_1^k n_i^3 - N^3}{\left\{N^2 - q\right\}^2}\right] \qquad ...(5)$$

ولايجاد حدود المقدار $\sum_1^k n_i^3$ نعتبر $x_i = n_i - N/K$ ونحصل على $q = \sum_1^k x_i^2 + N^2/K$ وكذلك

$$\sum_1^k n_i^3 = \gamma\left(\sum_1^k x_i^2\right)^{\frac{3}{2}} + \frac{3N}{K}\sum_1^k x_i^2 + \frac{N^3}{K^2} \qquad ...(6)$$

حيث ان $\gamma = \sum_1^k x_i^2 / \left(\sum_1^k x_i^2\right)^{3/2}$ يمكن اعتباره مقياس للالتواء بين $n_1 ...... n_k$.

الحدود الدنيا للمعامل Lower Bounds A

الحدود الدنيا للمعامل A تعتمد على الحدود العليا للمقدار $\sum_1^k n_i^3$ وفيما يتعلق بقيمة q وكذلك N . فاذا كان r = 0,1 يتم تحديد حد ادنى واحد وعندما تكون $r \geq 2$ فيمكن تحديد حدين للمعامل A والنظرية الاتية يمكن من خلالها تحديد الحد الادنى الاول للمعامل A

نظرية (1)

اذا كان

$$A \geq L_1(q) = 2\left[1 + \frac{(K-1)(k-2)N}{K^2}\frac{g(q)}{(N^2-q)^2}\right] \qquad ...(7)$$

حيث ان :

$$g(q) = 3\frac{K}{K-1}\left(q - N^2/K\right)N - 2\left(\frac{K}{K-1}\left(q - N^2/K\right)\right)^{3/2} - N^3$$

عندما $r \geq 2$ فان المعادلة (7) التي تمثل الحد الادنى الاول للمعامل A تكون غير كافية واذا كانت K كبيرة بدرجة ما فان المعادلة (7) تقود الى ايجاد قيمة يمكن اعتبارها اصغر من الحد الادنى للمعامل A المعطى في المعادلة (5) لقيم n_i طبقاً للمعادلة (4).

عندما يكون

$$n_2 = ... = n_k = m,\ n_1 = m + r \qquad ...(8)$$

يمكن ايجاد قيمة $q = \sum_1^k n_i^2 = q_{\min} + r(r-1)$

وهذا يؤدي الى التعامل مع النظرية التالية لايجاد الحد الادنى الثاني للمعامل A.

<u>نظرية (2):</u>

اذا كان $q_{\min} \leq q \leq q_m + r(r-1)$

فان :

$$A \geq L_2(q) = 2\left[1 + \frac{(q-q_{\min})b_1 - b_0 - f(q-q_{\min})}{(N^2-q)^2}N\right] \qquad ...(9)$$

حيث ان :

$$b_1 = 3[N - 2m - 1] = 3((K-2)m + r - 1)$$

$$b_0 = K(K-1)(K-2)m^3 + 3(K-1)(K-2)rm^2 + 3(K-2)r(r-1)m + r(r-1)(r-2)$$

(10)

والدالة :

$$f(y) = y\sqrt{4y+1}$$

<u>الحدود العليا للمعامل Upper Bounds A:</u> لاحظنا سابقا ان المعامل A في معادلة (5) متناقض مع $\sum_1^k n_i^2$ ، هناك كنا نحتاج الى تحديد الحد الادنى للمجموع التكعيبي ، فلو

افترضنا لبعض P حيث ان $p = 0.1.2.....k-2$ انه على الاقل ان اخر قيمة لقيم p من بين $n_1.....n_k$ تساوي 1 بحيث ان :

$$N = p + \sum_{1}^{K-p} n_i \quad .q = p + \sum_{1}^{K-p} n_i^2$$

وكذلك:

$$\sum_{1}^{k} n_i^3 = p + \sum_{1}^{k} n_i^3$$

ولو اعتبرنا ان $x_i = n_i - (N-K)/(K-p).i = 1.....K-p$ ونستخدم معادلة (6)

ولكن هنا القيم $n_1,....,n_{K-p}$ سنحصل على :

$$\sum_{1}^{K-p} n_i^3 = \sum_{1}^{K-p} x_i^2 + (N-p)^2/(K-p)$$

$$\sum_{1}^{K-p} n_i^3 = \gamma\, p\left(\sum_{1}^{K-p} x_i^2\right)^{\frac{3}{2}} + \frac{3(N-p)}{K-p)}\sum_{1}^{K-p} x_i^2 + \frac{(N-p)^3}{(K-p)^3} \qquad ...(11)$$

حيث ان :

$$\gamma\, p = \sum_{1}^{K-p} x_i^3 \bigg/ \left(\sum_{1}^{K-p} x_i^2\right)^{3/2}$$

وكما في المعادلة (6) فان الحد الاول والثاني في المعادلة (10) لاتظهران عندما $x_1 = ... = x_{k-p} = 0$ اي انه عندما $n_1 = ... = n_{k-p}$ فان الحد الاول من المعادلة اعلاه يزول عندما $p = k-2$ لان $x_1 = -x_2$ وبالتالي فان $x_1^3 + x_2^3 = 0$ في هذه الحالات.

يمكن استخدام النظرية التالية في ايجاد الحد الاعلى للمعامل A.

<u>نظرية (3)</u>

اذا كان :

$$q_{min} \leq q \leq_{max}$$

فان :

$$A \leq U(q) = 2\left[1 + \frac{3Nq - 2c(q) - N^3}{(N^2-q)^2} N\right] \qquad ...(12)$$

حيث ان :

$$C(q) = \max(C_{min}.C_p(q))$$

$$q_p \le q \le q_{p+1} \;,\; p = 0,......,k-2 \qquad ...(13)$$

وان

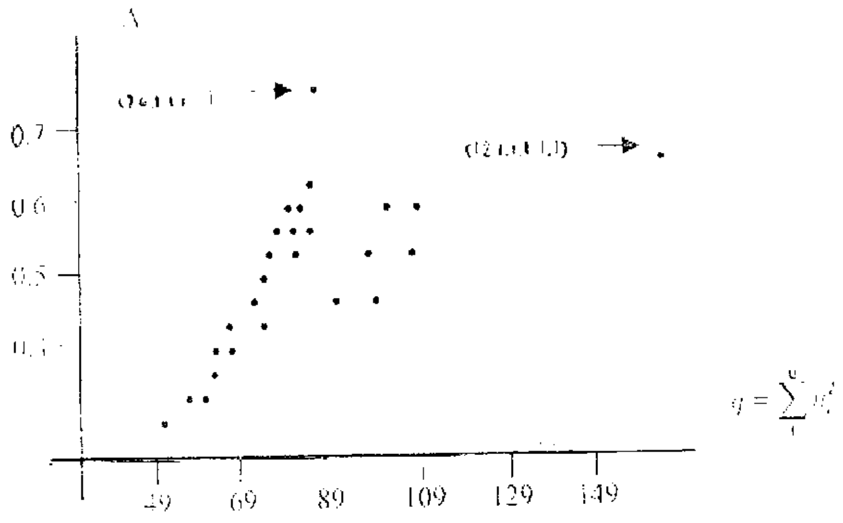
$$C_{min} = r(m+1)^3 + (\alpha - r)m^3$$

$$C_p(q) = p - \frac{(k-p-2)(q-q_p)^{3/2}}{\sqrt{(k-p)(k-p-1)}} + \frac{3(N-p)(q-q_p)}{K-p} + \frac{(N-P)^3}{(K-P)^3}$$

التطبيقات Application :

التطبيق الاول :

لـــو اخـــذنا التصـــميم الاول عنـــدما K=6 ، N=17 وان $n_1 \ge .... \ge n_6$ فيكـــون $m = 17/6 = 2$ وان $r = N - am$ فيمكن استخدام المعادلة (7) في ايجاد الحـد الادنـى الاول والثاني لان $r \ge 2$ ، طبقـاً للمعادلــة (4) فــان $q_{min} + r(r-1) = 69, q_{min} = 49$ واسـتخدام $L_2(q)$ في المعادلة فان الحد الادنى يكون بين $49 \le q \le 69$ ، وكما موضح من خلال المنحنـى المتقطع في شكل رقم (1) ادناه:



شكل رقم (1) قيم وحدود A و $\sum_{1}^{6} n_i^2$ للتصاميم مع $n_1 + ... + n_6 = 17$

وعدديا يمكن احتساب الحد الادنى الاول والثاني $L_1(q), L_2(q)$

$L_1(49) = 0.4049$. $L_2(49) = 0.4063$

وعدديا يمكن احتساب الحد الادنى الاول والثاني $L_1(q), L_2(q)$

$L_2(49) = 0.4063$ ، $L_1(49) = 0.4049$

وحيث ان $L_2(49)$ يساوي اقل قيمة من قيم A فيمكن ايضاً ايجاد النقاط الحرجة (Break-Point) q_p حيث ان :

$$q_0 = 48\frac{1}{6}, q_1 = 52\frac{1}{5}, q_2 = 58\frac{1}{4}, q_3 = 68\frac{1}{3}, q_4 = 88\frac{1}{2}, q_5 = 149$$

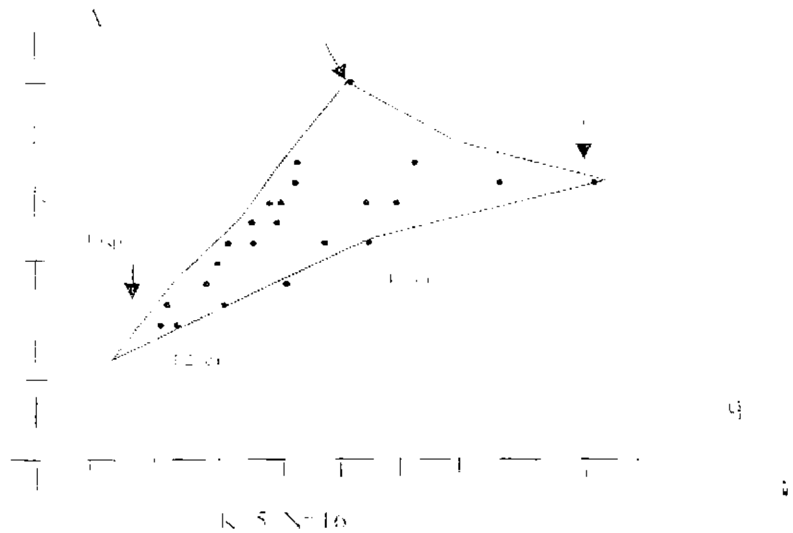
وكل هذه النقاط يمكن تشخيصها وتاشيرها على الحد الاعلى U(q). وان اعلى قيمة U(q)تكون عند $q_4 = 88\frac{1}{2}$ بالرغم من ان اعلى قيمة تكون طبقا للتصميم عندما $n_3 = ... = n_6 = 1, n_2 = 6, n_1 = 7$ وعندما $q = 89$.

التطبيق الثاني:

عندما K=5 ، N=16 فيكون m=16/5=3 وان r =1 ، وعندما $r(r-1) = 0$ فاننا لانحتاج الى ايجاد $L_2(q)$ حسب النظرية (2) فنحتاج الى ايجاد $L_1(q)$ وان A عند $q = q_{min} = 52$. اما النقاط الحرجة q_p للحد الاعلى U(q) يمكن ايجادها كما يلي :

$$q_0 = 5\frac{1}{5}, q_1 = 57\frac{1}{4}, q_2 = 67\frac{1}{3}, q_3 = 87\frac{1}{2}, q_4 = 148$$

ويمكن مشاهدة اخر اربعة قيم على U(q) في الشكل رقم (2)



من ملاحظة الشكل رقم (1) والشكل رقم (2) نجد انه يمكن تحسين وتطوير $L_1(q)$ وعندما $q \geq q_{min} + r(r-1)$.

في هذا التطبيق يمكن تحديد سلوك $\hat{\sigma}_t^2$ لقيم مختلفة من N ، اقل قيمة الى A للتصميم N=16، K=5 تعطى من خلال معادلة (2) تساوي 289/146 للتصميم (4،3،3،3،3) عندما q=52 ونفس الشيء باستخدام معادلة (3) يمكن ايجاد اقل قيم للمعاملين B وكذلك C ومن ثم استخدام المعادلة (1) لايجاد $Var(\sigma_t^2)$ كما يلي :

$$Var(\hat{\sigma}_t^2) = \frac{146}{289}\sigma_t^4 + \frac{16}{51}\sigma_t^2\sigma_e^2 + 0.0671\sigma_e^4$$
$$= 0.5052\sigma_t^4 + 0.3137\sigma_t^2\sigma_e^2 + 0.0671\sigma_e^4$$

فان اقل تباين عندما a=5 و N=15 يمكن الحصول على باستخدام تصميم متوازن **(Blanced)** مثل (3،3،3،3،3) من خلال :

$$Var(\hat{\sigma}_t^2) = \frac{1}{2}\sigma_t^4 + \frac{1}{3}\sigma_t^2\sigma_e^2 + \frac{7}{90}\sigma_e^4$$

فاذا اقتربت σ_t^2 من الصفر فان C وبعض من مدى B ستساهم في قيمة $Var(\sigma_t^2)$ ومن ثم التباين للتصميم N=16 يكون اصغر من تباين التصميم N=15 من جهة اخرى اذا كان σ_e^2 صغيرة مقارنة مع σ_t^2 فان A يصبح الحد الرئيسي وفي هذه الحالة يكون تباين التصميم N=15 اصغر من التصميم N=16 بالرغم من ان عدد المشاهدات في التصميم الاول تكون اقل، لتحديد مساهمة A في قيمة $Var(\sigma_t^2)$ عندما σ_e^2 تقترب من الصفر فلا يكون هناك تباين بين المجدموعات **(Within Groups)** ذلك فان **Mse=0** .

وان :

$$\hat{\sigma}_t^2 = \frac{Mst}{W} = \frac{1}{W(K-1)}\sum n_i(\bar{Y}_i - \bar{Y}..)^2$$

وعندما $Y_{i1} = Y_{i2} = ... = Y_{ini} = \bar{Y}_i$ وان $\sigma_e^2 = 0$ فان متوسطات العينة $\bar{Y}_i ..... \bar{Y}_K$ يمكن اعتبارها عينة بحجم K مسحوبة من مجتمع يتوزع طبيعياً بمتوسط μ وتباين σ_t^2 .

References:

1. Anderson, R.L., Grump (1967), "Comparisons of Design and Estimation Procedures for Estimating Parameters in a Tow-Stage nested Process", Technometrics, 9:4,499-516.
2. Mulholl and H.P. (1977), "On the Null Distribution of $\sqrt{b_1}$ for Sample of size at most 25 with tables", Biometric, 64, 401-409.
3. Searlw S.R, Casella, G., McCulloch, C.F (1992)," Variance Components" Wiley, New York.

....

....

....