تحليل التباين

د. سيف القحطاني

503

الإحصاء التربوي

تحليل التباين

- في كثير من الأحيان، تدفعنا الحاجة لمقارنة عدد من المجموعات.
- في السابق استخدمنا اختبار (ت) في مقارنة مجموعتين. فقمنا بحساب الفرق بين متوسطي المجموعتين ومقارنته بالتشتت المتوقع عن المعاينة العشوائية.
- الإشكالية تظهر عند وجود أكثر من مجموعتين نود التعرف على الفروق بينها.
- سنضطر لمقارنة زوجين من المتوسطات في كل مره.
- مثلا لوكان عندنا أربع مجموعات (أ، ب، ج، و د) سيتوجب علينا إجراء 6 عمليات مقارنة
- أ في مقابل ب، أ في مقابل ج، و أ في مقابل د، و ب في مقابل ج، و ب في مقابل د، و ج في مقابل د.

الإشكالية

- الاختبار الفردي يقوم على احتمالية أن الفرض الصفري صحيح، لذا في كل مرة نجري اختبار (ت) بين متوسطين نحن عرضة لارتكاب الخطأ من النوع الأول.
- مستوى الدلالة (α) الذي نضعه لأنفسنا كمحك لرفض الفرض الصفري هو مستوى المخاطرة الذي نسمح به.
- فإذا كان مستوى الدلالة 5% فذلك يعني أننا نقبل بالمخاطرة في رفض الفرض الصفري وهو صحيح 5 مرات من أصل 100مرة.
- بمعنى آخر من أصل 100 مرة سنقوم برفض الفرض الصفري والواجب قبوله 5 مرات.
- فإذا استخدمنا اختبار (ت) في مقارنات متعددة على مجموعات مستقلة في أدائها على متغير واحد ستتزايد نسبة الخطأ من النوع الأول بشكل كبير

حساب نسبة الخطأ

- وعليه فنسبة الخطأ من النوع الأول تحسب من المعادلة التالية:
- $1-(1-\alpha)^t$
- حيث تمثل α نسبة الدلالة عادة 5%
- T تمثل عدد المقارنات باستخدام اختبار (ت)
- فلوكان عندنا أربع مقارنات فنسبة الخطأ من النوع الأول لن تكون 5% كما نتوقع بل ستكون
- $1-(1-.05)^4$
- وتساوي 0.18 أي أننا سنكون عرضة لارتكاب الخطأ من النوع الأول 18 مرة من أصل 100 مرة!

- بدلا من إجراء عدة مقارنات وبالتالي زيادة المخاطرة في الوقوع في الخطأ من النوع الأول سنستخدم تحليل التباين.
- تحليل التباين أسلوب إحصائي يكشف عن الفروق بين المجموعات مع أخذه للتباين الكلي في الاعتبار.
- وعليه فتحليل التباين يقوم على أساس أن جميع المجموعات متساوية ومختارة عشوائيا من نفس المجتمع وأي فروق مشاهدة هي نتاج المعيانة العشوائية.
- ويحاول الإجابة عن السؤال التالي: هل تختلف متوسطات المجموعات عن بعضها؟

شروط استخدام تحليل التباين

- تحليل التباين اختبار معلمي (يشترط بعض الاشتراطات في شكل توزيع المجتمع)

1) المتغير التابع متغير من المستوى الكمي ويتبع التوزيع الاعتدالي الطبيعي
2) المتغير المستقل من المستوى الاسمي (ذكر / أنثى)
3) استقلالية المجموعات (لايمكن لنفس الشخص أن يكون في مجموعتين...فقط في مجموعة واحدة)
4) تجانس تباين المجموعات

تحليل التباين

- تقوم فكرة اختبار تحليل التباين على أن المجموعات تمثل مجتمعا واحدا، وعليه فإن التباين الكلي الناتج عن اعتبارهم مجموعة واحدة يساوي حاصل جمع التباين لكل مجموعة على حدة.
- بعبارة أخرى، نقوم بحساب التباين لكل مجموعة على حدة، ثم نجمع الأشخاص في مجموعة واحدة ونحسب التباين (التباين الكلي)
- فإذا وجدنا أن حاصل جمع التباين الخاص بكل مجموعة أقل التباين الكلي فذلك دليل على أن هناك فرق بين متوسط المجموعات (أي أن هناك تباين آخر سبب الفرق)
- التباين الكلي = التباين داخل المجموعات + التباين بين المجموعات

مثال

- لنفترض أن عندنا ثلاث طرق تدريس ونريد مقارنتها
- طريقة تدريس مبرمج / طريقة تدريس بالحاسب/ الطريقة التقليدية
- المتغير المستقل: طريقة التدريس (اسمي تصنيفي وفيه ثلاث مستويات)
- المتغير التابع: اختبار تحصيلي في مادة الرياضيات

البيانات التالية تمثل ثلاث عينات مختارة عشوائيا

التقليدية	الحاسب	المبرمج
16	29	12
8	19	20
.....	......	....
16	25	10

- لنفترض أننا حصلنا على الإحصاءات التالية:

الانحراف (S المعياري	$\bar{X}$	N	الطريقة
4.6	15	11	المبرمج
4.5	20	11	الحاسب
5.10	18	11	التقليدية

- يبدو من الجدول أن التعليم بالحاسب انتج أداء أفضل ثم التقليدي (قارن قيمة المتوسطات)
- لكن هذه الإحصاءات الثلاث مجرد عينة عشوائية من عدد كبير من العينات. فما احتمالية أن نحصل على مثل هذه النتائج (الفروق) لو استخدمنا عينات عشوائية أخرى.
- بمعنى آخر لو اخترت 33 شخصا (11+11+11) آخرين وقمت بتوزيعهم على الطرق بشكل عشوائي (11 في كل مجموعة) هل سأحصل على فرق مشابه؟ أم أن الفرق الحاصل هو نتاج معاينة عشوائية؟

الفروض

- الفرض الصفري يقول:
- $\mu_1 = \mu_2 = \mu_3$
- الفرض البديل يقول:
- على الأقل أحد المتوسطات مختلف

تحليل التباين

- لاحظ أن لكل مجموعة تباينها الخاص (1.00) لكل مجموعة في مثالنا
- وإذا اعتبرناهم كلهم مجموعة واحدة فسيكون هناك تباين كلي (1.225)...جربها.

		تقليدي	حاسب	مبرمج
مبرمج	6	8	7	6
	5	7	6	5
	4	6	5	4
حاسب	7			
	6			
	5			
تقليدي	8			
	7			
	6			

الإحصاء الوصفي

الطريقة	N	المتوسط	الانحراف المعياري
المبرمج	3	5	1
الحاسب	3	6	1
التقليدي	3	7	1
الجميع	9	6	1.225

التباين داخل المجموعات

- بمعرفة أن التباين

$$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$

- فإن مجموع المربعات يساوي التباين ضرب (N-1)
- فبإمكاننا حساب التشتت داخل المجموعات وذلك بجمع مربع الانحرافات (في مثالنا السابق: 1*(3-1)=2 للمجموعة الأولى و *(3-1)=2 للمجموعة الثانية و 1*(3-1)=2 للمجموعة الثالثة....المجموع 6

التباين الكلي

- التشتت الكلي يساوي التباين الكلي مضروبا في حجم العينة ناقص واحد. وهذا يسمى مجموع مربع الانحرافات الكلي (يعني انحراف كل قيمة عن المتوسط الأكبر)

$$s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$$

- التباين يمكن تقسيمه إلى:

1) تباين داخل المجموعات
2) تباين بين المجموعات
3) تباين كلي (ويساوي تباين داخل المجموعات + التباين بين المجموعات)
4) في مثالنا مجموع التشتت الكلي (12) = التشتت داخل المجموعات (6) + التشتت بين المجموعات (6)

مثال

المجموعات	A	B	C
1	1	7	5
2	3	4	8
3	2	2	6
4	2	3	9
5	3	9	7
6	5	4	9
7	7	4	10
8	4	8	8
9	2	6	7
10	1	5	10
المتوسط	3	5.2	7.9
الانحراف المعياري	1.89	2.25	1.66

اجمع القيم واقسمها على عددها لتحصل على المتوسط

مجموعة من الإحصاءات الوصفية

الإحصاء الوصفي

البيانات

	حجم العينة (ن)	المتوسط	الانحراف المعياري	الخطأ المعياري	95% فترة الثقة للمتوسط		القيمة الدنيا	القيمة القصوى
					الحد الأدنى	الحد الأعلى		
1.00	10	3.0000	1.88562	.59628	1.6511	4.3489	1.00	7.00
2.00	10	5.2000	2.25093	.71181	3.5898	6.8102	2.00	9.00
3.00	10	7.9000	1.66333	.52599	6.7101	9.0899	5.00	10.00
المجموع	30	5.3667	2.77282	.50624	4.3313	6.4021	1.00	10.00

شرط تجانس التباين

<u>للتأكد من شرط تجانس التباين نستخدم اختبار ليفين (Levene's Test)</u>

- الفرض الصفري:

- $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

- الفرض البحثي:

- $H_a: \sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$

- من الجدول يتضح أن القيمة الاحتمالية لاختبار ليفين لتساوي التباين للمجموعات أكبر من (α / 0.05) وعليه نقبل الفرض الصفري القائل بتجانس تباين المجموعات ويمكننا استخدام اختبار تحليل التباين

القيمة الاحتمالية
أكبر من 5%

Test of Homogeneity of Variances

VAR00001

Levene Statistic	df1	df2	Sig.
.669	2	27	.520

جدول تحليل التباين

القيمة الاحتمالية لقيمة (ف)
أقل من 5%

تحليل التباين الأحادي (ANOVA)

البيانات

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	القيمة الفائية	القيمة الاحتمالية
بين المجموعات	120.467	2	60.233	15.866	.000
داخل المجموعات	102.500	27	3.796		
المجموع	222.967	29			

نرفض الفرض الصفري ونقول على الأقل أحد المتوسطات مختلف

المقارنة البعدية

- تحليل التباين يخبرنا فقط وجود فرق بين اثنين من المجموعات على الأقل ولكن لا يخبرنا عن عدد الفروق ولا عن مصدرها. لذا نحتاج إلى اختبارات بعدية تحدد لنا أين مصدر الفروق
- ومن أشهر الاختبارات البعدية ما يسمى بشافيه (Scheffe)

Multiple Comparisons

VAR00001

Scheffe

(I) VAR00002	(J) VAR00002	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1. 00	2.00	-2.20000-	.87135	.057	-4.4568-	.0568
	3.00	-4.90000-*	.87135	.000	-7.1568-	-2.6432-
2.00	1.00	2.20000	.87135	.057	-.0568-	4.4568
	3.00	-2.70000-*	.87135	.016	-4.9568-	-.4432-
3.00	1.00	4.90000*	.87135	.000	2.6432	7.1568
	2.00	2.70000*	.87135	.016	.4432	4.9568

*. The mean difference is significant at the 0.05 level.

- في الصف الأول مقارنة بين المجموعة الثالثة والمجموعة الأولى

  الفرق دال إحصائيا أقل من 5% (0.0001)

- في الصف الثاني مقارنة بين المجموعة الثالثة والمجموعة الثانية

  الفرق دال إحصائيا أقل من 5% (0.016)