تحليل التباين
Analysis of Variance

مقدمة

تحليل التباين هو أحد الأدوات الإحصائية المهمة والتي تعنى بعملية دراسة العلاقة بين متغير كمي تابع مع متغير آخر أو عدة متغيرات مستقلة والتي عادة ما تكون وصفية. ويهتم تحليل التباين في عملية البحث أو تحديد مصادر الاختلاف بين المتوسطات ولا يهتم في تحديد نوع العلاقة بين المتغيرات التابعة والمتغيرات المستقلة. ويتدرج تحليل التباين من تحليل التباين بعامل واحد إلى تحليل التباين بعاملين ثم تحليل التباين للبيانات المتكررة ثم إلى أنواع أخرى لا يتسع المجال لذكرها. وسوف يتم خلال الفصل عرض تحليل التباين بصورة مبسطة ثم تطبيق مجموعة من التحليلات باستخدام برنامج SPSS.

تحليل التباين بعامل واحد One-Way Analysis of Variance

يستخدم تحليل التباين بعامل واحد في تحليل بيانات عن مجوعة من الحالات، حيث يوجد لكل حالة قياس لمتغيرين، متغير كمي وهو المتغير التابع ومتغير مستقل وهو متغير وصفي يصنف الحالات إلى مستويين أو أكثر. لذلك يعتبر تحليل التباين تعميم لاختبار T للمتوسطات من حيث أن تحليل التباين يختبر وجود اختلافات بين متوسطات لمتغير يمكن تصنيف مشاهداته تبعاً لمتغير تصنيفي أو عامل مؤثر على بيانات المتغير في حين أن اختبار T يختبر وجود اختلاف بين متغير يمكن تصنيفه لمستويين فقط. ويمكن صياغة النموذج الرياضي لتحليل التباين على النحو التالي:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

$$i = 1, 2, \ldots, r ; n = 1, 2, \ldots, n_i$$

حيث تمثل y_{ij} القيمة j للمتغير التابع في المستوى التصنيفي i، وتمثل μ_i قيمة متوسط المتغير في المستوى التصنيفي i، وتمثل ε_{ij} الخطأ العشوائي للمتغير التابع. ويتم من خلال تحليل التباين اختبار الفرضية التي تدعي عدم تساوي متوسطات المتغير. ويمكن صياغة فرضية العدم والفرضية البديلة على النحو التالي:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$$
$$H_a : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_r$$

بذلك فإن فرضية العدم تشير إلى تساوي جميع المتوسطات، في حين أن فرضية العدم تشير إلى عدم تساوي المتوسطات، إلا أنه يجب التنبيه هنا إلى أن عدم التساوي في الفرضية البديلة يعني وجود احد المتوسطات يختلف عن البقية ولا يعنى أن جميعها مختلفة عن بعضها البعض.

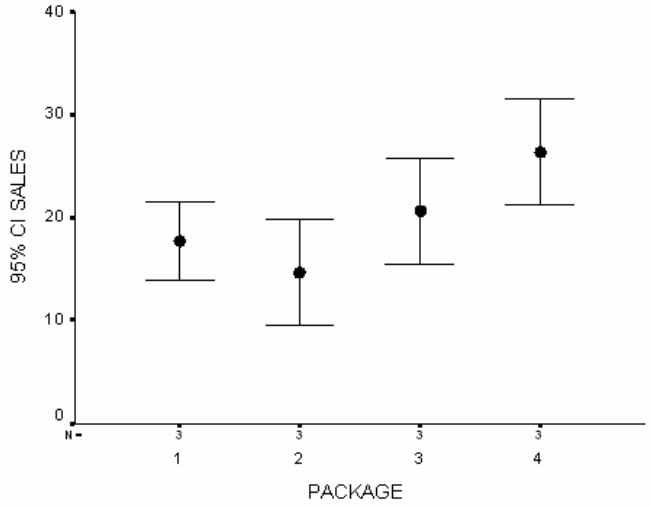
لنفرض أن لدينا بيانات عن مبيعات سلعة في ثلاث متاجر وقد تم تغليف السلعة بأربع طرق مختلفة، والسؤال هنا هو هل كانت طريقة التغليق غير مؤثر في المبيعات في المتاجر الأربعة.

		المتجر		
		1	2	3
طريقة التغليف	1	16	18	19
	2	14	17	13
	3	19	20	23
	4	24	28	27

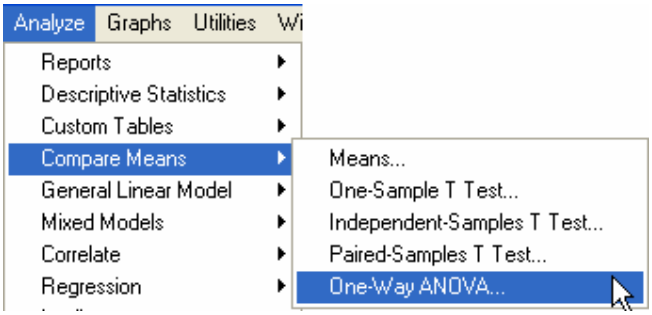
يمكن صياغة فرضية العدم والفرضية البديلة على النحو التالي:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

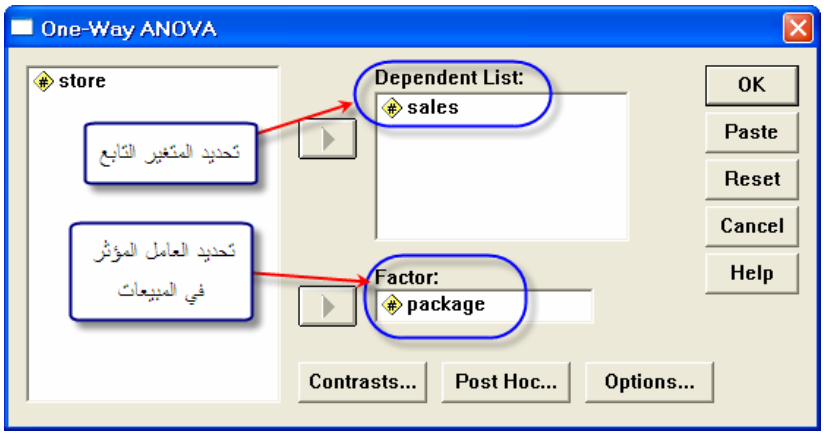
وبذلك فإن فرضية العدم تشير إلى تساوي متوسط المبيعات حسب طرائق التغليف، في حين تشير الفرضية البديلة إلى عدم تساوي متوسط المبيعات. وقد يكون من المناسب هنا عرض فترات ثقة للمتوسطات باستخدام الأمر Error Bar من قائمة Graphs.



ويلاحظ عدم التداخل بين فترات الثقة للطريقة الرابعة مع الطريقة 1 والطريقة 2، وهذه تعتبر إشارة لتأثير طريقة التغليف على المبيعات، إلا أن الفحص البياني لا يكفي ويجب التأكد من صحة الاستنتاجات وذلك باختبار الفرضية السابقة باستخدام اختبار F. ولاختبار الفرضية السابقة، نستخدم الأمر



وبذلك يظهر مربع الحوار التالي



وبذلك تم تحديد المتغير التابع وهو المبيعات والمتغير المستقل وهو طريقة التغليف. وعند النقر على OK تظهر النتائج التالية.

ANOVA

SALES

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	223.000	3	74.333	19.391	.000
Within Groups	30.667	8	3.833		
Total	253.667	11			

درجات الحرية

مجموع مربعات الفروق بين المجموعات

المجموع الكلي لمربعات الفروق

مجموع مربعات الفروق داخل المجموعات

حيث

مجموعات مربعات الفروق بين المجموعات هي $SSTR = \sum_{i=1}^{r} n_i \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2$

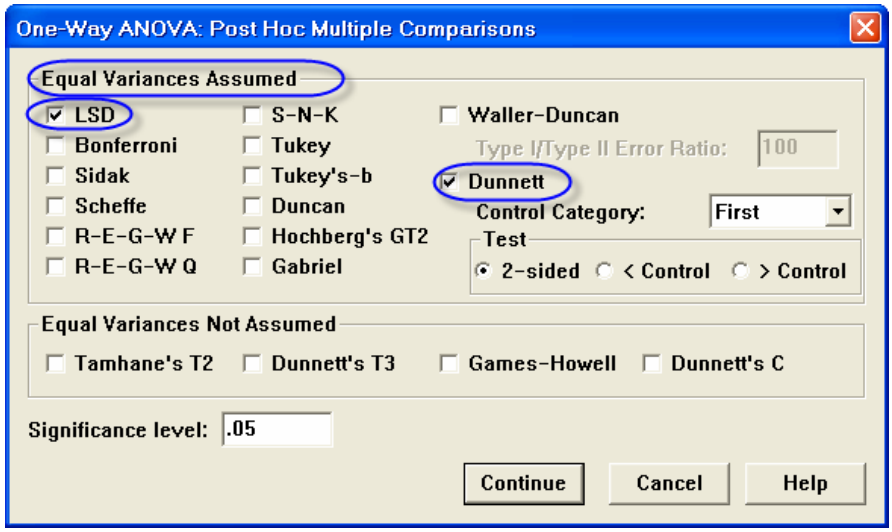
مجموع مربعات الفروق داخل المجموعات هي $SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2$

مجموع الكلي لمربعات الفروق هي $SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{..}\right)^2$

ودرجات الحرية هي $(r-1=3)$، $\left(\sum_{i=1}^{r} n_i - r\right)=8$ و $\left(\sum_{i=1}^{r} n_i - 1\right)=11$

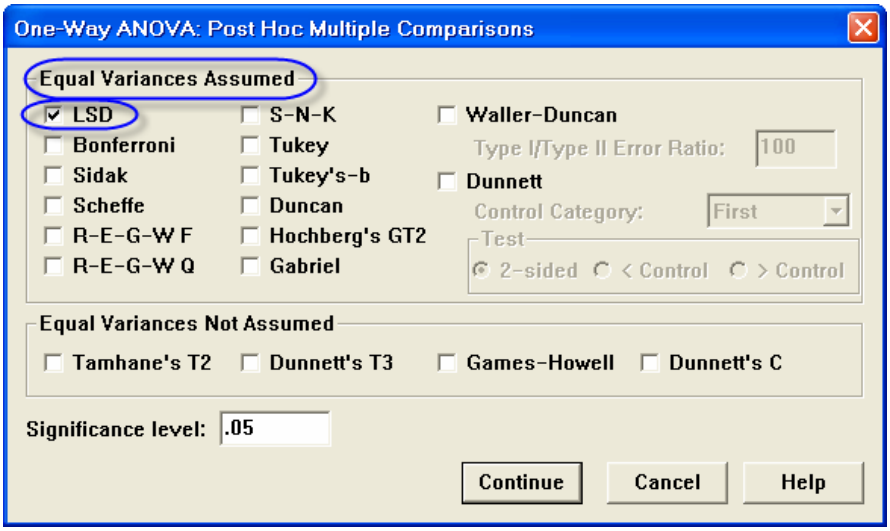
وباستخدام الجدول السابق يمكن رفض فرضية العدم عند استخدام مستوى معنوية 5% وقبول الفرضية البديلة والتي تشير إلى وجود اختلافات بين المبيعات كنتيجة لاختلاف طريقة التغليف. وتم اتخاذ القرار بالنظر إلى قيمة F والتي تشير إلى أن نسبة كبيرة من الاختلافات ناتجة من الاختلافات بين المتوسطات وكذلك قيمة P-Value والتي تساوي صفر.

والسؤال الذي يتبادر إلى الذهن هو هل جميع المتوسطات تختلف عن بعضها البعض؟ أم أن أحد المتوسطات يختلف عن البقية؟ ويمكن الإجابة على هذا التساؤل باستخدام المقارنات المتعددة Multiple Comparison والتي تعنى بالبحث عن الفروقات الجوهرية بين المتوسطات المختلفة. ويتم ذلك بالنقر على زر Post Hoc في مربع حوار One-Way ANOVA ليظهر مربع الحوار التالي



وبافتراض تساوي التباين للمجموعات، فإن يوجد العديد من الطرائق لاختبار عدم وجود اختلاف بين متوسطات المجوعات، ويمكن للقارئ النقر بزر بالفارة الأيمن على أي طريقة

للحصول على شرح مبسط عن الطريقة. ويوفر LSD مجموعة من اختبارات T لاختبار عدم وجود اختلاف بين جميع المتوسطات وذلك عند مقارنة أحدها بالآخر.



وبالنقر على زر Continue ثم زر OK، تظهر النتائج التالية.

Multiple Comparisons

Dependent Variable: SALES

LSD

(I) PACKAGE	(J) PACKAGE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	3.00	1.599	.097	-.69	6.69
	3	-3.00	1.599	.097	-6.69	.69
	4	-8.67*	1.599	.001	-12.35	-4.98
2	1	-3.00	1.599	.097	-6.69	.69
	3	-6.00*	1.599	.006	-9.69	-2.31
	4	-11.67*	1.599	.000	-15.35	-7.98
3	1	3.00	1.599	.097	-.69	6.69
	2	6.00*	1.599	.006	2.31	9.69
	4	-5.67*	1.599	.008	-9.35	-1.98
4	1	8.67*	1.599	.001	4.98	12.35
	2	11.67*	1.599	.000	7.98	15.35
	3	5.67*	1.599	.008	1.98	9.35

*. The mean difference is significant at the .05 level.

ويلاحظ من النتائج أن طريقة التغليف رقم 1 كان لها تأثير أقل في المبيعات من طريقة التغليف رقم 4، وكذلك كانت طريقة التغليف رقم 2 أقل تأثيراً من طريقتي التغليف رقم 3 ورقم 4، وكذلك كانت طريقة التغليف رقم 3 أقل تأثيراً من طريقة التغليف رقم 4. وبذلك فإن طريقة التغليف رقم 4 كانت أفضل طريقة تغليف.

تحليل التباين بعاملين Two-Way Analysis of Variance

يهتم تحليل التباين بعاملين في معرفة تأثير عاملين مستقلين على المتغير التابع، وبذلك يتطلب الأمر لتطبيق تحليل التباين بعاملين إلى وجود بيانات عن حالات بحيث يوجد لكل حالة قيمة للمتغير الكمي وهو المتغير التابع وقيمتين للمتغيرين التصنيفيين وهما المتغيرين المستقلين. ويستخدم اختبار F لاختبار تأثير كل عامل بصورة منفصلة (Main Effect) وكذلك اختبار تأثير تفاعل العاملين Interaction.

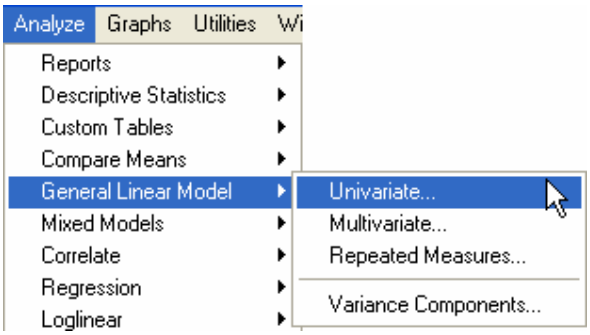
في البيانات السابقة تم اختبار تأثير طريقة التغليف على المبيعات، إلا أن البعض قد يتساءل عن إذا كان للمتجر تأثير ايجابي أو سلبي على المبيعات، وفي هذه الحالة يمكن بناء نموذج أخر واعتبار أن المبيعات هي المتغير التابع والمتغير الذي يمثل المتاجر هو المتغير المستقل. إلا أنه في كثير من الأحيان يكون التفاعل بين المتغير له تأثير كبير على النتائج، لذلك فإن استخدام أي من المتغيرين بصورة منفصلة عن الآخر قد يخفي معلومات مهمة عن الباحث. وسيتم استخدام بيانات مشابهة للبيانات المستخدمة في المثال السابق لبناء نموذج تحليل التباين بعاملين واختبار فرضيات حول النموذج.

طريقة التغليف \ المتجر	1	2	3
1	16 13	18 19	19 21
2	14 15	17 15	13 16
3	19 18	20 18	23 22
4	24 22	28 30	27 24

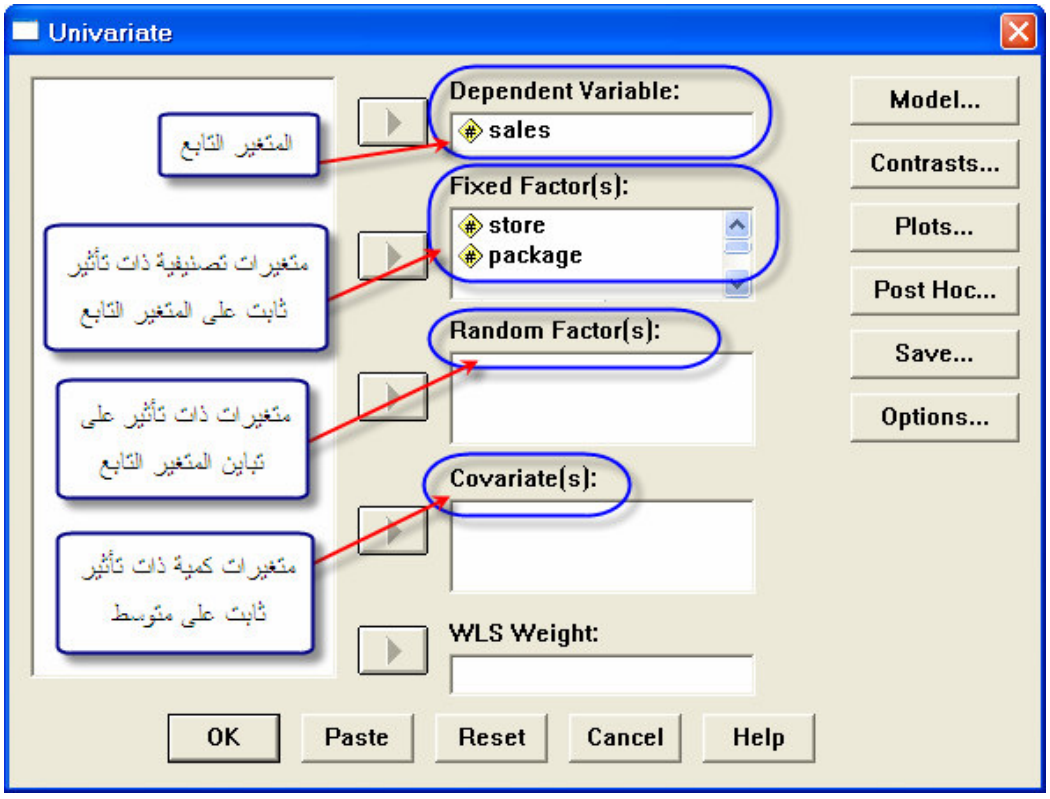
ويمكن صياغة النموذج على النحو التالي:

$$y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$$

وبهذا فإن النموذج يشير إلى أن المبيعات لها متوسط عام لا يتأثر بالمتجر ولا بنوعية التغليف وهو $\mu_{..}$، إلا هذا المتوسط العام يتأثر بنوعية التغليف α_i وبأي المتاجر β_j وبتفاعل نوعية التغليف مع المتجر δ_{ij}، وأخيراً بالخطأ العشوائي ε_{ijk}. ويتم بناء النموذج السابق باستخدام الأمر



ليظهر مربع الحوار التالي



وقد تم تحديد المتغير التابع وهو المبيعات والمتغيرات المستقلة وهي طريقة التغليف والمتاجر ضمن قائمة المتغيرات ذات التأثير الثابت على متوسط المبيعات Fixed Effect، ولا يوجد متغيرات ثابت تأثير على تباين المتغير التابع وكذلك لا يوجد متغيرات كمية ذات تأثير ثابت على متوسط المتغير التابع. وبالنقر على OK تظهر النتائج التالية.

Tests of Between-Subjects Effects

Dependent Variable: SALES

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	474.125[a]	11	43.102	20.283	.000
Intercept	9243.375	1	9243.375	4349.824	.000
PACKAGE	383.458	3	127.819	60.150	.000
STORE	48.000	2	24.000	11.294	.002
PACKAGE * STORE	42.667	6	7.111	3.346	.036
Error	25.500	12	2.125		
Total	9743.000	24			
Corrected Total	499.625	23			

a. R Squared = .949 (Adjusted R Squared = .902)

ويمكن باستخدام خصائص الجداول المحورية تنقيح الجدول السابق لاستبعاد تأثير المتوسط العام على مجموع مربعات الفروق.

Tests of Between-Subjects Effects

Dependent Variable: SALES

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
PACKAGE	383.458	3	127.819	60.150	.000
STORE	48.000	2	24.000	11.294	.002
PACKAGE * STORE	42.667	6	7.111	3.346	.036
Error	25.500	12	2.125		
Corrected Total	499.625	23			

ويختبر الجدول السابق الفرضيات التالية

$$H_0 : \alpha_i = 0$$
$$H_a : \alpha_i \neq 0$$

الفرضية الأولى

وبناء على قيمة (F=60.15) و(P-Value=0.00)، فإنه يمكن رفض فرضية العدم وبذلك فإن لطريقة التغليف عامل مؤثر على المبيعات.

$$H_0 : \beta_j = 0$$
$$H_a : \beta_j \neq 0$$

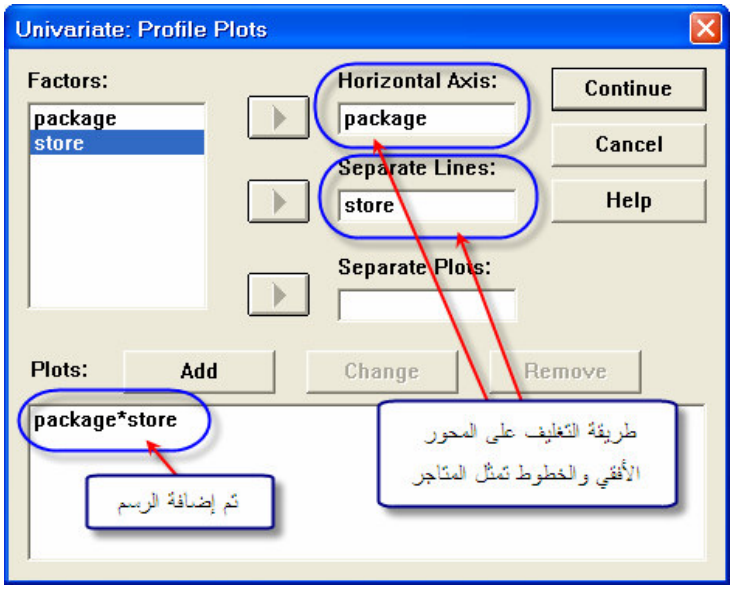
الفرضية الثانية

وبناء على قيمة (F=11.29) و(P-Value=0.002)، فإنه يمكن رفض فرضية العدم وبذلك فإن المتجر عامل مؤثر على المبيعات.

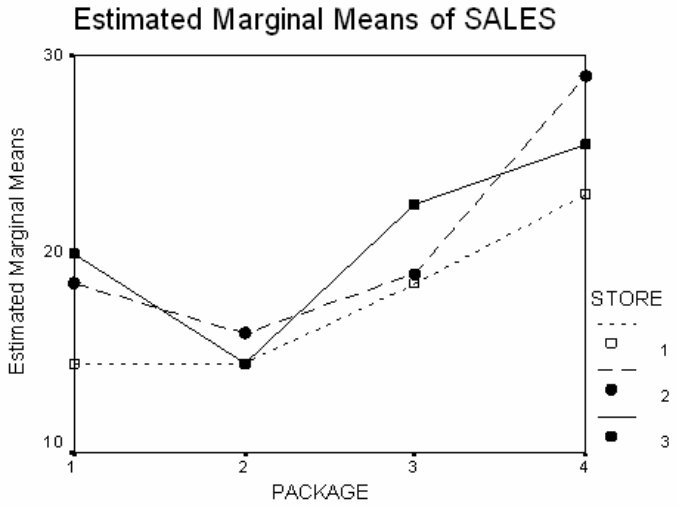
$$H_0 : \delta_{ij} = 0$$
$$H_a : \delta_{ij} \neq 0$$

الفرضية الثالثة

وبناء على قيمة (F=3.35) و(P-Value=0.036)، فإنه يمكن رفض فرضية العدم وبـــذلك فإنه يوجد تفاعل بين طريقة التغليف والمتجر كعامل مؤثر على المبيعات. وبافتراض صحة النموذج السابق، فإنه يمكن تمثيل المتوسطات الحدية بيانياً لعرض تـــأثير طريقـــة التغليـــف والمتجر وذلك بالنقر على زر Plots في مربع الحوار السابق ليظهر لنا مربع الحوار التالي.



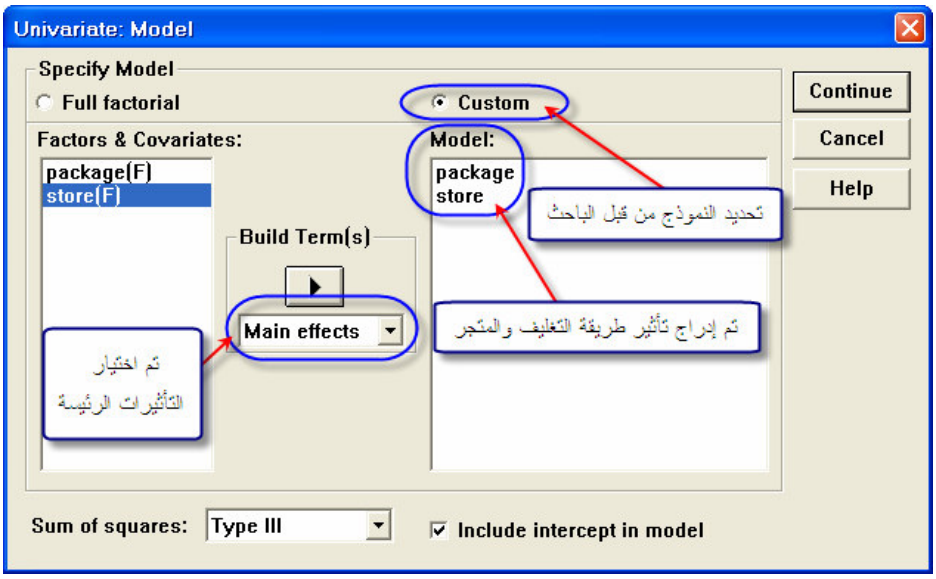
وبالنقر على Continue ثم OK تظهر النتائج التالية.



ويلاحظ التأثير لطريقة التغليف، إلا أن تأثير المتاجر ليس بنفس القوة وكذلك فإن التفاعل بين طريقة التغليف والمتاجر ضعيف.

ويلاحظ من جدول تحليل التباين السابق أنه لو تم استخدام مستوى معنوية $(\alpha = 0.01)$ فإنـــه سيتم قبول فرضية العدم لثالثة مما يعنى أنه لا يوجد تفاعل بين المتجر وطريقـــة التغليـــف.

ويمكن استبعاد التفاعل من النموذج السابق بالنقر على زر Model في مربع الحوار الرئيسي لتحليل التباين ليظهر المربع التالي.



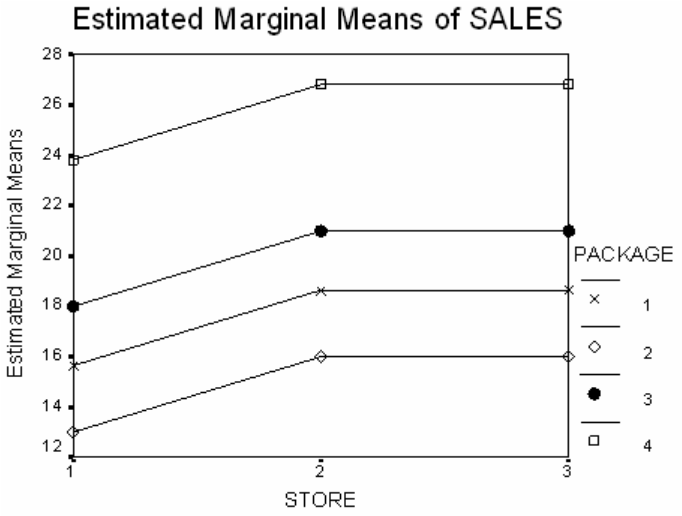
وبالنقر على Continue ثم OK تظهر النتائج التالية، وقد تم تنقيح الجدول لاستبعاد تأثير المتوسط العام على مجموع مربعات الفروق.

Tests of Between-Subjects Effects

Dependent Variable: SALES

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
PACKAGE	383.458	3	127.819	33.752	.000
STORE	48.000	2	24.000	6.337	.008
Error	68.167	18	3.787		
Corrected Total	499.625	23			

ويمكن استخدام النتائج السابقة لاختبار الفرضيتين السابقتين الأولى والثانية. وتشير النتائج إلى رفض فرضيتي العدم في كلا الفرضيتين. ويمكن عرض المتوسطات الحدية لهذا النموذج بنفس الطريقة السابقة.



ويمثل المحور الأفقي في هذا الرسم المتاجر، في حين أن الخطوط تمثل طريقـــة التغليـــف. ويلاحظ التأثير القوي طريقة التغليف، أما المتاجر فكان تأثيرها أقل.