المتغيرات الوهمية Dummy Variables

يتم تطبيق تحليل الانحدار بصورة عامة على المتغيرات الكمية (Quantitative)أي المتغيرات التي تقاس وتأخذ قيما معينة،بل بشكل صفات أو أنواع ، وتسمى بالمتغيرات النوعية(Quantitative variables) ومثال ذلك الجنس (ذكر أو أنثى)والحالة التعليمية (حاصل على الشهادة ثانوية ،كلية أو شهادة عليا)....الخ وغير ذلك من الصفات النوعية:

ويمكن تطبيق تحليل الانحدار على مثل هذه المتغيرات النوعية كما هو الحال في المتغيرات الكمية باستخدام المتغيرات الوهمية أو ما تسمى بالمتغيرات الدالة أو العبرة (Indicated Variables)والتي تعطى قيما متسلسلة(0،1،2،3،.....الخ)ولكن الطريقة الأسهل استخداما والأكثر وضوحا عند تفسيرها هي أعطاء الرقمين (0،1)للتعبير عن هذه المتغيرات

قاعدة عامة للتعبير عن المتغيرات النوعية:

يمكن التعبير عن المتغيرات النوعية بهيئة أستخدام كما يلي:

عندما يكون للمتغير النوعي k من المستويات (levels) أو الاقسام أو المجاميع أو الفئات (classes) فأنه يمكن تمثيله ب (k−1)من المتغيرات الدالة أو الوهمية ذات الفئات المختلفة وكما يلي:

أ- **عندما يكون للمتغير النوعي فئتان أو مجموعتان فقط**

نفرض ان لدينا متغير نوعي يعبر عن جنس رب الاسرة الذي قد يكون ذكرا أو أنثى وعليه فأن المتغير النوعي لدينا يتكون من مستويين (k=2) وبهذا يمكن التعبير عن هذا المتغير النوعي بمتغير كمي واحد(k=1) وكما يلي:

نرمز للجنس مثلا بx_i فتكون قيم هذا المتغير الكمي كما يلي:

$$x_i = \begin{cases} 1 \\ \\ 0 \end{cases}$$

1 إذا كان رب الأسرة ذكرا

0 إذا كان رب الأسرة أنثى

ب-عندما يكون للمتغير النوعي أكثر من فئتين:

نفرض أن لدينا ألان متغير نوعي يعبر عن حالة التدخين لشخص معين وبهذا يمكن تقسيم هذا الشخص إلى ثلاث فئات هي:

الفئات 1- لايدخن

2-يدخن أحيانا

3-يدخن كثيرا

ونلاحظ أن هذا المتغير النوعي يحتاج إلى متغيرين وهميين (k-1=2) فلو رمزنا لحالة التدخين للشخص بالرمز x_i فان المتغيرات الوهمية ستكون

$$x_i = \begin{cases} 1 \\ \\ 0 \end{cases}$$

1 إذا كان رب الأسرة يدخن أحيانا

0 إذا كان رب الأسرة غير ذلك

$$x_i = \begin{cases} 1 \\ \\ 0 \end{cases}$$

1 إذا كان رب الأسرة يدخن كثيرا

0 إذا كان رب الأسرة غير ذلك

ويتضح من المتغيرين الوهميين أعلاه أن الشخص الذي لايدخن أخذ كأساس للمقارنة لان قيمته كانت صفرا في كلا المتغيرين ،كما نلاحظ أن القيم التي تم أخذها (0،1) للمتغير الوهمي لم تهدف ال أعطاء الأهمية الكمية لفئات المتغير النوعي بل أن هدفها هو التمييز بين هذه الفئات ،حيث يمكن استخدام أرقام أخرى بدلا من الصفر والواحد .ولكن كما ذكرنا فان الصفر والواحد هي أكثر الأرقام استعمالا من غيرها وأبسطها في تفسير النتائج.

ومن الجدير بالذكر أننا لم نستخدم k من المتغيرات الدالة حتى لانحصل على مصفوفة مفردة (singular) وهي التي ليس لها معكوس (محددها يساوي)

أستخدام المتغير الوهمي عند وجود متغير مستقل نوعي واحد في المعادلة الانحدار

نفرض أن معادلة انحدار متعدد بمتغيرين وكما يلي :

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

حيث أن x_1 متغير مستقل كمي وx_2 متغير مستقل نوعي

ولهذا النوع من المعادلات فأننا من الممكن أن نواجه حالتين تعتمدان على فئات المتغير النوعي،وكل من هاتين الحالتين ممكن أن تتضمن أيضا وجود أو عدم وجود تداخل بين المتغير المستقل النوعي والكمي وكما هو واضح فيما يلي

أ–عندما يكون للمتغير النوعي فئتان أو مجموعتان فقط:

1–عند عدم وجود تداخل بين المتغير المستقل النوعي والمتغير المستقل الكمي:

نفرض أننا نريد دراسة العلاقة بين الراتب السنوي كمتغير معتمد y وسنسن الخدمة x_1 والخبرة الادارية x_2 لعدد من الموظفين في أحدى المؤسسات

فنلاحظ من هناك أن x_1 هو متغير كمي بينما x_2 هو متغير نوعي بفئتين (له الخبرة أو ليس له خبرة) ولكون هذا المتغير له فئتان فقط لذا فمن الممكن تمثيله بمتغير وهمي واحد (k–1=1) وهو

$$x_2 = \begin{cases} 1 \\ \\ 0 \end{cases}$$

1 الموظف الذي له خبرة

0 الموظف الذي ليس له خبرة

وستكون دالة الاستجابة في هذه الحالة كما يلي:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

وبهذا فان للموظف الذي ليست له خبرة أدارية ستكون الدالة كما يلي

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (0)$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1$$

وهذه الدالة عبارة عن خط مستقيم بمعمل انحدار $\hat{\beta}_1$ ونقطة تقاطع $\hat{\beta}_0$

أما الموظف الذي له خبرة أدارية ستكون الدالة كما يلي

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 (1)$$

$$\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 X_1$$

الدالة عبارة عن خط مستقيم بمعمل انحدار $\hat{\beta}_1$ولكن ونقطة تقاطع $\hat{\beta}_0 + \hat{\beta}_1$

ويمكن توضيح معلمات هذه المعادلة من خلال الرسم البياني التالي

ومن خلال الرسم أعلاه نلاحظ أن دالة الاستجابة $\hat{Y}_i$ هي دالة خطية لسنين الخدمة x1 بمعامل انحدار $\hat{\beta}_1$ في كلا الحالتين (له خبرة أو ليس خبرة) ،

أما الفرق بين موقع دالة الاستجابة للموظف الذي له خبرة او ليس له خبرة متمثلة $\hat{\beta}_2$ أي تمثل الفرق في التأثير بين الموظف الذي له خبرة أو الذي ليس له خبرة،وبمعنى أوضح فأن راتب الشخص الذي له خبرة أدارية يزيد بمقدار $\hat{\beta}_2$ عن راتب الشخص الذي ليس له خبرة أدارية.

أما بالنسبة ل x_1 فأنها تعني أن زيادة عدد سنين الخدمة للموظف سنة واحدة سيزيد مقدار راتبه بمقدار $\hat{\beta}_1$ (مع بقاء x_2 ثابتة)

وعند حساب حدود الثقة إلى $\hat{\beta}_2$ مثلا :

$$\beta_{2L} \le \beta_2 \le \beta_{2U}$$