الجمهورية العربية المتحدة

معهد التخطيط القومى

مذكرة رقم (٩٦٤)

مقدمة فى مبادئ التحليل الكمى الاحصائى

اعداد

دكتور / كمال أحمد الجنزورى

مايو ١٩٧٠

القاهرة

٣ شارع محمد مظهر - بالزمالك

الآراء التي وردت في هذه المذكـــــــــــــــــــــرة
تمثل رأى الكاتب ولاتمثل رأى المعهد ذاتـــــــــــــــه

تمهيـــــــد

يهتم الدارسون فى مجال البحوث التطبيقيه بتحديد العلاقه بين المتغيرات المختلفـــــه التى تحكم أو تؤثر على ظاهرة معنيه ، فمثلا فى مجال البحوث الاقتصاديه فانهم يهتمون بتحديـد العلاقه بين المتغيرات الاقتصاديه التى تحكم ظاهره اقتصاديه معينه ، كمتغيرات أسعار سلعه معينـه وأسعار السلع المتصله (البديله أو المكملة) ودخول الأفراد .... الخ التى تحكم ظاهـره زيادة أو نقص الطلب على هذه السلعه خلال فتره معينه ، أو كمتغيرات أسعار العناصر المتغـــيره واسعار العناصر الثابته ..... الخ التى تؤثر على زيادة أو نقص التكاليف الانتاجيه . ولا يقف الاهتمام فى مجال البحوث الاقتصاديه عند تحديد العلاقه بين المتغيرات الاقتصاديه فحسب ، بل قد يذهـب الى تحديد العلاقه بين المتغيرات الاقتصاديه والطبيعيه والتكنولوجيه و .... الخ فمثلا قد يتجه الاهتمام الى تحديد العلاقه بين المتغيرات الاقتصاديه والتكنولوجيه والجويـــــه التى تحكم ظاهره ارتفاع أو تدهور غله محصول زراعى معين .

وتعتبر الأدوات التى يستخدمها الدارسون فى تحليلهم وتقديرهم للعلاقات بين المتغـــيرات المختلفه التى تحكم بعض الظواهر عديده ، ومن بينها تحليل الانحدار (Regression analysis) وتحليل الارتباط (Correlation analysis) وتحليل التبايـــــن (Analysis of variance)

تحليل الانحـــدار
Regression Analysis

(١) تعريف :

يرتكز تحليل الانحدار على دراسة التوزيع الاحصائى او التوزيع التكرارى (Frequency distribution) لمتغير معين عند مستويات مختلفه لمتغير آخـــر . ومن أمثله ذلك دراســــه التوزيع الاحصائى لغله محصول زراعى معين عند استخدام كميات مختلفه من الأسمده ، أو لصلابـــه

ماده البلاستيك عند معاملتها حراريا لفترات زمنيه مختلفه ، أو للانفاق الاستهلاكى للأسر عند مستويات مختلفه من الدخل المتاح لهذه الأسر .

وبدراسة التوزيع الاحصائى لقيم أحد المتغيرات عند مستويات مختلفه لمتغير آخر ، يمكن التنبوء بقيم هذا المتغير . أى أن الباحثين غالبا مايلجأون الى استخدام قيم أحد المتغيرات فى التقدير أو التنبوء بقيم أحد المتغيرات الأخرى . ويطلق على المتغير الأول " المتغير المستقل Indepen-dent variable " وعلى المتغير الآخر " المتغير التابع Depend-ent variable " لأن قيمه تتحدد على أساس قيم المتغير المستقل .

فاذا فرض مثلا أننا نرغب فى دراسة التوزيع الاحصائى للانفاق الاستهلاكى لمجتمع معين من الأسر بالنسبه للدخول المتاحه لهذه الأسر ، فاننا نقسم هذا المجتمع الى مجموعات على أى يتساوى الدخل المتاح للأسره داخل المجموعه الواحدة . وعلى هذا فانه يوجد توزيع للانفاق الاستهلاكى للأسر المختلفه عند أى مستوى معين من الدخل المتاح ، ويمثل متوسط هذا التوزيع متوسط الانفاق الاستهلاكى لكل الأسر داخل المجموعة .

ويمكن تعريف انحدار الانفاق الاستهلاكى للأسر على الدخل المتاح لهذه الأسر على أنه متوسط توزيع الانفاق الاستهلاكى لكل الأسر التى لها دخل مماثل . ونظرا لأن توزيع الانفاق الاستهلاكى للأسر يعتمد على الدخل المتاح لهذه الأسر ، فان الانفاق الاستهلاكى يشار اليه كمتغير تابع ويرمز له بالحرف (ى) ، بينما يشار الى الدخل المتاح كمتغير مستقل ويرمز له بالحرف (س) . كما نرمز لمتوسط توزيع الانفاق الاستهلاكى (ى) عند مستوى معين من الدخل المتاح (س) بالحرف ($م_{ى.س}$) ، ولتباين هذا التوزيع بالحرف ($ت^{2}_{ى.س}$) . ويعتبر كل من ($م_{ى.س}$) ، ($ت^{2}_{ى.س}$) ثابتا عند مستوى معين من الدخل (س) ولكنهما ربما يختلفان من توزيع لآخر للانفاق الاستهلاكى عند مستويات مختلفه من الدخل .

ونظرا لانه فى كثير من تطبيقات نظريه الانحدار يفترض أن منحنى انحدار المتغير التابع على المتغير المستقل خط مستقيم ، وعلى هذا اذا فرضنا ان العلاقه بين الانفاق الاستهلاكى والدخل المتاح خطيه ، فانه يمكن التعبير عن متوسط التوزيع الاحصائى للانفاق الاستهلاكى عند مستوى معين من الدخل المتاح فى المعادله التالية :

$$\text{ل}_{\text{ى . س}} = \text{أ} + \text{ب} \left(\text{س} - \bar{\text{س}} \right) \qquad (١)$$

حيث أن كل من (أ) ، (ب) يعتبر ثابتا غير معروف Unknown constant

ويطلق عليهما " معاملات الانحدار Regression coefficients "

وتشير (س) الى قيمه الدخل المتاح للأسره ، بينما تشير ($\bar{\text{س}}$) الى متوسط قيم الدخل المتاح لمختلف أسر المجتمع ·

(٢) <u>طريقه المربعات الصغرى</u> Method of least squares

تعتبر قيمه كل من (أ) ، (ب) فى المعادله (١) ثابته مجهوله لانها قيمه للمجتمع كله ، لذا يلزم لتقدير قيمه كل منهما أن نختار عينه عشوائيه من أسر المجتمع موضع الدراسة ونقسمها الى مجموعات ، على أن تكون كل الأسر داخل المجموعة الواحده لها دخل مماثل · وعلى هذا يمكن تقدير قيمتى (أ) ، (ب) ، ويرمز للقيمه المقدره (أ) بالحرف ($\hat{\text{أ}}$) والقيمه المقدره (ب) بالحرف ($\hat{\text{ب}}$) · وبالطبع فانه بتقدير كل من ($\hat{\text{أ}}$) ، ($\hat{\text{ب}}$) يمكن تقدير متوسط توزيع الانفاق الاستهلاكى للأسر التى لها دخل مماثل، ويرمز لهذا المتوسط المقدر - الذى يمثل القيمه المقدره للمتوسط $\text{ل}_{\text{ى . س}}$ - بالحرف ($\hat{\text{ى}}$) · ويقدر المتوسط ($\hat{\text{ى}}$) طبقا للمعادلة التاليـــة :

$$\hat{\text{ى}} = \hat{\text{أ}} + \hat{\text{ب}} \left(\text{س} - \bar{\text{س}} \right) \qquad (٢)$$

ويمكن تقدير قيمه كل من ($\hat{\text{أ}}$) ، ($\hat{\text{ب}}$) باستخدام طريقه المربعات الصغرى على النحو التالى :

$$\hat{\text{أ}} = \frac{\text{مجـ ى}}{\text{ن}} \qquad (٣)$$

$$\hat{\text{ب}} = \frac{\text{مجـ ى}\ (\text{س} - \bar{\text{س}})}{\text{مجـ}\ (\text{س} - \bar{\text{س}})^{٢}} \qquad (٤)$$

ويلاحظ أن (مجـ) تشير الى المجموع ، أي أن (مجـ ى) تشير الى مجموع قيم (ى) كما يلاحظ أن قيمه البسط فى المعادله (٤) تتساوى مع القيمه التاليه :

$$\text{مجـ ى}\ (\text{س} - \bar{\text{س}}) = \text{مجـ س ى} - \bar{\text{س}}\ \text{مجـ س}$$

$$= \text{مجـ س ى} - \text{ن}\ \bar{\text{س}}\ \text{ى}$$

$$= \text{مجـ س ى} - \frac{١}{\text{ن}}\ (\text{مجـ س})\ (\text{مجـ ى})$$

وأن قيمه المقام فى المعادلة (٤) تتساوى مع القيمه التالية :

$$\text{مجـ}\ (\text{س} - \bar{\text{س}})^{٢} = \text{مجـ س}^{٢} - \frac{١}{\text{ن}}\ (\text{مجـ س})^{٢}$$

وعلى هذا فان قيمه ($\hat{\text{ب}}$) فى المعادله (٤) يمكن التعبير عنها فى الصوره التاليه :

$$\hat{\text{ب}} = \frac{\text{مجـ س ى} - \frac{١}{\text{ن}}\ (\text{مجـ س})\ (\text{مجـ ى})}{\text{مجـ س}^{٢} - \frac{١}{\text{ن}}\ (\text{مجـ س})^{٢}} \qquad (٥)$$

مثال :

اذا فرض أننا نرغب فى دراسة التوزيع الاحصائى للانفاق الاستهلاكى (ى) لمجتمع معيين من الأسر عند مستويات مختلفه من الدخل المتاح (س) ، وأن العلاقه بين الانفاق الاستهلاكى والدخل المتاح خطيه . وأننا اخترنا عينه عشوائيه من أسر المجتمع موضع الدراسة وكانت قيم (ى) ، (س) فى هذه العينه تطابق القيم الافتراضيه المدونه فى الجدول رقم (١) .

جدول رقم (١)

القيم الافتراضيه للدخل المتاح والانفاق الاستهلاكى لعينه عشوائيه من الأسر

قيمه الدخل المتاح للأسرة بالجنيه (س)	قيمه الانفاق الاستهلاكى للأسره بالجنيه (ى)
١٠	٧
١٠	٨
١٢	٩
١٢	١٠
١٢	١١
١٤	١٢
١٤	١١
١٦	١٣
١٦	١٤
١٨	١٥

احسب من هذه القيم قيمتى $\hat{أ}$ ، $\hat{ب}$ باستخدام المعادلتين السابقتين (٣) ،(٥) على التوالى .

الحل :

١ــ لتقدير $\hat{\text{أ}}$ ، $\hat{\text{ب}}$ وبالتالى $\hat{\text{ى}}$ لابد من تقدير القيم الخمس التاليــــة :
(مجـ س ، مجـ ى ، مجـ س^{٢} ، مجـ ى^{٢} ، مجـ س ى) ، والتى تقدر كما يتبين فى الجدول رقم (٢)

جدول رقم (٢)

س	ى	س^{٢}	ى^{٢}	س ى
١٠	٧	١٠٠	٤٩	٧٠
١٠	٨	١٠٠	٦٤	٨٠
١٢	٩	١٤٤	٨١	١٠٨
١٢	١٠	١٤٤	١٠٠	١٢٠
١٢	١١	١٤٤	١٢١	١٣٢
١٤	١٢	١٩٦	١٤٤	١٦٨
١٤	١١	١٩٦	١٢١	١٥٤
١٦	١٣	٢٥٦	١٦٩	٢٠٨
١٦	١٤	٢٥٦	١٩٦	٢٢٤
١٨	١٥	٣٢٤	٢٢٥	٢٧٠
المجموع = ١٣٤	١١٠	١٨٦٠	١٢٧٠	١٥٣٤

٢ــ وباستخدام هذه المجاميع يمكن تقدير كل من $\hat{\text{أ}}$ ، $\hat{\text{ب}}$ على النحو التالى :

$$\hat{\text{أ}} = \frac{\text{مج ى}}{\text{ن}} = \frac{\text{١١٠}}{\text{١٠}} = \text{١١}$$

$$\hat{\text{ب}} = \frac{\text{مج س ى} - \frac{\text{١}}{\text{ن}}(\text{مج س})(\text{مج ى})}{\text{مج س}^{\text{٢}} - \frac{\text{١}}{\text{ن}}(\text{مج س})^{\text{٢}}}$$

$$= \frac{\text{١٥٣٤} - \frac{\text{١}}{\text{١٠}}(\text{١٣٤})(\text{١١٠})}{\text{١٨٦٠} - \frac{\text{١}}{\text{١٠}}(\text{١٣٤})^{\text{٢}}}$$

$$= \frac{\text{١٥٣٤} - \frac{\text{١}}{\text{١٠}}(\text{١٤٧٤٠})}{\text{١٨٦٠} - \frac{\text{١}}{\text{١٠}}(\text{١٧٩٥٦})}$$

$$= \text{٩٣٢ ر ٠}$$

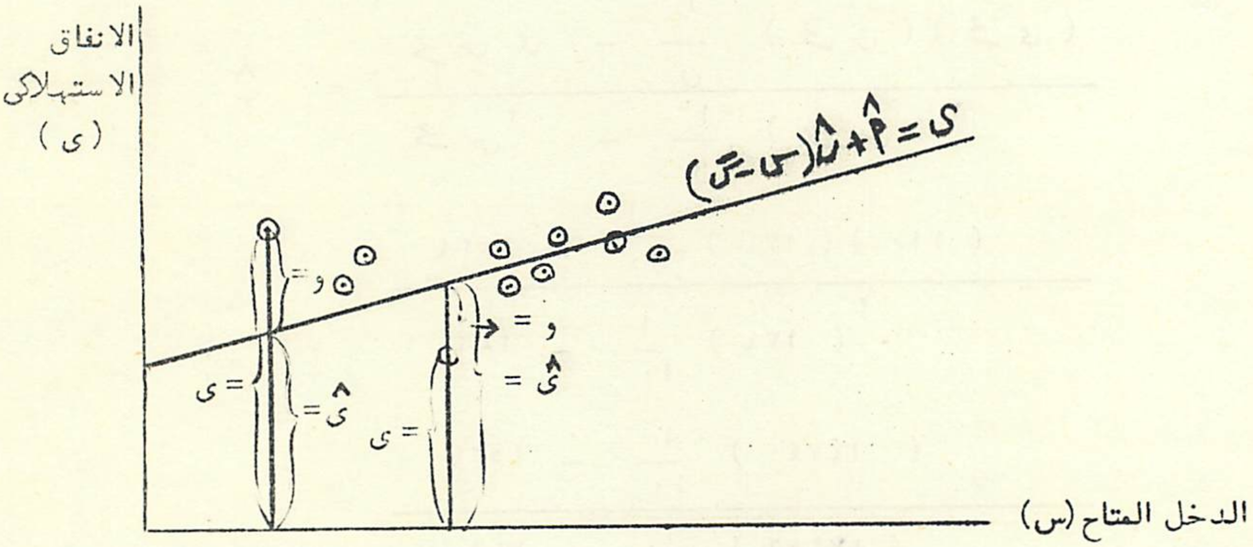
٣- وعلى هذا فان قيمه متوسط توزيع الانفاق الاستهلاكى ($\hat{\text{ى}}$) عند مستوى معين من الدخل (س) يقدر كما يلى :

$$\hat{\text{ى}} = \hat{\text{أ}} + \hat{\text{ب}}\,(\text{س} - \bar{\text{س}})$$

$$= \text{١١} + \text{٩٣٢ر٠}\,(\text{س} - \text{٤ ر ١٣})$$

ويمكن ايضاح القيم المقدرة لمتوسطات التوزيعات التكراريه للانفاق الاستهلاكى للأسر عند مستويات مختلفه من الدخل فى الشكل البيانى رقم (١)

شكل رقم (١)



ويتضح من الشكل البيانى رقم (١) الجوانب التالية :

١ـ أن الخط ($\hat{ى} = \hat{أ} + \hat{ب}(س - \bar{س})$) الذى يشير الى خط انحدار الانفاق الاستهلاكى (ى) على الدخل المتاح (س) ، يمثل فى الحقيقه قيم المتوسطات لتوزيعات المتغير التابع (ى) عند مستويات مختلفه من المتغير المستقل (س) ٠

٢ـ انه عند أى نقطه أو مشاهده فى هذا الشكل ، فان المسافه الرأسيه بين هذه النقط أو المشاهدات والمحور الأفقى تشير الى الانفاق الاستهلاكى الفعلى ، بينما تمثل المسافه بين خط الانحدار والمحور الأفقى القيمه المتوسطه للانفاق الاستهلاكى لأسر المجموعه التى لها نفس الدخل ٠ أما المسافه التى بين النقطه او المشاهده وخط الانحدار فتشير الى الانحراف او الفرق بين الانفاق الفعلى والقيمه المتوسطه لتوزيع

الانفاق . وبديهى فان هذا الفرق اما ان يكون موجبا اذا كانت النقطه او المشاهده تقع أعلى خط الانحدار أو سالبا اذا كانت المشاهده تقع أسفل خط الانحدار . ويطلق على هذه الفروق " المتغيرات العشوائيه المستقله " The independent random variables وهى تعتمد على الفروض الأساسيه التاليه :

أ ـ أن توزيع هذه المتغيرات معتدل Normal distribution بمتوسط يساوى صفرا وتباين ثابت ومستقل عن قيم المتغير المستقل (س) .

ب ـ أن قيمه كل من هذه المتغيرات العشوائيه مستقل عن قيمه المتغير الآخر .

يتضح مما تقدم أنه بطريقه المربعات الصغرى يمكن تقدير خط الانحدار $\hat{ى} = \hat{أ} + \hat{ب}\,(س - \bar{س})$ بحيث يكون مجموع مربعات الانحرافات أو المسافات الرأسيه بين النقط او المشاهدات وهذا الخط أصغر مايمكن . فاذا رمزنا لهذه المسافات الرأسيه والتى تمثل المتغيرات العشوائيه المستقله بالحروف $و_١$ ، $و_٢$ ، $و_٣$ ، ...... ، $و_ن$ ، فانه طبقا لطريقه المربعات الصغرى فان :

$$مج\,(و_١^٢ + و_٢^٢ + و_٣^٢ + \cdots\cdots + و_ن^٢)$$

يعتبر أصغر ما يمكن ،أو بمعنى آخر فان طريقه المربعات الصغرى المستخدمه فى تقدير قيمتى $\hat{أ}$ ، $\hat{ب}$ من العينه ، تؤدى الى أن القيمه التاليه :

$$مج\,(ى - \hat{ى})^٢ = \left[ى - \underbrace{\hat{أ} + \hat{ب}\,(س - \bar{س})}_{\hat{ى}} \right]^٢ \qquad (٦)$$

تكون أصغر مايمكن ، ويمكن رياضيا تتبع تحديد قيمه كل من $\hat{\text{أ}}$ ، $\hat{\text{ب}}$ والتى تحقق أصغر قيمه لمجموع مربعات المسافات بين النقط أو المشاهدات وخط الانحدار المقدر ، بالحصول على التفاضل الجزئى لقيم المعادله (٦) بالنسبة ($\hat{\text{أ}}$) ، ($\hat{\text{ب}}$) على التوالى ثم مساواتها بالصفر كما يلى :

$$\text{مجـ}\ ٢ \left[\text{ى} - \hat{\text{أ}} - \hat{\text{ب}} \left(\text{س} - \bar{\text{س}} \right) \right] \left(-١ \right) = \text{صفر} \qquad (٧)$$

$$\text{مجـ}\ ٢ \left[\text{ى} - \hat{\text{أ}} - \hat{\text{ب}} \left(\text{س} - \bar{\text{س}} \right) \right] \left(\text{س} - \bar{\text{س}} \right) = \text{صفر} \qquad (٨)$$

أى أن :

$$\text{مجـ ى} - \text{ن}\ \hat{\text{أ}} - \hat{\text{ب}}\ \overbrace{\text{مجـ} \left(\text{س} - \bar{\text{س}} \right)}^{\text{صفر}} = \text{صفر} \qquad (٩)$$

$$\text{مجـ ى} \left(\text{س} - \bar{\text{س}} \right) - \hat{\text{أ}}\ \overbrace{\text{مجـ} \left(\text{س} - \bar{\text{س}} \right)}^{\text{صفر}} - \hat{\text{ب}} \left(\text{س} - \bar{\text{س}} \right)^{٢} = \text{صفر} \qquad (١٠)$$

وعلى هذا تقدر كل من ($\hat{\text{أ}}$) ، ($\hat{\text{ب}}$) من المعادلتين (٩) ، (١٠) على التوالى كما يلى :

$$\hat{\text{أ}} = \frac{\text{مجـ ى}}{\text{ن}} \qquad (١١)$$

$$\hat{\text{ب}} = \frac{\text{مجـ ى} \left(\text{س} - \bar{\text{س}} \right)}{\text{مجـ} \left(\text{س} - \bar{\text{س}} \right)^{٢}} \qquad (١٢)$$

وبديهى فان المعادله (١١) تطابق المعادله (٣) ، وأيضا المعادله (١٢) تطابق المعادله (٤) السابق استخدامها فى تقدير ($\hat{\text{أ}}$) ، ($\hat{\text{ب}}$) وبالتالى فى تقدير

$$\left[\hat{\text{ى}} = \hat{\text{أ}} + \hat{\text{ب}} \left(\text{س} - \bar{\text{س}} \right) \right] .$$

مثال :

اذا فرض أننا نرغب فى دراسة التوزيع الاحصائى للوزن فى مجتمع معين من الاشخاص بالنسبه لأطوال هؤلاء الاشخاص ، وأننا اخترنا عينه عشوائيه حجمها ١٢ شخصا ، وكانت قيم الأوزان والأطوال تطابق القيم الافتراضيه المدونه فى الجدول رقم (٢) .

جدول رقم (٢)

القيم الافتراضيه للأوزان والأطوال لعينه عشوائيه من الاشخاص

الاطوال بالبوصة (س)	الأوزان بالرطل (ى)
٦٠	١١٠
٦٠	١٣٥
٦٠	١٢٠
٦٢	١٢٠
٦٢	١٤٠
٦٢	١٣٠
٦٢	١٣٥
٦٤	١٥٠
٦٤	١٤٥
٧٠	١٧٠
٧٠	١٨٠
٧٠	١٦٠

احسب من هذه القيم ، قيمه كل من $\hat{أ}$ ، $\hat{ب}$ ، $\hat{ى}$.

الحل :

١ـ تقدر الاحصاءات اللازمة لتقدير كل من $\hat{\text{أ}}$ ، $\hat{\text{ب}}$ ، $\hat{\text{ى}}$ على النحو التالـــــى :

$$\text{مجـ س} = \text{٧٦٦} \qquad \text{مجـ ى} = \text{١٧٠٠}$$

$$\text{مجـ س}^{\text{٢}} = \text{٤٩٠٦٨} \qquad \text{مجـ ى}^{\text{٢}} = \text{٢٤٦١٠٠}$$

$$\text{مجـ س ى} = \text{١٠٩٣٨٠}$$

٢ـ ويقدر كل من متوسط الأطوال ($\bar{\text{س}}$) ومتوسط الاوزان ($\bar{\text{ى}}$) كما يلى :

$$\bar{\text{س}} = \frac{\text{مجـ س}}{\text{ن}} = \frac{\text{٧٦٦}}{\text{١٢}} = \text{٦٣ر٨٣}$$

$$\bar{\text{ى}} = \frac{\text{مجـ ى}}{\text{ن}} = \frac{\text{١٧٠٠}}{\text{١٢}} = \text{١٤١ر٦٧}$$

٣ـ ويقدر كل من $\hat{\text{أ}}$ ، $\hat{\text{ب}}$ ، $\hat{\text{ى}}$ على النحو التالى :

$$\hat{\text{أ}} = \frac{\text{مجـ ى}}{\text{ن}} = \frac{\text{١٧٠٠}}{\text{١٢}} = \text{١٤١ر٦٧}$$

$$\hat{\text{ب}} = \frac{\text{مجـ س ى} - \frac{\text{١}}{\text{ن}}\,(\text{مجـ س})\,(\text{مجـ ى})}{\text{مجـ س}^{\text{٢}} - \frac{\text{١}}{\text{ن}}\,(\text{مجـ س})^{\text{٢}}}$$

$$= \frac{\text{١٠٩٣٨٠} - \frac{\text{١}}{\text{١٢}}\,(\text{٧٦٦})\,(\text{١٧٠٠})}{\text{٤٩٠٦٨} - \frac{\text{١}}{\text{١٢}}\,(\text{٧٦٦})^{\text{٢}}}$$