الانحدار الخطى البسيط

ثانيا:

الانحدار الخطي البسيط

معامل الانحدار

مصادر الاختلاف في الانحدار الخطي

اختبارات الفروض

1- أختبار t

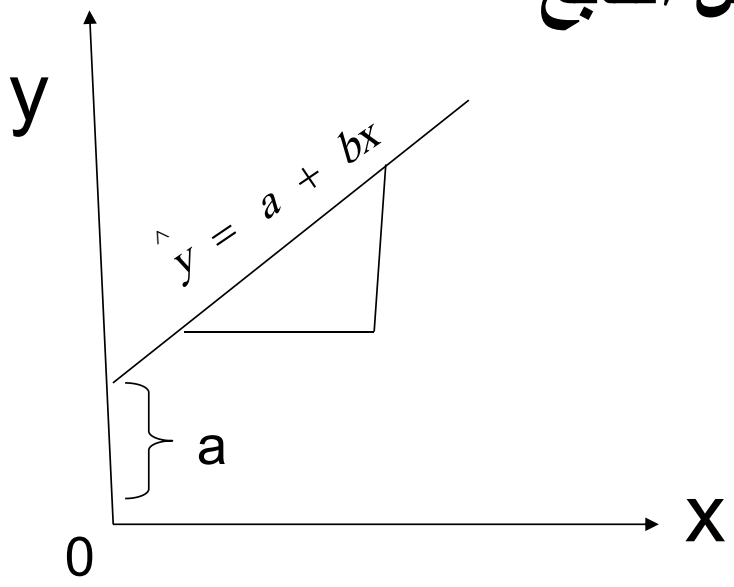
2- اختبار F

معامل التقدير

الانحدار الخطي البسيط
SIMPLE LINEAR REGRESSION

الإنحدار Regression : يقيس العلاقة بين متغيرين أحدهما مستقلX والآخر تابع Y

معامل الانحدار: هومقدار التغير في العامل التابعY نتيجة التغير بـوحدة واحدة من المتغير المستقل ويميز بوحدات قياس العامل التابع



إيجاد الخط الذي يمثل العلاقة بين المتغيرين Y, X

$$b = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{SP}{SSX}$$

$$a = \frac{1}{n}\left(\sum Y_i - b\sum X_i\right) = \overline{Y} - b\overline{X}$$

$$\hat{y} = a + bx$$

$$= \overline{Y} + b\,(X - \overline{X})$$

مثال 1

البيانات في الجدول التالي تبين متوسط الوزن بالكيلو جرام X واستهلاك الغذاء Y بالكيلو لعشرة حيوانات احسب معادلة خط الانحدار!!

X	Y	XY	$\hat{Y}$
4.6	87.1	400.66	90.634
5.1	93.1	-	-
4.8	89.8	-	-
4.4	91.4	-	-
5.9	99.5	-	-
4.7	92.1	-	-
5.1	95.5	-	-
5.2	99.3	-	-
4.9	93.4	-	-
5.1	94.4	481.44	94.479
$\sum X = 49.8$ $\overline{X} = 4.98$	$\sum Y = 935.6$ $\overline{Y} = 93.56$	$\sum XY = 4671.1$	

n=10

1- نحسب SSx مجموع المربعات للعامل المستقل X SSx=1.536

SSy مجموع المربعات للعامل التابع Y SSy=135.604

SP مجموع حواصل ضرب المتغيرين

$$SP = 4671.1 - \frac{(49.8)(935.6)}{10} = 11.812$$

وهو يعني أنه يلزم 69.7 كيلو جرام من الغذاء لكل كيلو من وزن الحيوان

2- نحسب معامل الانحدار b :

$$b = \frac{SP}{SS_X} = \frac{11.812}{1.536} = 7.69$$

3- نحسب a : $a = \overline{Y} - b\overline{X} = 93.56 - 7369(4398) = 55.26$

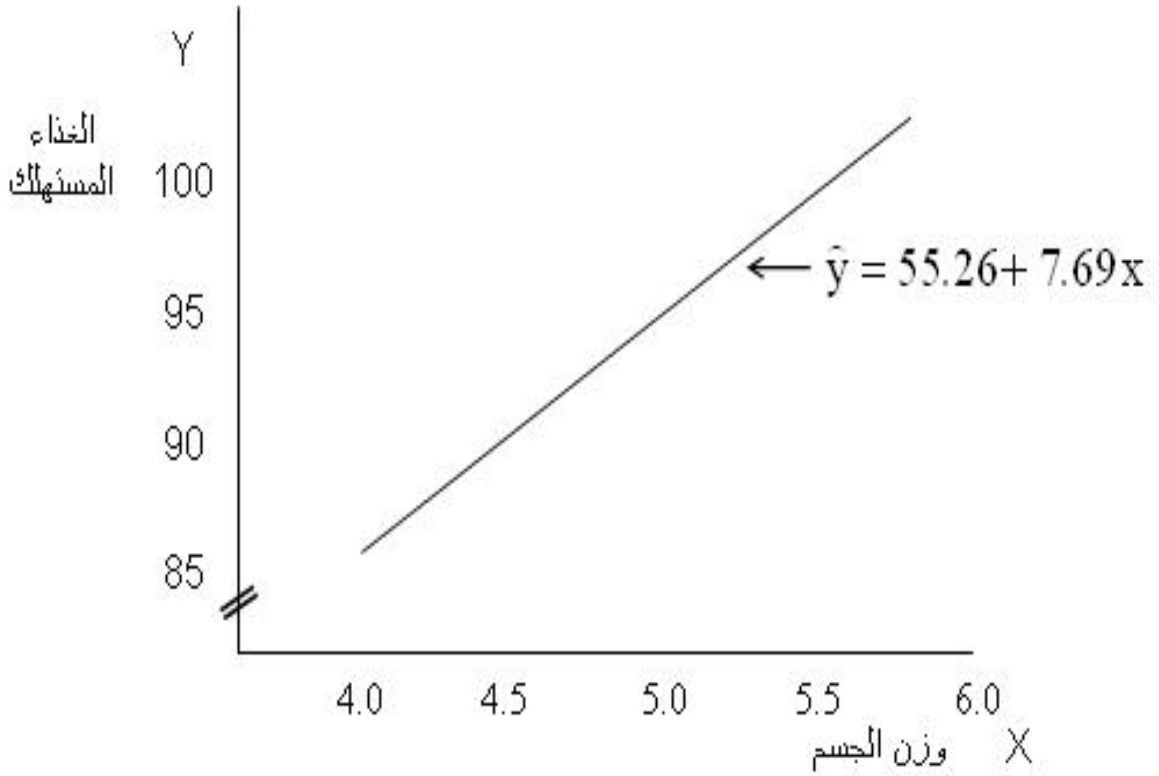
4- تكون معادلة خط الانحدار :

$$\hat{Y} = a + bx$$

$$\hat{Y} = 55.26 + 7.69X$$

5 - بالتعويض في معادلة خـط الانحدار عن قيم X يمكن حساب القيم التقديرية للعامل التابع Y (وهى تمثل متوسط قيم Y التي لها نفس قيمة X)

6- من القيم التقديرية يمكن رسم خط الانحدار الذي يمثل العلاقة بين المتغيرين X,Y كما بالشكل التالي


Y
الغذاء
المستهلك
100
95
90
85
$\hat{y} = 55.26 + 7.69x$
4.0
4.5
5.0
5.5
6.0
وزن الجسم
X

مصادر الاختلاف في الانحدار الخطي
Sources of Variation in Linear Regression

وبمفهوم العينة في الانحدار فإن المشاهدة Y مركبة من

أ- متوسط عينة $\bar{Y}$ وجزء خاص بالانحدار $b(x-\bar{x})$ بدورها تتكون من $\bar{Y}$ متوسط العينة و $\hat{Y}$ المحدد من خط الانحدار $\hat{Y} = \bar{Y} + b(x-\bar{x})$

ب- الانحرافات عن هذا المتوسط (أو residual): $e = Y - \hat{Y}$

- وهو تقدير للانحرافات العشوائية ε ويمكن كتابته أيضا $e_{y.x}$ اى ان : $Y = \hat{Y} + e$

ومنها

$$Y - \overline{Y} = (\hat{Y} - \overline{Y}) + (Y - \hat{Y})$$

$$= b(X - \overline{X}) + e_{Y.X}$$

وبالتالي فإن مجموع مربعات الانحرافات الكلي للعامل التابع SS_y يمكن تقسيمه حسب هذه المصادر كما يلي:

1- جزء يرجع إلى وجود الانحدار SS due Regression

2- جزء يرجع للانحراف عن خط الانحدار SS due Residual

ويمكن تلخيص ما سبق في المعادلة:

Total SS = Reg. SS + Res.SS

$$\sum (Y - \overline{Y})^2 = \sum (\hat{Y} - \overline{Y})^2 + \sum (Y - \hat{Y})^2$$

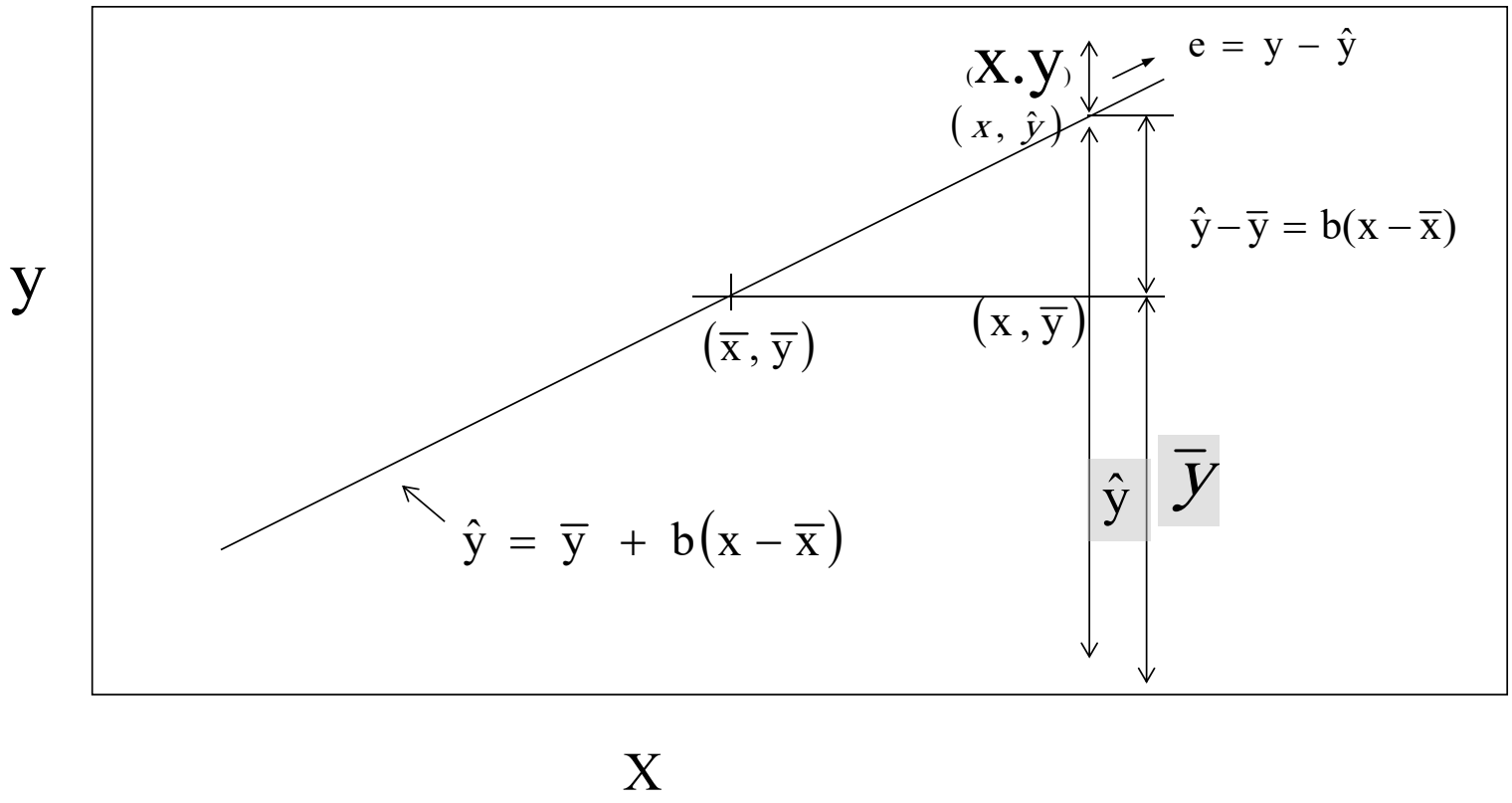
Or: $$\sum Y^2 - \frac{(\sum Y)^2}{n} = \frac{[\sum (X - \overline{X})(Y - \overline{Y})]^2}{\sum (X - \overline{X})^2} + Res.SS$$

$$SS_y = \frac{(SP)^2}{SS_x} + \text{Res. } SS$$

df: n-1 = 1 + n-2

مجموع المربعات المتبقى Res. SS يحسب بالطرح وتباينه هو تقدير للخطأ التجريبى σ^2 .

ويمكن توضيح التقسيم السابق فى المعادلة من الشكل التالى


y
x
(x.y)
$(x, \hat{y})$
e = y − ŷ
ŷ − ȳ = b(x − x̄)
$(\bar{x}, \bar{y})$
$(x, \bar{y})$
ŷ = ȳ + b(x − x̄)
ŷ
$\bar{y}$

اختبارات الفروض
Tests of Hypotheses

اختبار النظرية الفرضية $(\beta = \beta o)$ أو $(\beta = 0)$

اختبار t :

$$t = \frac{(b - 0)}{S_b} = \frac{b}{S_b}$$

$$S_b = \frac{S_{y.x}}{\sqrt{SS_x}} = \sqrt{\frac{S^2_{y.x}}{SS_x}}$$

$(SSy - ((SP^2)/SSx)) / n-2$

وبتطبيق الخطوات على بيانات جدول (1) السابقة

$$S^2_{y.x} = \frac{135.604 - (11.812)^2 / 1.536}{8} = \frac{44.77}{8} = 5.60$$

$$S_b = \sqrt{\frac{5.60}{1.536}} = \sqrt{3.656} = 1.91$$

$$t = \frac{7.69}{1.91} = 4.026$$

وحيث أن t المحسوبة أكبر من t.01 (8)= 3.355

هناك انحدار معنوى من استهلاك الغذاء على الوزن.

اختبار F :

جدول يوضح المعادلات و تحليل التباين لإنحدار

Source	df	**Symbolic SS**
Due to Reg. Residual	1 n-2	$(SP)^2/SS_x=b^2.SS_x$ by subtraction
Total	n-1	$\mathbf{SS_y}$

جدول (2) يوضح المعادلات و تحليل التباين لبيانات جدول (1) السابق

Source	df	Symbolic SS	df	SS	MS	F
Due to Reg.	1	$(SP)^2/SSx=b^2.SS_x$	1	90.836	90.836	6.22**
Residual	n-2	by subtraction	8	44.768	5.60	
Total	**n-1**	SS_y	**9**	**135.604**		

اختبار F هنا يختبر النظرية الفرضية " أن الاختلافات فى X لا تسهم فى الاختلافات فى Y ".

ورفض النظرية الفرضية(معنوية F) يعنى أن الاختلافات فى المتغير Y تعتمد على الاختلافات فى المتغير X .

. $F\alpha(1,n-2)$) حرية لدرجات الجدولية Fبقيمة المحسوبة F تقارن

معامل التقدير Coefficient of Determination

- عبارة عن النسبة بين مجموع المربعات الراجعة للانحدار إلى مجموع مربعات الانحرافات الكلية ويرمز له r^2 ويمكن تمثيله رياضيا كالآتى

$$r^2 = \frac{SS_{\text{Re } g\cdot}}{SS_{Total}}$$

- وتتراوح قيم r^2 بين صفر ، 1 وكلما زادت كلما كانت Y أكثر اعتمادا على X

من المثال $r^2 = \frac{90.836}{135.604} = .67$

. X قيم فى الاختلافات مصدرها Y قيم فى الاختلافات من 67% أن أى

العلاقة بين معامل الإنحدار و معامل الإرتباط:

- إذا كان X,Y متغيرين فإن ; $b_{x.y}=\frac{SP}{SS_y}$ $b_{y.x}=\frac{SP}{SS_x}$

حيث r^2 هى مربع معامل الإرتباط $r^2=\frac{(SP)^2}{SS_x.SS_y}$

لذلك فإن: $(b_{y.x})(b_{x.y})=\frac{SP}{SS_x}.\frac{SP}{SS_y}=\frac{(SP)^2}{SS_x.SS_y}$

أى أن $r^2=(b_{y.x})(b_{x.y})$

أى أن معامل الإرتباط لمتغيرين هو المتوسط الهندسي لمعاملي الإنحدار الذين يمكن حسابهما من هذين المتغيرين

$$r=\sqrt{(b_{y.x})(b_{x.y})}$$



مثال :- في والزيادة معين عمر في (x) للدجاجة الأساسي الوزن بين العلاقة دراسة عينة استخدمت محدودة لفترة البروتين من عالية عليقة استخدام نتيجة (y) الوزن في عشوائية n = 18 وتم الحصول علي التقديرات التاليــة :

$\bar{x} = 0.5$ $\bar{y} = 0.8\,kg$ $\sqrt{SS_x \cdot SS_y} = 2.0$ $r = 0.8$ $b_{y.x} = 1.6$

الانحدار خط معادلة أوجدأ- !!!!

$$r = \frac{SP}{\sqrt{SS_x \; SS_y}} \qquad 0.8 = \frac{SP}{20} \qquad SP = (0.8)(20) = 16.0$$

$$a = \bar{y} - b\,\bar{x}$$

$$= 0.8 - (1.6)(0.5) = 0.8 - 0.8 = 0$$

$$\hat{y} = a + b\,x$$

$$\hat{y} = 0 + 1.6\,x$$

$$\hat{y} = 1.6\,x$$

ب- أن علمت إذا by.x الانحدار ومعامل r التلازم معامل من كلا معنوية أختبر

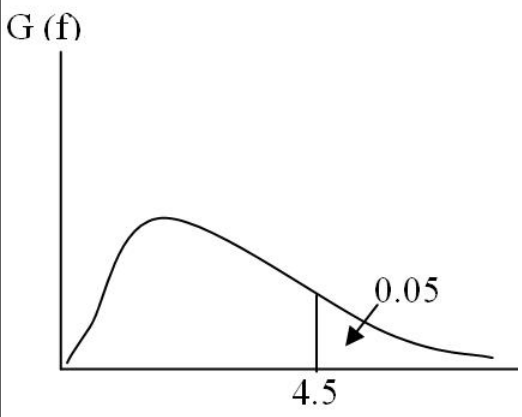
$$b_{y.x} = \frac{SP}{SS_x} \qquad 1.6 = \frac{16}{SS_x}$$

$$SS_x = \frac{16}{1.6} = 10$$

$$\sqrt{SS_y \cdot SS_x} = \sqrt{10\ SS_y} = 20$$

بتربيع الطرفين

$$10\ SS_y = 400$$
$$SS_y = 40$$



S. V	d f	SS	S^2	F	$F_{1.16}$
Reg	1	$b^2 SS_x = (1.6)^2\ 10 = 25.6$	25.6	$\frac{25.6}{0.9} = 28.44^*$	4.5
Error	n-2 = 16	$SS_{Total} - SS_{Reg} = 14.4$	0.9		
Total	n-1 = 17	$SS_y = 40$			

F المحسوبة > F الجدولية

∴ نرفض النظرية الفرضية $H_0 : B = 0$ وتقبل الفرض البديل $H_1 : B \neq 0$

أي لا توجد علاقة اعتماد لـ y عن x . أي أن الأختلافات في x لا تهم في الأختلافات في y