إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

تحليل الارتباط والانحدار الخطي البسيط

1. مقـــدمة

تم عرض أهم مقاييس الإحصاء الوصفي وتشمل مقاييس النزعة المركزية، والتشتت، ومقاييس الالتواء والتفرطح، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات الخاصة بمتغير أوصفة من الصفات المدروسة. ولكن يلاحظ أن دراسة وتحليل بيانات كل متغير من متغيرات الدراسة من خلال هذه المقاييس لا تكفي لمعرفة سلوكيات هذه المتغيرات ومسبباتها. ففي كثير من النواحي التطبيقية يهتم الباحثين بدراسة وتحليل العلاقة بين متغيرات الدراسة، ومن ثم يجب الأخذ في الاعتبار أساليب وطرق التحليل الإحصائي التي تهتم بدراسة وتحليل العلاقة بين متغيرين أو أكثر من ناحية، ودراسة تحليل أثر متغير أو مجموعة من متغيرات الدراسة تسمى بالمتغيرات المفسرة أو المستقلة على سلوكيات متغير يسمى بالمتغير التابع.

يتناول هذا الفصل عرض بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، لتحليل العلاقة بين المتغيرات، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار الخطي البسيط ، ومن الأمثلة على ذلك:

1- الإنفاق، والدخل العائلي.
2- سعر السلعة، والكمية المطلوبة منها.
3- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
4- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
5- كمية الغذاء اليومي من البروتين الذي يتناوله الطفل والزيادة في الوزن.
6- الوزن وضغط الدم

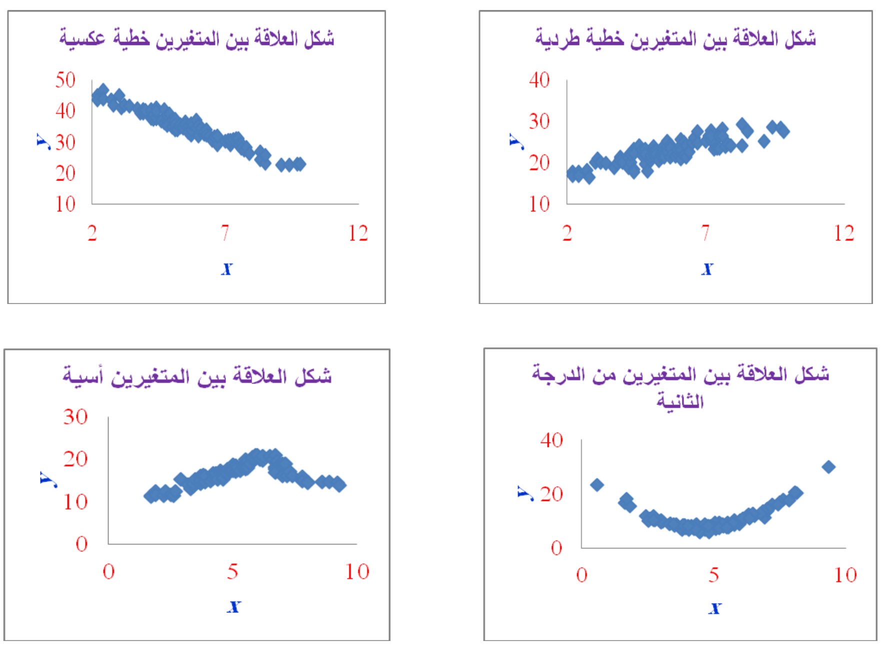
والأمثلة على ذلك في النواحي التطبيقية كثيرة لا يتسع المجال لحصرها.

2. نقط الانتشار واستكشاف شكل العلاقة

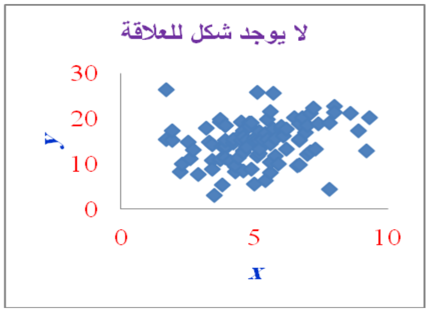
بفرض أن الباحث يهتم بدراسة شكل العلاقة بين المتغيرين $(x \; , \; y)$، وتوافرت لديه مشاهدات عن أزواج القيم $(x \; , \; y)$ ، يمكن تمثيلها بيانيا فيما يسمى بنقط انتشار، ومن خلالها يمكن استكشاف شكل العلاقة المتوقعة بين المتغيرين $(x \; , \; y)$، ويبين الشكل (1) الصور المختلفة للعلاقة بين المتغيرين $(x \; , \; y)$.

شكل (1)

بعض الصور المختلفة للعلاقة بين المتغيرين $(x \; , \; y)$



إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa



3. الارتباط الخطى البسيط Simple Correlation

يهتم هذا الفصل بدراسة العلاقة بين متغيرين من خلال أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية Quantitative، أما الحالة التي تكون فيها البيانات وصفية مقاسة بمعيار رتبي Ordinal سوف يتم تناولها في فصول قادمة.

3.1- الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط تحديد نوع وقوة العلاقة بين متغيرين من ناحية، ومعنوية هذه العلاقة من ناحية أخرى. وبفرض أن معامل الارتباط الخطي البسيط في المجتمع ويرمز بالرمز ρ (رو) غير معلوم ، فإنه يمكن استخدام معامل الارتباط الخطي البسيط المحسوب من بيانات العينة ويرمز له بالرمز r كتقدير لمعامل الارتباط في المجتمع ρ ، ومن التحديد السابق للغرض من تحليل الارتباط سوف يتم دراسة اتجاه وقوة العلاقة بين متغيرين، كما يتم عرض طريقة التقدير الإحصائي المستخدمة في تقدير معامل الإرتباط في المجتمع وكذلك تقدير فترة ثقة واختبار معنوية هذا المعامل.

- تحديد اتجاه العلاقة بين المتغيرين، وتأخذ ثلاث اتجاهات هي:

1- إذا كانت إشارة معامل الارتباط سالبة $(r < 0)$ توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة قيم أحد المتغيرين يصاحبه انخفاض في القيم التي يأخذها المتغير الثاني، والعكس صحيح.

2- إذا كانت إشارة معامل الارتباط موجبة $(r > 0)$ توجد علاقة طردية بين المتغيرين،

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

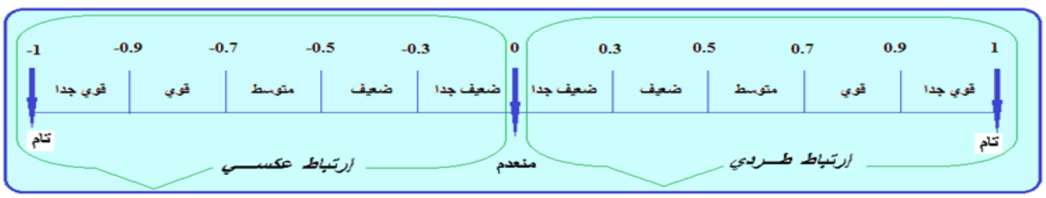
بمعنى أن زيادة قيم أحد المتغيرين يصاحبه زيادة في قيم المتغير الثاني، والعكس صحيح .

3- إذا كان معامل الارتباط قيمته صفرا ***(r = 0)*** دل ذلك على انعدام العلاقة بين المتغيرين.

- تحديد قوة العلاقة، إذ يمكن الحكم علي قوة العلاقة من خلال مدى قرب قيمة معامل الإرتباط من (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى ***(-1 < r < 1)***، وقد اجتهد بعض الإحصائيين في تحديد درجات قوة العلاقة بين متغيرين، وهي على النحو المبين بالشكل التالي:

شكل (2)

درجات الارتباط



3.2 التقدير الإحصائي لمعامل الارتباط الخطى البسيط

بفرض أن المتغير X له توزيع طبيعي متوسطه μ_x، وتباينه σ_x^2، وأن المتغير Y له أيضا توزيع طبيعي متوسطه μ_Y، وتباينه σ_Y^2، فإن التوزيع الطبيعي الثنائي للمتغيرين (X, Y) يعبر عنه بالصورة التالية:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xu}^2}} EXP\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} \quad (1)$$

حيث أن ρ_{XY} يعبر عن معامل الارتباط في المجتمع بين المتغيرين (x, y) ويعبر عنه بالصورة التالية:

malderiny@ksu.edu.sa إعداد د: محمود محمد الدريني

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad (2)$$

كما يعبر σ_{xy} عن التغاير بين المتغيرين (x, y)، وحيث أنه في كثيرمن الحالات يكون هذا المعامل غير معلوم، فإنه يمكن استخدام طريقة العزوم لبيرسون وتقديره من بيانات عينة. ومن ثم إذا كانت $\{x, y : (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ مشاهدات عينة عشوائية من أزواج القيم حجمها n فإن معامل الارتباط في العينة بين المتغيرين (x, y) ويرمز له بالرمز r_{xy} يمكن حسابة باستخدام طريقة بيرسون من خلال تطبيق المعاددلة التالية:

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\dfrac{\sum (x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum (x-\bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum (y-\bar{y})^2}{(n-1)}}} \qquad (3)$$

حيث أن :

$S_{xy} = \sum (x-\bar{x})(y-\bar{y}) \big/ (n-1)$: هو تقدير التغاير σ_{xy} بين $(x \;,\; y)$،

$S_x = \sqrt{\sum (x-\bar{x})^2 \big/ (n-1)}$: هو تقدير الانحراف المعياري لقيم (x)،

$S_y = \sqrt{\sum (y-\bar{y})^2 \big/ (n-1)}$: هو تقدير الانحراف المعياري لقيم (y).

ويمكن تبسيط الصيغة التعريفية السابقة على النحو التالي:

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} \qquad (4)$$

تطبيق (1)

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

أراد باحث أن يدرس العلاقة بين المساحة المنزرعة بالأعلاف الخضراء، وإنتاج اللحوم خلال الفترة من 1995 حتى عام 2002م، قام بجمع بيانات سلسلة زمنية عن المساحة المنزرعة بالألف هكتار، وكمية اللحوم المنتجة بالألف طن خلال تلك الفترة.

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	297	289	233	214	240	217
الكمية	592	603	662	607	635	699	719	747

والمطلوب:

1- احسب معامل الارتباط بين المساحة المنزرعة وكمية الإنتاج من اللحوم.

2- ما الذي يمكنك أن تبينه للباحث تجاه العلاقة بين المساحة وكمية اللحوم المنتجة؟

حل التطبيق

1- حساب معامل الارتباط بين المساحة المنزرعة وكمية الإنتاج من اللحوم.

بفرض أن المتغير x يعبر هي المساحة المنزرعة، المتغير y يعبر عن الكمية المنتجة.

ولحساب معامل الارتباط بين $(x \ , \ y)$ يتم تطبيق المعادلة (4) وذلك على النحو التالي:

- حساب الوسط الحسابي لكل من المساحة، والكمية $(\bar{x} \ , \ \bar{y})$.

$$\bar{x} = \frac{\sum x}{n} = \frac{2108}{8} = 263.5 \quad , \quad \bar{y} = \frac{\sum y}{n} = \frac{5264}{8} = 658$$

- حساب المجاميع

$$\sum (x - \bar{x})^2 = 12040 \quad , \sum (y - \bar{y})^2 = 23850 \quad , \sum (x - \bar{x})(y - \bar{y}) = -13528$$

x	y	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	**5264**	**0**	**12040**	**0**	**23850**	**-13528**

إذا معامل الارتباط قيمته هي:

$$r_{xy} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} = \frac{-13528}{\sqrt{12040}\sqrt{23850}}$$

$$= \frac{-13528}{(109.727)(154.434)} = \frac{-13528}{16945.619} = -0.798$$

2- يدل معامل الارتباط على وجود علاقة عكسية قوية بين المساحة المنزرعة، وكمية إنتاج اللحوم، فمع مرور الزمن تقل المساحة المنزرعة بالأعلاف الخضراء ويصاحبها تزايد في كمية اللحوم المنتجة، ويستدل من ذلك على أن المملكة اتجهت إلى سياسة تقليل مساحة الجزء المخصص لزراعة الأعلاف والاعتماد على سياسة الاستيراد.

تبسيط العمليات الحسابية:

في كثير من الحالات التطبيقية يصعب على الباحث تطبيق الصيغة التعريفية (4) لحساب معامل الإرتباط، وقد يترتب حدوث بعض الأخطاء خاصة إذا لازم العمليات الحسابية قيما كسرية. وباستخدام بعض العمليات الحسابية البسيطة يمكن تبسيط الصيغة(4) إلى صيغة أسهل تعتمد على مجموع القيم ومربعاتها ولا تعتمد على مجموع انحرافات القيم عن وسطها

الحسابي، وهذه الصيغة هي:

$$r_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)}} \qquad (5)$$

وباستخدام بيانات التطبيق السابق يمكن حساب معامل الارتباط بين المساحة المنزرعة وكمية الإنتاج من اللحوم باستخدام الصيغة الحسابية في (5) كما يلي:

- حساب المجاميع

x	y	xy	x^2	y^2
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244
289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009
2108	5264	1373536	567498	3487562

المجاميع المطلوبة

$\sum x = 2108 \quad , \quad \sum y = 5264$

$\sum xy = 1373536$

$\sum x^2 = 567498$

$\sum y^2 = 3487562$

- حساب معامل الارتباط:

بالتطبيق على المعادلة (16.5) أعلاه، نجد أن معامل الارتباط قيمته هي:

$$r_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left(567498 - \frac{(2108)^2}{8}\right)\left(3487562 - \frac{(5264)^2}{8}\right)}}$$

$$= \frac{-13528}{\sqrt{(12040)(23850)}} = \frac{-13528}{16945.619} = -0.798$$

وهي نفس النتيجة السابقة:

3.3 اختبار معنويته معامل الارتباط في المجتمع ρ_{xy}

في كثير من النواحي التطبيقية يكون هدف الباحث الوصول إلى قرار بخصوص الدلالة الإحصائية للعلاقة بين المتغيرين (x, y)، ويمكن إجراء ذلك من خلال اختبار معنوية معامل الإرتباط فيى المجتمع ρ_{xy} ، ويقصد به اختبار ما إذا كان المعامل ρ_{xy} يساوي صفرا أو يختلف عن الصفر، ولإجراء هذا الإختبار تتبع الخطوات التالية:

- صياغة الفرض العدم والفرض البديل.

الفرض العدم: لا توجد علاقة معنوية بين المتغيرين $H_0 : \rho_{xy} = 0$

الفرض البديل: توجد علاقة معنوية بين المتغيرين $H_1 : \rho_{xy} \neq 0$

- حساب معامل الإرتباط في العينة r_{xy} واستخدمها في حساب إحصائية الاختبار t،حيث أن:

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}} \qquad (6)$$

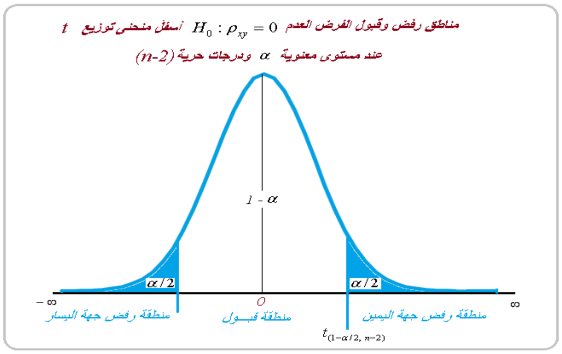
وتحت صحة الفرض العدم $H_0 : \rho_{xy} = 0$ ، فإن الإحصاء t أعلاه يتبع توزيع t بدرجات

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

حرية $(n-2)$.

- تحديد مناطق الرفض والقبول وذلك عند مستوى المعنوية المحدد α ودرجات حرية $(n-2)$. حيث تستخرج القيمة الجدولية $t_{(1-\alpha/2,\, n-2)}$ من جدول مئويات توزيع t، والرسم التالي يبين مناطق الرفض والقبول للفرض العدم $H_0 : \rho_{xy} = 0$.

شكل (3)



- اتخاذ القرار: إذا كانت القيمة المطلقة لإحصائية الاختبار $|t|$ تزيد عن القيمة الجدولية $t_{(1-\alpha/2,\, n-2)}$، أي تقع في منطقة الرفض جهة اليمين، يرفض فرض العدم $H_0 : \rho_{xy} = 0$ ويقبل الفرض البديل $H_1 : \rho_{xy} \neq 0$ ويستدل من ذلك على معنوية العلاقة بين المتغيرين (x, y).

3.4 تقدير فترة ثقة $100(1-\alpha)\%$ لمعامل الارتباط في المجتمع.

لتقدير فترة ثقة لمعامل الارتباط في المجتمع يمكن استخدام تحويلة "فيشر" للمتغير الطبيعي القياسي Z حيث يتم إيجاد فترة ثقة للمعلمة $\{(1/2)\, Log_e[(1+\rho_{xy})/(1-\rho_{xy})]\}$، ومنها يتم حساب الحدين الأدنى والأعلى للثقة للمعامل ρ_{xy} بطريق غير مباشر وذلك باتباع الخطوات التالية:

- حساب حدي الثقة للمعلمة $\{(1/2)\, Log_e[(1+\rho_{xy})/(1-\rho_{xy})]\}$ وهما:

$$\frac{1}{2} Log_e\left(\frac{1+r_{xy}}{1-r_{xy}}\right) - \frac{Z_{(1-\alpha/2)}}{\sqrt{n-3}} < \frac{1}{2} Log_e\left(\frac{1+\rho_{xy}}{1-\rho_{xy}}\right) < \frac{1}{2} Log_e\left(\frac{1+r_{xy}}{1-r_{xy}}\right) + \frac{Z_{(1-\alpha/2)}}{\sqrt{n-3}} \qquad (7)$$

حيث أن Log_e هو اللوغريتم للأساس e ويعرف باللوغاريتم الطبيعي، $Z_{(1-\alpha/2)}$ هي قيمة المتغير الطبيعي القياسي التي يقل عنها مساحة قدرها $(1-\alpha/2)$، وتستخرج من جدول التوزيع الطبيعي القياسي.

ومن ثم يعبر عن الحدين الأدنى والأعلى L_z، U_z على التوالي كما يلي:

$$L_z = \frac{1}{2} Log_e\left(\frac{1+r_{xy}}{1-r_{xy}}\right) - \frac{Z_{(1-\alpha/2)}}{\sqrt{n-3}}, \quad U_z = \frac{1}{2} Log_e\left(\frac{1+r_{xy}}{1-r_{xy}}\right) + \frac{Z_{(1-\alpha/2)}}{\sqrt{n-3}} \qquad (8)$$

- ويحسب الحدين الأدنى والأعلى للثقة لمعامل الارتباط ρ_{xy} كما يلي:

$$\frac{e^{2L_z}-1)}{e^{2L_z}+1} < \rho_{xy} < \frac{e^{2L_U}-1)}{e^{2L_U}+1}$$
$$L_\rho = \frac{e^{2L_z}-1)}{e^{2L_z}+1}, \quad U_\rho = \frac{e^{2L_U}-1)}{e^{2L_U}+1} \qquad (9)$$

تطبيق (2)

لدراسة العلاقة بين مستوى الطالب عند إلحاقه بإحدى الكليات و المعدل التراكمي في نهاية السنة الأولى، تم اختيار عينة عشوائية من الطلاب الذين إلتحقوا بها حجمها 20 طالب، وسجل لكل طالب درجة إختبار دخوله والمعدل التراكمي في نهاية السنة الأولى ولخصت البيانات في الجدول التالي:

الطالب	1	2	3	4	5	6	7	8	9	10
درجة الاختبار	5.5	4.8	4.7	3.9	4.5	6.2	6.0	5.2	4.7	4.3
المعدل التراكمي	3.1	2.3	3.0	1.9	2.5	3.7	3.4	2.6	2.8	1.6

الطالب	11	12	13	14	15	16	17	18	19	20
درجة الاختبار	4.9	5.4	5.0	6.3	4.6	4.3	5.0	5.9	4.1	4.7
المعدل التراكمي	2.0	2.9	2.3	3.2	1.8	1.4	2.0	3.8	2.2	1.5

بفرض أن x تعبر عن درجة اختبار الطالب عند الدخول، y تعبر عن المعدل التراكمي،

وإذا إستخدمت البيانات أعلاه لحساب نتائج العمليات التالية:

$$\sum x_i y_i = 257.66, \quad \sum y_i^2 = 134.84, \quad \sum x_i^2 = 509.12 \ , \ \sum y_i = 50, \quad \sum x_i = 100$$

أ- احسب معامل الارتباط بين درجة اختبار الدخول والمعدل التراكمي، وما هو مدلوله؟

ب- اختبر معنوية العلاقة بين درجة اختبار الدخول ومعدله التراكمي، حيث أن $\alpha = 0.05$.

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

جـ - تقدر فترة ثقة 95% لمعامل الارتباط بين درجة اختبار الدخول والمعدل التراكمي ثم فسره؟

حل التطبيق:

أ- حساب معامل الارتباط بين درجة اختبار الدخول x والمعدل التراكمي y: من النتائج أعلاه يمكن تطبيق المعادلة (16.5) كما يلي:

$$r_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)}}$$

$$r_{xy} = \frac{257.66 - \frac{(100)(50)}{20}}{\sqrt{\left(509.12 - \frac{(100)^2}{20}\right)\left(134.84 - \frac{(50)^2}{20}\right)}}$$

$$= \frac{7.66}{(3.019934)\quad(3.136877)} = 0.809$$

يوجد ارتباط طردي قوي بين درجة اختبار الدخول والمعدل التراكمي في نهاية السنة الأولى.

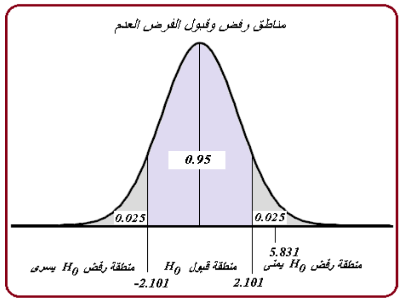
ب- اختبار معنوية العلاقة بين درجة اختبار الدخول والمعدل التراكمي عند مستوى معنوية $\alpha = 0.05$، وذلك باتباع الخطوات التالية:

- $H_0 : \rho_{xy} = 0$ (لا توجد علاقة معنوية بين درجة اختبار الدخول والمعدل التراكمي)
- $H_1 : \rho_{xy} \neq 0$ (توجد علاقة معنوية بين درجة اختبار الدخول والمعدل التراكمي)
- معامل الإرتباط في العينة قيمته هي: $r_{xy} = 0.809$، إذا إحصائية الاختبار t هي:

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}} = \frac{0.808\sqrt{20-2}}{\sqrt{1 - 0.808^2}} = \frac{3.4306}{0.588} = 5.831$$

- القيمة الجدولية (الحرجة) وتستخرج من جدول توزيع t عند مستوى معنوية

$\alpha = 0.05$ ودرجات حرية $(n - 2 = 18)$ وهي: $\{t_{(1-\alpha/2,\, n-2)} = t_{(0.975\,,\,18)} = 2.101\}$، ومن ثم تحدد مناطق الرفض والقبول كما في الشكل التالي:



- بما أن قيمة إحصائية الاختبار الموجبة $|t| = 5.831$ تزيد عن القيمة الجدولية $\{t_{(0.975\,,\,18)} = 2.101\}$ أي تقع في منطقة الرفض، إذا يرفض فرض العدم ويقبل الفرض البديل، ويستدل من ذلك على وجود علاقة ذات دلالة بين درجة اختبار دخول الطالب ومعدله التراكمي في نهاية السنة الأولى.

الاحتمال المشاهد .Sig

المعنوية المحسوبة .Sig أو ما يسمى بالاحتمال المشاهد $(p - Value)$ هي ضعف المساحة أسفل منحنى توزيع t على يمين القيمة المطلقة لإحصائية الاختبار $|t|$، أي أن:

$$p - Value = 2[\,prob\,(t_{(df)} > |t|)] \qquad (10)$$

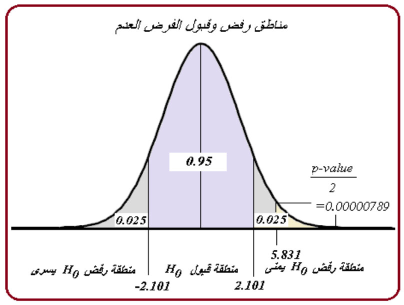
ويمكن اتخاذ قرار بخصوص الفرض العدم من خلال مقارنة الاحتمال المشاهد $(p - Value)$ بمستوى المعنوية المحدد α، فإذا كان الاحتمال المشاهد $(p - Value)$ أقل من مستوى المعنوية يمكن رفض الفرض العدم وقبول الفرض البديل، وفي التطبيق السابق نجد أن هذه القيمة هي:

$$p - Value = 2[\,prob\,(t_{(18)} > 5.831)] = 2(0.00000798) = 0.000016$$

وبمقارنة قيمة الاحتمال المشاهد أعلاه $p - Value = [0.000016]$ بمستوى المعنوية المحدد $\alpha = 0.05$، نجد أن الاحتمال المشاهد أقل من مستوى المعنوية أي أن $(p - Value < \alpha)$ ومن ثم نرفض الفرض العدم $H_0 : \rho_{xy} = 0$ ونقبل الفرض البديل $H_1 : \rho_{xy} \neq 0$.

ويمكن بيان ذلك على الشكل التالي:

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa



جـ - تقدير فترة ثقة 95% لمعامل الارتباط بين درجة اختبار الدخول والمعدل التراكمي:

لتقدير فترة ثقة لمعامل الارتباط في المجتمع يمكن استخدام تحويلة "فيشر" للمتغير الطبيعي القياسي z، وذلك بتطبيق المعادلات (8)، (9) كما يلي:

- حساب حدي الثقة للمعلمة $\{(1/2)\,Log_e[(1+\rho_{xy})/(1-\rho_{xy})]\}$

- $\alpha = 0.05,\ (1-\alpha) = 0.95$، القيمة الجدولية: $Z_{(1-\alpha/2)} = Z_{0.975} = 1.96$، $r_{xy} = 0.809$.

- $$L_Z = \frac{1}{2} Log_e\left(\frac{1+r_{xy}}{1-r_{xy}}\right) - \frac{Z_{(1-\alpha/2)}}{\sqrt{n-3}} = \frac{1}{2} Log_e\left(\frac{1+0.809}{1-0.809}\right) - \frac{1.96}{\sqrt{17}}$$

$$= \frac{1}{2} Log_e(9.471204) - 0.47537 = 1.124128 - 0.47537 = 0.648758$$

- $$U_Z = \frac{1}{2} Log_e\left(\frac{1+r_{xy}}{1-r_{xy}}\right) + \frac{Z_{(1-\alpha/2)}}{\sqrt{n-3}} = 1.124128 + 0.47537 = 1.599498$$

- ويحسب الحدين الأدنى والأعلى للثقة لمعامل الارتباط ρ_{xy} بتطبيق المتباينة كما يلي:

$$L_\rho = \frac{e^{2L_z} - 1)}{e^{2L_z} + 1} = \frac{e^{2(0.648758)} - 1)}{e^{2(0.648758)} + 1} = \frac{2.660195}{4.660195} = 0.571$$
$$U_\rho = \frac{e^{2L_U} - 1)}{e^{2L_U} + 1} = \frac{e^{2(1.599498)} - 1)}{e^{2(1.599498)} + 1} = \frac{23.5079}{25.5079} = 0.922 \qquad (11)$$

- أي أن معامل الارتباط بين درجة الطالب عند الدخول والمعدل التراكمي، سوف يتراوح بين حد أدنى 0.571 ، وحد أعلى 0.922 باحتمال 95%.

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

4. تحليل الانحدار الخطي البسيط Simple Regression

يستخدم أسلوب تحليل الانحدار الخطي عند دراسة وتحليل أثر متغيرات كمية تسمى بالمتغيرات المستقلة على متغير كمي آخر يسمى بالمتغير التابع، فإذا شمل التحليل متغير واحد مستقل أطلق عليه بتحليل الانحدار الخطي البسيط، وأما الحالة التي يشمل التحليل فيها عدد من المتغيرات المستقلة سمي بتحليل الانحدار الخطي المتعدد. وسف يتناول هذا الفصل عرض أسلوب تحليل الانحدالر الخطي البسيط، بينما يتناول الفصل القادم شرح أسلوب تحليل الانحدار الخطي المتعدد.

4.1 نموذج الانحدار الخطي

يهدف تحليل الانحدار الخطي البسيط إلى دراسة أثر أحد متغيرين كميين على المتغير الآخر، والمتغير الذي يهتم الباحث بدراسة أثره يسمى بالمتغير المستقل independent أو المفسر explanatory وفي بعض الحالات يطلق عليه بالمتغير المتنبأ منه، بينما يطلق على المتغير الآخر الذي يتأثر بالمتغير التابع أو المتنبأ به. والأمثله على ذلك كثيره منها:

- دراسة أثر كمية السماد على إنتاجية الدونم من محصول زراعي معين .
- دراسة أثر الكمية المستخدمة من عنصر إنتاجي معين على كمية الإنتاج.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- دراسة أثر الدخل العائلي على الإنفاق الاستهلاكي العائلي.
- دراسة أثر السعر على الكمية المطلوبة.
- دراسة أثر الوزن على ضغط الدم.

فإذا فرض أن المشاهدة رقم i على المتغير التابع هي y_i ، وعلى المتغير المستقل هي x_i ، فإن المتوسط الشرطي للمتغير التابع بمعلومية المشاهدة على المتغير المستقل $\mu_{Y|x_i}$ يمكن التعبير عنه بمعادلة خط مستقيم تأخذ الصورة التالية:

$$\mu_{Y|x_i} = (\beta_0 + \beta_1 x_i) \qquad (12)$$

كما يمكن تمثيل المشاهدة y_i بمعلومية المشاهدة x_i بمعادلة خطية تأخذ الصورة التالية:

إعداد د: محمود محمد الدريني
malderiny@ksu.edu.sa

$$\begin{aligned} y_i &= \mu_{Y|x_i} + \varepsilon_i \\ &= \beta_0 + \beta_1 x_i + \varepsilon_i \end{aligned} \tag{13}$$

وتسمى بمعادلة الانحدار الخطي البسيط، حيث أن:

y_i : هي المشاهدة رقم i على المتغير التابع (المتنبأ به).

x_i :تعبر عن المشاهدة رقم i على المتغير المستقل (المتنبأ منه).

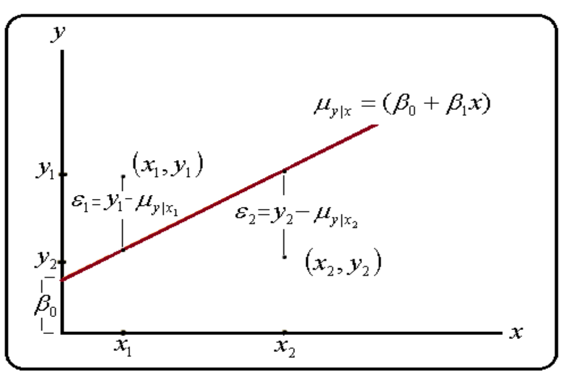
β_0 :ويعبر عن ثابت الانحدار أو القاطع intercept، وهو الجزء المقطوع من المحور الرأسي y ، كما يعكس قيمة المتغير التابع y في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $x = 0$.

β_1 : ويعبر عن ميل الخط المستقيم الموضح بالمعادلة (12)، ويعكس هذا المعامل مقدار التغير في المتغير التابع y إذا حدث تغير في المتغير المستقل x بوحدة واحدة، ويسمى بمعامل الانحدار.

ε_i : يمثل الخطأ العشوائي للمشاهدة رقم i، وهو مقدار انحراف القيمة الفعلية y_i عن متوسطها $[\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i)]$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.

شكل (4-16)

الخطأ العشوائي



ويحتوى النموذج (13) على معاملين هما (β_0, β_1) يمكن تقديرها باستخدام بعض طرق التقدير الإحصائي كطريقة المربعات الصغرى العادي Ordinary Least Squares (OLS) ، وطريقة الإمكانية العظمى Maximum Likelihood (ML).

4.2 افترضات نموذج الانحدار الخطي

لتقدير معاملي الانحدار (β_0, β_1) هناك بعض الشروط التي يجب تحققها في نموذج الانحدار، وهي كالتالي:

- أن المتغير التابع Y متغير كمي مستمر، له توزيع طبيعي متوسطه الشرطي عند القيمة المحددة x_i هو: $\mu_{Y|x_i}$ ويعبر عنه بمعادلة خط مستقيم تأخذ الشكل (12) وتباينه الشرطي $\sigma^2_{Y|x_i}$ ثابت من مشاهدة لأخرى. كما أن المشاهدات على المتغير التابع $y_i^{,s}$ مستقلة.
- أن المشاهدة على المتغير المستقل x_i محددة fixed ومعطاه.
- عند كل قيمة معطاة للمتغير المستقل X توجد قيمة واحدة مناظرة للمتغير التابع Y.
- أن الخطأ العشوائي ε_i يتبع توزيع طبيعي متوسطه صفر وتباينه الشرطي عند القيمة المحددة x_i هو $\sigma^2_{Y|x_i}$ وهو ثابت من مشاهدة لأخرى، أي أن:

$$\sigma^2_{Y|x_1} = \sigma^2_{Y|x_2} = \dots = \sigma^2$$

وهو شرط تجانس التباينات.

كما أن الأخطاء العشوائية مستقلة، ويعبر عن ذلك رياضيا بالصورة التالية.

$$E(\varepsilon_i, \varepsilon_{i`}) = \begin{cases} \sigma^2 & if \quad i = i` \\ 0 & if \quad i \neq i` \end{cases} \qquad (14)$$

وهو شرط استقلال الأخطاء العشوائية.

- يوجد استقلال خطي بين الأخطاء العشوائية، والمشاهدات المستقلة، أي أن:

$$E(\varepsilon_i, x_i) = 0 \qquad (15)$$

4.3 تقدير المربعات الصغرى العادي (OLS) لمعاملي الانحدار (β_0, β_1)

تعبر المعادلة (13) عن نموذج الانحدار في المجتمع، ومن ثم يمكن تقدير معاملي الانحدار $(\beta_1 \ , \ \beta_0)$ في النموذج باستخدام طريقة المربعات الصغرى العادية (OLS)، والغرض من تطبيق هذه الطريقة هو الحصول على قيمة للمعاملين $(\beta_1 \ , \ \beta_0)$ التي تجعل مجموع مربعات الأخطاء العشوائية

$$\left[SSE = \sum_i \varepsilon_i^2 = \sum_i \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2 \right] \tag{16}$$

أقل ما يمكن، فإذا كانت $\{x, y : (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ مشاهدات عينة عشوائية من أزواج القيم (x, y) حجمها n فإن تقديرات (OLS) للمعاملين $(\beta_1 \ , \ \beta_0)$ ينتج من حل المعادلتين الناتجتين من إيجاد المشتقة الجزئية الأولى للمعاملين $(\beta_1 \ , \ \beta_0)$ ومساواتها بالصفر، وهاتين المعادلتين هما:

$$\begin{aligned} n\hat{\beta}_0 + \left(\sum_{i=1}^{n} x_i \right) \hat{\beta}_1 &= \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i \right) \hat{\beta}_0 + \left(\sum_{i=1}^{n} x_i^2 \right) \hat{\beta}_1 &= \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{17}$$

وبحل هاتين المعادلتين يمكن الحصول على تقدير المربعات الصغرى للمعاملين وهو:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i \right)\left(\sum_{i=1}^{n} y_i \right)}{n}}{\left(\sum_{i=1}^{n} x_i^2 \right) - \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n}} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \end{aligned} \tag{18}$$

حيث أن $\left(\bar{x} = \sum_{i=1}^{n} x_i / n \right)$ هو الوسط الحسابي للمشاهدات المستقلة، $\left(\bar{y} = \sum_{i=1}^{n} y_i / n \right)$ هو الوسط الحسابي للمشاهدات التابعة، كما أن $\hat{\beta}_1$ يعبر عن تقدير معامل الانحدار β_1، $\hat{\beta}_0$ يعبر

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

عن تقدير ثابت الانحدار β_0 .

تطبيق (3)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع أقل من شهرين، ومقدار وزن العجل بعد مرور شهر بالكجم، وذلك لعينة من العجول الرضيعة حجمها **10**.

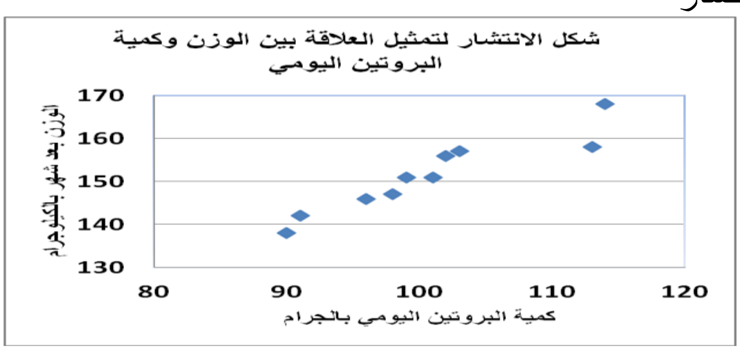
كمية البروتين	**90**	**91**	**96**	**98**	**99**	**101**	**102**	**103**	**113**	**114**
الوزن بعد شهر	**138**	**142**	**146**	**147**	**151**	**151**	**156**	**157**	**158**	**168**

والمطلوب :

1- ارسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة ؟
2- قدر معادلة انحدار الوزن على كمية البروتين.
3- فسر معادلة الانحدار.
4- ما هو مقدار الوزن عند إعطاء عجل رضيع أقل من شهرين **101** جرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟
5- ارسم معادلة الانحدار على نقط الانتشار في المطلوب (1) .

الحـل

1- رسم نقط الانتشار



من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على الوزن.

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

2- تقدير معادلة انحدار الوزن y على كمية البروتين x.

بفرض أن x هي كمية البروتين الذي يتناوله العجل الرضيع يوميا بالجرام، y هي الوزن بعد شهر بالكيلو جرام، يمكن تطبيق المعادلتين في (6)، ومن ثم يتم حساب المجاميع التالية:

كمية البروتين x	الزيادة في الوزن y	$x\ y$	x^2
90	**138**	**12420**	**8100**
91	**142**	**12922**	**8281**
96	**146**	**14016**	**9216**
98	**147**	**14406**	**9604**
99	**151**	**14949**	**9801**
101	**151**	**15251**	**10201**
102	**156**	**15912**	**10404**
103	**157**	**16171**	**10609**
113	**158**	**17854**	**12769**
114	**168**	**19152**	**12996**
1007	**1514**	**153053**	**101981**

المجاميع المطلوبة

$\sum x = 1007$

$\sum y = 1514$

$\sum xy = 153053$

$\sum x^2 = 101981$

إذا الوسط الحسابي:

$$\bar{x} = \frac{\sum x}{n} = \frac{1007}{10} = 100.7$$

$$\bar{y} = \frac{\sum x}{n} = \frac{1514}{10} = 151.4$$

- بتطبيق المعادلة الأولى في (6) يمكن حساب $\hat{\beta}_1$ كما يلي:

$$\hat{\beta}_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{(153053) - \frac{(1007)(1514)}{10}}{(101981) - \frac{(1007)^2}{10}}$$

$$= \frac{593.2}{576.1} = 1.03$$

- بتطبيق المعادلة الثانية في (6) يمكن حساب $\hat{\beta}_0$ كما يلي:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 100.7 - (1.03)(151.4) = 47.71$$

- إذا تقدير معادلة انحدار الوزن على كمية البروتين هي:

$$\hat{y} = 47.71 + 1.03x$$

3- تفسير المعادلة:

- الثابت $\hat{\beta}_0 = 47.71$: يدل على أنه في حالة عدم استخدام كميات من البروتين في التغذية، فإن الوزن يثبت عند 47.71.
- معامل الانحدار $\hat{\beta}_1 = 1.03$: يدل على أنه كلما زادت كمية البروتين جرام واحد، يكون من المتوقع حدوث زيادة في وزن العجل الرضيع بمقدار 1.03 كجم بعد مرور شهر.

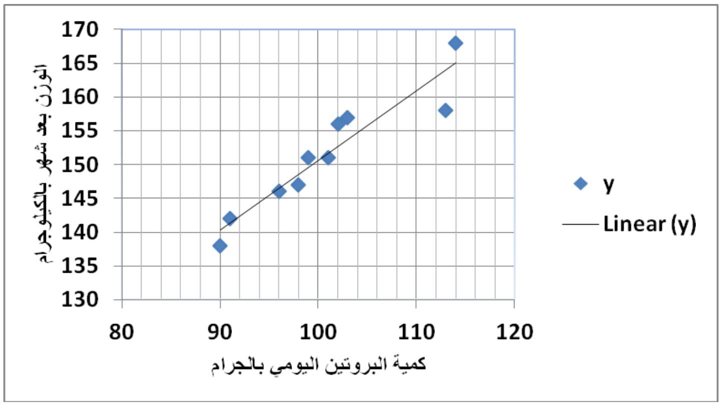
4- مقدار الزيادة في الوزن عند $x = 101$ هو:

$$\hat{y} = 47.71 + 1.03(101) = 151.7$$

وأما مقدار الخطأ العشوائي هو:

$$\hat{\varepsilon}_{x=101} = y_{x=101} - \hat{y}_{x=101} = 151 - 151.7 = -0.7$$

5- رسم معادلة الانحدار على نقط الانتشار.
يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم. إذا الشكل المطلوب هو:



4.4 مؤشرات جودة النموذج

- **معامل التحديد**

$$\sum \left(Y_i - \overline{Y}\right)^2 = \sum \left(\hat{Y} - \overline{Y}\right)^2 + \sum \varepsilon_i^2$$

$$SSY \quad = \quad SSR \quad + \quad SSE$$

$$R^2 = \frac{SSR}{SSY}$$

R^2 : معامل التحديد، يبين نسبة ما يشرحه أو يفسره المتغير المستقل من التغيرات الكلية في المتغير التابع.

- **الخطأ المعياري للتقدير** S

و هو الجذر التربيعي لتباين الخطأ العشوائي

$$S = \sqrt{S^2} = \sqrt{SSE \ /(n-2)}$$

تعبر 2 عن عدد الثوابت في المعادلة (β_0, β_1)

وسوف يتم التطبيق على برنامج **SPSS** للحصول على كافة النتائج الممكنة.

4.5 اختبار صلاحية النموذج:

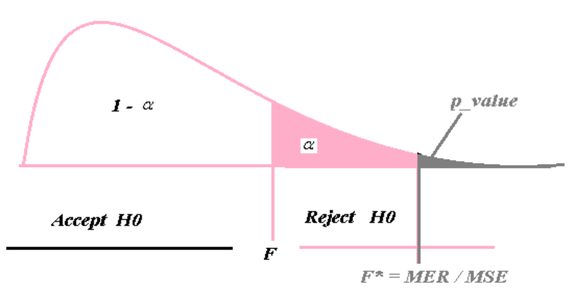
- الفرض العدم: النموذج الخطي غير مناسب $H_0 : \beta_1 = 0$
  الفرض البديل : النموذج الخطي مناسب $H_1 : \beta_1 \neq 0$
- مستوى المعنوية ، $\alpha = 0.05$
- إحصائية الاختبار (تكوين جدول تحليل التباين لحساب الإحصائي)

$$F^*_{(\alpha,(1,n-2))} = \frac{MSR}{MSE} = \frac{SSR\ /(1)}{SSE\ /(n-2}$$

- المعنوية المحسوبة **Sig.**

$$Sig. = \Pr\left(F_{(1,n-2)}\ Distribution > F^*\right)$$

- القرار: إذا كانت المعنوية المحسوبة أقل من مستوى المعنوية α نرفض الفرض العدم: النموذج الخطي غير مناسب $H_0 : \beta_1 = 0$، ونقبل الفرض البديل، ويستدل من ذلك على أن النموذج الخطي المفترض هو تمثيل جيد للعلاقة بين المتغيرين



4.6 اختبار معنوية معامل الانحدار:

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

ويقصد به اختبار ما إذا كان للمتغير المستقل x أثر معنوي وذو دلالة على المتغير التابع y أم لا.
وفيما يلي خطوات الاختبار بالنسبة للمعامل β_1.

- الفرض العدم:أثر المتغير المستقل x على المتغير التابع غير معنوي $H_0 : \beta_1 = 0$

  الفرض البديل : للمتغير المستقل x أثر معنوي على المتغير التابع $H_1 : \beta_1 \neq 0$

- مستوى المعنوية ، $\alpha = 0.05$.

- إحصائية الاختبار :

$$t^* = \frac{\hat{\beta}_1}{S.E_{\hat{\beta}_1}}$$

- المعنوية المحسوبة .Sig

$$Sig. = \Pr\left(t_{(n-2)}\ distribution > t^*\right)$$

- القرار: إذا كانت المعنوية المحسوبة أقل من مستوى المعنوية α نرفض الفرض العدم: $H_0 : \beta_1 = 0$ ، ونقبل الفرض البديل، ويستدل من ذلك على أن المتغير المستقل له أثر معنوي.

إعداد د: محمود محمد الدريني malderiny@ksu.edu.sa

4.7 التنبؤ Prediction

يمكن التنبؤ بقيمة ومتوسط المتغير التابع y عند قيمة محددة للمتغير المستقل x_0

	التنبؤ بالقيمة $Y \mid x_o$	التنبؤ بالمتوسط $\mu_{y\mid x_o}$
القيمة المتنبأ بها	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_o$	$\hat{\mu}_{y\mid x_o} = \hat{\beta}_0 + \hat{\beta}_1 x_o = \hat{Y}$
فترة تنبؤ $(1-\alpha)\%$	$\hat{Y} \pm \left(S.E_{\hat{y}}\right)\left(t_{(\alpha/2,\, df_{error})}\right)$ $S.E_{\hat{y}} \sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_o-\bar{x})^2}{(n-1)S_x^2}\right)}$	$\hat{Y} \pm \left(S.E_{\mu_{\hat{y}}}\right)\left(t_{(\alpha/2,\, df_{error})}\right)$ $S.E_{\mu_{\hat{y}}} = \sqrt{MSE\left(\frac{1}{n}+\frac{(x_o-\bar{x})^2}{(n-1)S_x^2}\right)}$