

مقدمة في التحليل الإحصائي



إعداد

د. انتصار الجباوي

رئيس قسم بحوث الشوندر السكري

الهيئة العامة للبحوث العلمية الزراعية

البريد الإلكتروني:

dr.entessara@gmail.com

المحتويات

١- مصطلحات إحصائية

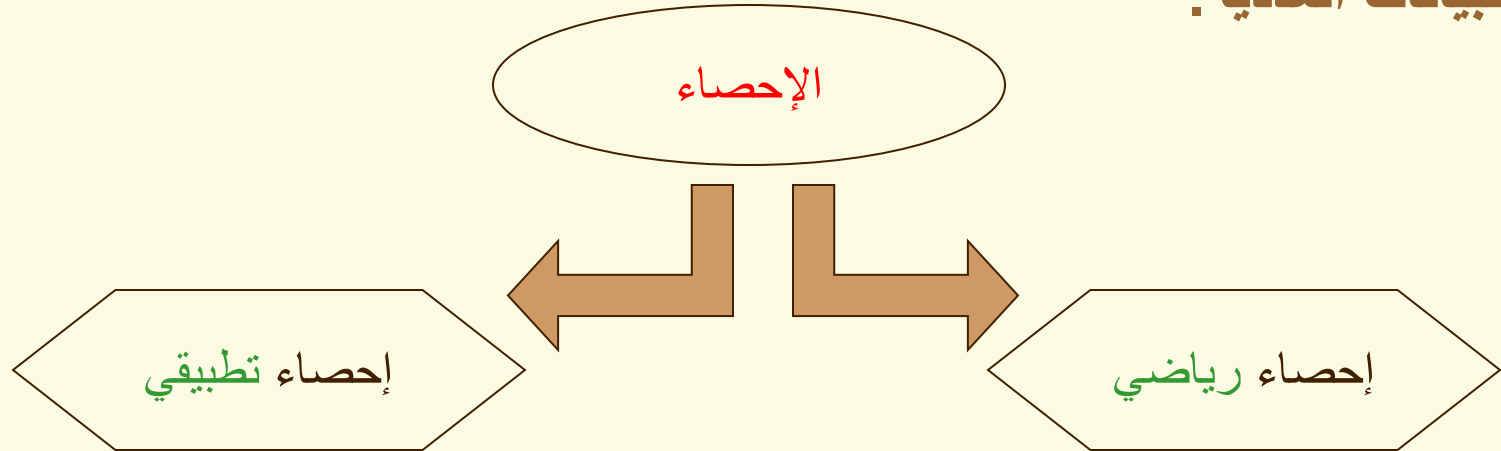
٢- مقاييس النزعة المركزية

٣- مقاييس التشتت

٤- تمارين

١- مصطلحات إحصائية Terminology

الإحصاء Statistic: هو علم تجميع وتنظيم وعرض وتحليل وتفسير البيانات العددية.



يشغل به متخصصين
على دراية بالرياضيات
وعلم الإحصاء
ومجال تطبيقه

علم نظري
يشغل به علماء الرياضيات
المتخصصين في الإحصاء

١- مصطلحات إحصائية Terminology

الإحصاء التطبيقي: وحيث أن مجال التطبيق متعدد فإن المشتغلين بالإحصاء التطبيقي تتعدد صورهم أيضاً فمنهم الزراعيين والأطباء والمهندسين والاجتماعيين والاقتصاديين والمهتمين بالدراسات السكانية والاستيراد والتصدير...إلخ

□ لا يوجد شخص واحد خبير في جميع التخصصات التي يطبق فيها الإحصاء

□ المبادئ الإحصائية الأساسية المستخدمة في العلوم البيولوجية والاقتصادية والاجتماعية لا تتغير

□ يمكن استعمال طرق معينة في فرع من العلوم دون الفرع الآخر نشوء فروع كثيرة لعلوم الإحصاء كإحصاء الاقتصادي والإحصاء الاجتماعي...إلخ

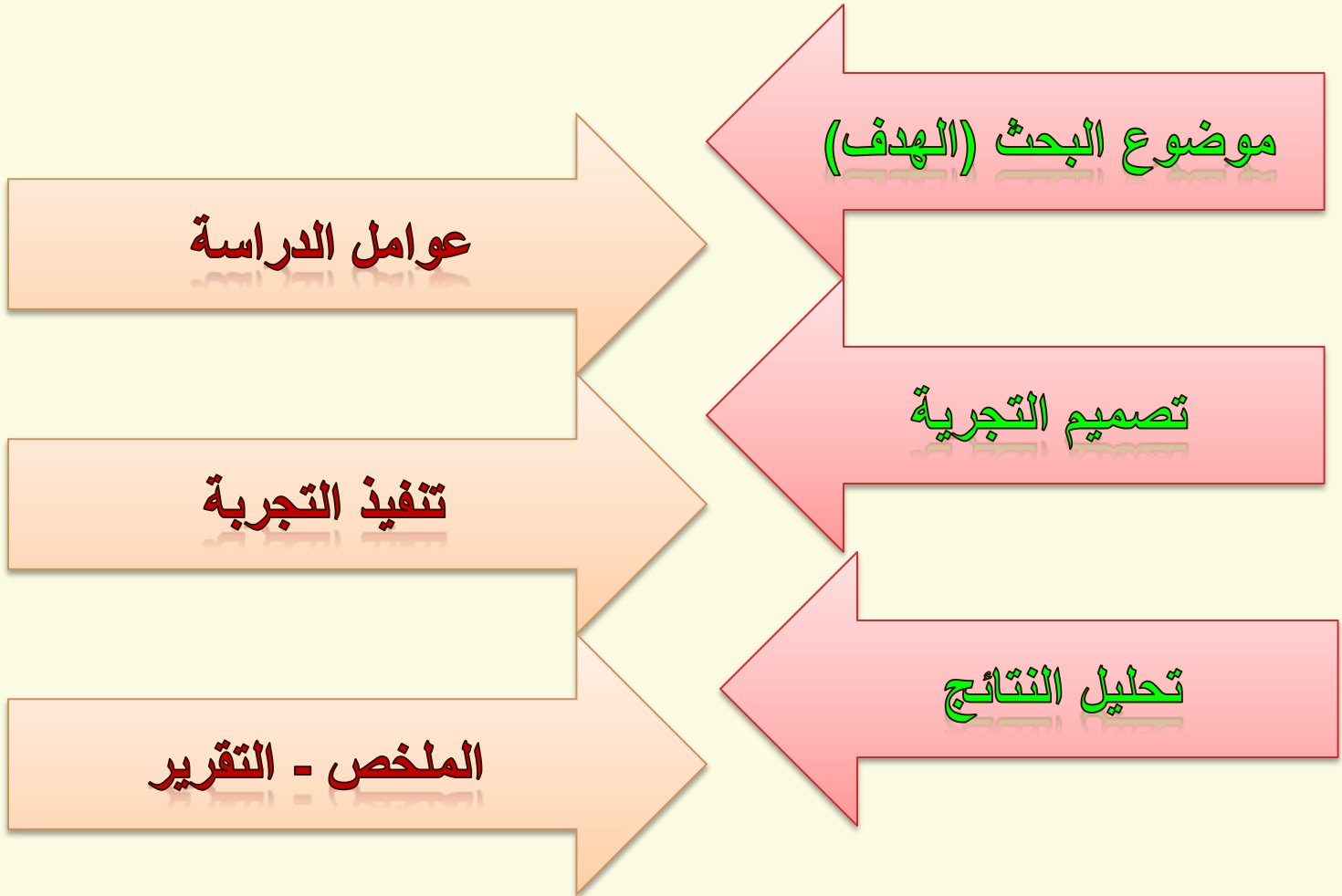
□ ويهتما في مجال العلوم الزراعية دراسة الإحصاء التجريبي (التطبيقي) وتصميم وتحليل التجارب الزراعية

١- مصطلحات إحصائية Terminology

الطريقة العلمية Scientific method: خطوات معينة لابد من اتباعها لابد من اتباعها في أي دراسة علمية ويأخذ بها العلماء في كل مكان.
خطوات الطريقة العلمية:

١. تحديد موضوع البحث وأهدافه **Definition and Objectives**
٢. بناء النظرية الفرضية (فرضية العدم): **Null Hypothesis**
٣. تصميم التجربة لاختبار الفرض **Experimental Design**
٤. القيم بتنفيذ التجربة **Execution**
٥. تحليل النتائج **Analysis**
٦. كتابة ملخص بالنتائج المتحصل عليها من التجربة للإفادة منها في التجارب المستقبلية **Conclusion**

خطوات الطريقة العلمية



١- مصطلحات إحصائية Terminology

المجتمع Population: هو مجموعة من الأفراد أو المفردات والأشياء أو القياسات تشترك فيما بينها في صفة متغيرة واحدة أو أكثر وقد يكون عدده محدود (السكان مثلاً) أو غير محدود (الأسماك).

المجتمعات المحدودة يمكن دراسة المجتمع كله أما المجتمعات غير المحدودة يستحيل دراسة جميع مفرداتها فليس أمامنا سوى العينة في دراستها.

العينة Sample: جزء صغير من المجتمع يؤخذ بطريقة معينة لتمثيل المجتمع وبشكل عام كلما كبر حجم العينة كلما زادت دقتها.
يجب أن تتصف العينة بما يلي:

١. تمثل المجتمع بشكل جيد **Representative**

٢. أن تؤخذ بشكل عشوائي **Random**

٣. أن تؤخذ بعدة مكررات **Convenience**

مصطلحات إحصائية Terminology

المتغيرات Variables: هي الصفة القابلة للتغير في النوع أو الكم من فرد لآخر من أفراد العينة أو المجتمع، وعادة تتصف هذه المتغيرات بما يلي:

١. إمكانية قياسها

٢. متغيرة

٣. بياناتها غير كسرية

Variables

Dependent/Response

Independent/Predictor

• من الصعب التحكم به خلال تنفيذ التجربة (مثل الغلّة)

• يسمى أيضاً Factor حيث يتم التحكم به من خلال التجربة (المعاملة) (مثل معدلات السماد)

مصطلحات إحصائية Terminology

القيمة الإحصائية Variate: هي قيمة المتغير التي تخص كل فرد من أفراد المجتمع، وعندما تكون قيم المتغيرات **مرتبة في اتجاه واحد** يرمز للقيمة الإحصائية بالرمز X_i وتقرأ **X -sub i** : أي أن X تأخذ قيم مختلفة من فرد لآخر. وإذا أردنا جمع هذه القيم المتغيرة يرفق الرمز Σ فتكتب:

n

$$\Sigma X_i = X_1 + X_2 + \dots + X_n$$

i=1

وتقرأ Summation of X_i

حيث $i = 1$ إلى n

وتدل على أن القيم المطلوب جمعها هي من X_1 إلى X_n

مصطلحات إحصائية Terminology

أما عندما تكون قيم المتغيرات مرتبة في اتجاهين (صفوف وأعمدة) يرمز للقيمة الإحصائية بالرمز X_{ij}

يدل i على رقم الصف، ويدل j على رقم العمود الذي توجد به القيمة.
مثال: X_{12} تدل على القيمة في الصف الأول والعمود الثاني
وإذا أردنا جمع هذه القيم المتغيرة فتكتب:

3 2

$$\sum_{i=1}^3 \sum_{j=1}^2 X_{ij} = X_{11} + X_{12} + \dots + X_{32}$$

$i=1 \quad j=1$

$i \backslash j$	1	2	Total
1	5	9	$X_{1.}=14$
2	8	10	$X_{2.}=18$
3	7	12	$X_{3.}=19$
Total	$X_{.1}=20$	$X_{.2}$	$X_{..}=51$

Measures of Central Tendency مقاييس النزعة المركزية

الوسط الحسابي Arithmetic mean

أكثر المتوسطات استعمالاً وأسهلها فهماً لعامة الناس حيث يطلقون عليه المتوسط

$$\bar{X} = \sum X_i/n$$

Measures of Central Tendency مقاييس النزعة المركزية

خواص الوسط الحسابي

□ مجموع انحرافات المفردات عن الوسط الحسابي = صفر

$$\sum X - \bar{X} = 0$$

□ مجموع مربعات الانحرافات للمفردات عن الوسط الحسابي = أقل ما يمكن

$$\sum (X - \bar{X})^2 = \text{Minimum}$$

مقاييس النزعة المركزية Measures of Central Tendency

عيوب الوسط الحسابي

- إذا وجدت بالعينة قيم متطرفة أو شاذة فإن قيمة الوسط الحسابي لا تمثل البيانات تمثيلاً سليماً
- يتعذر حساب الوسط الحسابي إذا كان الجدول التكراري مفتوحاً
- لا يصلح في حالة المتغيرات الوصفية أو النوعية

Measures of Dispersion

مقاييس التشتت

مثال:

أعلنت شركة عن حاجتها إلى وظيفة كيميائي فتقدم ٣ أشخاص لشغل الوظيفة. تم اختبارهم بإعطاء كل منهم ٥ عينات لتقدير نسبة الكالسيوم به وكانت النتائج:

رقم المتقدم	كمية الكالسيوم مغ	المجموع	المتوسط
1	4 4 4 4 4	20	4
2	1 11 3 2 3	20	4
3	3 5 3 4 5	20	4

Measures of Dispersion

مقاييس التشتت

التشتت: هو الدرجة التي تتجه بها البيانات الرقمية للمفردات
للانتشار حول المتوسط

أهم مقاييس التشتت:

المدى – الانحراف المتوسط – الانحراف المعياري – الخطأ
المعياري – معامل الاختلاف

Measures of Dispersion

مقاييس التشتت

المدى Range

❖ تتأثر قيمته بالقيم المتطرفة أو الشاذة \leq انطباع غير سليم???

لأنه يعتمد على أكبر مفردة وأصغر مفردة (ليس على كل المفردات)

❖ لا يمكن حسابه من الجداول التكرارية المفتوحة

Measures of Dispersion

مقاييس التشتت

Mean of Absolute Deviation الانحراف المتوسط

هو متوسط الانحرافات المطلقة لقيم المفردات عن وسطها الحسابي ويرمز له بالرمز MD

$$MD = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

Measures of Dispersion

مقاييس التشتت

مثال:

	A المتقدم		B المتقدم		C المتقدم	
samples	Ca /mg	$ X_i - \bar{X} $	Ca /mg	$ X_i - \bar{X} $	Ca /mg	$ X_i - \bar{X} $
1	4	0	1	3	3	1
2	4	0	11	7	5	1
3	4	0	3	1	3	1
4	4	0	2	2	4	0
5	4	0	3	1	5	1
المجموع	20	0	20	14	20	4
المتوسط	4	0	4	2.8	4	0.8

Measures of Dispersion

مقاييس التشتت

Mean of Absolute Deviation الانحراف المتوسط

- ❖ تتأثر قيمته بالقيم المتطرفة أو الشاذة
- ❖ لا يمكن حسابه من الجداول التكرارية المفتوحة
- ❖ لا يمكن التعامل معه جبرياً نظراً لإهمال الإشارات

Measures of Dispersion

مقاييس التشتت

Standard Deviation الانحراف المعياري

من أكثر مقاييس التشتت شيوعاً واستخداماً
يرمز للانحراف المعياري للمجتمع بالرمز σ
بينما يرمز للعينة بالرمز S

نظراً لأن الانحراف المتوسط لا يخضع للعمليات الجبرية
ونظراً لإمكانية التخلص من الإشارات السالبة عن طريق تربيع
الانحرافات، نشأت فكرة الانحراف المعياري

Measures of Dispersion

مقاييس التشتت

حساب تباين المجتمع σ^2

١. حساب الوسط الحسابي للمجتمع μ
٢. انحرافات مفردات المجتمع عن الوسط الحسابي
٣. تربيع الفقرة ٢ (مربع الانحرافات عن الوسط الحسابي)
٤. مجموع الفقرة ٣ (مجموع مربع الانحرافات عن الوسط الحسابي) أي

Sum of Squares of Deviation وتختصر SS

$$SS = \sum_{i=1}^n (X_i - \mu)^2$$

٥. تقسيم SS على عدد مفردات المجتمع N = متوسط مجموع الانحرافات Mean Sum of Squares (MS) أي (التباين Variance)

$$\sigma^2 = \sum_{i=1}^n (X_i - \mu)^2 / N$$

Measures of Dispersion

مقاييس التشتت

حساب تباين العينة S^2

١. حساب الوسط الحسابي للعينة \bar{X}
٢. انحرافات مفردات العينة عن الوسط الحسابي
٣. تربيع الفقرة ٢ (مربع الانحرافات عن الوسط الحسابي)
٤. مجموع الفقرة ٣ (مجموع مربع الانحرافات عن الوسط الحسابي) أي

Sum of Squares of Deviation وتختصر SS

$$SS = \sum_{i=1}^n (X_i - \bar{X})^2$$

٥. تقسيم SS على عدد مفردات العينة n = متوسط مجموع الانحرافات Mean Sum of Squares (MS) أي (التباين Variance)

$$S^2 = \sum_{i=1}^n (X_i - \mu)^2 / n - 1$$

Measures of Dispersion

مقاييس التشتت

بأخذ الجذر التربيعي للتباين
(الفقرة ٥)



الانحراف المعياري
Standard Deviation

$$S = \sqrt{S^2}$$

Measures of Dispersion

مقاييس التشتت

Standard Deviation الانحراف المعياري

لا يستخدم الانحراف المعياري في مقارنة عينتان تختلفان في وحدات القياس للتغلب على هذه الصعوبة نستخدم مقياس نسبي للتشتت

Measures of Dispersion

مقاييس التشتت

معامل الاختلاف Coefficient of variation

❖ هو تحويل الانحراف المعياري إلى نسبة مئوية من الوسط

الحسابي لنفس البيانات ويرمز له CV

❖ هو مقياس مطلق غير مميز بوحدات قياس ويحسب:

$$CV = (S/\bar{X}) * 100$$

Measures of Dispersion

مقاييس التشتت

معامل الاختلاف Coefficient of variation

تمرين:

قارن بين التشتت في عينتين الأولى من الكتاكيت والأخرى من النباتات حيث قدر في الأولى الوزن وفي الثانية الطول، علماً بأن:

$$\bar{X}_1=150, S_1=15, \bar{X}_2= 70, S_2=10$$