

محاضرات في أساسيات البحث العلمي

أنواع البيانات والمتغيرات والوصف الإحصائي

الثلاثاء 15/6/2021

الدكتور معاذ الشرفاوي الجزائري

النائب العلمي للمعهد العالي للتنمية الإدارية

منسق محور العلوم الاقتصادية والإدارية والقانونية

ورشة جامعة دمشق للبحث العلمي نحو تشاركية بحثية تطبيقية - حزيران 2021

مخرجات المحاضرة

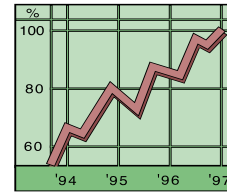
- تمييز أنواع البيانات و طرائق جمعها.
- تمييز أنواع المتغيرات.
- أهمية التعريف الواضح لمتغير البحث ولمستوى قياسه.
- التمييز بين الأزواج الآتية من الاصطلاحات الإحصائية:
 - الإحصاء الوصفي والإحصاء الاستدلالي
 - البيانات الأولية و البيانات الثانوية.
 - العينة و المجتمع (الإحصائي).
 - البيانات الكمية و البيانات النوعية.
 - بيانات السلاسل الزمنية و البيانات المقطعية.
- خطوات جمع البيانات الأولية باستخدام أداة الاستبانة ومحاذيرها.
- ² وصف الطرائق الأساسية للمعاينة.

الإحصاء الوصفي والإحصاء الاستدلالي:

□ الإحصاء الوصفي:

■ جمع ← عرض ← وصف

$$\frac{\sum x_i}{n}$$



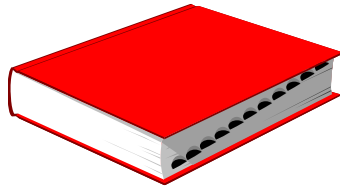
□ الإحصاء الاستدلالي:

■ معلومات من العينة ← استنتاجات عن المجتمع.

مصادر البيانات

جمع البيانات
الثانوية

بيانات منشورة:
مطبوعة / إلكترونية



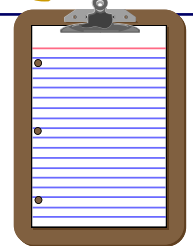
جمع البيانات
الأولية

مشاهدة



تجارب

مسح



البيانات الأولية: خطوات تصميم الاستبيان

- حدد المشكلة بوضوح.
- حدد المجتمع المستهدف.
- قم بصياغة أسئلة الاستبيان (في ضوء النموذج الإحصائي)
- اختر الاستبيان قبل تطبيقه.
- حدد حجم العينة و طريقة المعاينة.
- قم باختيار العينة و نفذ المسح ميدانياً.

أنواع الأسئلة:

□ الأسئلة المغلقة.

- الاختيار من قائمة من الخيارات المعرفة سلفاً.

□ الأسئلة المفتوحة.

- نمط حر في الإجابة تقبل في أية عبارة / قيمة / ...

□ الديموغرافيات:

- عن الخصائص الشخصية للمبحوث.

المجتمعات و العينات

□ المجتمع هو كافة العناصر أو الأفراد محل الاهتمام

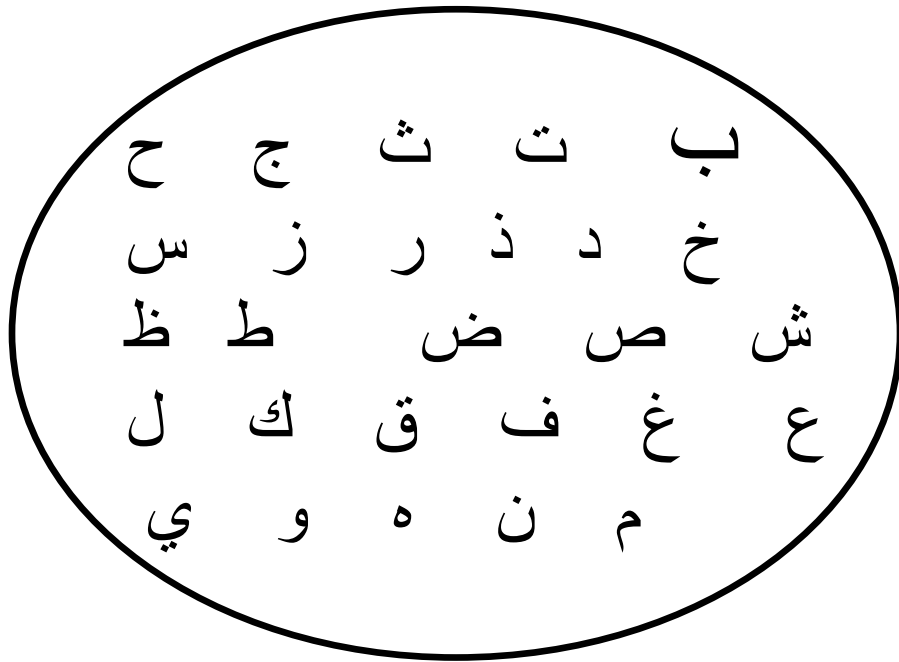
■ مثال : كافة الناخبين المحتملين في الانتخابات القادمة.
كافة القطع المنتجة في الشهر الماضي.

□ العينة هي مجموعة جزئية من المجتمع.

■ مثال : ألف ناخب تم اختيارهم عشوائياً لإجراء مقابلات لاستطلاع الرأي.
أربعين قطعة من إنتاج الشهر الماضي تم سحبها لإجراء الاختبارات.

المجتمع والعينة

مجتمع



المعلم (البارامتر)

مجهول على مستوى المجتمع

عينة



الإحصاءة

معلومة من العينة

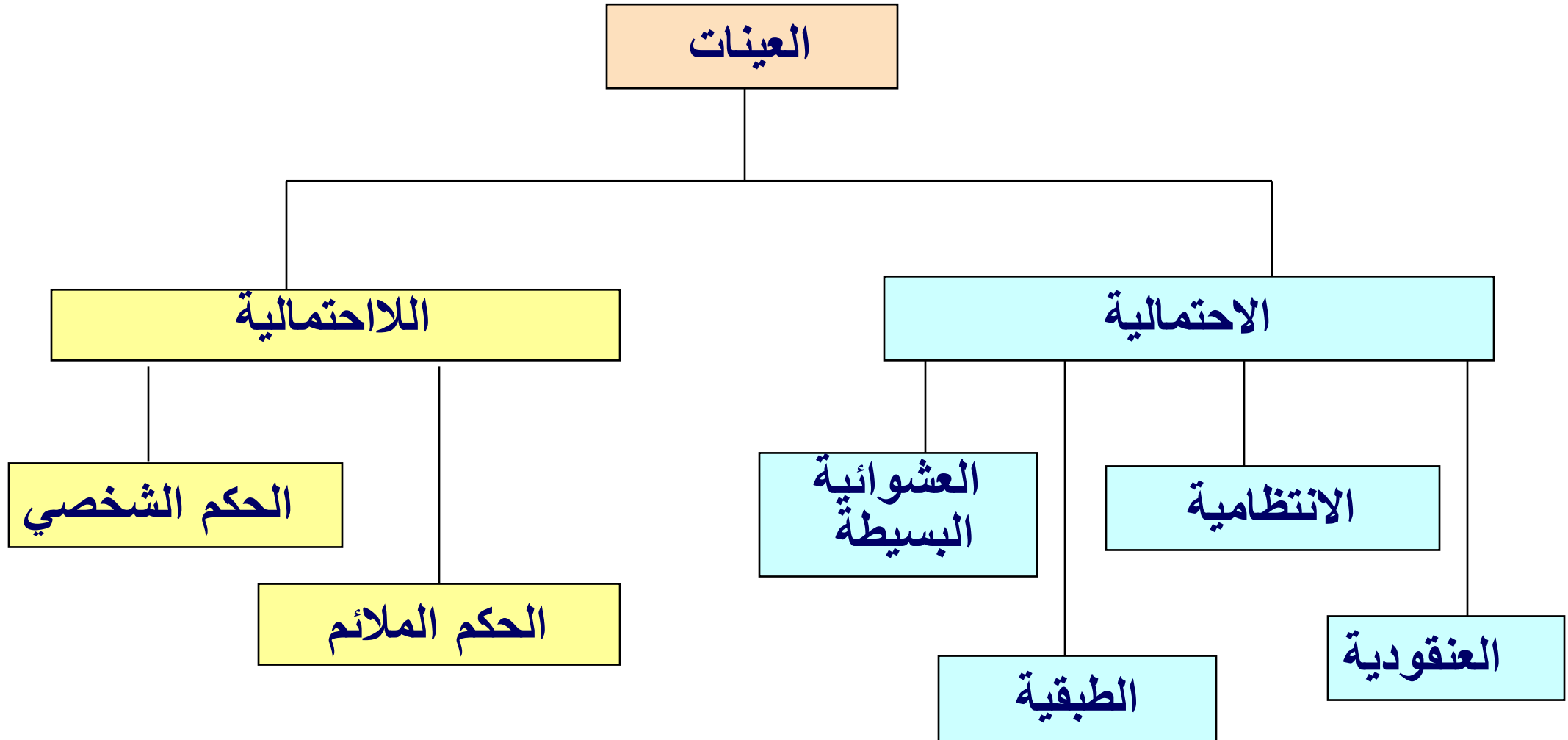
لماذا نسحب عينة؟

□ **الوقت:** هل لديك الوقت الكافي لقياس كافة مفردات المجتمع؟

□ **التكلفة:** هل لديك الميزانية اللازمة لمسح المجتمع بأكمله؟

□ **الدقة:** هل من داعٍ لمسح المجتمع إذا كانت العينة تعطي تقديرات مجتمعية دقيقة إلى حد مرضٍ؟

تقنيات المعاينة الأكثر شيوعاً



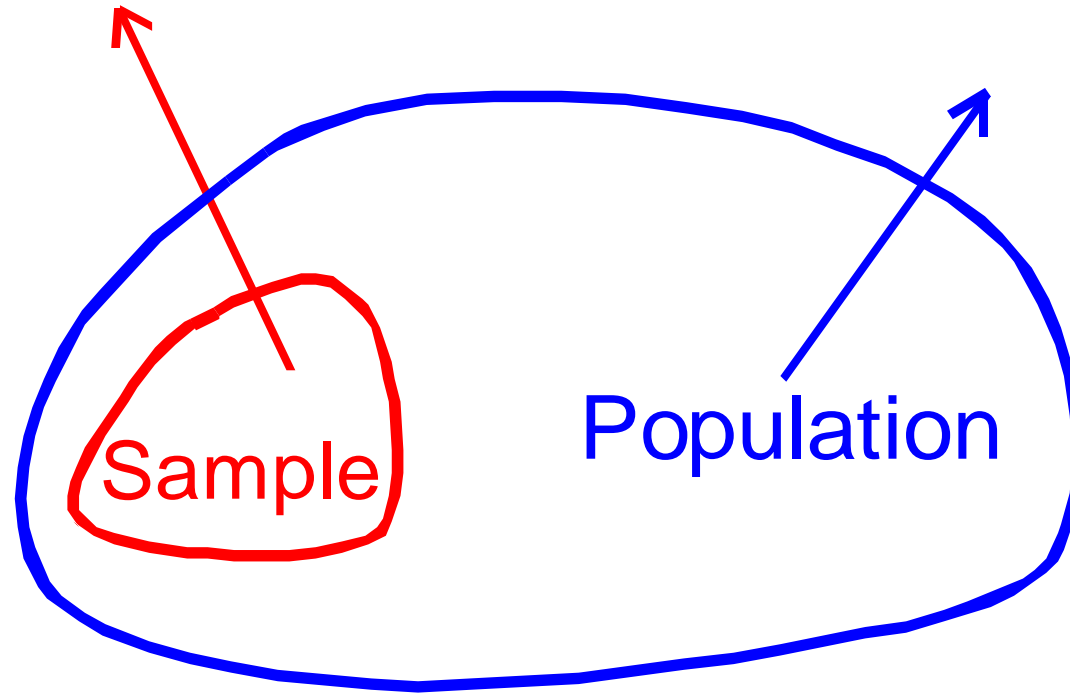
الإحصاء الاستدلالي

□ فحص نتائج المعاينة بغرض الوصول لاستنتاجات حول المجتمع

إحصاءات العينة
(معلومة)



بارامترات المجتمع
(مجهولة و لكن بالإمكان تقديرها)



استدلال

الإحصاء الاستدلالي

الوصول لاستنتاجات و / أو صنع قرارات بخصوص المجتمع بناءً على النتائج المستخلصة من العينة.

□ التقدير:

■ تقدير وسطي وزن طلاب المرحلة الإعدادية من الذكور على مستوى محافظة دمشق بأكملها.

□ اختبار الفرضية:

■ استخدام قرائن من العينة المسحوبة عشوائياً من طلاب الإعدادية (خمسمائة طالب من مدارس دمشق) لاختبار الزعم القائل بأن متوسط الوزن لا يزيد عن 55 كغ.

أنواع البيانات: الكمية و الكيفية.

البيانات

كيفية

كمية

أمثلة:

لون العين، الحالة المدنية.
فئات محددة بتعريف
(واضح)

منقطعة

أمثلة:

عدد الأطفال، عدد الأعطال في
اليوم (أشياء للعد)

مستمرة

أمثلة:

الوزن، الفولت، الطول
(خصائص للقياس)

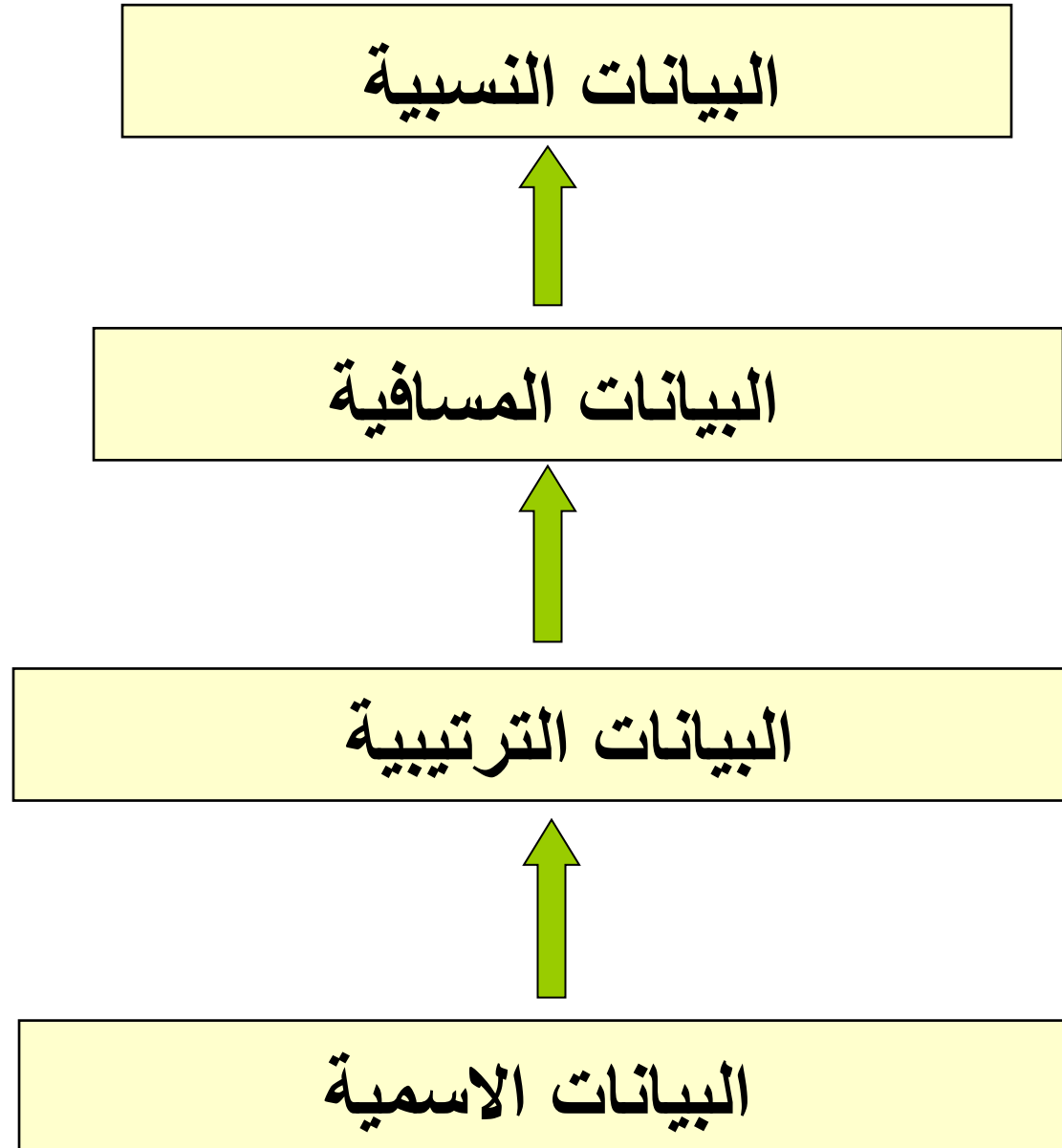
أنواع البيانات: الزمنية و المقطعية والوحية

حجم الإنتاج الموزع على المحافظات (ملايين الليرات السورية)				
2007	2006	2005	2004	
435	460	475	490	دمشق
320	345	375	395	حلب
405	390	410	395	المنطقة الوسطى
260	270	285	280	المنطقة الساحلية

بيانات
سلسلة
زمنية

بيانات مقطعية

مستويات قياس البيانات



مستويات قياس البيانات

البيانات النسبية



البيانات المسافية

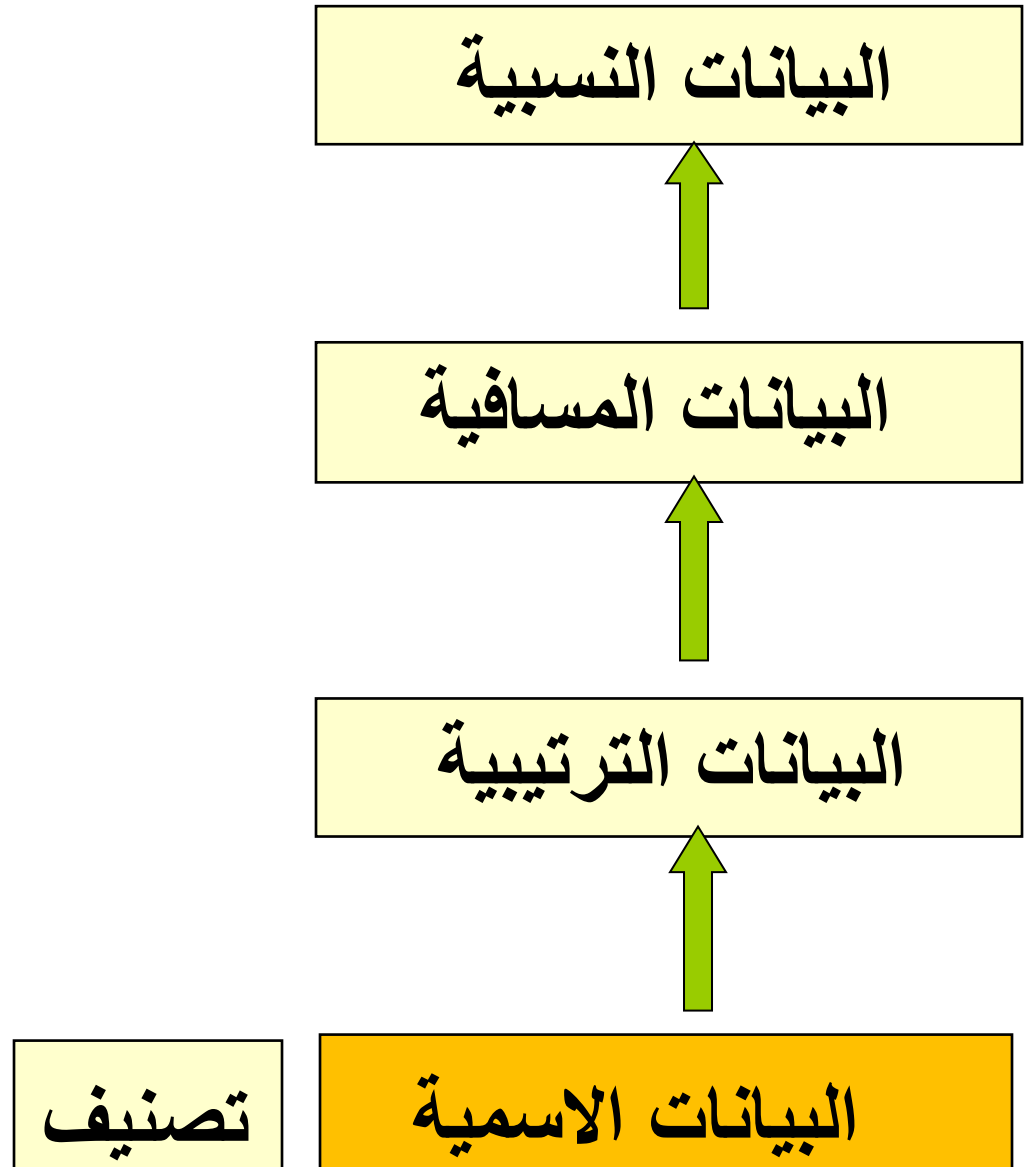


البيانات الترتيبية

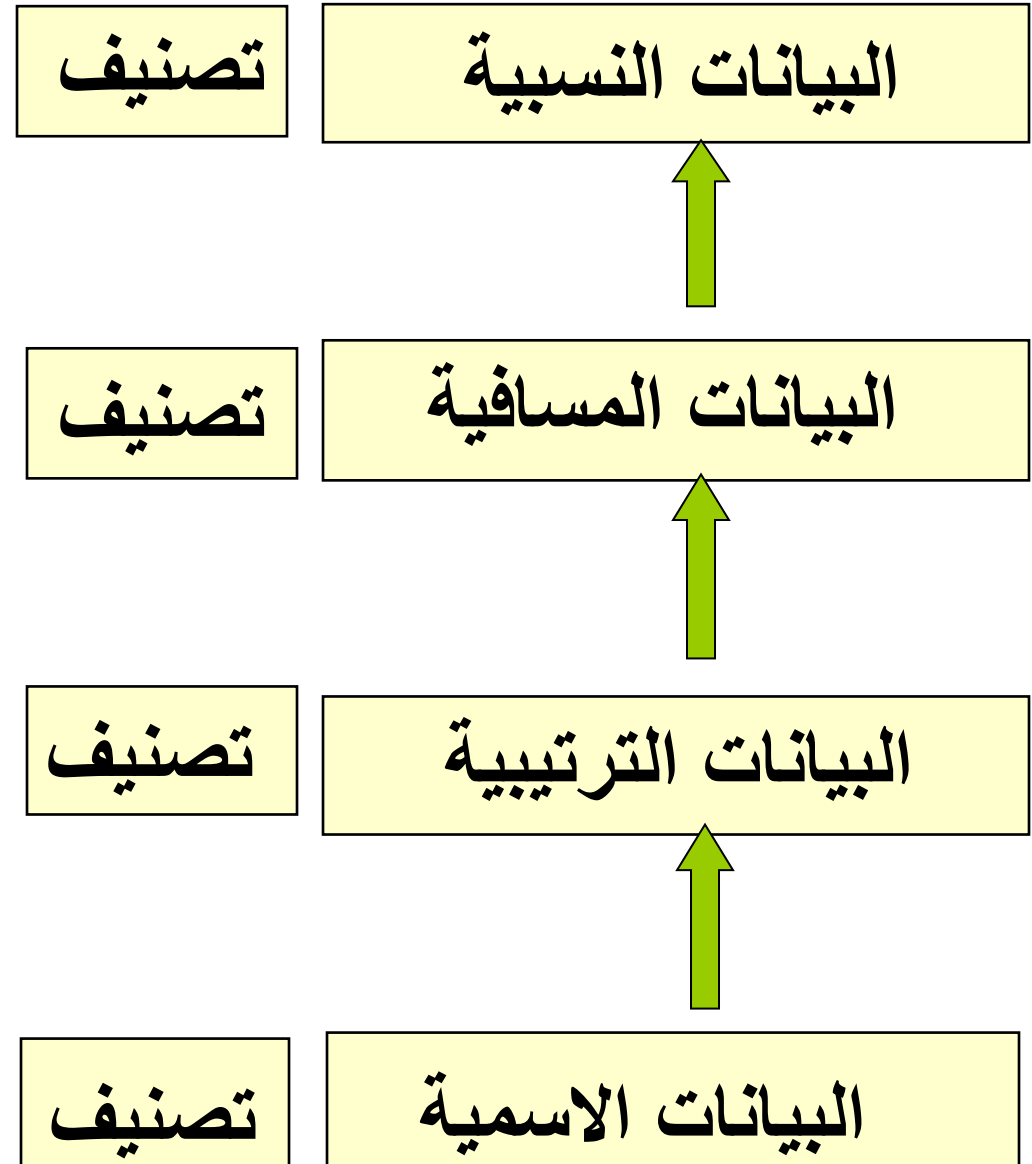


البيانات الاسمية

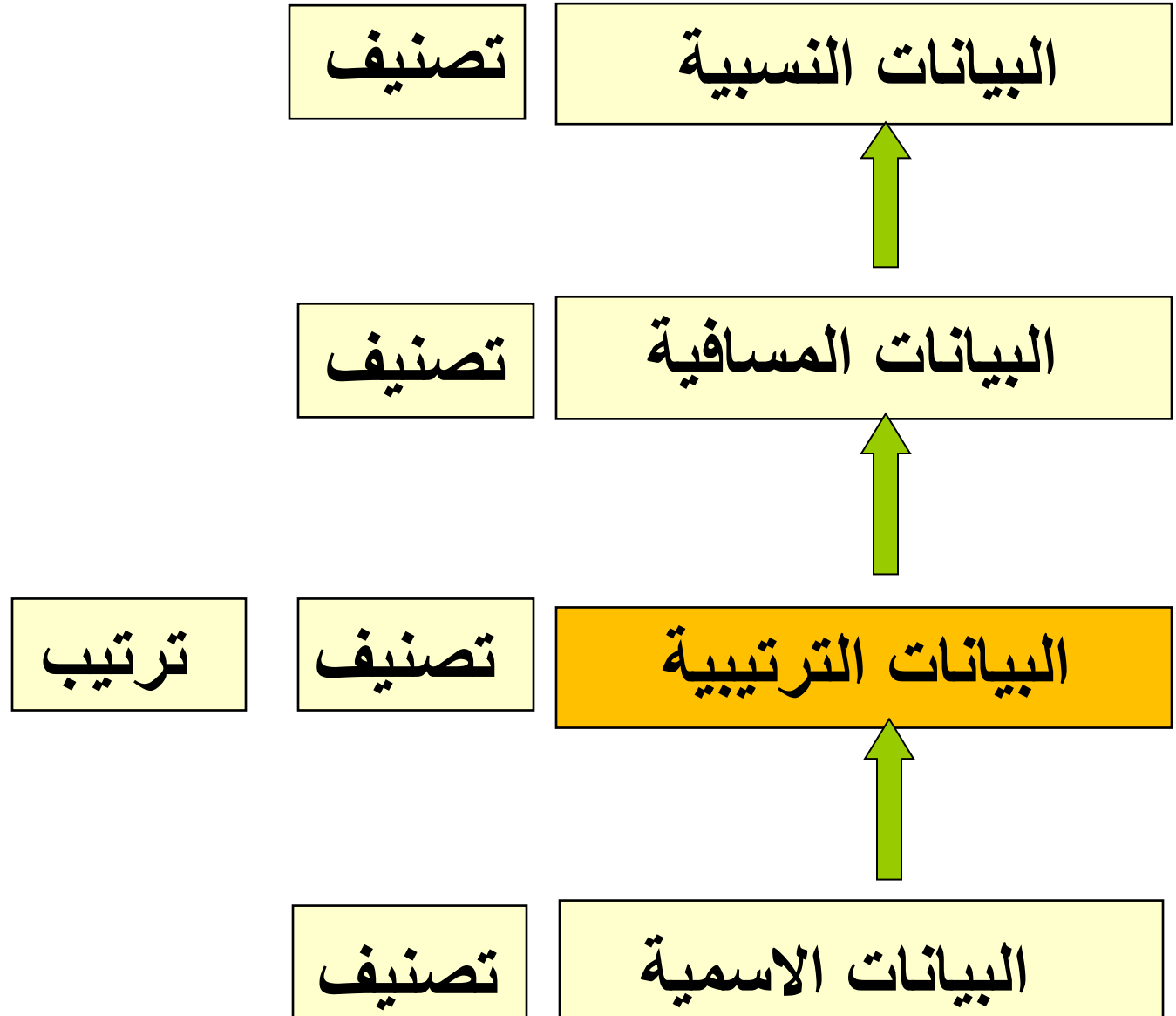
مستويات قياس البيانات



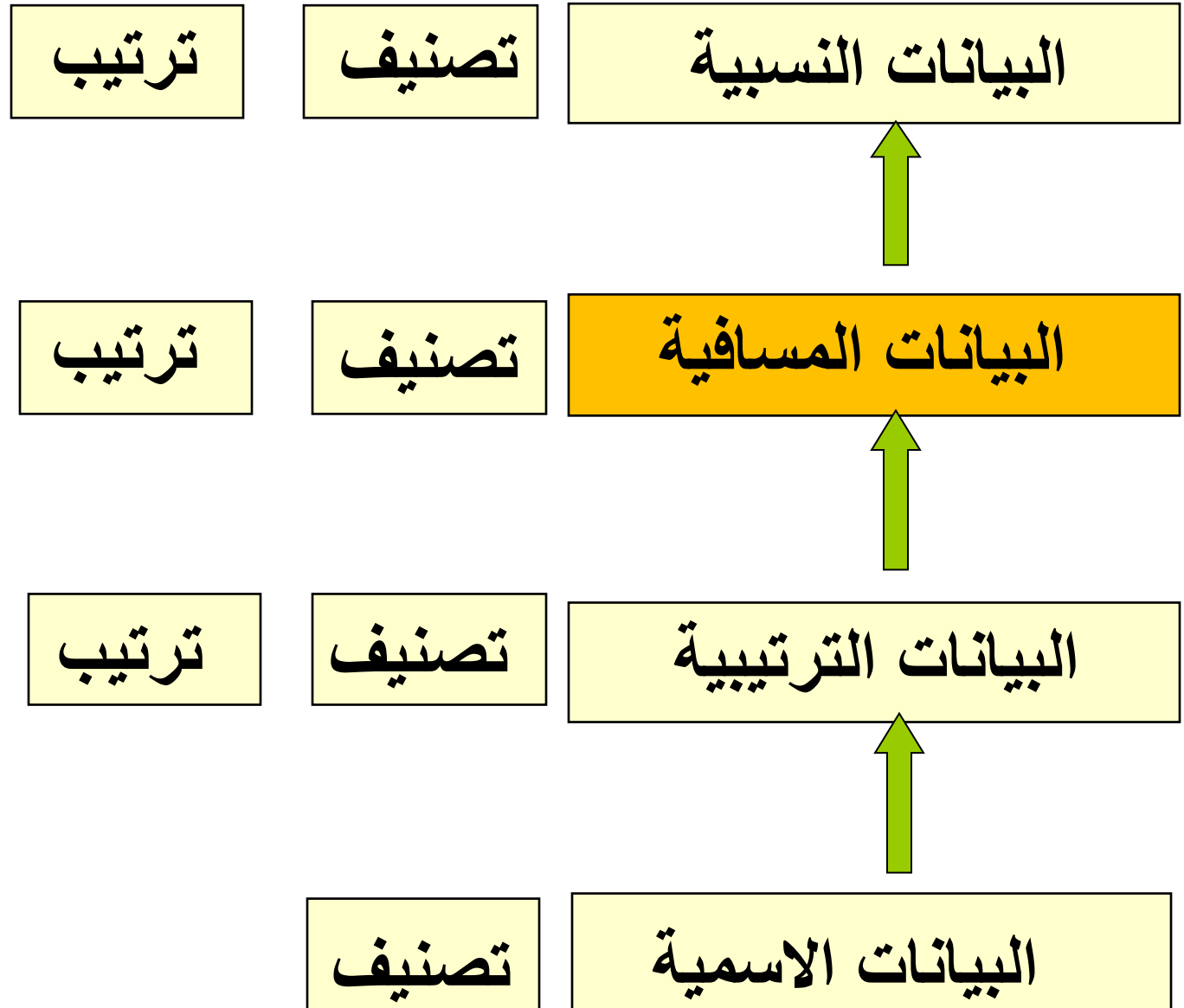
مستويات قياس البيانات



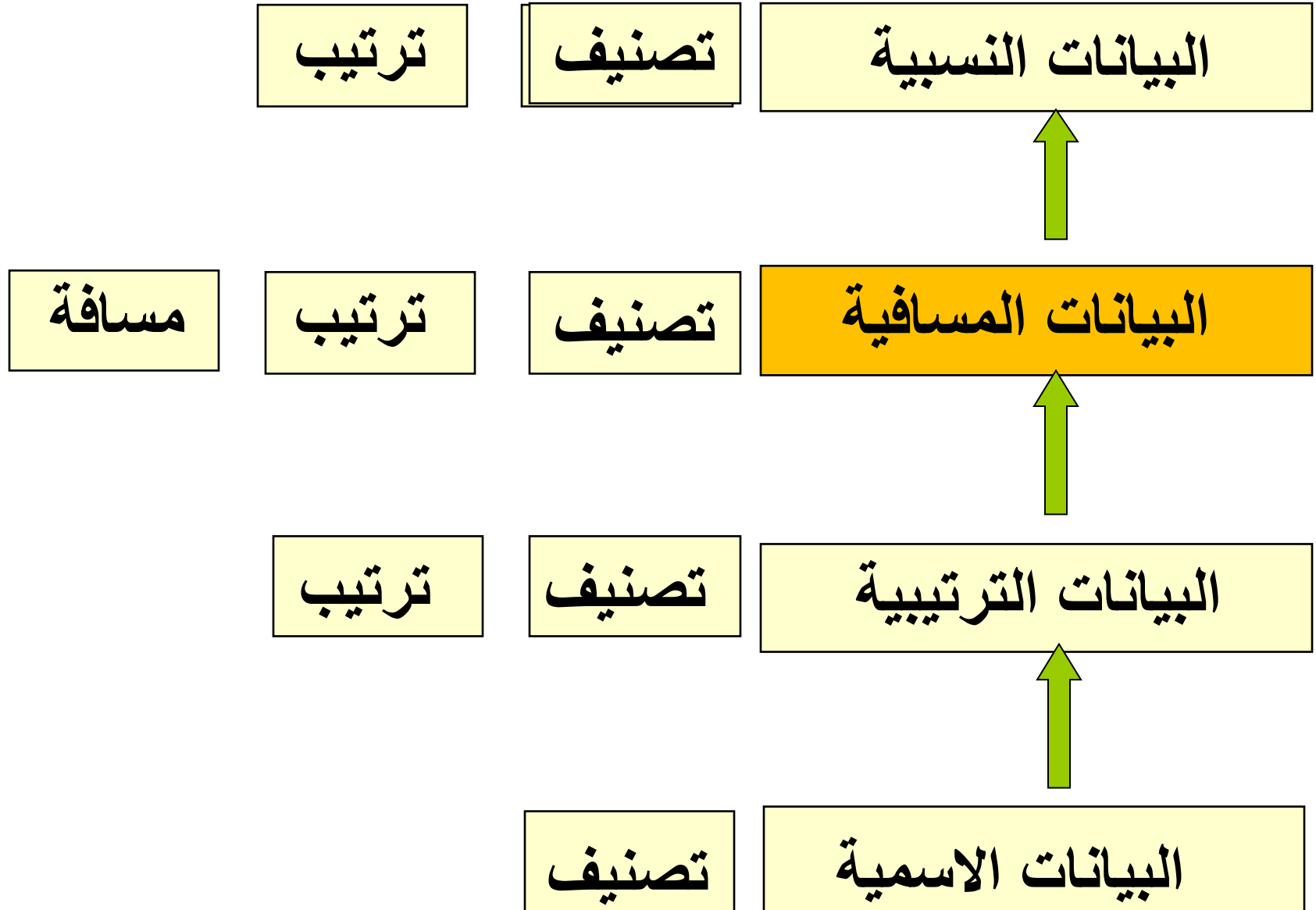
مستويات قياس البيانات



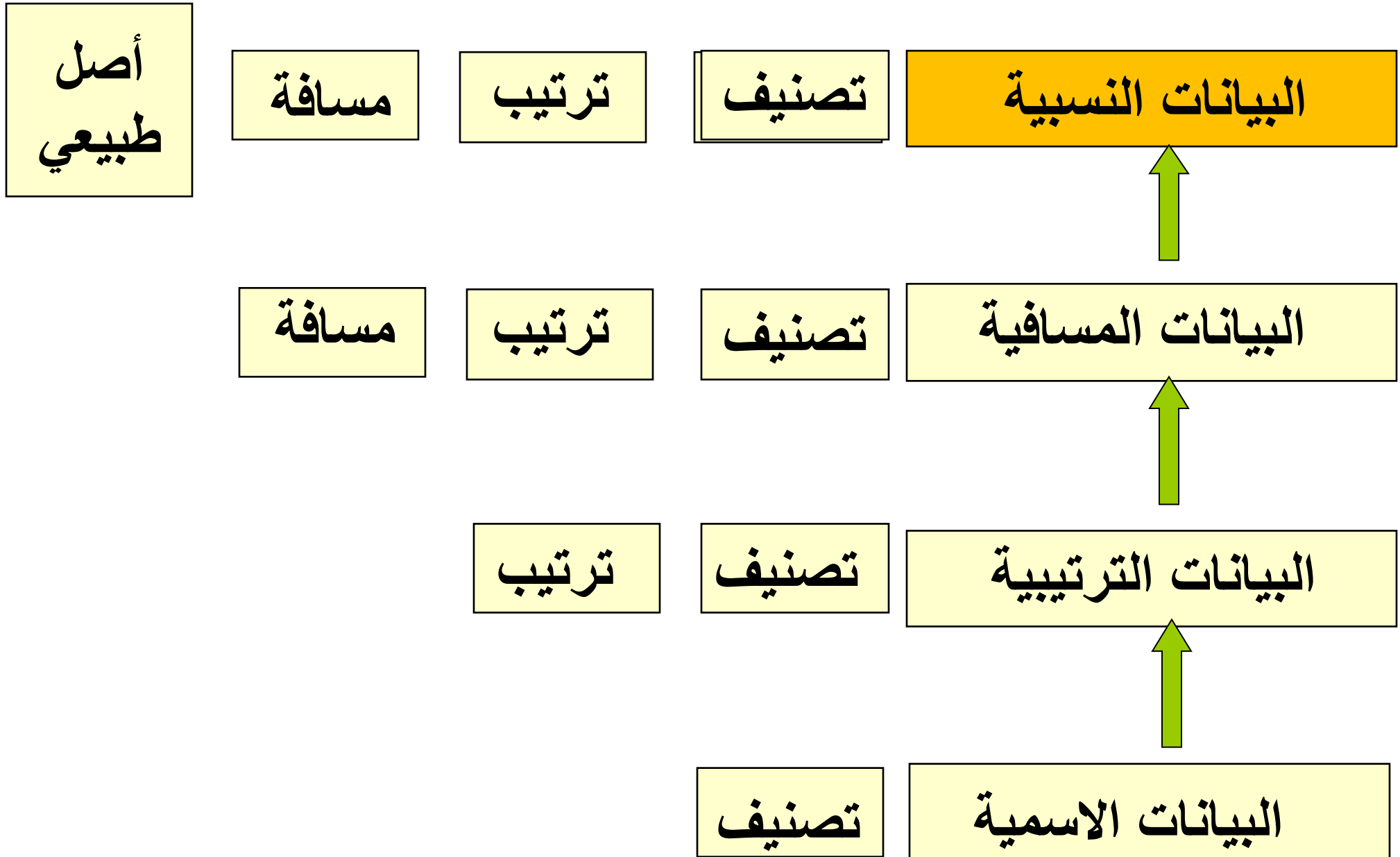
مستويات قياس البيانات



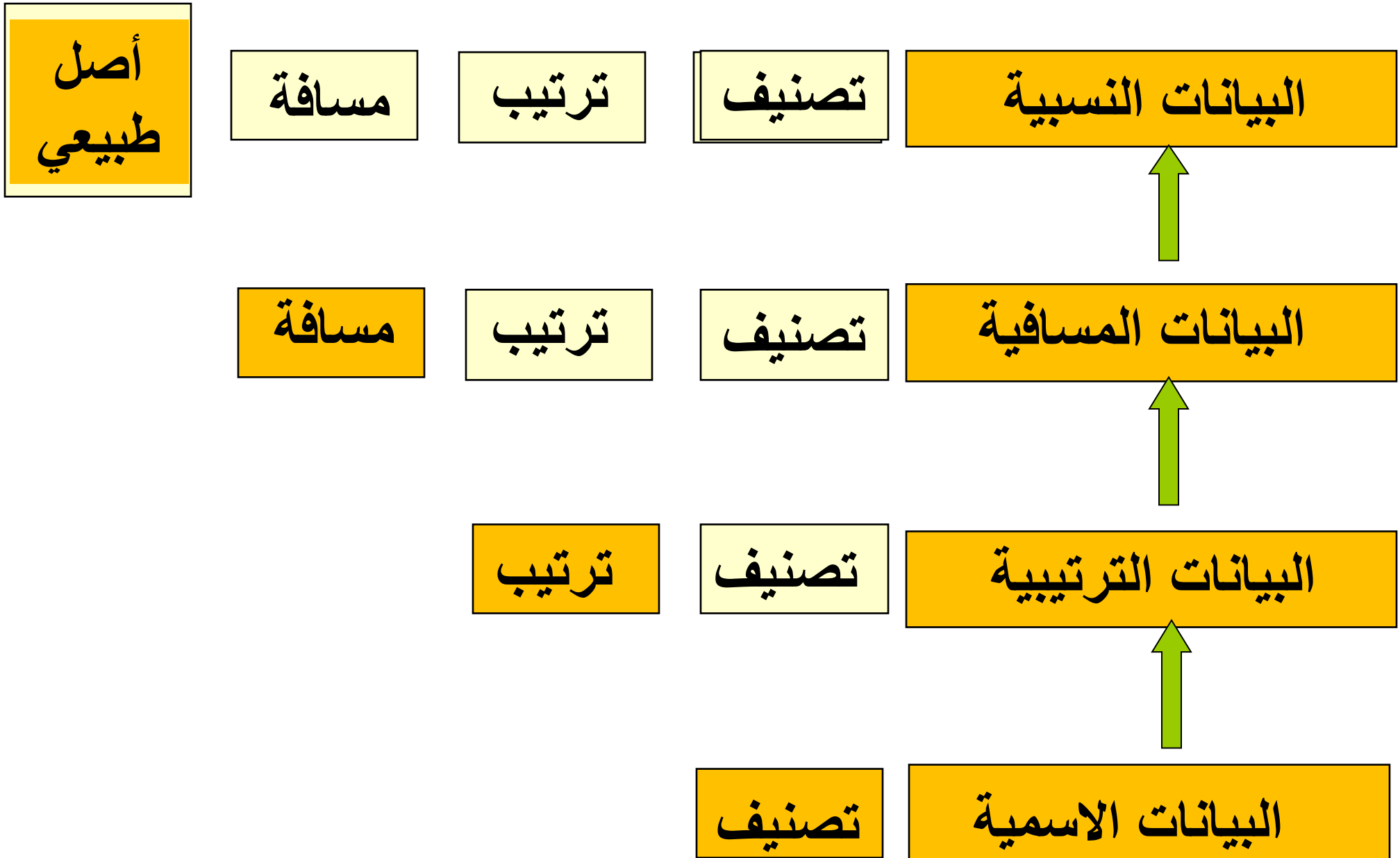
مستويات قياس البيانات



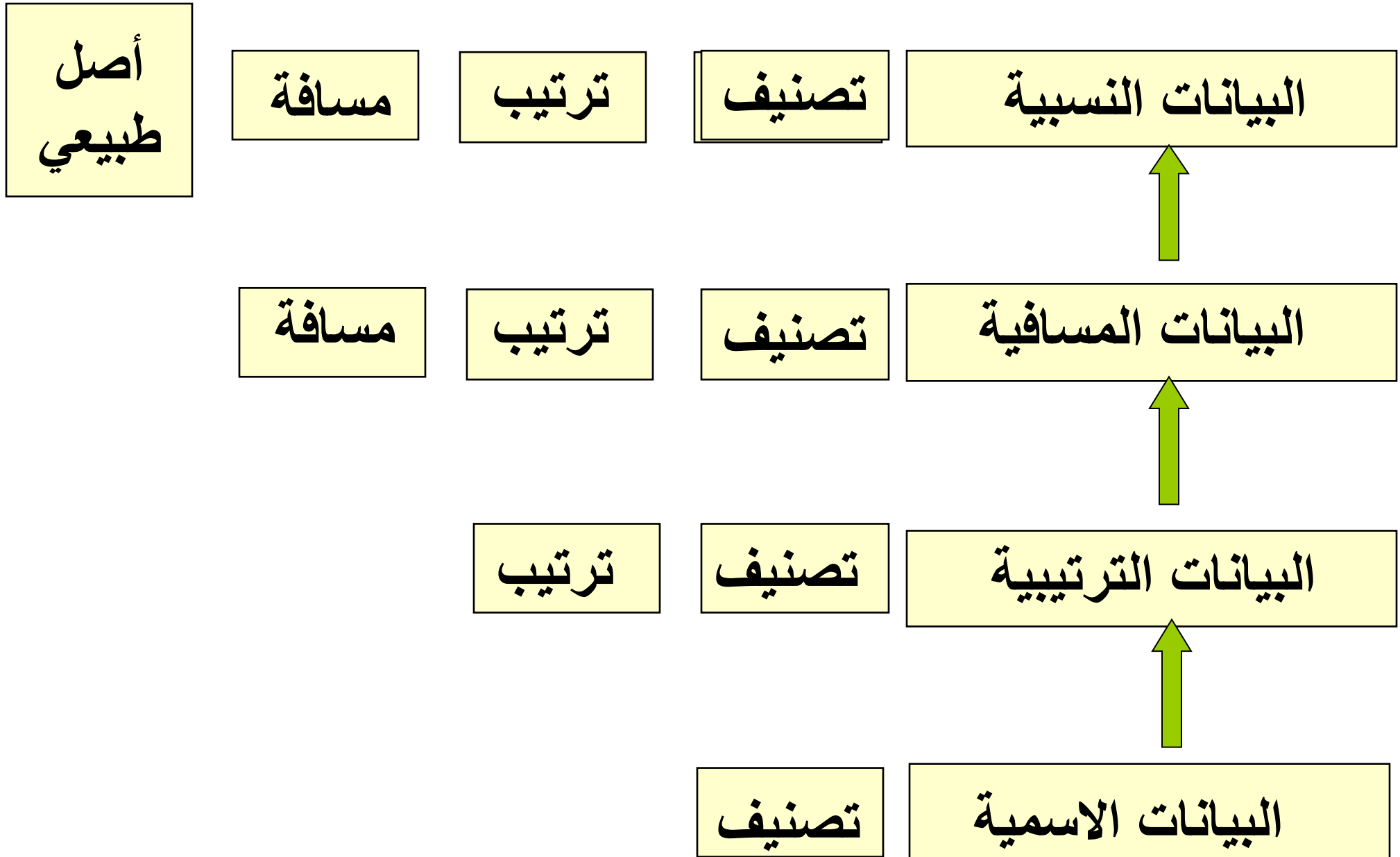
مستويات قياس البيانات



مستويات قياس البيانات



مستويات قياس البيانات



وصف البيانات: التوزيعات التكرارية

ما هي؟

هي عبارة عن **قائمة/جدول** تحتوي **قيم متغير** أو مجموعة من الفئات التي تقع فيها البيانات، مع **التكرارات الموافقة** لقيم المتغير أو للفئات التي تقع داخلها البيانات.

□ لم نستخدمها؟

- تلخص البيانات.
- تكثف البيانات الخام في صيغ أكثر فائدة.
- تساعد في الوصول إلى تفسير مرئي يوفر الوقت و الجهد.

التوزيعات التكرارية: البيانات المنقطعة

□ هل تذكر ما هي البيانات المتقطعة؟ (المحاضرة الأولى!)

عدد الأيام	التكرارات
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
الإجمالي	200

مثال:

قامت جريدة يومية تصدر في هونولولو
بسؤال مائتي مواطن عن عدد المرات التي
يقوم القارئ فيها بشراء الجريدة في الأسبوع،
و قد تم تلخيص نتائج الاستبيان في الجدول
إلى اليسار.

التكرار النسبي

تابع المثال السابق:

$$\frac{44}{200} = .22$$



التكرار النسبي	التكرار	عدد الأيام
.22	44	0
.12	24	1
.09	18	2
.08	16	3
.10	20	4
.11	22	5
.13	26	6
.15	30	7
1.00	200	Total

22% من القراء المشمولين بالعينة أفادوا بأنهم لا يقرؤون الجريدة و لا حتى مرة واحدة في الأسبوع.

مثال على التوزيعات التكرارية

رتب البيانات تصاعدياً

22, 22, 23, 27, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 42, 43, 43, 44, 44

التوزيع التكراري		
التكرار النسبي	التكرار	الفئة
.45	9	تساوي/أكبر من 20 و أصغر من 30
.25	5	تساوي/أكبر من 30 و أصغر من 40
.30	6	تساوي/أكبر من 40 و أصغر من 50
1.00	20	الإجمالي

وصف البيانات مرئياً: الهستوغرام

إذا كان لديك مخططاً بالمواصفات التالية:

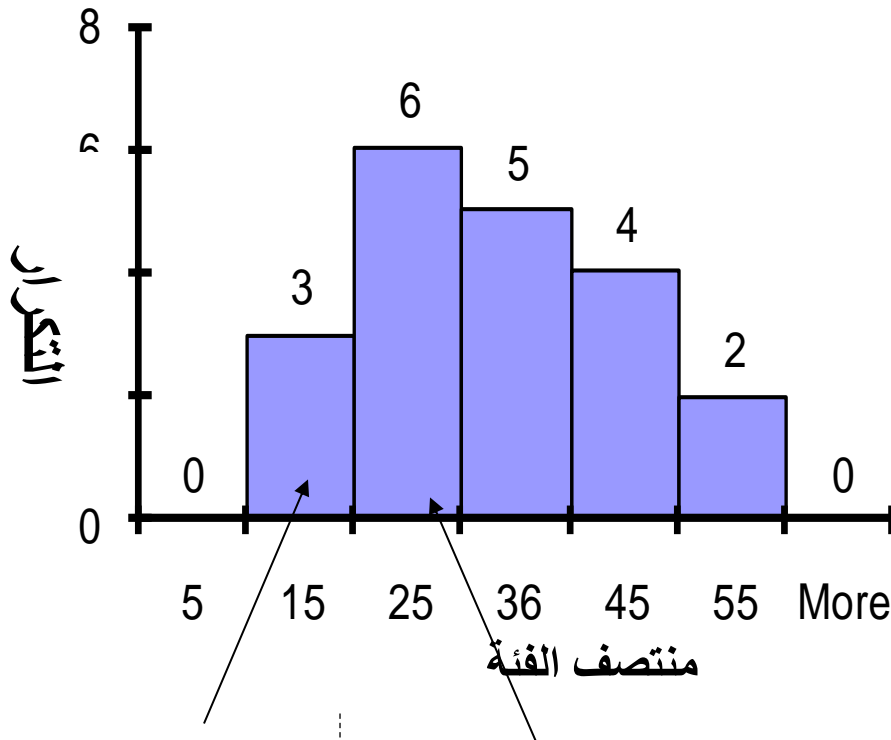
□ الفئات على المحور الأفقي.

□ التكرارات على المحور العمودي.

□ أطوال الأعمدة تمثل عدد المشاهدات في الفئات.

فالمخطط الذي لديك هو "هستوغرام"

هستوغرام بعلامات عشرين طالباً في مادة الإحصاء!



مثال: علامات الطلاب في مادة الإحصاء!

رتب علامات الطلاب تصاعدياً ثم قسمهم إلى فئات

22, 22, 23, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

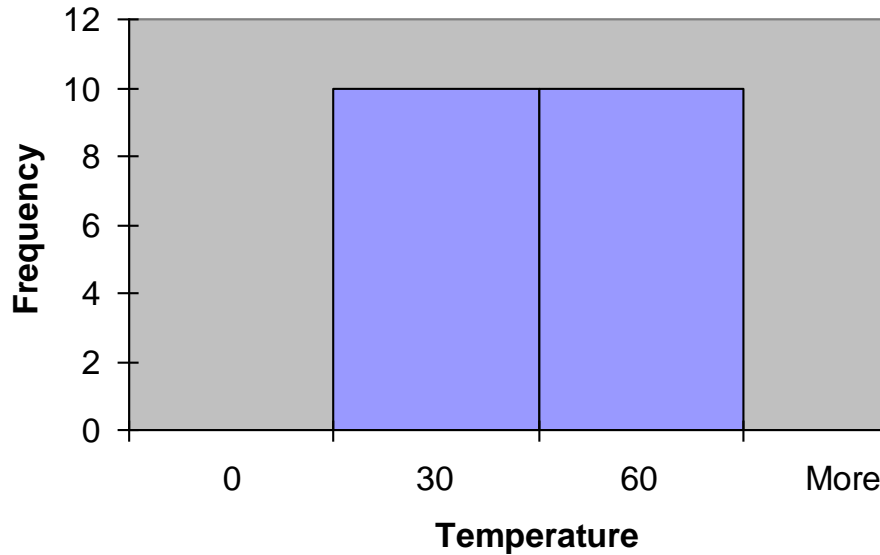
أسئلة عن تجميع البيانات في فئات:

1. كم هو عرض الفئة؟ كيف نحدده؟

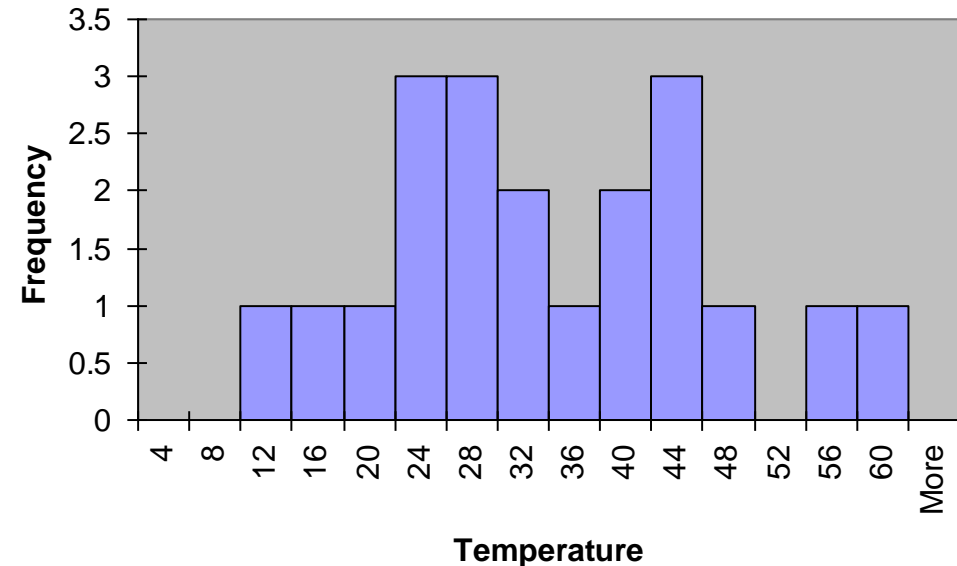
1. (بالتالي، كم هو عدد الفئات الواجب إنشاؤها)

2. كيف نحدد النقاط الحدية للفئة؟

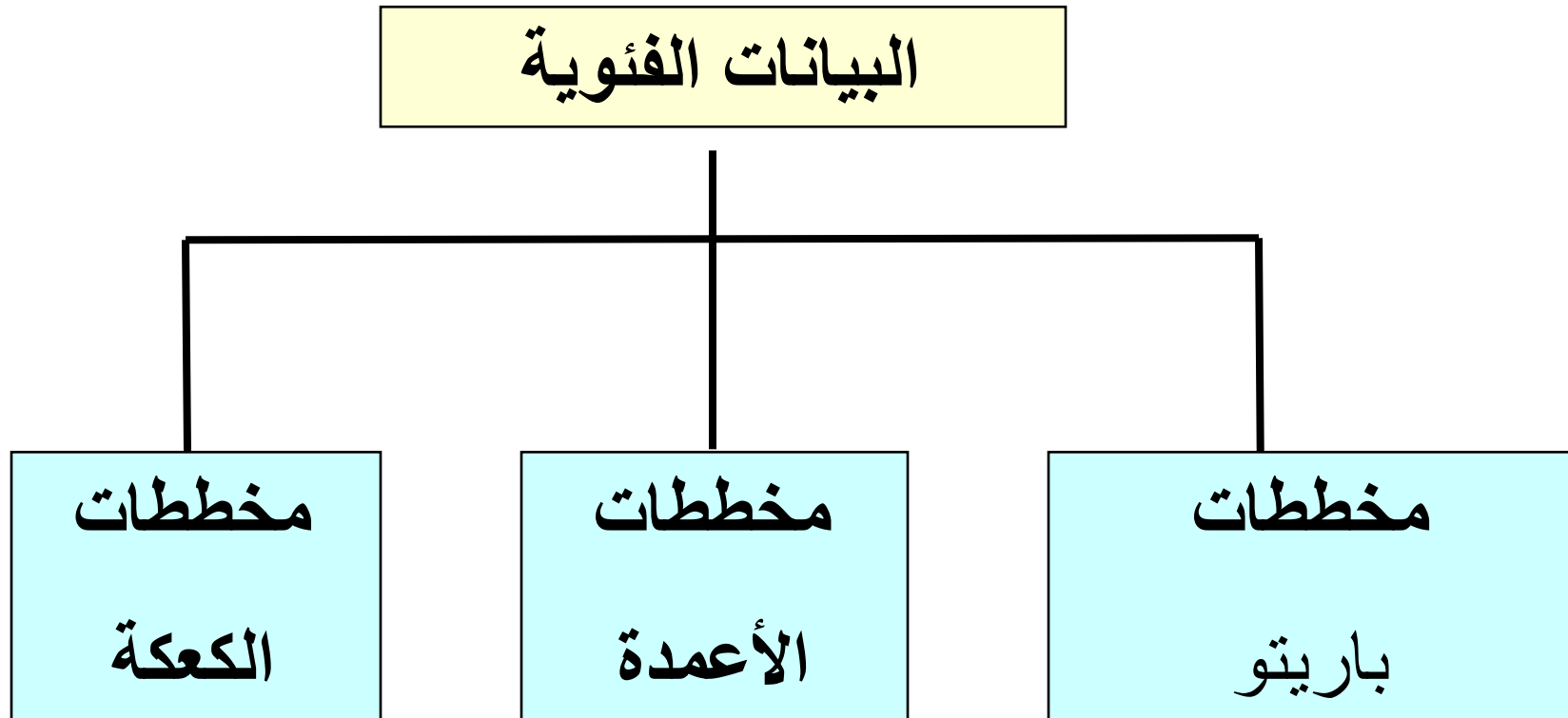
أم هذه؟



هذه؟

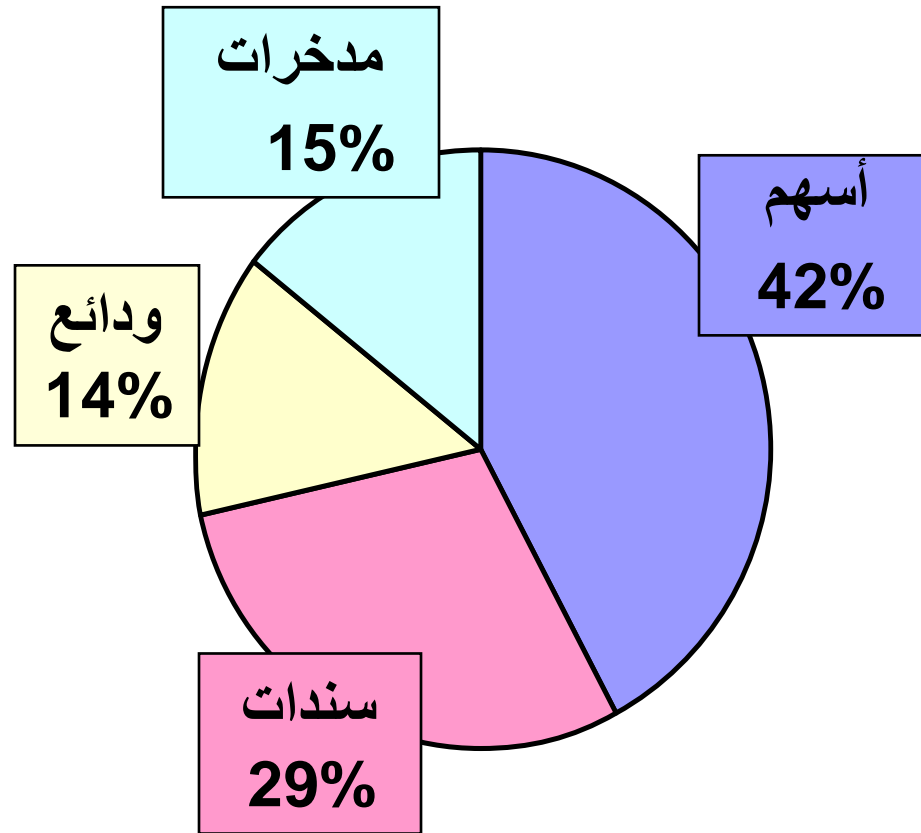


التمثيل البياني للبيانات الفئوية



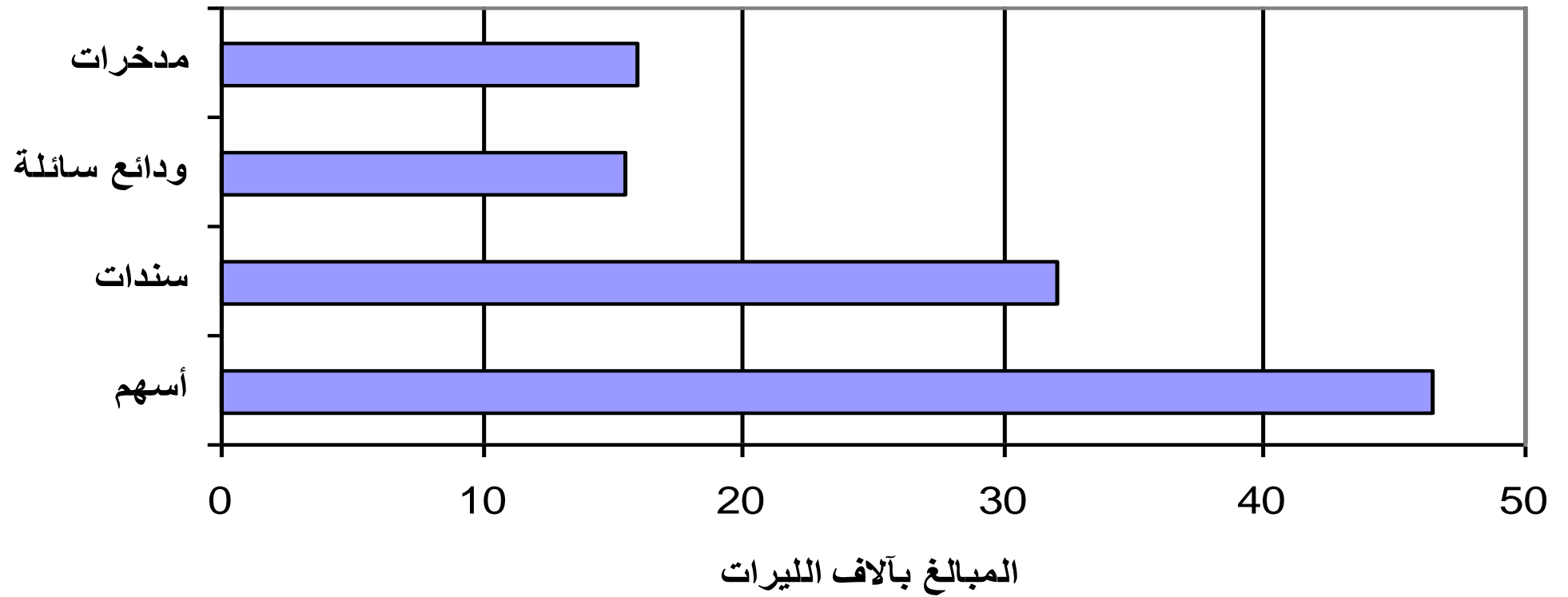
مثال: مخطط كعكة

مثال: الحافظة الاستثمارية الحالية

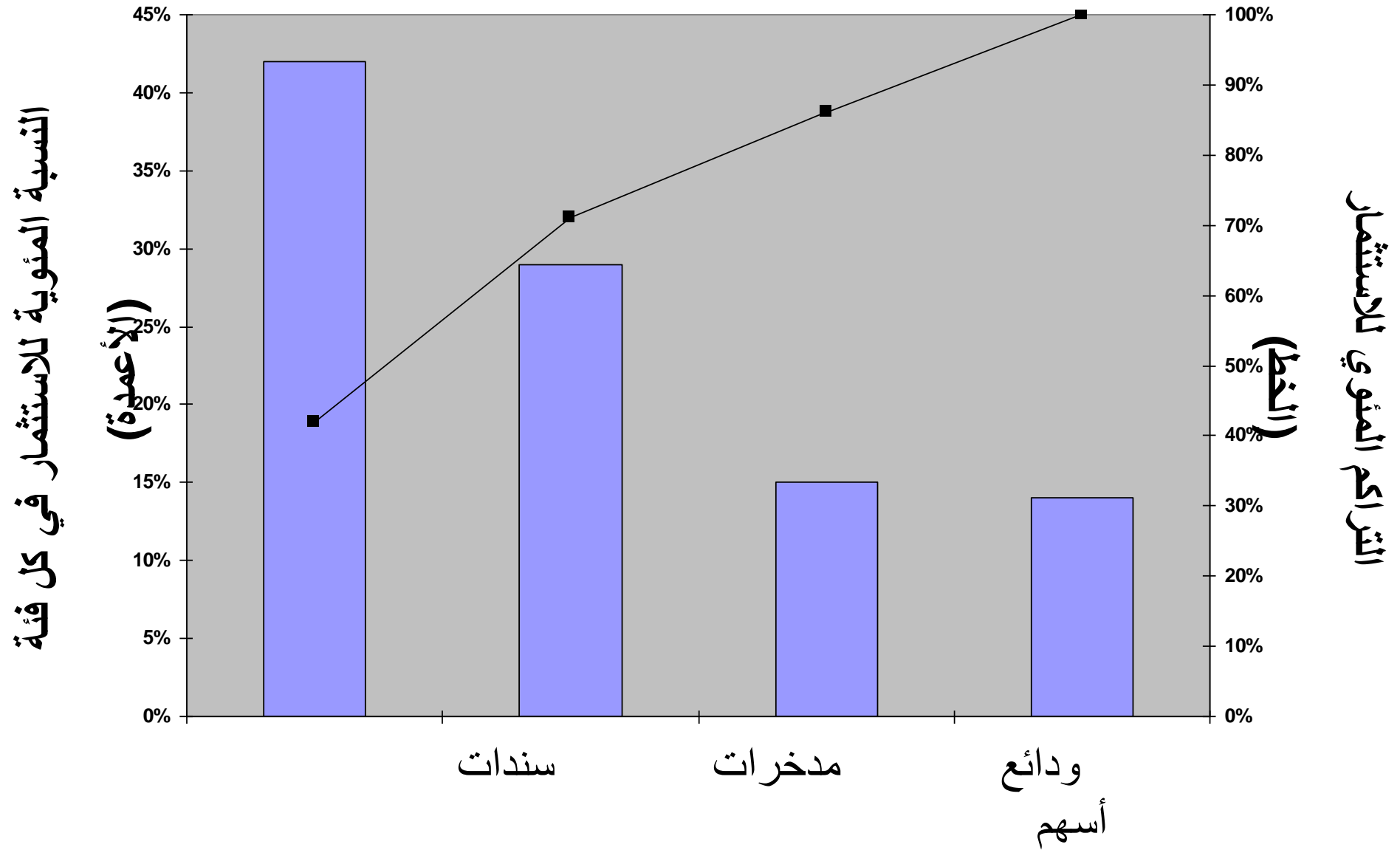


مثال: مخطط BAR

حافطة المستثمر

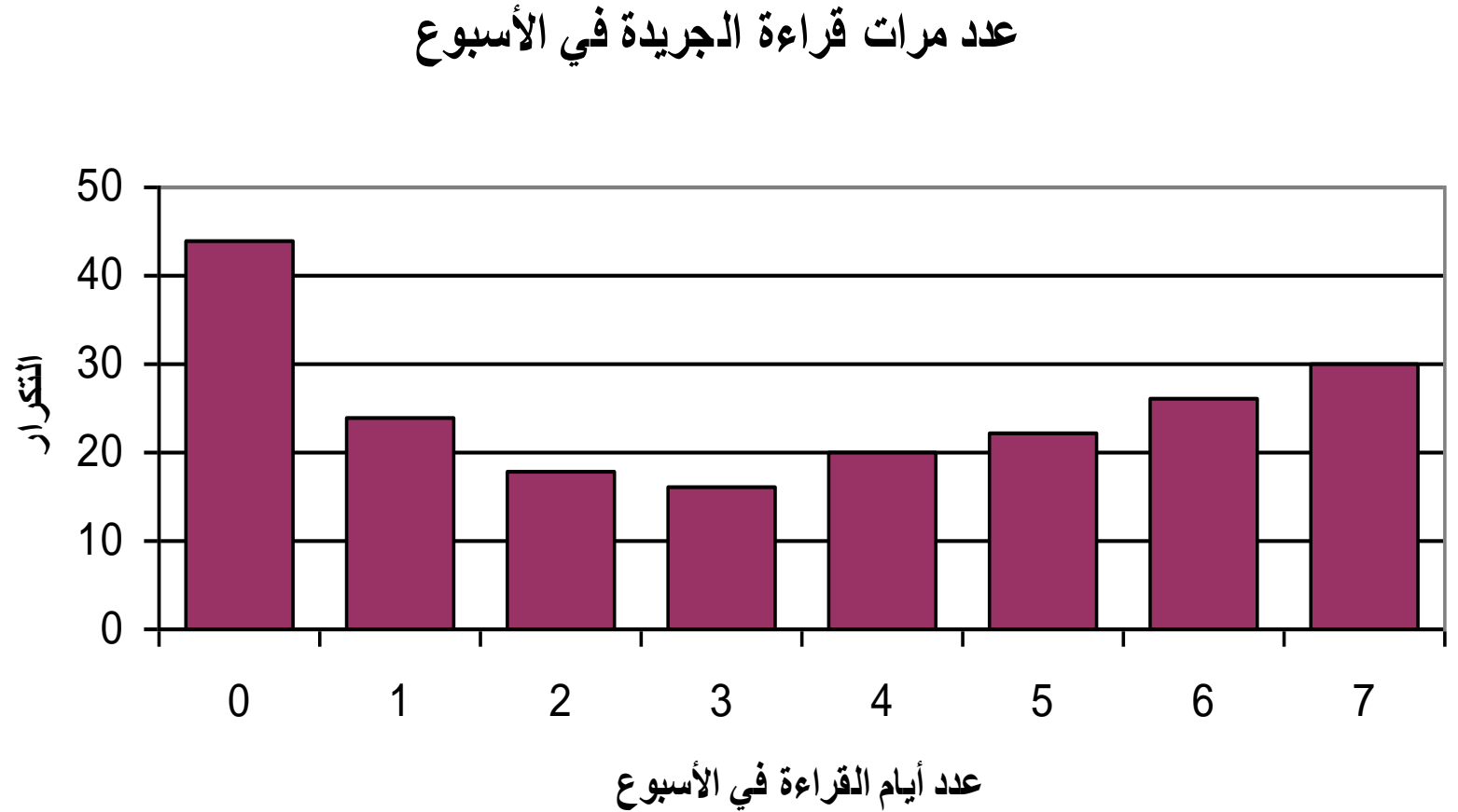


مثال: مخطط باريتو



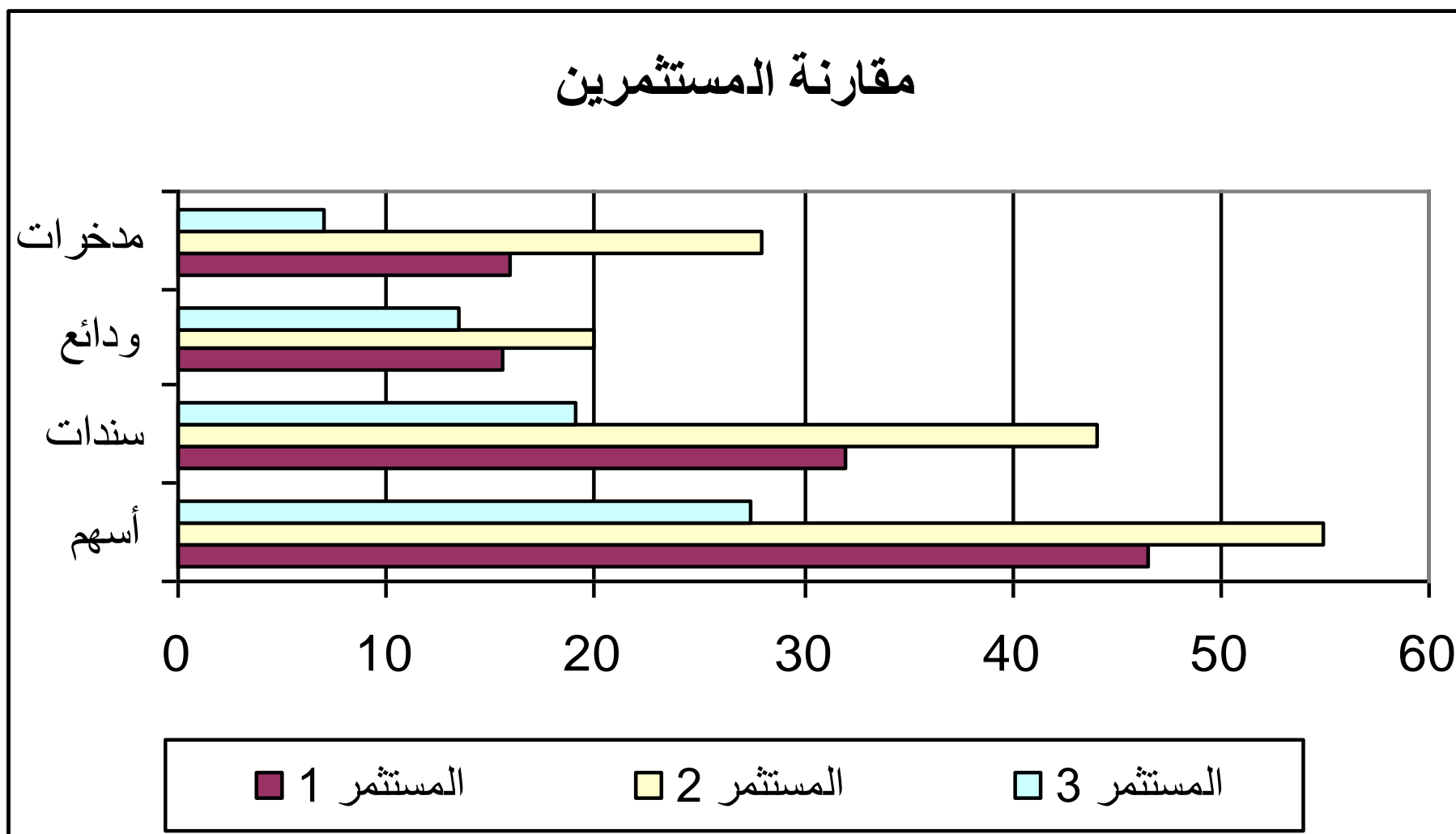
مثال: مخطط أعمدة بسيط

عدد الأيام	التكرارات
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
الإجمالي	200

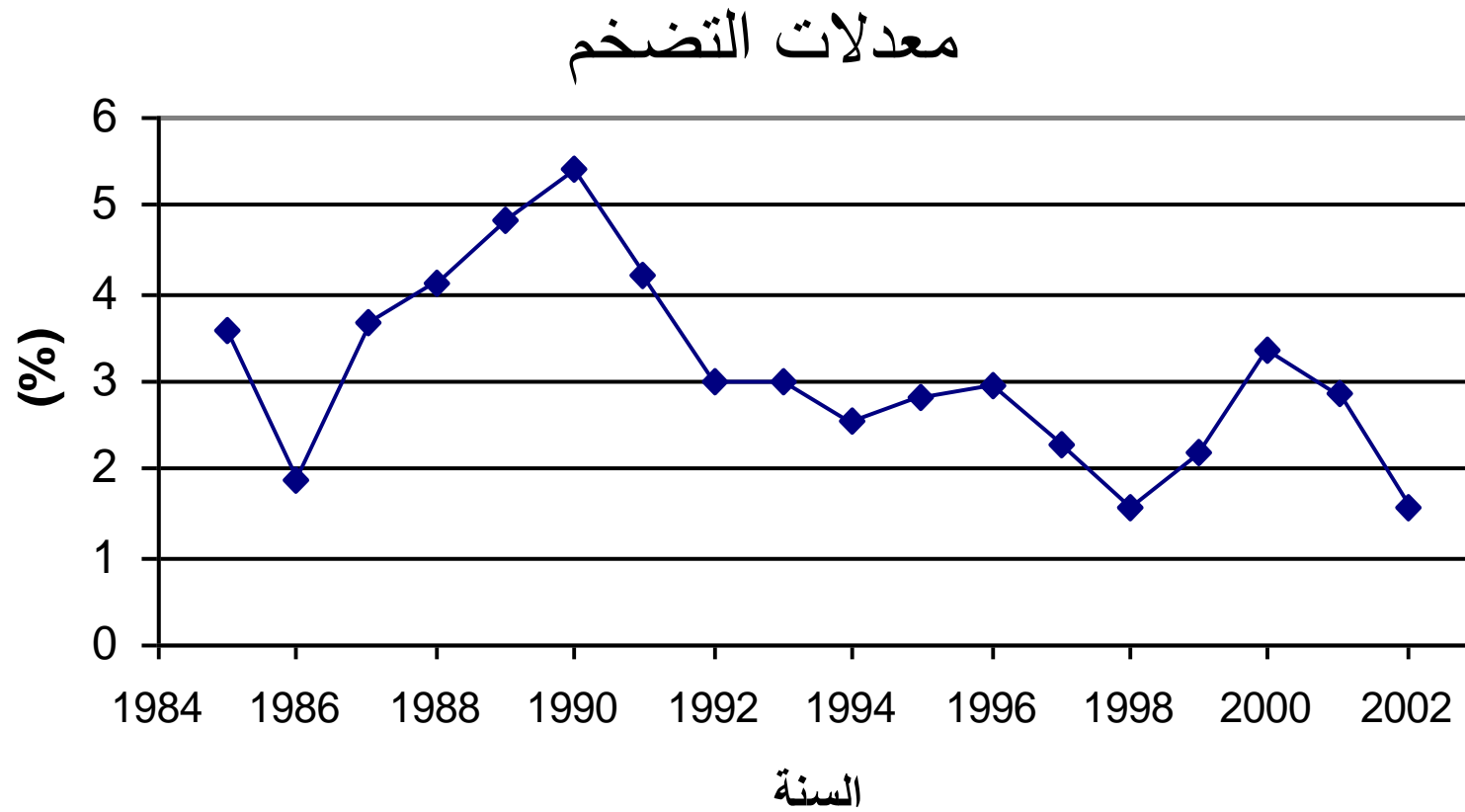


مثال: مخطط Multiple Bar

□ "سايد باي سايد تشارت"



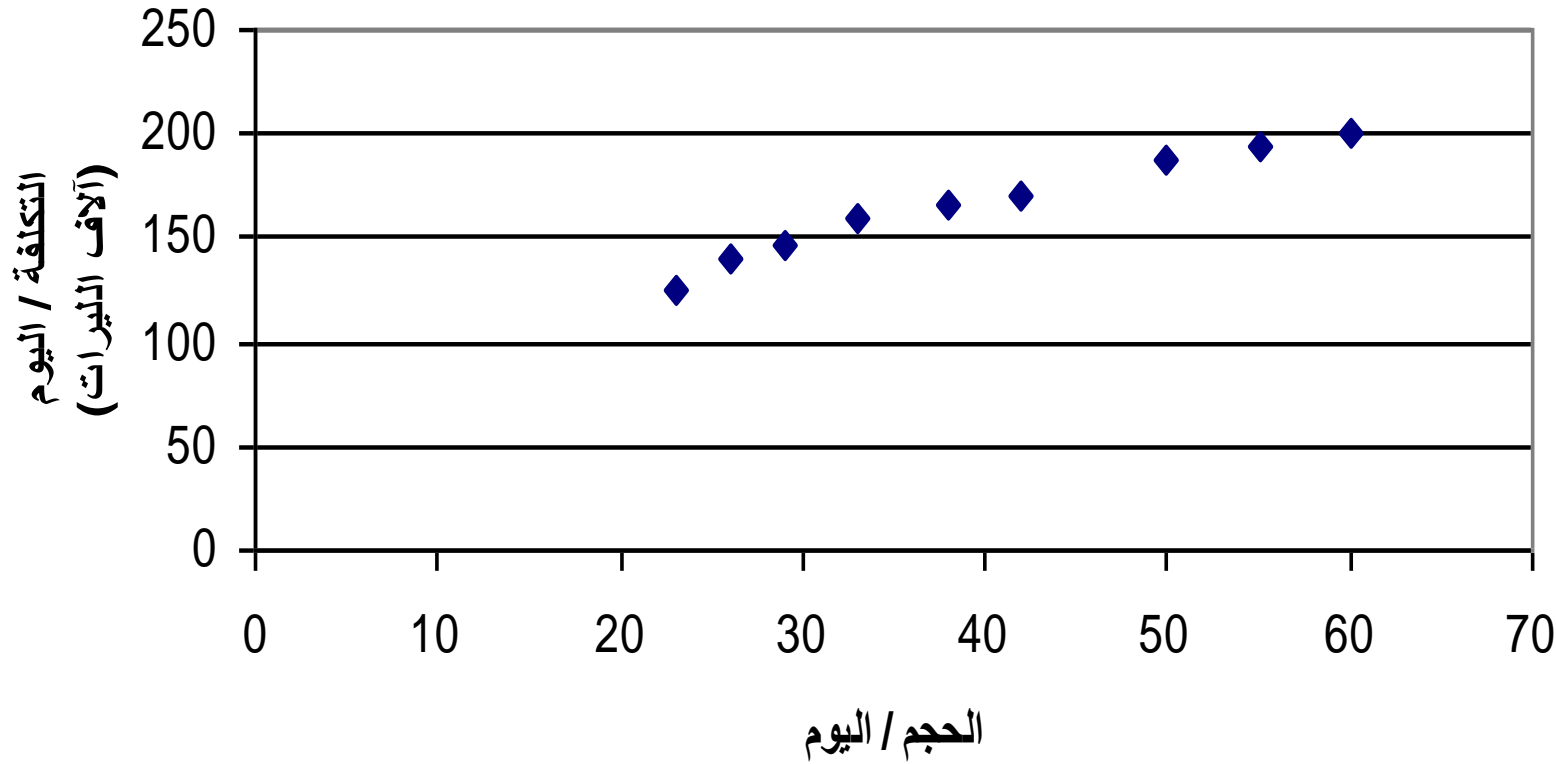
مثال: مخطط خطي



السنة	معدل التضخم
1985	3.56
1986	1.86
1987	3.65
1988	4.14
1989	4.82
1990	5.40
1991	4.21
1992	3.01
1993	2.99
1994	2.56
1995	2.83
1996	2.95
1997	2.29
1998	1.56
1999	2.21
2000	3.36
2001	2.85
2002	1.58

مثال: مخط تبعثر

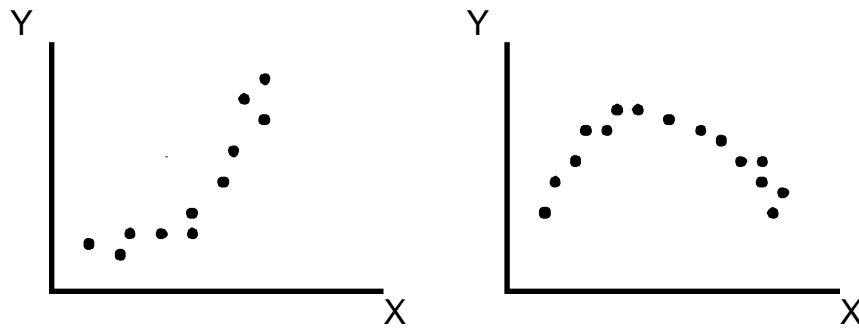
الحجم اليومي للإنتاج إزاء التكلفة اليومية المرافقة



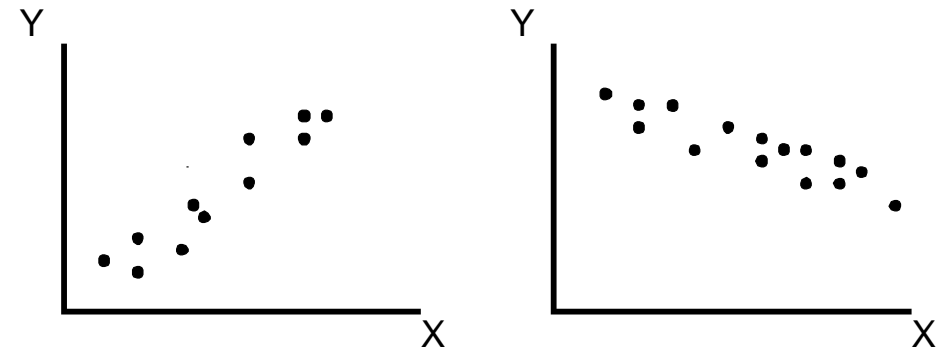
الحجم اليومي للإنتاج	التكلفة اليومية
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

أنواع العلاقات:

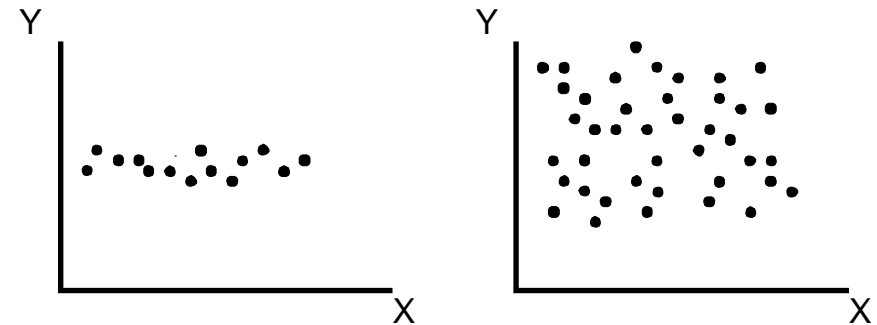
العلاقات "الانحناخية"



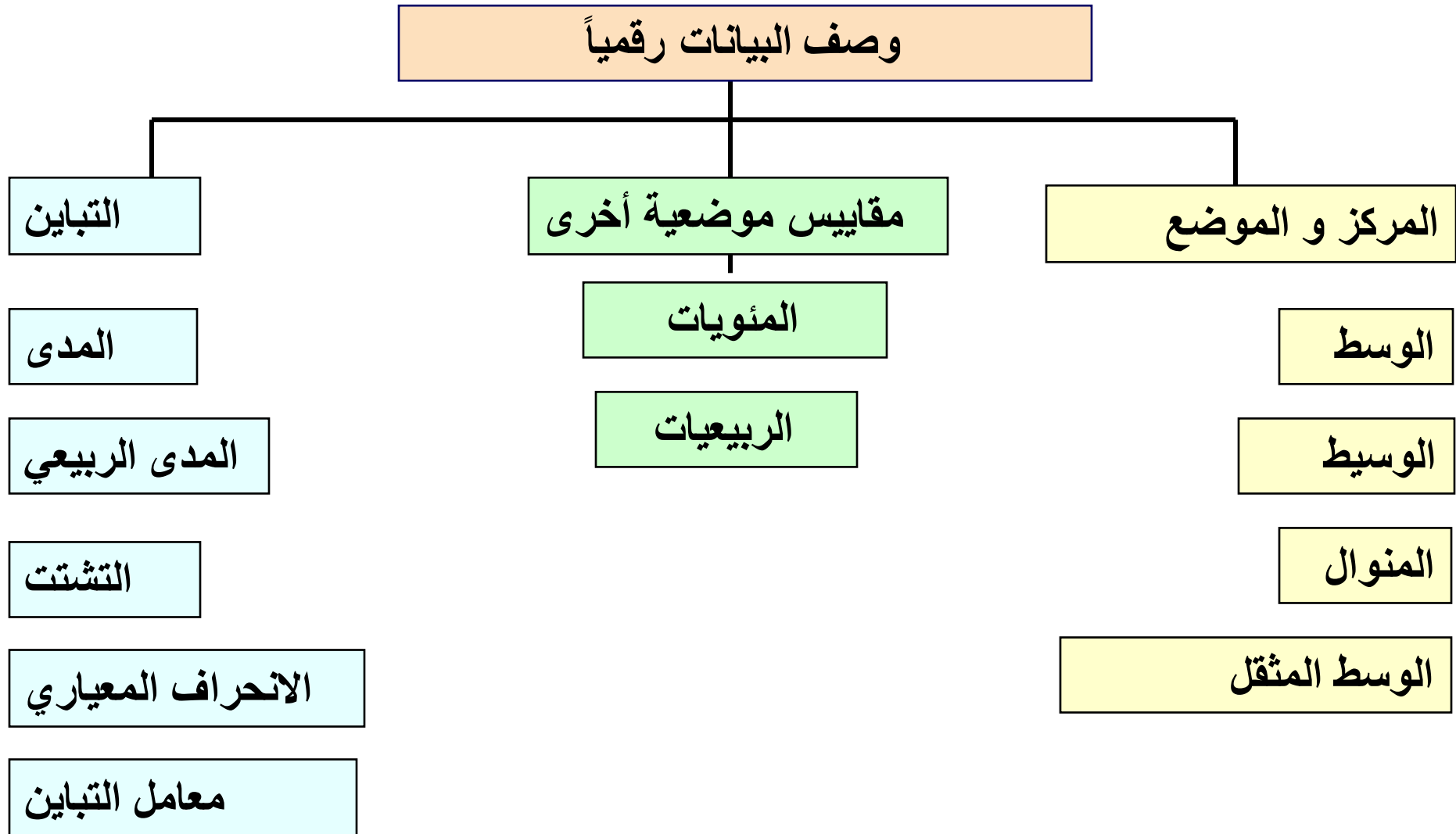
العلاقات الخطية



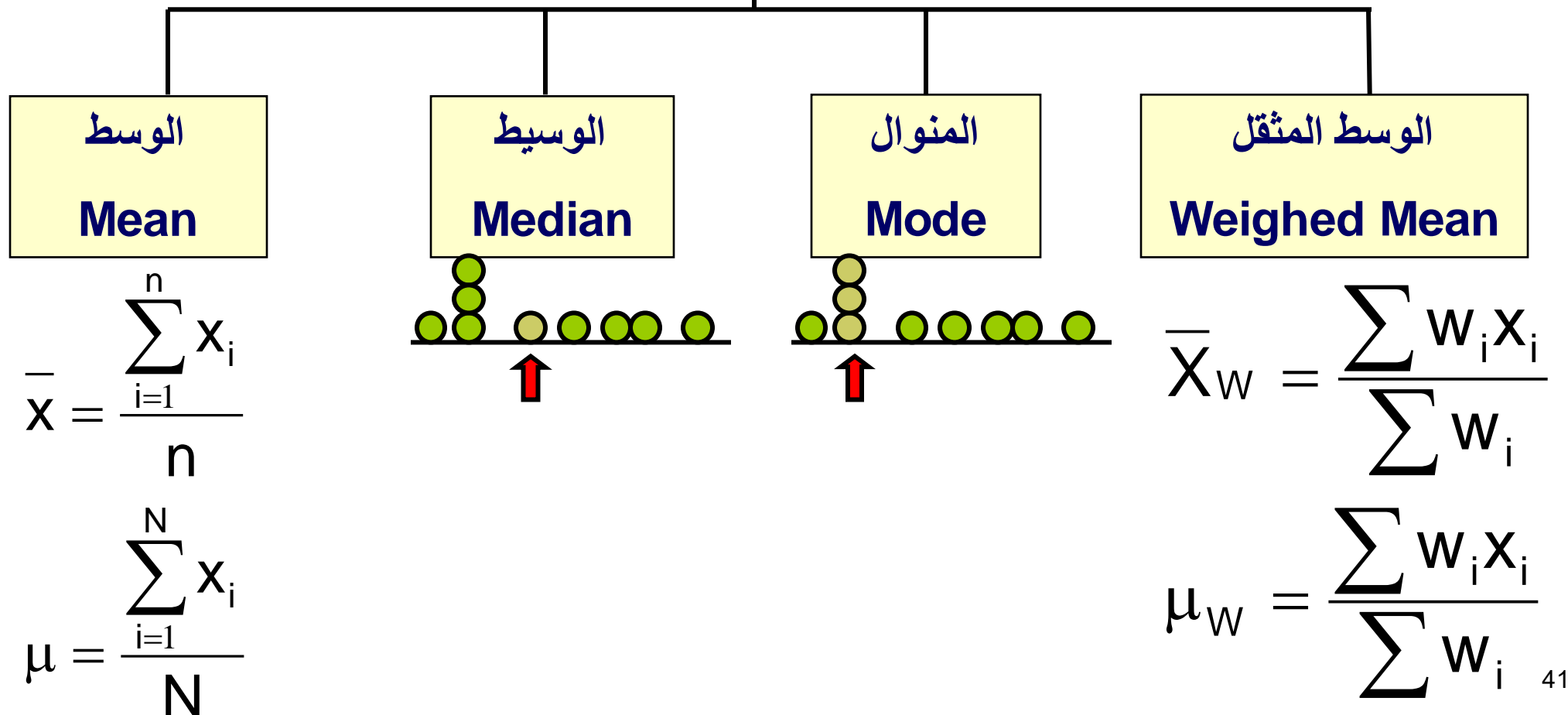
انعدام العلاقة



الوصف الإحصائي للبيانات



المركز و الموقع



الوسط (الوسط الحسابي) 1

□ **الوسط** هو ببساطة الوسط الحسابي لمجموعة من البيانات!

■ **وسط العينة**

حجم العينة
 $n =$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

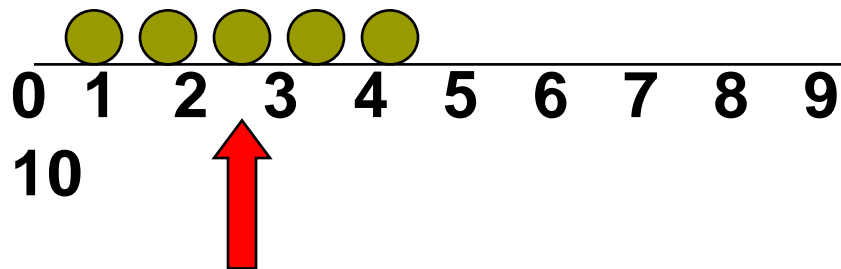
■ **وسط المجتمع**

حجم المجتمع
 $N =$

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

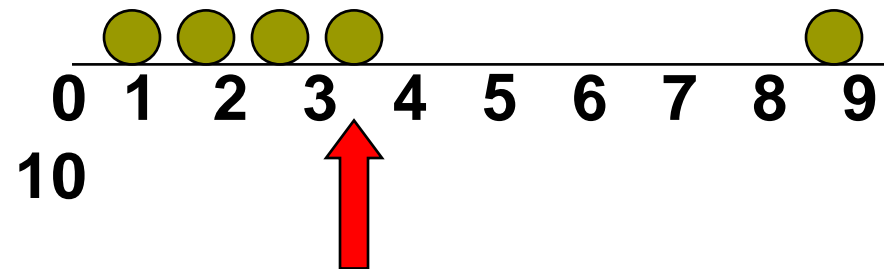
الوسط (الوسط الحسابي) 2

- الأكثر شيوعاً بين مقاييس النزعة المركزية.
- يساوي مجموع القيم مقسوماً على عددها.
- يتأثر بالقيم الشاذة.



الوسط = 3

$$\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$



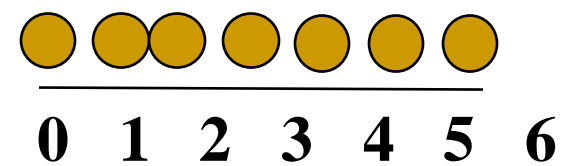
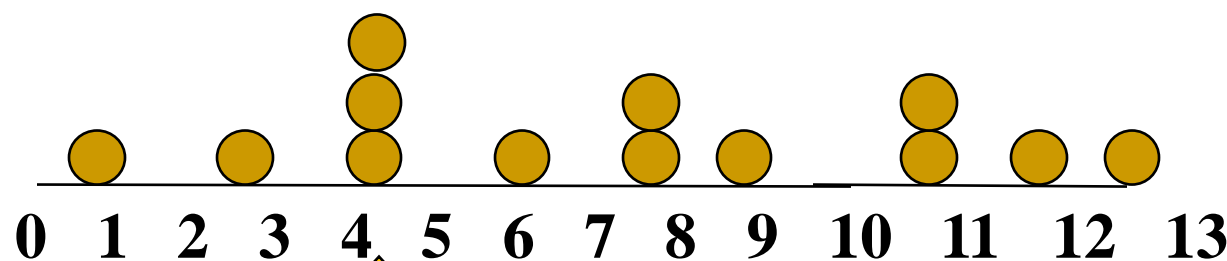
الوسط = 4

$$\frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

Mode

المنوال

- هو مقياس للنزعة المركزية.
- يمثل القيمة **الأكثر ظهوراً** (حدوثاً أو مشاهدةً)
- لا يتأثر بالقيم الشاذة
- يمكن استخدامه للبيانات الرقمية و الفئوية.
- لا يشترط وجوده بالضرورة.
- و قد يوجد من أكثر من واحد.



14
المنوال = 5

لا يوجد منوال

شكل التوزيع وشكل مخطط بوكس و فيكسر

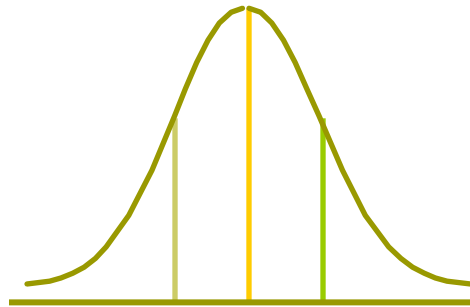
ملتو يساراً



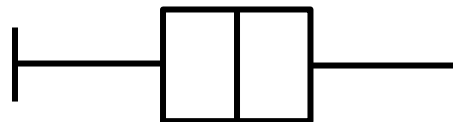
Q1 Q2 Q3



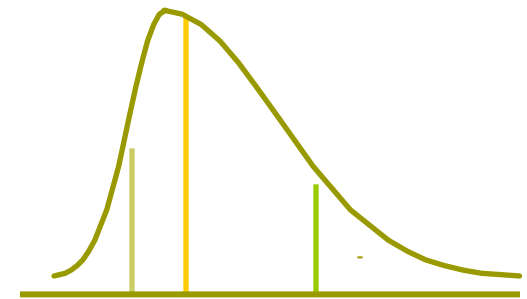
متناظر



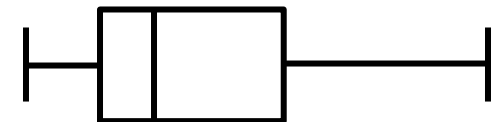
Q1 Q2 Q3



ملتو يميناً

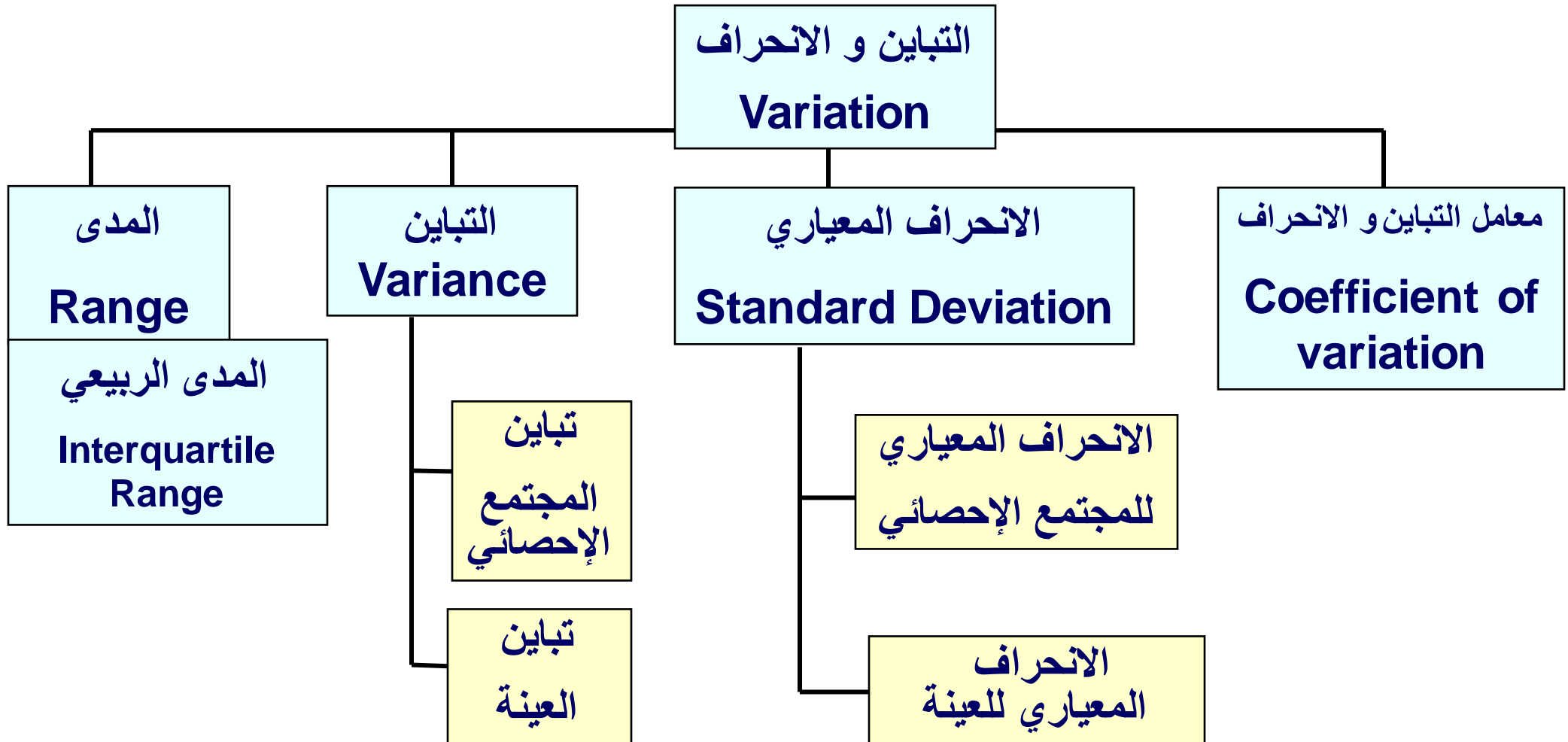


Q1 Q2 Q3



Variation Measures

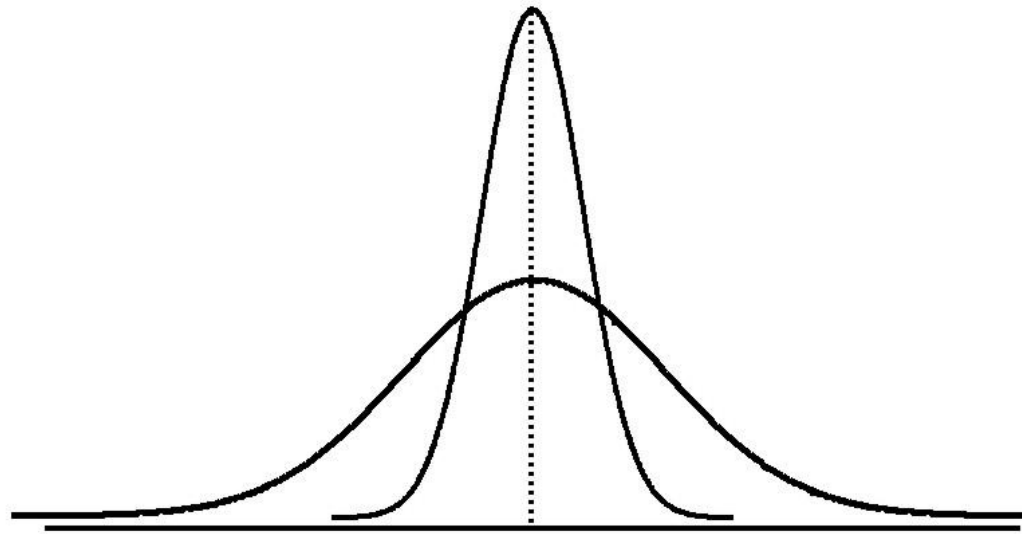
مقاييس التباين و الانحراف



Variance

التباين

□ يعطي معلومات عن انتشار و تباين البيانات



المركز هو نفسه

أما التباين فهو مختلف

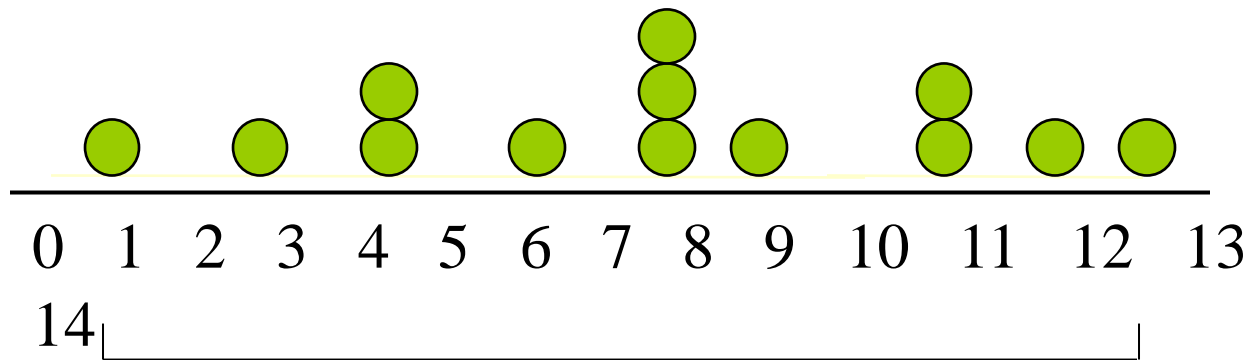
Range

المدى

□ أبسط مقاييس الانتشار و التباين و يعرف بأنه الفرق بين المشاهدة ذات القيمة العليا و المشاهدة ذات القيمة الدنيا

$$\text{Range} = X_{\text{maximum}} - X_{\text{minimum}}$$

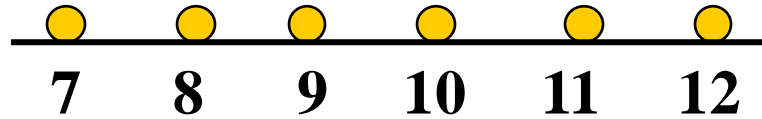
المدى = القيمة العليا - القيمة الدنيا



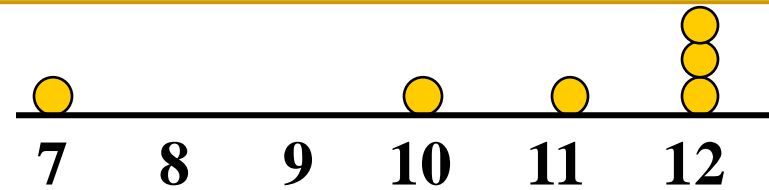
$$\text{Range} = 14 - 1 = 13$$

مثالب مقياس المدى

□ يتجاهل توزيع البيانات:



$$\text{Range} = 12 - 7 = 5$$



$$\text{Range} = 12 - 7 = 5$$

□ حساس للقيم المتطرفة:

1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 5

$$\text{Range} = 5 - 1 = 4$$

1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 120

$$\text{Range} = 120 - 1 = 119$$

التباين

□ معدل مربعات الانحرافات عن الوسيط

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

■ تباين العينة:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

■ تباين المجتمع:

الانحراف المعياري

- أكثر مقاييس التباين استخداماً.
- يظهر التباينات حول الوسيط.
- له نفس وحدات البيانات الأصلية.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

■ الانحراف المعياري للعينة:

■ الانحراف المعياري للمجتمع:

مثال حسابي: الانحراف المعياري للعينة

بيانات

العينة (X_i):

10 12 14 15 17 18 18 24

$n = 8$

الوسط = $\bar{x} = 16$

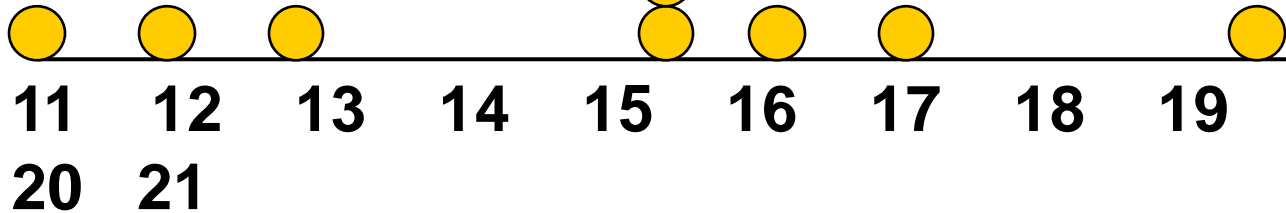
$$s = \sqrt{\frac{(10 - \bar{x})^2 + (12 - \bar{x})^2 + (14 - \bar{x})^2 + \dots + (24 - \bar{x})^2}{n - 1}}$$

$$= \sqrt{\frac{(10 - 16)^2 + (12 - 16)^2 + (14 - 16)^2 + \dots + (24 - 16)^2}{8 - 1}}$$

$$= \sqrt{\frac{126}{7}} = 4.2426$$

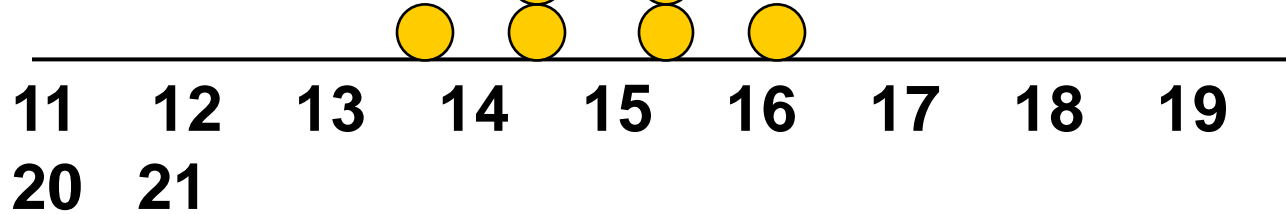
مقارنة الانحرافات المعيارية

البيانات أ



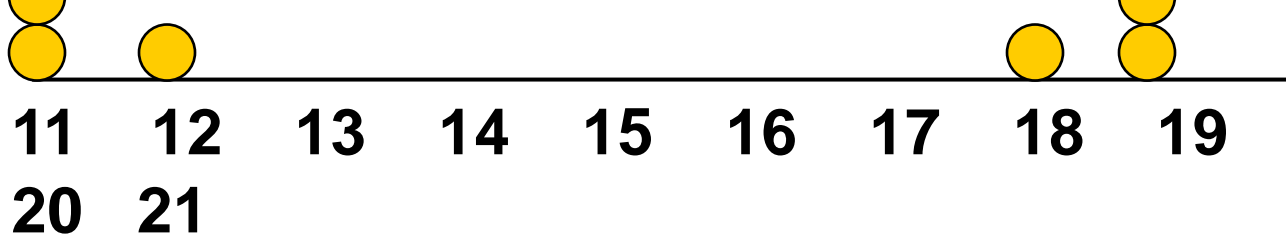
Mean = 15.5
S = 3.338

البيانات ب



Mean = 15.5
S = .9258

البيانات ج



Mean = 15.5
S = 4.57

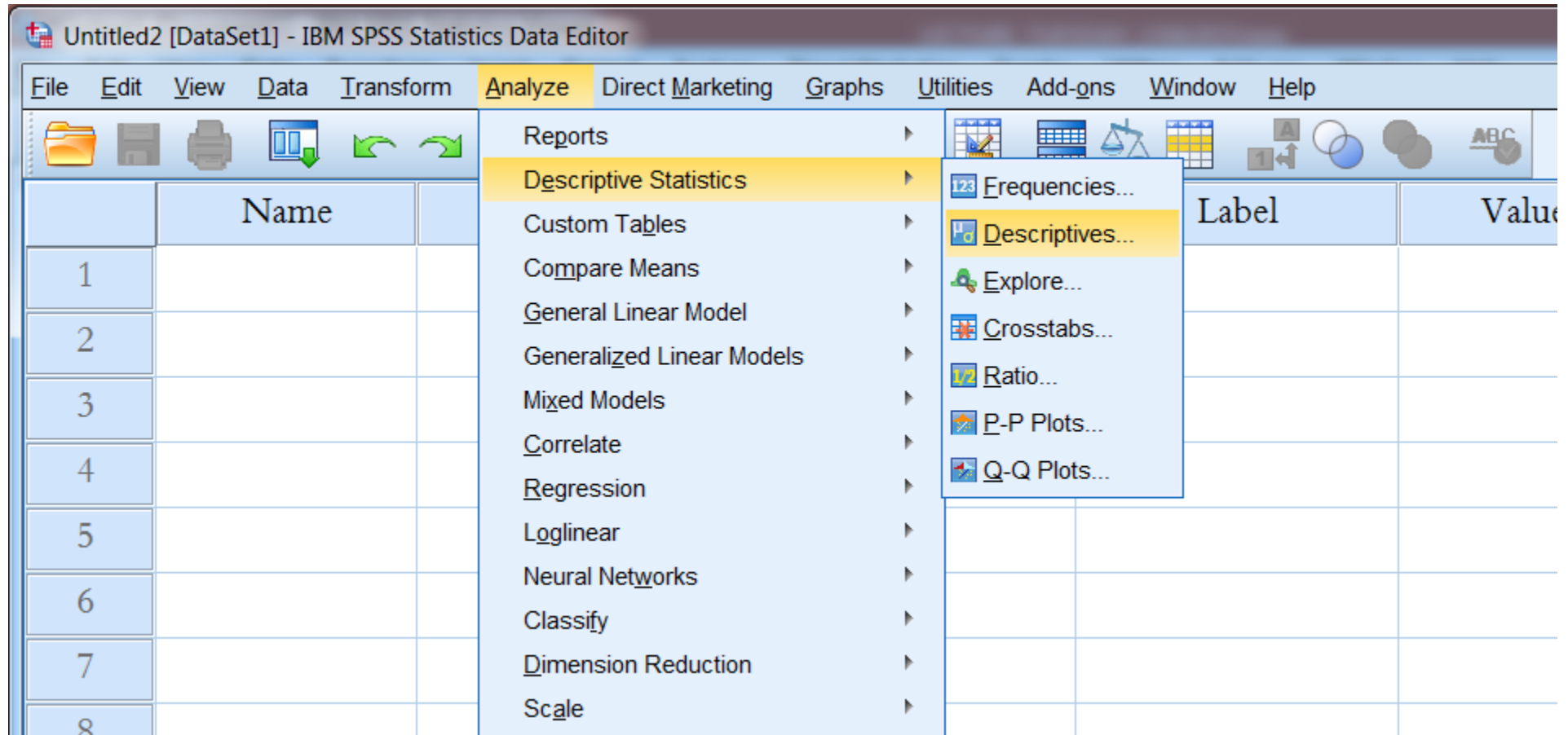
توليد بيانات الإحصاء الوصفي (Excel)

The image shows a screenshot of the Microsoft Excel interface. The title bar indicates the file is 'Book1 - Excel' and the user is 'Mouaaz Sherfawi'. The 'Data' tab is active in the ribbon. A red arrow points to the 'Data Analysis' button in the 'Analyze' group of the ribbon. Below the ribbon, the 'Data Analysis' task pane is open, displaying a list of analysis tools. A red arrow points to the 'Descriptive Statistics' option, which is currently selected. The list of tools includes: Correlation, Covariance, Descriptive Statistics, Exponential Smoothing, F-Test Two-Sample for Variances, Fourier Analysis, Histogram, Moving Average, Random Number Generation, and Rank and Percentile. The 'OK', 'Cancel', and 'Help' buttons are visible on the right side of the task pane.

توليد بيانات الإحصاء الوصفي (Excel)

	income_2010	income_2011	income_2012	income_2013	income_2014
Mean	45231.35	2548177.92	2549278.73	58774.60	65099.65
Standard Error	3126.61	2498767.68	2498739.79	5924.65	7165.82
Median	42229.15	44187.82	46426.56	53448.25	54831.23
Standard Deviation	19774.42	15803594.43	15803418.04	37470.79	45320.63
Kurtosis	-0.85	40.00	40.00	5.83	4.62
Skewness	0.08	6.32	6.32	1.89	1.93
Range	74239.31	99992834.78	99993034.88	202930.18	224610.67
Minimum	6072.4	7162.22	6963.12	6914.7	7283.2
Maximum	80311.71	99999997.00	99999998.00	209844.88	231893.87
Count	40	40	40	40	40

توليد بيانات الإحصاء الوصفي (SPSS)



توليد بيانات الإحصاء الوصفي (SPSS)

Descriptive Statistics

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Gross total income over 2010	40	\$45,231.3520	19774.42346	.077	.374	-.847	.733
Gross total income over 2011	40	2548177.922	15803594.43	6.325	.374	40.000	.733
Gross total income over 2012	40	2549278.732	15803418.04	6.325	.374	40.000	.733
Gross total income over 2013	40	\$58,774.6030	37470.78514	1.890	.374	5.830	.733
Gross total income over 2014	40	\$65,099.6463	45320.63410	1.926	.374	4.615	.733
Valid N (listwise)	40						

Descriptive Statistics

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Gross total income over 2010	40	\$45,231.3520	19774.42346	.077	.374	-.847	.733
Gross total income over 2011	40	2548177.922	15803594.43	6.325	.374	40.000	.733
Gross total income over 2012	40	2549278.732	15803418.04	6.325	.374	40.000	.733
Gross total income over 2013	40	\$58,774.6030	37470.78514	1.890	.374	5.830	.733
Gross total income over 2014	40	\$65,099.6463	45320.63410	1.926	.374	4.615	.733
Valid N (listwise)	40						

توليد بيانات الإحصاء الوصفي (Stata)

The screenshot displays the Stata software interface. The 'Statistics' menu is open, and the path to 'Summary statistics' is highlighted with red circles and arrows. The 'Summary statistics' option in the final submenu is also circled in red.

- Statistics
- User
- Window
- Help
- Summaries, tables, and tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Summary and descriptive statistics
- Frequency tables
- Other tables
- Classical tests of hypotheses
- Nonparametric tests of hypotheses
- Distributional plots and tests
- Multivariate test of means, covariances, and normality
- Summary statistics
- Means
- Proportions
- Ratios
- Totals
- Pairwise comparisons of means
- Confidence intervals
- Normal mean CI calculator
- Poisson mean CI calculator
- Proportion CI calculator

توليد بيانات الإحصاء الوصفي (Stata)

```
. sum income_2010 income_2011 income_2012 income_2013 income_2014
```

Variable	Obs	Mean	Std. Dev.	Min	Max
income_2010	40	45231.35	19774.42	6072.4	80311.71
income_2011	40	2548178	1.58e+07	7162.22	1.00e+08
income_2012	40	2549279	1.58e+07	6963.12	1.00e+08
income_2013	40	58774.6	37470.79	6914.7	209844.9
income_2014	40	65099.65	45320.63	7283.2	231893.9