

صحافة البيانات – أساسيات في التعامل مع الأرقام

نسبة التغير في قيمة معينة = ((القيمة الجديدة – القيمة القديمة) / القيمة القديمة) × 100

مثال: إذا كان عدد الطلاب المقبولين في دائرة الصحافة لعام 2016 هو 65 وأصبح 72 عام 2017 يمكننا القول إن القبول في الدائرة ارتفع أو زاد بنسبة 10.8% كما يلي:

$$\text{التغير} = 72 - 65 = 7$$

النسبة = $\% = (7 \div 65) \times 100 = 10.769\%$ أو 10.8% بالتقريب لمنزلة عشرية واحدة.

أما لو انخفض العدد عام 2017 إلى 55 فنقول إن القبول انخفض أو نقص بنسبة 15.4% (يرجى التحقق).

مؤشرات إحصائية: يمكن استخدام المقاييس الإحصائية التالية لتلخيص قائمة بيانات. لنفترض أن الجدول التالي (ورقة إكسل) هو للطول والكتلة لعشرة أشخاص. هناك خمسة مقاييس إحصائية تلخص الجدول وهي:

القيمة العظمى أو أكبر قيمة (Max)، القيمة الدنيا أو أقل قيمة (Min) وهاتان القيمتان تحددان مدى البيانات (تتراوح من ... إلى ...)، المعدل أو المتوسط الحسابي (Average) وهو مجموع القيم مقسوماً على عددها، الوسيط (Median) وهو القيمة التي تقع في الوسط إذا رتبنا القيم تصاعدياً أو تنازلياً (أي أن نصف القيم أقل من الوسيط والنصف الآخر أكبر منه)، والانحراف المعياري (STDEV.P) وهو مقياس لتشتت القيم (بعدها عن المتوسط الحسابي). ويوضح الجدول كيفية حساب كل منها للطول والكتلة (الوزن).

	A	B	C	D	E	F	G	H
1		الاسم	الطول (سم)	الوزن (كغم)	الجنس		Height	Weight
2		الشخص الأول	170	58	أنثى	Max	=MAX(B2:B11)	=MAX(C2:C11)
3		الشخص الثاني	174	65	ذكر	Min	=MIN(B2:B11)	=MIN(C2:C11)
4		الشخص الثالث	180	85	ذكر	Average	=AVERAGE(B2:B11)	=AVERAGE(C2:C11)
5		الشخص الرابع	172	79	ذكر	Median	=MEDIAN(B2:B11)	=MEDIAN(C2:C11)
6		الشخص الخامس	165	70	أنثى	St.Dev	=STDEV.P(B2:B11)	=STDEV.P(C2:C11)
7		الشخص السادس	155	56	أنثى			
8		الشخص السابع	181	78	ذكر	Correlation	=CORREL(B2:B11,C2:C11)	
9		الشخص الثامن	168	74	أنثى			
10		الشخص التاسع	178	86	ذكر			
11		الشخص العاشر	185	106	ذكر			
12								

ملاحظة 1: إذا كان عدد القيم زوجياً، كما في هذه الحالة، يؤخذ متوسط القيمتين في الوسط (بعد الترتيب)، أي مجموع القيمتين مقسوماً على 2.

ملاحظة 2: إذا كان هناك فرق واضح بين المتوسط الحسابي والوسيط فإن ذلك يدل على وجود قيم متطرفة في قائمة البيانات

القاعدة الذهبية للانحراف المعياري: في التوزيع الطبيعي (مثال: معدلات التوجيهي) يقع 68% من القيم تقريباً ضمن انحراف معياري واحد عن المتوسط الحسابي (أكبر أو أقل) و 95% من القيم ضمن انحرافين معياريين و 99.7% (أي معظمها) ضمن ثلاثة انحرافات معيارية عن المتوسط الحسابي.

مثال: كان المتوسط الحسابي لاختبار التوظيف لعام 2017 هو 54.3 والانحراف المعياري 13.5 وعليه فإن 99.7% (تقريباً) من المتقدمين حصلوا على علامة ما بين: $54.3 - 3 \times 13.5$ و $54.3 + 3 \times 13.5$ أي ما بين 13.8 و 94.8 وإذا كان عدد المتقدمين 39326 فهذه النسبة = 0.997×39326 أي 39208 (الرقم الحقيقي حسب البيانات هو 39313)

الارتباط (العلاقة) بين قائمتي بيانات (Correlation): يحدد كيفية وكمية تغير قيم القائمة الثانية بالنسبة للأولى (طردى ما بين صفر وواحد أو عكسي ما بين -1 وصفر). في المثال علاقة الوزن مع الطول طردية بنسبة 80% أي قوية نوعاً ما. والارتباط لا يعني بالضرورة السببية (الطول ليس بالضرورة سبباً لزيادة الوزن).

ملاحظة: تم إظهار الصيغ (المعادلات) في الجدول أعلاه بدل القيم الناتجة باستخدام خاصية إظهار الصيغ (Menu: Formulas -> Show Formulas)

للمزيد حول صحافة البيانات: <https://github.com/abedkhooli/ddj>