

تكليف البحث عن المصطلحات مادة تتقّب بيانات

أ/نورا

حنين القباطي

CS4_AM / 21_16_0076

1. Skewness

مصطلح "الانحراف" أو "التشويه" هو مفهوم إحصائي يستخدم لتوصيف توزيع البيانات في مجموعة من القيم. هذا المصطلح يقيس درجة انحراف التوزيع عن التوزيع الطبيعي (المتماثل)، وبالتالي يعطي فكرة عن شكل الانحراف أو التوجه في البيانات.

يعتبر التوزيع الطبيعي متماثلاً حيث يكون لديه القيم الأعلى والأقل وسيطين متساويين ومتماثلين حول المتوسط. ومع ذلك، يمكن أن تحدث تشوهات في التوزيع تجاه اليمين أو اليسار، وهذا ما يُمثله مصطلح الانحراف.

إذا كانت قيمة الانحراف أكبر من الصفر، فإن ذلك يعني وجود توجه في التوزيع نحو اليمين (تشوه إيجابي)، بينما إذا كانت قيمة الانحراف أقل من الصفر، فإن ذلك يشير إلى وجود توجه في التوزيع نحو اليسار (تشوه سلبي). وعندما تكون قيمة الانحراف قريبة من الصفر، فإن ذلك يعني أن التوزيع متقارب للتوزيع الطبيعي.

تُستخدم قيمة الانحراف في التحليل الإحصائي لتقدير شكل التوزيع وتحديد مدى انحرافه عن التوزيع الطبيعي. يتم استخدام عدة طرق لحساب قيمة الانحراف، وأحد أكثر الطرق شيوعاً هو استخدام الانحراف المعياري والمتوسط الحسابي.

إنه من الممكن استخدام قيمة الانحراف في تنقيب البيانات (Data Mining)، حيث قيمة الانحراف يمكن أن تكون ضمن المعلومات التي يُمكن استخدامها في هذا السياق.

هذه بعض الطرق التي يُمكن استخدام قيمة الانحراف في تنقيب البيانات:

- اكتشاف الأنماط والتوجهات.
- الفصل بين البيانات.
- تحديد القيم الشاذة.
- الاستنتاجات الإحصائية.

يوجد طريقتان لحساب الانحراف القياسي هما:

الطريقة الأولى:

١. قم بحساب المتوسط الحسابي لمجموعة البيانات.
٢. احسب فروق كل قيمة في المجموعة عن المتوسط.
٣. اربع الفروق.
٤. احسب المتوسط الحسابي للأعداد المربعة.
٥. خذ الجذر التربيعي للمتوسط الحسابي السابق للحصول على الانحراف القياسي.

الطريقة الثانية (الأكثر شيوعاً):

١. قم بحساب المتوسط الحسابي لمجموعة البيانات.
٢. احسب فروق كل قيمة في المجموعة عن المتوسط.
٣. اربع الفروق.
٤. قسم مجموع الأعداد المربعة على عدد القيم في المجموعة.
٥. خذ الجذر التربيعي للناتج للحصول على الانحراف القياسي.

ما يلي مثال لحساب الانحراف القياسي:

نستخدم مجموعة الأعمار هذه (٢٠، ٢٢، ٢٥، ٢٨، ٣٠) لحساب الانحراف القياسي:

١. حساب المتوسط الحسابي:

$$\text{متوسط} = 0 / (20 + 22 + 25 + 28 + 30) = 25$$

٢. حساب الفروق:

$$\text{الفروق} = (20 - 25, 22 - 25, 25 - 25, 28 - 25, 30 - 25) = (-5, -3, 0, 3, 5)$$

٣. اربع الفروق:

$$\text{الأعداد المربعة} = (25, 9, 0, 9, 25)$$

٤. حساب المتوسط الحسابي للأعداد المربعة:

$$\text{متوسط الأعداد المربعة} = (25 + 9 + 0 + 9 + 25) / 5 = 13,6$$

٥. خذ الجذر التربيعي:

$$\text{الانحراف القياسي} = \sqrt{13,6} \approx 3,68$$

إذاً الانحراف القياسي لمجموعة الأعمار هو تقريباً ٣,٦٨.

يُلاحظ أن الانحراف القياسي يعكس قدرة البيانات على التشتت حول المتوسط. إذا كانت القيمة أعلى، فهذا يشير إلى وجود تشتت أكبر في البيانات، في حين أن القيمة المنخفضة تشير إلى تجمع البيانات بشكل أكبر حول المتوسط.

2. Variance

كلمة "variance" هي مصطلح "التباين" أو "التشتت". في الإحصاء وعلم الاحتمالات، حيث يُستخدم لقياس مدى تشتت القيم حول المتوسط الحسابي في مجموعة من البيانات.

التباين يتم حسابه عن طريق مقارنة كل قيمة في مجموعة البيانات مع المتوسط الحسابي، ثم يتم رفع هذا الفرق إلى الأس الثاني وجمعه مع الفروق الأخرى وتقسيمها على عدد القيم في المجموعة. يتم استخدام الصيغة التالية لحساب التباين:

$$\text{التباين} = \frac{\sum (\text{قيمة} - \text{متوسط})^2}{\text{عدد القيم}}$$

حيث:

- \sum تمثل علامة المجموع لجميع القيم في المجموعة.
- قيمة هي القيمة الفردية في المجموعة.
- متوسط هو المتوسط الحسابي للمجموعة.
- عدد القيم يُمثل عدد القيم في المجموعة.

تستخدم قيمة التباين لعدة أغراض، بما في ذلك:

١. قياس تشتت البيانات.
٢. تحليل الاختلافات.
٣. حساب الانحراف المعياري.
٤. تحليل الاستقرار.

إن قيمة التباين تعتبر مقياساً لتشتت البيانات أو انحرافها عن المتوسط. يمكن استخدام قيمتها في تحليل البيانات وتنقيبها في العديد من الطرق، وهذه بعض الأمثلة:

١. تحليل الانحراف القياسي.
٢. اكتشاف القيم الشاذة.
٣. تقييم الاستقرار والتغير.
٤. تحديد النمط والترابط.

تختلف الطرق والتقنيات المستخدمة باختلاف المجال والغرض من تحليل البيانات.

ما يلي مثال بسيط لحساب التباين:

لنفترض أن لدينا مجموعة من الأعمار لخمس أشخاص: ٢٠، ٢٢، ٢٥، ٢٨، ٣٠.

أولاً، نحسب المتوسط الحسابي لهذه البيانات:

$$\text{متوسط} = 20 / (30 + 28 + 25 + 22 + 20) = 25$$

ثم، نستخدم هذا المتوسط لحساب التباين. نقوم بمربع الفروق بين كل قيمة والمتوسط، ثم نجمعها ونقسمها على عدد القيم في المجموعة:

$$\text{تباين} = 20 / ((28-20)^2 + 28(25-28)^2 + 28(25-25)^2 + 28(25-22)^2 + 28(25-20)^2)$$

$$= 20 / (28^2 + 28 \cdot 9 + 28 \cdot 0 + 28 \cdot 9 + 28 \cdot 25)$$

$$= 20 / (25 + 9 + 0 + 9 + 25)$$

$$= 20 / 68$$

$$= 13,6$$

إذاً التباين في هذا المثال هو ١٣,٦.

يُلاحظ أن قيمة التباين تعبر عن التشتت أو الانحراف عن المتوسط. في هذا المثال، القيم تتفاوت حول المتوسط بشكل نسبي، وهذا يعكسه قيمة التباين.

3. Deviation

إن مصطلح "deviation" في سياق الإحصاء والتحليل الاحتمالي مرتبطاً بمفهوم الانحراف عن المتوسط. يشير إلى مقدار الاختلاف بين قيمة معينة والمتوسط الحسابي لمجموعة البيانات.

يتم استخدام مصطلح "deviation" في عدة سياقات، وهذه بعض منها:

١. (الانحراف المعياري) Standard Deviation.
٢. (الانحراف المطلق) Mean Deviation.
٣. (نقاط الانحراف) Deviation Score.

إن قيمة deviation أو الانحراف يمكن استخدامها في تحليل البيانات وتنقيبها بطرق عدة. ما يلي بعض الأمثلة على كيفية استخدام قيمة deviation في تنقيب البيانات:

١. تحديد القيم الشاذة.
٢. تحليل الانحرافات المقارنة.
٣. تقييم الثبات والتغير.
٤. تحسين النماذج التنبؤية.

تعتمد الطريقة المناسبة على نوع المجموعة البيانات والأهداف المحددة للتحليل.

ما يلي مثال لكيفية حساب نقاط الانحراف:

نفترض أن لدينا مجموعة بيانات تتكون من درجات طلاب في اختبار الرياضيات، ونريد حساب نقاط الانحراف لكل طالب عن المتوسط العام للفصل.

١. نقوم بجمع جميع الدرجات للطلاب في الاختبار ونحسب المتوسط الحسابي للدرجات. ونفترض أن المتوسط العام هو ٨٠.
٢. نحسب فروق كل درجة طالب عن المتوسط العام. على سبيل المثال، إذا كانت درجة الطالب الأول ٨٥، فإن الفرق هو $٨٥ - ٨٠ = ٥$.
٣. نستمر في حساب فروق الدرجات عن المتوسط العام لجميع الطلاب في المجموعة.
٤. قد يكون لدينا الآن قائمة بالفروق لكل طالب. على سبيل المثال، قد يكون لدينا الفروق التالية: [٥، ٢، ٣، ٠، ١].

٥. يمكننا استخدام هذه الفروق لحساب نقاط الانحراف. ببساطة، يتم طرح كل درجة من المتوسط العام للحصول على نقطة الانحراف. في هذا المثال، ستكون نقاط الانحراف على النحو التالي:

$$[٨٥ - ٨٠, ٧٨ - ٨٠, ٨٣ - ٨٠, ٨٠ - ٨٠, ٨٠ - ٧٩] = [٥, ٢, ٣, ٠, ١]$$

باستخدام هذه العملية، يمكننا حساب نقاط الانحراف لكل طالب في المجموعة. هذه النقاط توضح كم يختلف أداء كل طالب عن المتوسط العام للفصل في اختبار الرياضيات. إذا كان لدى الطالب نقطة انحراف إيجابية، فهذا يعني أن أدائه أعلى من المتوسط، وإذا كانت لديه نقطة انحراف سلبية، فهذا يعني أن أدائه أقل من المتوسط.