ملخص الوحدة 10: تحليل البيانات

البرمجة والذكاء الاصطناعي، أولى ثانوي

مهندس رحيم

10-1 أنواع البيانات والتحليل

- **البيانات الكمية (Quantitative Data)**
  - **البيانات الفاصلة (Interval):** الفرق بين رقمين ليه معنى
  - **البيانات النسبية (Ratio):** فيها **صفر مطلق**
- **البيانات النوعية (Qualitative Data)**
  - **البيانات الاسمية (Nominal):** مجرد أسماء تفرّق بين الفئات
  - **البيانات الترتيبية (Ordinal): الترتيب ليه معنى** بس المسافات بين الدرجات لأ
- **إزاي تفرّق بسرعة؟**
  - اسأل: **ينفع أجمع الرقم أو أطرحه؟** لو لأ (زي رقم التليفون) يبقى **نوعي**.
  - نوعي: **فيه ترتيب؟** أيوه = ترتيبي
  - كمي: **الصفر معناه "مفيش"؟** أيوه = نسبي (0 كجم = مفيش وزن)
- **طرق تحليل البيانات**
  - **الإحصاء الوصفي (Descriptive Statistics):** بيلخّص وينظّم بيانات موجودة عشان نفهم شكل العينة.
  - **مقاييس النزعة المركزية:** **المتوسط
  - **مقاييس التشتت:** **المدى
  - **الإحصاء الاستدلالي (Inferential Statistics):** بنستنتج ونتنبأ ونعمّم على **المجتمع كله** من بيانات **عينة**

Quantitative data	البيانات الكمية: أرقام قابلة للقياس وحجمها ليه معنى
Qualitative data	البيانات النوعية: بتوصف صفة أو تصنيف
Nominal	البيانات الاسمية: فئات من غير ترتيب (فصيلة الدم)
Ordinal	البيانات الترتيبية: فئات ليها ترتيب (ترتيب الامتحان)
Interval	البيانات الفاصلة: أرقام من غير صفر مطلق (درجة الحرارة)
Ratio	البيانات النسبية: أرقام بصفر مطلق والنسب ليها معنى (الوزن)
Descriptive statistics	الإحصاء الوصفي: تلخيص وتنظيم البيانات الموجودة
Inferential statistics	الإحصاء الاستدلالي: استنتاج خصائص المجتمع من عينة
Measures of central tendency	مقاييس النزعة المركزية: المتوسط والوسيط والمنوال
Measures of spread	مقاييس التشتت: المدى والتباين والانحراف المعياري

10-2 تحليل البيانات [1]

- **جدول التوزيع التكراري (Frequency distribution table)**
  - **الفئة (Class):** المسافة نفسها
  - **عرض الفئة (Class width):** طول المسافة = الحد الأعلى − الحد الأدنى (19 − 16 = 3).
  - **علامة الفئة (Class mark):** منتصف الفئة = (الحد الأدنى + الحد الأعلى) ÷ 2
  - **التكرار (Frequency):** عدد القيم اللي وقعت في الفئة
- **المدرج التكراري (Histogram)**: هو جدول التوزيع التكراري مرسوم **أعمدة**: المحور الأفقي الفئات والرأسي التكرار
- **مقاييس النزعة المركزية**

- **المتوسط (Mean):** مجموع القيم ÷ عددها
- **الوسيط (Median):** رتّب البيانات **تصاعدي** وخد اللي في النص
- **المنوال (Mode):** القيمة **الأكتر تكراراً**

- **مثال محلول**
  - **المتوسط:** المجموع 84 ÷ 12 = **7**.
  - **الوسيط:** بعد الترتيب 3
  - **المنوال:** 6 اتكررت 3 مرات = **6**.

Frequency distribution table	جدول التوزيع التكراري: بيعدّ القيم في كل فئة
Class	الفئة: مسافة من مسافات الجدول
Class width	عرض الفئة: الحد الأعلى − الحد الأدنى
Class mark	علامة الفئة: منتصف الفئة
Frequency	التكرار: عدد القيم في الفئة
Histogram	المدرج التكراري: الجدول التكراري مرسوم أعمدة
Mean	المتوسط: المجموع ÷ العدد (AVERAGE)
Median	الوسيط: القيمة اللي في النص بعد الترتيب (MEDIAN)
Mode	المنوال: القيمة الأكتر تكراراً (MODE)

10-3 تحليل البيانات [2]

- **الرباعيات (Quartiles)**: أرقام بتقسم البيانات المرتبة تصاعدي لـ **4 أجزاء متساوية**: **الرباعي الأول (Q1)**
- **خطوات حساب الرباعيات**
  - **1):** رتّب البيانات **تصاعدي** وهات **الوسيط** (ده Q2).
  - **2):** قسّم البيانات عند الوسيط لنصّين: **النص الصغير** و**النص الكبير**.
  - **3):** **وسيط النص الصغير = Q1**
- **مخطط الصندوق والشوارب (Box-and-whisker plot)**
  - **1):** حدّد الخمس أرقام وارسم عند كل واحد خط صغير.
  - **2):** ارسم **صندوق** من **Q1** لـ **Q3**
  - **3):** ارسم **الشوارب**: خط من طرف الصندوق لـ **القيمة الدنيا** ومن الطرف التاني لـ **القيمة العليا**.
  - ممكن يترسم أفقي أو رأسي.
- **قراية المخطط**
  - **الصندوق:** فيه تقريباً **نص البيانات اللي في النص**
  - لو عندك **31 قيمة**: الوسيط هو القيمة رقم **16**
  - لو **الوسيط أقل من 60** يبقى **نص الطلبة على الأقل** جابوا 60 أو أقل.
  - **المدى = القيمة العليا − القيمة الدنيا:** (طول المخطط كله).

Quartile	الرباعي: أرقام بتقسم البيانات المرتبة لـ 4 أجزاء متساوية
First quartile	الرباعي الأول (Q1): وسيط النص الصغير
Second quartile	الرباعي الثاني (Q2): هو الوسيط
Third quartile	الرباعي الثالث (Q3): وسيط النص الكبير
Box-and-whisker plot	مخطط الصندوق والشوارب: يلخص الخمس أرقام في رسمة
Minimum value	القيمة الدنيا: أصغر قيمة (طرف الشارب)
Maximum value	القيمة العليا: أكبر قيمة (طرف الشارب التاني)

10-4 تحليل البيانات [3]

- **ليه التباين والانحراف المعياري؟**: فصلين ممكن يكون ليهم نفس **المتوسط**
- **التعريفات**
  - **الانحراف (Deviation):** القيمة – المتوسط
  - **التباين (Variance): متوسط مربعات الانحرافات**
  - **الانحراف المعياري (Standard deviation): الجذر التربيعي الموجب للتباين** s = √التباين
- **خطوات الحساب (اعمل جدول)**
  - **1):** احسب **المتوسط**.
  - **2):** لكل قيمة احسب **الانحراف** = القيمة – المتوسط.
  - **3):** **ربّع** كل انحراف (عشان السالب والموجب ميلغوش بعض).
  - **4): التباين** = متوسط المربعات (مجموعها ÷ عدد القيم).
  - **5): الانحراف المعياري** = √التباين.
- **مثال محلول**
  - المتوسط = 40 ÷ 5 = **8**.
  - الانحرافات: −4
  - التباين = 40 ÷ 5 = **8**

Deviation	الانحراف: القيمة – المتوسط
Variance	التباين: متوسط مربعات الانحرافات (VAR.P)
Standard deviation	الانحراف المعياري: الجذر التربيعي للتباين (STDEV.P)
Spread	التشتت: البيانات متفرّقة حوالين المتوسط قد إيه
VAR.P	دالة التباين في Excel (وفي Google Sheets اسمها VARP)
STDEV.P	دالة الانحراف المعياري في برامج جداول البيانات

10-5 تحليل البيانات [4]

- **المخطط المبعثر (Scatter plot)**
  - **ارتباط إيجابي (Positive):** x يزيد ← y يزيد
  - **ارتباط سلبي (Negative):** x يزيد ← y يقل
  - **لا يوجد ارتباط (No correlation):** النقط متفرّقة من غير اتجاه.
- **معامل الارتباط r (Correlation coefficient)**
  - قيمته دايماً **−1 ≤ r ≤ 1**.
  - **قريب من 1:** = ارتباط إيجابي **قوي**
  - **قريب من 0:** = مفيش ارتباط تقريباً
  - في جداول البيانات: الدالة `CORREL` .
- **السببية والارتباط الزائف**
  - **علاقة سببية (Causal relationship):** حاجة **سبب** والتانية **نتيجة**
  - **الارتباط الزائف (Pseudo-correlation):** حاجتين **باينين** إن بينهم سبب ونتيجة وهو مش موجود
  - يعني **الارتباط لوحده مش دليل على السببية**.
- **الجدولة المتقاطعة (Cross tabulation)**: جدول بيقارن البيانات بين **فئتين أو أكتر** في نفس الوقت (زي النوع × الرأي)

Scatter plot	مخطط مبعثر: نقط لأزواج القيم (x, y)
Positive correlation	ارتباط إيجابي: x يزيد و y يزيد
Negative correlation	ارتباط سلبي: x يزيد و y يقل

No correlation	لا يوجد ارتباط: مفيش اتجاه واضح
Correlation coefficient	معامل الارتباط r: بين −1 و 1 ويقيس قوة الارتباط (CORREL)
Causal relationship	علاقة سببية: حاجة سبب والتانية نتيجة
Pseudo-correlation	الارتباط الزائف: يبان فيه سببية وهي مش موجودة
Cross tabulation	الجدولة المتقاطعة: مقارنة البيانات بين فئتين أو أكتر

10-6 تحليل البيانات [5]

- **تحليل الانحدار (Regression analysis)**
  - **تحليل الانحدار البسيط (Simple regression analysis):** بنتوقع **المتغير الناتج y** من **متغير سببي واحد x**.
- **خط الانحدار (Regression line)**
  - **a:** = الميل: لو **موجب** الخط طالع (ارتباط إيجابي)
  - **b:** = قيمة y لما x = 0.
- **طريقة المربعات الصغرى (Least squares method): المتبقي (Residual):** المسافة الرأسية بين النقطة الحقيقية والخط
- **التوقع بالمعادلة**
  - y = 0.0287 × 5,000 + 297.86 = 143.5 + 297.86 = 441.36 ≈ **441 دقيقة**.

Regression analysis	تحليل الانحدار: دراسة العلاقة بين قيمة سببية وقيمة ناتجة
Simple regression analysis	تحليل الانحدار البسيط: توقع y من متغير سببي واحد x
Regression line	خط الانحدار: الخط اللي بنجيب منه القيم المتوقعة
Regression line equation	معادلة خط الانحدار: y = ax + b
Least squares method	طريقة المربعات الصغرى: تقلّل المتبقيات لأقل حاجة
Residual	المتبقي: المسافة الرأسية بين النقطة والخط

ذاكر المادة كلها مجاناً

شرح كل الدروس، أسئلة بالإجابات، بنك على نمط الوزارة، وامتحان بشكل الوزارة

zerooneacademy.com

مهندس رحيم