

教材11:テスト結果の基本統計量を求めよう

学習目標

生徒の点数について、

- 平均値
- 中央値
- 最高点
- 最低点
- 標準偏差

をExcel関数で計算します。

元データ **score.xlsx**:

名 前	点 数
田 中	80
鈴 木	65
佐 藤	92
山 田	73
高 橋	88
伊 藤	54
渡 辺	76
中 村	84

小 69
林

加 95
藤

Pythonプログラム

None

```
from openpyxl import load_workbook

wb = load_workbook("score.xlsx")
ws = wb.active

ws["D1"] = "統計項目"
ws["E1"] = "結果"

ws["D2"] = "平均"
ws["E2"] = "=AVERAGE(B2:B11)"

ws["D3"] = "中央値"
ws["E3"] = "=MEDIAN(B2:B11)"

ws["D4"] = "最高点"
ws["E4"] = "=MAX(B2:B11)"

ws["D5"] = "最低点"
ws["E5"] = "=MIN(B2:B11)"

ws["D6"] = "標準偏差"
ws["E6"] = "=STDEV(B2:B11)"

wb.save("statistics.xlsx")
```

Excelで開くと、それぞれの統計値が計算されます。

学習ポイント

「平均値だけではデータの特徴を十分に説明できない」ません。

例えば、

None

Aクラス : 60, 70, 80, 90, 100

Bクラス : 78, 79, 80, 81, 82

はどちらも平均**80**点ですが、ばらつきが全く違います。

そこで標準偏差が必要になってきます。

教材12: 点数のばらつきを詳しく分析しよう

今度は、

- 分散
- 標準偏差
- 第1四分位数
- 第3四分位数

を求めます。

Pythonプログラム

None

```
from openpyxl import load_workbook
```

```
wb = load_workbook("score.xlsx")
```

```
ws = wb.active
```

```
ws["D1"] = "統計項目"
```

```
ws["E1"] = "結果"
```

```
ws["D2"] = "分散"
```

```
ws["E2"] = "=VAR(B2:B11)"
```

```
ws["D3"] = "標準偏差"
```

```
ws["E3"] = "=STDEV(B2:B11)"
```

```
ws["D4"] = "第1四分位数"
```

```
ws["E4"] = "=QUARTILE(B2:B11,1)"
```

```
ws["D5"] = "中央値"
```

```
ws["E5"] = "=MEDIAN(B2:B11)"
```

```
ws["D6"] = "第3四分位数"
```

```
ws["E6"] = "=QUARTILE(B2:B11,3)"
```

```
ws["D7"] = "四分位範囲"  
ws["E7"] = "=E6-E4"  
  
wb.save("dispersion.xlsx")  
  
print("dispersion.xlsx を作成しました")
```

下記の2行で四分位範囲を求めています。

```
None  
ws["D7"] = "四分位範囲"  
ws["E7"] = "=E6-E4"
```

つまり、

IQR=Q3-Q1

です。

この教材で教えたこと

平均だけを見るのではなく、

```
None  
中心 → 平均・中央値  
  
ばらつき → 分散・標準偏差  
  
分布 → 四分位数
```

という統計データの見方を学習できます。

教材13: 数学と英語の点数には関係があるか調べよう

subjects.xlsx を次のようにします。

名前	数学	英語
----	----	----

田中	85	80
----	----	----

鈴木	70	65
----	----	----

佐藤	95	90
----	----	----

山田	60	70
----	----	----

高橋	88	84
----	----	----

伊藤	55	50
----	----	----

渡辺	75	78
----	----	----

中村	92	87
----	----	----

小林	68	72
----	----	----

加藤	82	85
----	----	----

プログラムは次のようにします。

None

```
from openpyxl import load_workbook

wb = load_workbook("subjects.xlsx")

ws = wb.active

ws["E1"] = "分析項目"

ws["F1"] = "結果"

ws["E2"] = "数学平均"

ws["F2"] = "=AVERAGE(B2:B11)"

ws["E3"] = "英語平均"

ws["F3"] = "=AVERAGE(C2:C11)"

ws["E4"] = "数学の標準偏差"

ws["F4"] = "=STDEV(B2:B11)"

ws["E5"] = "英語の標準偏差"

ws["F5"] = "=STDEV(C2:C11)"

ws["E6"] = "相関係数"

ws["F6"] = "=CORREL(B2:B11,C2:C11)"


wb.save("correlation.xlsx")

print("correlation.xlsx を作成しました")
```

ここでは、あえて標準偏差も追加しました。

これによって、

1. それぞれの平均を調べる
2. ばらつきを調べる
3. 2つのデータの関係を調べる

という統計分析らしい流れになります。

CORRELの結果

相関係数は

$$-1 \leq r \leq 1$$

の範囲になります。

例えば、

None

0.9

に近ければ強い正の相関、

None

0

に近ければ相関が弱い、

None

-0.9

に近ければ強い負の相関です。

教材14: 条件ごとに成績を分析しよう

こちらは基本的に前回のプログラムで問題ありません。

元データ:

名前	クラス	性別	点数
田中	A	男	80
鈴木	A	女	65
佐藤	B	女	92
山田	A	男	73
高橋	B	男	88
伊藤	B	女	54
渡辺	A	女	76
中村	B	男	84

プログラム:

None

```
from openpyxl import load_workbook

wb = load_workbook("class_score.xlsx")
ws = wb.active

ws["F1"] = "統計項目"
ws["G1"] = "結果"
ws["F2"] = "Aクラス平均"
ws["G2"] = '=AVERAGEIF(B2:B9,"A",D2:D9)'
ws["F3"] = "Bクラス平均"
ws["G3"] = '=AVERAGEIF(B2:B9,"B",D2:D9)'
```

```
ws["F4"] = "男子平均"
ws["G4"] = '=AVERAGEIF(C2:C9,"男",D2:D9)'
ws["F5"] = "女子平均"
ws["G5"] = '=AVERAGEIF(C2:C9,"女",D2:D9)'
ws["F6"] = "80点以上の人数"
ws["G6"] = '=COUNTIF(D2:D9,">=80')
ws["F7"] = "Aクラス男子平均"
ws["G7"] = '=AVERAGEIFS(D2:D9,B2:B9,"A",C2:C9,"男")'
ws["F8"] = "Bクラス女子平均"
ws["G8"] = '=AVERAGEIFS(D2:D9,B2:B9,"B",C2:C9,"女")'
wb.save("group_analysis.xlsx")
print("group_analysis.xlsx を作成しました")
```

この教材では、

None

AVERAGEIF

と

None

AVERAGEIFS

の違いを説明します。

AVERAGEIF

条件が1つです。

None

=AVERAGEIF(B2:B9,"A",D2:D9)

「Aクラスの点数の平均」

です。

AVERAGEIFS

条件を複数指定できます。

None

```
=AVERAGEIFS(D2:D9, B2:B9, "A", C2:C9, "男")
```

これは、

Aクラス かつ 男子

という2条件です。

プログラミングの

None

```
if class_name == "A" and gender == "男":
```

との比較もできるため、教材として面白いところです。

教材15: 広告費から売上を予測しよう

最後は回帰分析まで進めます。

例えば企業のデータとして、

月	広告費(万円)	売上(万円)
1月	10	105
2月	15	130
3月	20	160
4月	25	180
5月	30	215
6月	35	240
7月	40	270
8月	45	300

とします。

「広告費が増えると売上はどの程度増えるのか」を分析します。

Pythonプログラム

None

```
from openpyxl import load_workbook

wb = load_workbook("sales_data.xlsx")
ws = wb.active

ws["E1"] = "回帰分析"
ws["F1"] = "結果"
ws["E2"] = "傾き"
ws["F2"] = "=SLOPE(C2:C9,B2:B9)"
ws["E3"] = "切片"
ws["F3"] = "=INTERCEPT(C2:C9,B2:B9)"
ws["E4"] = "決定係数"
```

```
ws["F4"] = "=RSQ(C2:C9,B2:B9)"
ws["E6"] = "広告費"
ws["F6"] = 50
ws["E7"] = "予測売上"
ws["F7"] = "=FORECAST(F6,C2:C9,B2:B9)"
wb.save("regression.xlsx")

print("regression.xlsx を作成しました")
```

この教材の重要ポイント

例えば傾きが

None

5.5

なら、

広告費が1万円増えると、売上が平均約**5.5**万円増える関係がある

と解釈できます。

そして、

None

=RSQ(C2:C9,B2:B9)

では決定係数 **R²** を求めます。

例えば、

None

R² = 0.97

なら、このデータでは広告費と売上の直線的関係がかなり強いことを示します。

さらに、

None

=FORECAST.LINEAR(50,C2:C9,B2:B9)

によって、

広告費を**50**万円にすると、売上はいくらと予測されるか
まで計算できます。