第７回　標準偏差の理解

■標準偏差の大きさのイメージ

Ｔ得点（正規偏差値）：　平均50，標準偏差10の正規分布
ＩＱ（ウェクスラー系知能検査）：　平均100，標準偏差15の分布
ＩＱ（ビネー系知能検査）：　　平均100，標準偏差16の分布
　偏差値６０は，平均から１ＳＤ上で，上位１６％くらい．
　偏差値４０は，平均から１ＳＤ下で，下位１６％くらい．
　ＩＱ　８０は，下位１０％くらい．
　ＩＱ１２０は，上位１０％くらい．
　ＩＱ　７０は，平均から２ＳＤくらい下で，下位２．５％くらい．

５段階評定相対評価
平均＋1.5SD ～　　　　　　　の範囲に，約 7%のデータが存在する．　5
平均＋0.5SD ～ 平均＋1.5SDの範囲に，約24%のデータが存在する．　4
平均－0.5SD ～ 平均＋0.5SDの範囲に，約38%のデータが存在する．　3
平均－1.5SD ～ 平均－0.5SDの範囲に，約24%のデータが存在する．　2
　　　　　　～ 平均－1.5SDの範囲に，約 7%のデータが存在する．　1

正規分布はいろいろな統計分析の基礎となる．
　数学的に非常に取り扱いやすい．
　平均と標準偏差というわずか２つのパラメタで分布を記述することができる．
　平均から離れた値ほど少なくなるというのは，現実にもあてはまりの良いモデルである．
→　データは正規分布からの無作為標本であるというモデルを立てて分析．

■正規分布（normal distribution，ガウス分布）

確率分布：離散的な値をとるデータにおいて，ある値の発生のしやすさ（確率）を表す分布．確率を合計すると１になる．
　丁半ばくち…偶数が出る確率0.5，奇数が出る確率0.5
確率密度分布：連続的な値をとるデータにおいて，ある値の発生のしやすさ（確率密度）を表す分布．確率密度を積分すると１になる．
　ルーレットダーツ…面積が広い領域ほど矢が刺さりやすい
正規分布：平均（μ）と標準偏差(σ）{または分散（σ^2)}の値が決まれば，あるデータ値ｘが発生する確率密度yが一意に定まるような確率密度分布の１つ．
　　1次関数 y = ax + b は， a と b が決まれば，あるxに対するyの値が定まる．
　　2次関数 y = ax^2 + bx + cは，a,b,cの値が決まれば，あるxに対するyの値が定まる．
　　同様に正規分布は，平均μと標準偏差σ（または分散）の値が決まれば，あるxに対するyの値が定まる．
確率密度関数：

$$y = \frac{1}{\sqrt{(2\pi)}\sigma} e^{\left\{-\frac{1}{2} \times \left(\frac{x-\mu}{\sigma}\right)^2\right\}}$$

　π：円周率（3.1415…），ｅ：自然対数の底（2.71828…）

平均μ，標準偏差σがどんな値でも，正規分布の性質は同一
　釣り鐘型（ベル型）の分布．
　平均に近いほど発生確率が高く，平均から遠い値ほど発生確率が小くなる．
　単峰．左右対称　→　平均値＝中央値＝最頻値　歪度０
　理論上 －∞から＋∞の値を取る．
　±1SDのところが変曲点（上に凸と下に凸の境目）になる．　尖度　０
標準正規分布：平均０，標準偏差１の正規分布．

平均±１SDの範囲に，68.26%のデータが存在する．　片側34.13%
平均±２SDの範囲に，95.44%のデータが存在する．　片側47.72%　△13.59%
平均±３SDの範囲に，99.72%のデータが存在する．　片側49.86%　△ 2.14%
　→　平均±３SDより外にあるデータは外れ値として除外（３シグマ法）