

■データの種類

- 数量データ：試験の得点など数値で表されるデータで、データ間の大小関係を比較できる
- カテゴリーデータ：血液型などのデータで、比較することができない

■数値データの基本統計量

- 標本(sample)：データのカタマリを指す
- 標本数：データの数 n
- 最大：データの最大値
- 最小：データの最小値
- 合計：データの合計 $T = \sum_{i=1}^n x_i$
- 平均(mean)：合計を標本数で割った値 $\mu = \frac{T}{n}$
- 中央値(median)：数値の大きい順に並べたとき、ちょうど真ん中にある値
 例えば 5 人の年齢の中央値は、上からあるいは下から 3 番目の人の年齢になる。6 人のように中央の値がないので、3 番目と 4 番目の人の年齢の平均値になる。
- 最頻値：データの中で最も多く存在する数値
- 分散(variance)：データのバラツキを表す値。データの平均値と値の差の自乗和平均

$$V = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}, \quad ((n-1) \text{で割ることもある})$$

- 標準偏差(standard deviation)：分散の平方根 $\sigma = \sqrt{V}$
- 偏差値：偏差値 = $\frac{(x_i - \mu)}{\text{標準偏差 } \sigma} \times 10 + 50$
- 変動係数：データのバラツキを表す値。標準偏差を平均値で割った値 σ / μ

判断例

変動係数	変動
0~0.2	ほとんどない
0.2~0.5	ややあり
0.5~1.0	ある
1.0以上	かなりあり

- 尖度（せんど）：正規分布が上下方向にどの程度偏っているか？

尖度	上下方向の偏り
H>0	尖っている
H=0	正規分布と同じ形状
H<0	扁平

- 歪度（わいど）：正規分布が左右方向にどの程度偏っているか？

歪度	左右方向の偏り
G>0	峰が左寄り
G=0	峰が中央にある
G<0	峰が右寄り

■ 単相関係数：数量データ間の関係を明らかにする統計量

2つのデータに直線的な関係があるかどうかを表す。

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$$

ここに、偏差平方和 S_{xx} , S_{yy} と積和 S_{xy} であり、

$$S_{xx} = \sum (x_i - \bar{x})^2, \quad S_{yy} = \sum (y_i - \bar{y})^2, \quad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

+1~-1 の値となり、0だと全く相関がなく、±1だと正あるいは負の相関がある

(判断例)

単相関係数	相関
0.9以上	非常に強い
0.7以上	やや強い
0.5以上	やや弱い
0.5未満	非常に弱い

■ 回帰分析

- ・ 直線回帰式

$$y = ax + b, \quad a = \frac{S_{xy}}{S_{xx}}, \quad b = \bar{y} - a\bar{x}$$

- ・ 決定係数 R^2 : $R^2 = 1 - \frac{S_e}{S_{yy}}$

ここに、 S_e は残差平方和（回帰式とデータの差の自乗和）であり、

$$S_e = \sum (y - y_i)^2 = \sum \{(ax_i + b) - y_i\}^2$$

(判断例)

決定係数R2	相関
0.8以上	非常によい
0.5以上	良い
0.25以上	やや良い
0.25未満	良くない

- ・ 曲線回帰式（変数変換により直線式になる例）

$$y = ax + b, \quad y = a/x + b, \quad y = a\sqrt{x} + b, \quad y = a \log x + b$$

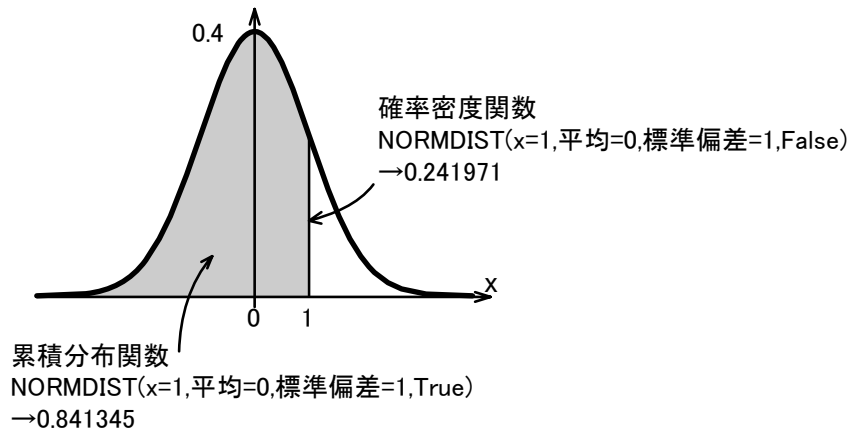
$$y = ax^b (\log y = b \log x + \log a), \quad y = ab^x (\log y = \log b \times x + \log a)$$

■ 正規分布

確率密度関数 = NORMDIST(x, 平均 μ , 標準偏差 σ , False)

$$\frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$$

累積分布関数 = NORMDIST(x, 平均 μ , 標準偏差 σ , True)

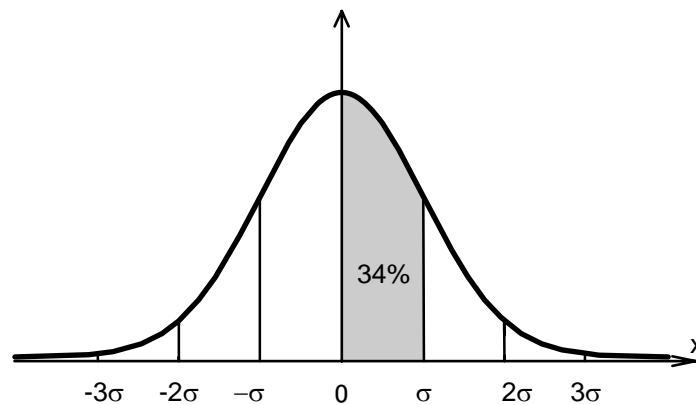


標準正規分布

標準偏差 s が 1.0, 平均値が 0.0 のときの正規分布

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

[問題] 平均値からのバラツキは標準偏差 σ で表されることが多い。例えば, $0 \sim \sigma$ 間は全体の約 34% になる。下図の各間隔の割合を求めよ。



[問題] 5段階評価を偏差値で下記のように付けた場合, それぞれの評定の割合を求めよ。

評定	偏差値	割合
1	~ 35	約7%
2	35 ~ 45	
3	45 ~ 55	
4	55 ~ 65	
5	65 ~	

=NORMDIST(35,50,10,TRUE)

[問題] 平均点が 60 点, 標準偏差が 10 点の試験で, A 君は 80 点の成績だった。受験者が 1000 人だったとすると, A 君の成績は, 上から何番目くらいだと考えられるか?

[問題] 合格率が 20% の試験がある。平均点が 60 点, 標準偏差が 10 点だったとすると, 合格点は何点か?