

ランダムとデータ要約について

B3 60181052 宮崎智己

資料

- ▶ 医薬研究者の視点からみた道具としての統計学
- ▶ 著者：奥田千恵子
- ▶ 範囲
 - ▶ 第3章 ランダム
 - ▶ 第4章 データの要約

目次

▶ 第3章

- ▶ 統計学におけるランダム
- ▶ 乱数表
- ▶ ガウスの誤差の法則

▶ 第4章

- ▶ データの中心性を示す代表値
- ▶ 標準偏差
- ▶ 標準偏差の性質と変動係数
- ▶ その他のばらつきの指標

統計学におけるランダム

▶ ランダム(random)

- ▶ 次に起こることが、それ以前に起きたこととまったく無関係に起こることをさす。
- ▶ 無作為、偶然などと訳される。

▶ 無作為化(randomization)

- ▶ 標本の抽出や、処置の割付をランダムにすること。
- ▶ 確率化ともいう
- ▶ 一般的に乱数表を用いて行う

乱数表

	出現個数	偶然誤差
0	104	4
1	94	-6
2	97	-3
3	105	5
4	83	-17
5	96	-4
6	103	3
7	108	8
8	102	2
9	108	8

表 3-1

▶ 乱数

- ▶ 一定の範囲で、均一で同じ数字がもう一度出てくるまでの間隔に偏りがない数列

▶ 偶然誤差

- ▶ 乱数の出現個数－期待値＝偶然誤差となる。

ガウスの誤差の法則

- ▶ 正負同程度に起こる
- ▶ 絶対値の大きな誤差ほど生じにくい
- ▶ データの平均値が真の値の推定値になる

連続量のデータの中心性を示す代表値 (1)

- ▶ 連続量・・・長さ、重さなど
- ▶ 相加平均値
 - ▶ データの総和で割った値
 - ▶ 算術平均値、あるいは単に平均値とも
- ▶ 中央値
 - ▶ 大きさの順に並べたときちょうど中央にある値
- ▶ 最頻値
 - ▶ データをいくつかの範囲に区切って集計したとき、もっとも頻繁に出現する値

連続量のデータの中心性を示す代表値 (2)

▶ 相乗平均値

- ▶ N 個のデータの積の N 乗根をとったもの
- ▶ 相乗平均値の対数は相加平均値に等しくなる

▶ 調和平均値

- ▶ データの逆数の総和をデータ数で割って、その逆数を取ったもの

▶ 調整平均値

- ▶ データを大きさの順に並べ、中央値に近い部分のみを用いて平均値を求めたもの

標準偏差

- ▶ データの標準値 $\bar{x}$ を用いたとき、個々のデータ、 $x_1, x_2, \dots, x_n$ の平均値からのずれ $x_i - \bar{x}$ のことを偏差という
- ▶ 標準偏差(SD)はデータのばらつきの程度を表す。
- ▶ 公式

$$SD = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}}$$

上の公式では、自由度が n ではなく $n-1$ を用いて計算している。
これは平均値を求める際に自由度 n を用いており、そのために $n-1$ 個のデータが決まれば n 番目のデータも必然的に決まるという考えからである。

標準偏差の性質と変動係数

- ▶ もともと小数のデータを扱っている場合、データ数が増えれば平均値に近い値が増えるため標準偏差が小さくなることもある。
- ▶ 平均値から大きくかけ離れた値が出る確率が0でないため、その値が含まれると標準偏差が逆に大きくなってしまう。
- ▶ 多くの値を扱う場合、標準偏差ではなく変動係数(CV)を用いるほうがよい。
- ▶ 公式

$$CV = \frac{\sqrt{\frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_n - \bar{X})^2}{n-1}}}{\bar{X}}$$

その他のばらつきの指標

- ▶ 正規分布に従わない場合、標準偏差は用いることができない
- ▶ その場合、四分位範囲を用いる
- ▶ 表し方は、箱ひげ図を用いる

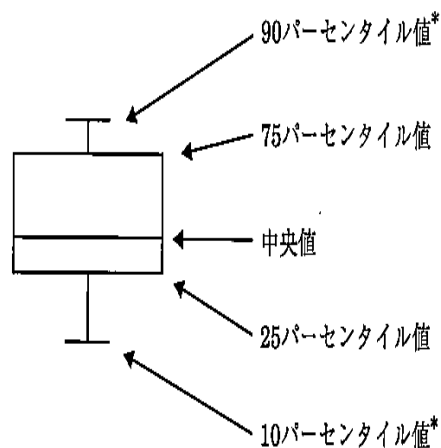


図4-2