第２章　確率・統計

統計学は、独自の専門用語に加えて、数学の言葉で記述されているため、
これらをよめるようになるのが、目標

1) 集合

集合：ものの集まり

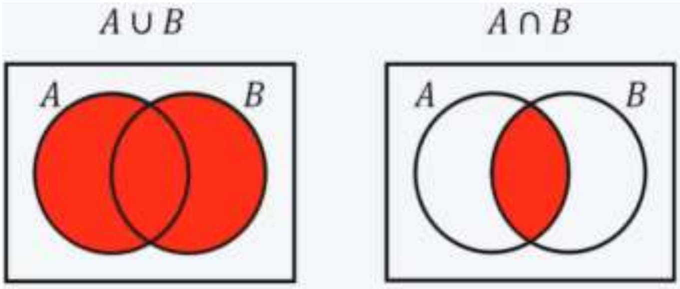
S = {a,b,c,d,e}　と表す。

要素の集まりで成り立つ。要素は明確に区別ができる。

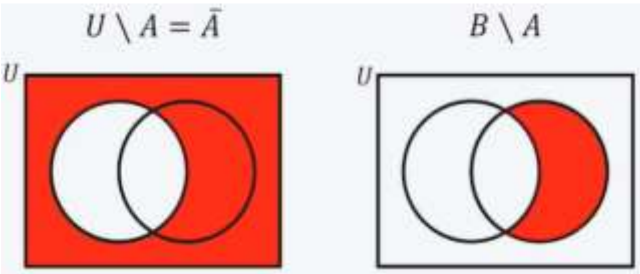
aは、Sの要素であることを　$a \in S$　と表す

Sの中に、M={d,e}がある場合、　$M \subset S$　と表す

和集合と共通部分



絶対補と相対補



2) 確率

確率：確かさの率。下記の2つの考え方がある。

頻度確率(客観確率)：発生する頻度

ベイズ確率(主観確率)：信念の度合い

確率は、 $P(A) = \frac{n(A)}{n(U)} = \frac{\text{事象}A\text{が起こる数}}{\text{すべての事象の数}}$ で表す

3) 条件付き確率

ある事象Bが与えられた下で、Aとなる確率

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{n(A \cap B)}{n(B)}$$

4) 独立な事象の同時確率

お互いの発生には因果関係のない事象Aと事象Bが同時に発生する確率

$$P(A \cap B) = P(A)P(B|A)$$
$$= P(A)P(B)$$

5) ベイズ則

「ある事象 A が起こったという条件のもとでの事象 B の確率 P(B|A)」を使って
「ある事象 B が起こったという条件のもとでの事象 A の確率 P(A|B)」を求めようというもの

ベイズの定理

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

※$P(B) > 0$

講義動画では、これを応用して式を置き換えることで、問題を解いている。

$$P(A)P(B|A) = P(B)P(A|B)$$

6) 統計

記述統計：集団の性質を要約し記述する
推測統計：集団から一部を取り出し、元の集団の性質を推測する
確率変数：事象と結びつけられた数値
確率分布：事象の発生する確率の分布
期待値：分布における確率変数の平均の値orありえそうな値

$$\text{期待値}E(f) = \sum_{k=1}^{n} P(X = x_k) f(X = x_k)$$

と表す

分散：データの散らばり具合
データのそれぞれの値が期待値からどれだけズレているか平均したもの
共分散：２つのデータ系列の傾向の違い
標準偏差：分散を逆演算して求めたもの

6) 確率分布

ベルヌーイ分布：結果が2種類しかないような事象の分布
マルチヌーイ分布：結果が3種類以上ある事象の分布
二項分布：ベルヌーイ分布の多施行版
ガウス分布：釣鐘型の連続分布

7) 推定

推定統計：集団から一部を取り出し、元の集団の性質を推測する
点推定：平均値などを１つの値に推定すること
区間推定：平均値などが存在する範囲を推定すること
推定量：パラメータを推定するために利用する数値の計算方法や計算式のこと
推定値：実際に施行を行った結果から計算した値

標本平均：母集団から取り出した標本の平均値

標本分散：母集団から取り出した標本の分散。母分散より小さくなりやすい。

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

不偏分散：標本分散にn/(n-1)をかけることによって、母分散に近づける

$$s^2 = \frac{n}{n-1} \times \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$