◆ 記述統計と推測統計

検討しているデータの見方が異なる２つの立場があることに注意！

・記述統計：　検討しているデータ ＝ 母集団　とみなす．
要約統計 ともいう．検討しているデータを代表値で要約して記述．

・推測統計：　検討しているデータ ＝ 母集団の一部 ＝ 標本 とみなす．
欲しいものは，検討しているデータの特性そのものではない．
欲しいものは，母集団の特性である．

> 統計量（statistic）= 標本の関数（未知母数を含まない）
> パラメトリック解析とノンパラメトリック解析
> 標本平均と標本分散
> 好ましい推定（一致性，有効性，漸近正規性，不偏性など）

- 標本平均： $\overline{X} = \frac{1}{N} \sum_{i=1 \sim N} X_i$

- 標本分散： $s^2 = \frac{1}{N-1} \sum_{i=1 \sim N} \left(X_i - \overline{X} \right)^2$

cf. いわゆる分散： $\frac{1}{N} \sum_{i=1 \sim N} \left(X_i - \overline{X} \right)^2$

？ 疑問：標本平均や いわゆる分散 は，N （= 標本サイズ）で除すのに対し，**なぜ，標本分散は，N - 1 で除すのか？**

その背景には，**不偏性** と **自由度** という概念があるのです．

◆ 標本（確率変数）の計算（2）と不偏分散

同一の分布関数 F から独立に抽出した標本（サイズ＝ N ）$X_{i=1\sim N}$ を用いる．

12)

$$E\sum_{i=1\sim N}\left(X_i-\overline{X}\right)^2=\sum_{i=1\sim N}E\left(X_i^{\,2}\right)-2\sum_{i=1\sim N}E\left(X_i\overline{X}\right)+NE\left(\overline{X}^{\,2}\right)$$
$$=N\left\{E\left(X^2\right)-E\left(\overline{X}^{\,2}\right)\right\}$$
$$=N\left\{\left(\mu^2+\sigma^2\right)-\left(\mu^2+\frac{\sigma^2}{N}\right)\right\}=(N-1)\sigma^2$$

なお，上では，6), 9) および 10) の関係を用いている．
上の関係を整理することにより，以下が得られる．

13) $$E\left(s^2\right)=\sigma^2 \quad \text{for} \quad s^2=\frac{1}{N-1}\sum_{i=1\sim N}\left(X_i-\overline{X}\right)^2$$

言い換えれば，$\sum_{i=1\sim N}\left(X_i-\overline{X}\right)^2$ の変動の大きさは，平均的にみて，
X_i の変動（$=\left(X_i-\mu\right)^2$）の大きさの $(N-1)$ コ分（自由度）に相当する．

モンテカルロで確認 (4)

```
> vals <- rnorm(20) # test any other distribution: ex. rexp(12)
> vals
 [1]  -0.15086814 -0.53577110 -1.92426062 -1.02552999  ...

> (m.vals <- mean(vals))
[1] 0.1490854
> sum((vals - mean(vals))^2)
[1] 23.43915
> sum((vals - mean(vals))^2)/20
[1] 1.171957
> sum((vals - mean(vals))^2)/(20 - 1)
[1] 1.233639
> var(vals) # built-in function
[1] 1.233639
>
> variances <- numeric(10000)
> for(j in 1:10000) {
+   vals <- rnorm(20)
+   variances[j] <- sum((vals - mean(vals))^2)}
> mean(variances)
[1] 18.96152
>
```