

スライド・黒板が見やすい位置に着席してください  
本日のスライド・エクセルをCES HPにアップしました

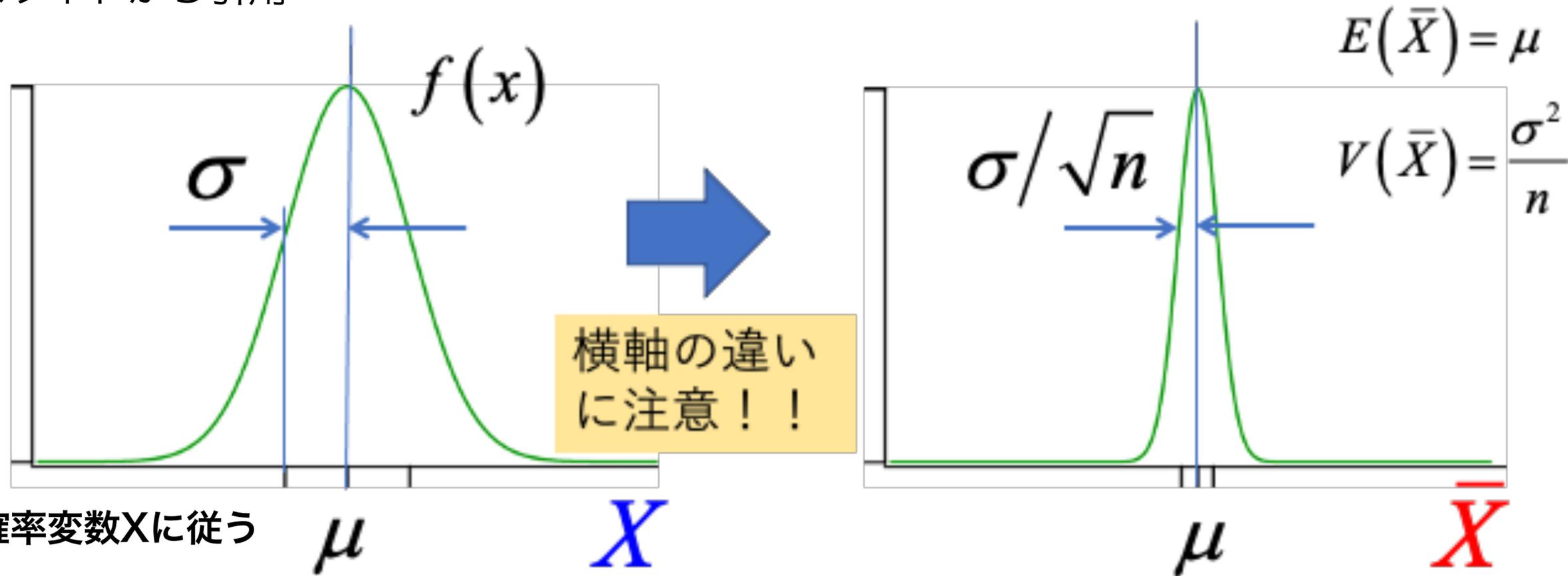
# 第5回DSB演習

2022/06/27(月)

担当：熊野英和先生、小山翔子

# 確率変数の平均の期待値と分散

熊野先生概説スライドから引用



m個のデータが確率変数 $X$ に従う  
(分布は不問)

n個の標本を抽出し、  
標本平均を計算する  $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$

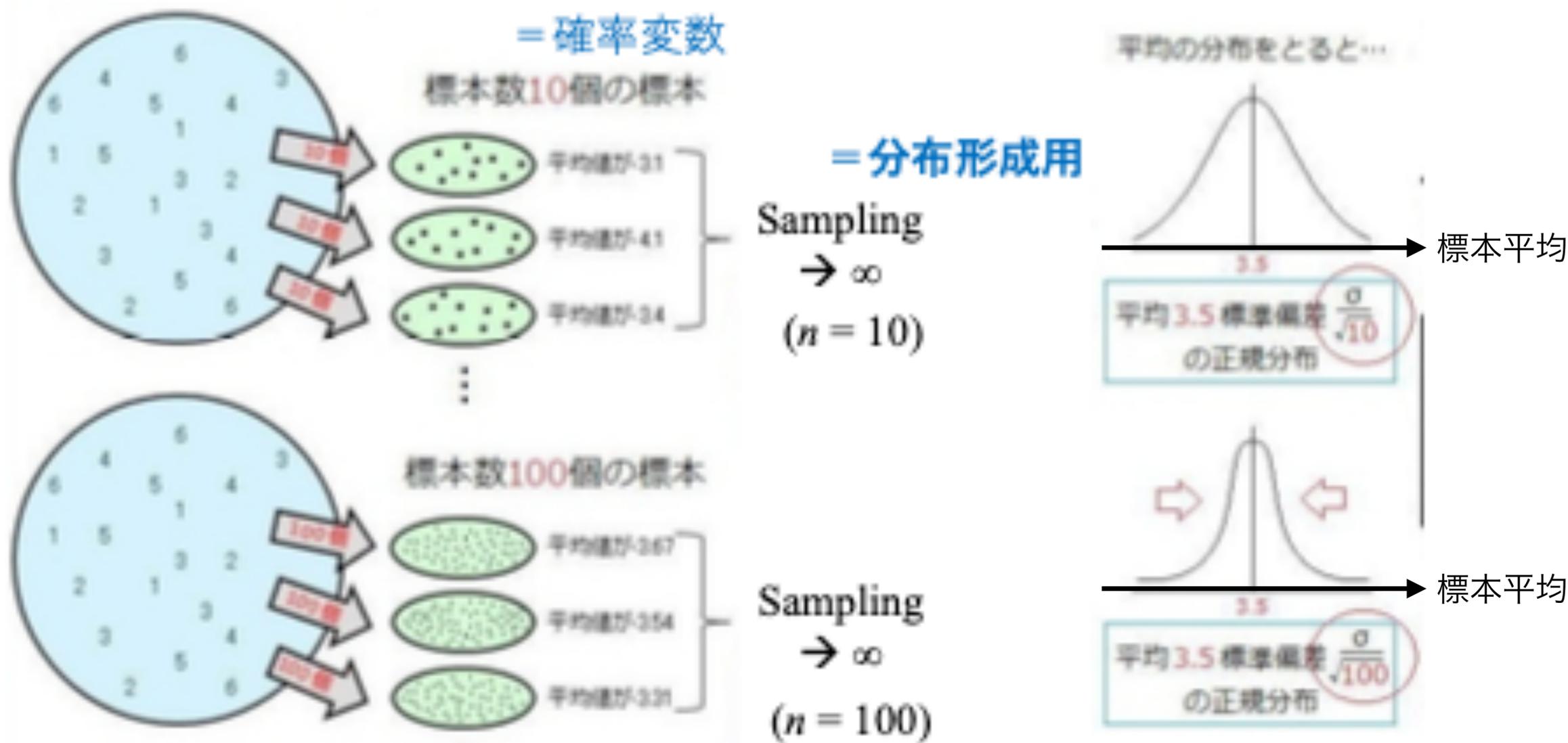
これを繰り返すと

母平均： $\mu$   
母分散： $\sigma^2$



標本平均の平均： $\mu$   
標本平均の分散： $\sigma^2/n$

# 大数の法則：まとめ



標本数  $n$  を増やせば、標本平均の分布幅が狭まり、標本平均が母平均  $\mu$  にいくらでも近づく。

この事実を用いれば、1回の  $n$  の大きな実験を行った結果 = 母平均としてよい。

# 演習1 0~9の乱数を100個生成: 母集団? 母平均? 母分散?

## 5つランダムに取り出す作業を繰り返す: 標本抽出

- 1回取り出した時の標本 [1, 4, 9, 1, 4]
  - 標本平均: 標本分散: 不偏分散:
- 108回取り出した時の標本→スプレッドシート1
  - 標本平均: 標本分散: 不偏分散:
  - 結果をスプレッドシートに書き込みましょう
  - 全員の記入が終わったらファイル→ダウンロード→excelを選択
  - 各自excelで以下を計算してみましょう
    - 標本平均の平均 (→母平均推定量)
    - 不偏分散の平均 (→母分散推定量)

計算してみましょう!

一人3回分計算してみましょう!  
各自値をコピーしてエクセルで  
計算した後、結果をスプレッドシート  
に記入してください

# 演習2 0~9の乱数を100個生成: 母集団? 母平均? 母分散?

## 20個ランダムに取り出す作業を繰り返す: 標本抽出

- 1回取り出した時の標本  
[3,8,7,4,4,2,9,4,3,1,3,1,1,0,1,0,8,5,0,2]
  - 標本平均: 標本分散: 不偏分散:
- 108回取り出した時の標本→スプレッドシート2
  - 標本平均: 標本分散: 不偏分散:
  - 結果をスプレッドシートに書き込みましょう
  - 全員の記入が終わったらファイル→ダウンロード→excelを選択
  - 各自excelで以下を計算してみましょう
    - 標本平均の平均 (→母平均推定量)
    - 不偏分散の平均 (→母分散推定量)

計算してみましょう!

一人3回分計算してみましょう!  
各自値をコピーしてエクセルで  
計算した後、結果をスプレッドシート  
に記入してください

# 結果から読み取れること

- ・ 母平均:  $\mu = 4.00$  と標本平均  $\bar{X}$ 
  - ・ 標本抽出1回: 標本数5の標本平均  $\bar{X} = 3.80$ , 標本数20の標本平均  $\bar{X} = 3.30$
  - ・ 標本抽出108回: 標本平均の平均(3.97, 4.03)は母平均に近い( $\bar{X} \rightarrow \mu$ )  
→ 標本抽出を繰り返すと 「標本平均の平均」から母平均を推定できる **\*大数の法則\***
- ・ 母分散:  $\sigma^2 = 9.30$  と標本分散  $\hat{\sigma}^2$  と不偏分散  $s^2$ 
  - ・ 標本抽出1回: 標本数5の標本分散  $\hat{\sigma}^2 = 8.56$ , 標本数20の標本分散  $\hat{\sigma}^2 = 7.61$   
標本数5の不偏分散  $s^2 = n\hat{\sigma}^2/(n-1) = 10.7$ , 標本数20の不偏分散  $\hat{s}^2 = 8.01$   
→ 不偏分散から母分散を推定できる
  - ・ 標本抽出108回: 不偏分散  $s^2$  の平均(9.187, 9.501)は(標本分散の平均より)母分散  $\sigma^2$  に近い  
→ 標本抽出を繰り返すと 「不偏分散の平均」から母分散を推定できる
  - ・ 標本抽出108回: 標本平均  $\bar{X}$  の標準偏差(1.346, 0.693)  
→ 標本抽出を繰り返すと標本平均  $\bar{X}$  の標準偏差 → 母標準偏差  $\sigma$  / 標本数  $\sqrt{n}$   
→ 1回の標本数が大きいほど標本平均の分布幅が狭まる

# 例題1

統計検定2級2019.11.問7

母平均 $\mu$ 、母分散 $\sigma^2$ をもつ母集団から、大きさ $n=100$ の標本を無作為抽出し、標本平均 $\bar{x} = 40.0$ および(標本の)不偏分散 $s^2 = 16.0$ を得たとする。このとき標本平均の標準偏差(=標準誤差)はいくらか？

# 例題2

熊野先生例題1 母分布：未知、母分散：既知、標本サイズ：大標本  
正規分布表かエクセルを使ってください

新潟県の小学校新入生（6歳児）男子の平均身長を調べたい。そこで、36人  
を無作為抽出したところ、その平均は  $\bar{x} = 116.4$  cmであった。

また経験的に、本県小学校新入生男子の身長の標準偏差は例年  $\sigma = 5.0$  cm  
あることがわかっている。このとき、母平均  $\mu$  の信頼度95%の信頼区間を求  
めよ。また、信頼度99%の信頼区間はどうか？

# 例題3

熊野先生例題 母分布：未知、母分散：未知、標本サイズ：大標本

正規分布表かエクセルを使ってください

**ある県内の農園から収穫したルレクチェ120個を無作為抽出したところ、平均468g, 標準偏差26gであった。母平均の95%信頼区間を求めよ。**

## 例題4 テキスト問題8-2

遺跡から貝殻が12個出土したので、その質量のデータを使って現存種の貝と同種のものであるかどうかを検定したい。データは次の通り(単位はg)。  
11.78, 12.92, 7.55, 14.52, 12.05, 19.0, 11.29, 11.81, 15.38,  
9.62, 14.19, 12.62

これと比較したい現存種の貝の質量の平均値は12.6 g 標準偏差は1.9 gである。この母集団は正規分布しているものとする。

平均質量の分布を考えて検定を行ったとき、この貝殻は現存種のものである、という仮説を危険度5%で棄却できるか。

# 例題5 統計検定2級2019.11.問18-1

次の表は、2017年の2人以上の労働者世帯について、47都道府県庁所在地市別に1世帯当たり1ヶ月間の収入と支出をまとめたものである(単位：万円)。なお、以下の表における世帯主収入の合計は、定期収入と賞与の和である。資料：総務省「2017年家計調査年報」

	世帯主収入			消費支出
	定期収入	賞与	合計	
札幌市	34.8	7.9	42.7	30.7
青森市	28.1	5.3	33.4	26.9
盛岡市	35.4	6.6	42	30.7
仙台市	30.6	5.3	35.9	30.9
...				
大分市	36.6	8.0	44.6	32.2
宮崎市	29.9	5.6	35.5	30.3
鹿児島市	33.5	6.4	39.9	30.9
那覇市	27.6	4.4	32	26.4

消費支出が定期収入及び賞与で説明できるかどうかを検証するため、次の重回帰モデルを考える。

$$\text{消費支出} = \alpha_0 + \alpha_1 \times \text{定期収入} + \alpha_2 \times \text{賞与} + u$$

ここで、誤差項 $u$ は互いに独立に正規分布 $N(0, \sigma_u^2)$ に従うとする。定期収入、賞与にそれぞれ対応する変数をincome, bonusとして、上記の重回帰モデルを統計ソフトウェアによって最小二乗法で推定したところ、次の出力結果が得られた(次スライド)。なお、出力結果の一部を加工している。また、出力結果のInterceptは定数項 $\alpha_0$ を表している

この重回帰モデルに対する解析結果の解釈で以下は正しいか？

- 賞与を一定としたときに、定期収入が1万円増えると消費支出が約0.39万円増える傾向がある。
- 賞与と定期収入が同時に1万円増えると消費支出が約0.39万円増える傾向がある。

## 重回帰モデルの出力結果

## Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	14.58851	2.49814	5.840	5.80e-07
income	0.39461	0.08944	4.412	6.54e-05
bonus	0.47247	0.24370	1.939	0.059

---

Residual standard error: 1.898 on 44 degrees of freedom  
Multiple R-squared: 0.5371, Adjusted R-squared: 0.5161  
F-statistic: 25.53 on 2 and 44 DF, p-value: 4.371e-08

# 例題6 テキスト問題8-1

とうもろこしの中には黄色と白の2色の実が混ざっていついているものがある。これはバイカラーと呼ばれ、優勢の黄色の実の純系品種と劣勢の白い実の純系品種の1代後輩であるために、その子（純系品種からみたら孫）である種子が持つ2個の遺伝子の黄色、白の組み合わせによって色の違いが生じたものである。この場合、最初の品種が純系であれば黄色と白の実の数の比の期待値はメンデルの法則に従って3:1となるが、そうでないと、例えば花粉が別の品種の花から飛んできたものだったりすると、比率の期待値は変化することになる。

今、あるとうもろこしの実を調べたところ、144粒の実のうち黄色の実が119個、白が25であった。この実が純系種の1代交配から作られたという仮説は棄却できるか、危険率を5%と1%にとって検定しなさい。