ひかり統計塾

統計検定２級の解説 サンプル動画

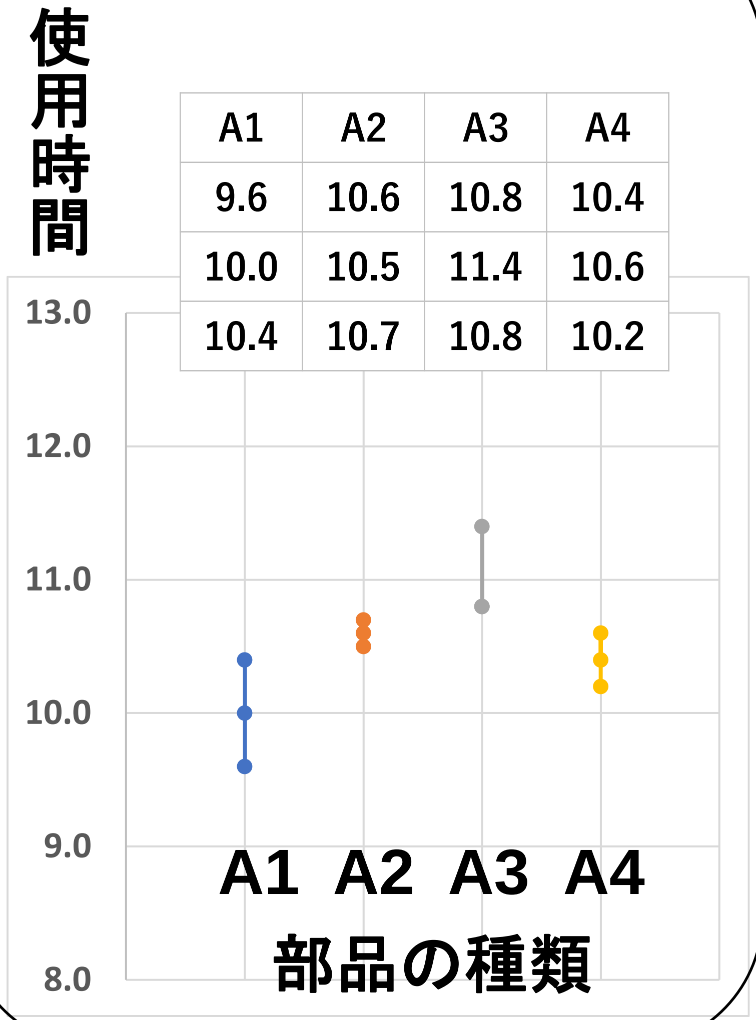
分散分析の基礎

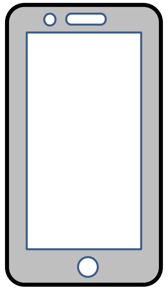
(関連問題)公式問題集(CBT対応版)

p160～p173，問1～問4

(p160.1a) [C10-2]分散分析の基礎: 対象とする問題

A1	A2	A3	A4
9.6	10.6	10.8	10.4
10.0	10.5	11.4	10.6
10.4	10.7	10.8	10.2



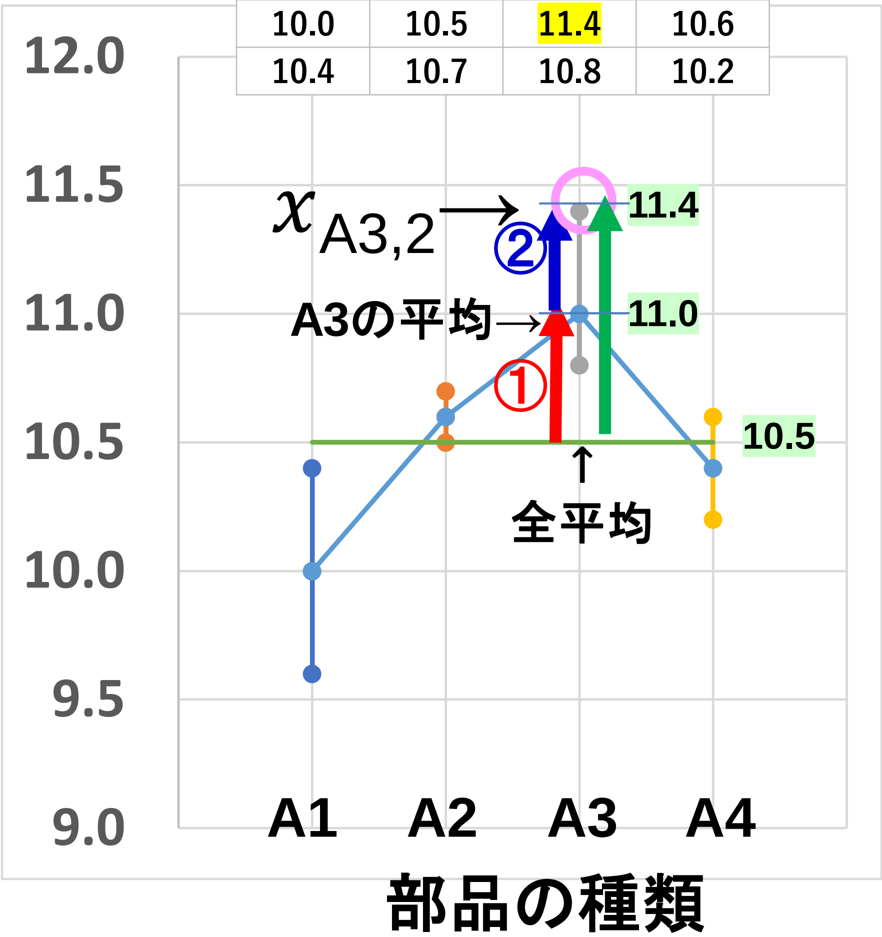


例題:スマホの1充電当たりの使用時間を長くする検討を行っています。ある部品に4種(*)を使い、使用時間を調べました。
(*)部品A1: 旧部品、 部品A2,A3,A4:新部品の候補

それぞれ、3個のサンプルを作成し、使用時間を調べました。
この部品の種類により、使用時間が異なると言えますか？

(p160.1b) [C10-2]分散分析の基礎: ばらつきの分解

A1	A2	A3	A4
9.6	10.6	10.8	10.4
10.0	10.5	11.4	10.6
10.4	10.7	10.8	10.2



ばらつきの総和(偏差平方和, 変動)は、

$$S = (11.4 - 10.5)^2 + \cdots \text{(残り11個分)}$$

$$= \{(11.4 - 11.0) + (11.0 - 10.5)\}^2 + \cdots \text{(残り11個分)}$$

②:ばらつき　①:部品の効果

総平方和

$$S = S_T = S_E + S_A$$

$$\begin{cases} S_E = (11.4 - 11.0)^2 + \cdots \text{(残り11個分)} \\ S_A = (11.0 - 10.5)^2 + \cdots \text{(残り11個分)} \end{cases}$$

②:ばらつき
誤差平方和

①:部品の効果
A間平方和

のように、総平方和を①と②に分けることができました。

$$(x + y)^2 = x^2 + y^2 + 2xy$$

(注) $2 \times (11.4 - 11.0) \times (11.0 - 10.5) + \cdots$ (残り11個分)
が出てきますが、この総和は0になります。(「入門統計解析法」p119[注6.1])

(p160.1c) [C10-2]分散分析の基礎: 自由度

母平均の検定･推定
(母分散:未知)

$$\text{不偏分散} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{\text{偏差平方和}}{\text{自由度}}$$

今の例では、

データ総数: $4 \times 3 = 12$
全自由度= データ総数 − 1
$\phi_T = n - 1 = 12 - 1 = 11$

A1	A2	A3	A4
9.6	10.6	10.8	10.4
10.0	10.5	11.4	10.6
10.4	10.7	10.8	10.2

4水準
⇒因子Aの自由度= 水準数 − 1
$\phi_A = 4 - 1 = 3$

⇒誤差の自由度= 全自由度 − 因子Aの自由度
$\phi_E = \phi_T - \phi_A$
$= 11 - 3 = 8$

公式: 全自由度(ϕ_T) = データ総数 − 1
因子Aの自由度(ϕ_A) = 水準数 − 1
誤差の自由度(ϕ_E)= $\phi_T - \phi_A$

のように、3つの自由度(ϕ_T,ϕ_A,ϕ_E)を求めます。

(p160.1d) [C10-2]分散分析の基礎: 自由度(練習)

灰色部分には値が有り、白色部分には値が無いとします

3種の自由度を求めてください

繰返し↓

A1	A2	A3	A4	A5
値:有り				

全自由度: $\phi_T = 5 \times 4 - 1 = 19$

因子Aの自由度: $\phi_A = 5 - 1 = 4$　(A: 5水準だから)

誤差の自由度: $\phi_E = 19 - 4 = 15$

繰返し↓

A1	A2	A3	A4	A5
値:有り				
	値:無し		値:無し	

全自由度: $\phi_T = (5 \times 4 - 2) - 1 = 17$

因子Aの自由度: $\phi_A = 5 - 1 = 4$

誤差の自由度: $\phi_E = 17 - 4 = 13$

公式:
全自由度(ϕ_T) = データ総数 − 1
因子Aの自由度(ϕ_A) = 水準数 − 1
誤差の自由度(ϕ_E)= 全自由度 − 因子Aの自由度 = $\phi_T - \phi_A$

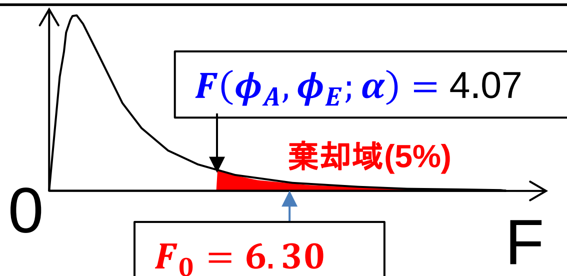
(p160.1e) [C10-2]分散分析の基礎: 検定の手順

(1): 2つの仮説を立てます
値の説明:(スマホの使用可能時間)要因A:(部品の種類)
帰無仮説 H_0: 要因Aの効果はない
対立仮説 H_1: 要因Aの効果はある

理論:帰無仮説が正しい時、検定統計量: $F_0 = \frac{V_A}{V_E}$は自由度(ϕ_A, ϕ_E)のF分布に従う

(2): 有意水準αを定めます: α= 0.05

A1	A2	A3	A4
9.6	10.6	10.8	10.4
10.0	10.5	11.4	10.6
10.4	10.7	10.8	10.2



(3):棄却域を定めます
データ総数(n)=(12),要因Aの水準数=(4)
全自由度: $\phi_T = 12 - 1 = 11$
因子Aの自由度: $\phi_A = 4 - 1 = 3$
誤差の自由度: $\phi_E = 11 - 3 = 8$
棄却域: $F > F(\phi_A, \phi_E; \alpha) = F(3, 8; 0.05) = 4.07$

(5): H_0を棄却する・しないの判定
F_0 =(6.30) は棄却域に(ある ない)

どっちですか？

(●) $F = F_0$が棄却域にあるので、H_0を棄却する ⇒ H_1は正しいと言える
( ) $F = F_0$が棄却域にないので、H_0は棄却できない ⇒ H_1は正しいかどうか何も言えない

(4)分散分析表を作成し、検定統計量を求めます。

★超重要

要因	(a)平方和 S	(b)自由度 φ	(c)平均平方(分散) V=S/φ	(d)F_0 =V_A/V_E	(d),(e)の大小比較	(e)棄却域の下限
A(因子)	1.56 ÷	3 =	0.52	6.30	>	4.07
E(誤差)	0.66 ÷	8 =	0.0825			
計	2.22	11				

検定統計量