



데이터 분석

데이터 수식화

[데이터 구조식]

$$X_{ij} : \mu + a_i + b_j + e_{ij}$$

$$\begin{array}{ll} i = 1, \dots, l & \sum_i^l a_i = 0 \\ j = 1, \dots, m & \sum_j^m b_j = 0 \\ e_{ij} \sim N(0, \delta E^2) & \end{array}$$

확률 실험을 위한 가정

$$F_0 > F_\alpha (\phi_A, \phi_E) \Rightarrow H_0 \text{ 기각}$$

$$F_0 < F_\alpha (\phi_A, \phi_E) \Rightarrow H_0 \text{ 채택}$$

데이터 구조

	A_1	A_2	A_3	A_4	
B_1	95.74	99.37	101.14	98.17	$\overline{x_{.1}}=98.60$ 394.4
B_2	103.88	100.43	100.47	102.24	$\overline{x_{.2}}=101.75$ 407
B_3	115.82	103.2	105.35	114.79	$\overline{x_{.3}}=109.79$ 439.16
	$\overline{x_{1.}} = 105.14$ 315.42	$\overline{x_{2.}} = 101$ 303	$\overline{x_{3.}} = 102.32$ 309.96	$\overline{x_{4.}} = 105.06$ 315.18	$\overline{\overline{x}} = 103.38$

A_1 = 식료품 및 비주류음료
 A_2 = 주류 및 담배
 A_3 = 오락 및 문화
 A_4 = 음식 및 숙박

B_1 = 코로나 이전
 (2018.01~2019.12)

B_2 = 코로나 진행
 (2020.01~2022.04)

B_3 = 코로나 이후
 (2022.05~2023.10)

$$\sum x = 1240.6$$

$$\sum x^2 = 128671.55$$

$$CT = \frac{T^2}{lm} = 128257.36$$

$$S_A = \frac{\sum T_{i.}^2}{m} - CT = S_A = \sum^l \sum^m (\overline{x_{i.}} - \overline{\overline{x}})^2 = \frac{384861.65}{3} - CT$$

$$S_B = \frac{\sum T_{.j}^2}{l} - CT = S_A = \sum^l \sum^m (\overline{x_{.j}} - \overline{\overline{x}})^2 = \frac{384861.65}{3} - CT$$

$$S_T = \sum^l \sum^m (\overline{x_{ij}} - \overline{\overline{x}})^2 = 414.19$$

$$S_E = S_T - S_A - S_B = 126.22$$

분산분석표

	S	ϕ	V	F_0	$F_\alpha (\phi_A, \phi_E)$
A	29.86	3	9.95	0.472	$F_{0.05}(3, 6) = 4.76$
B	258.11	2	129.055	6.13	$F_{0.05}(2, 6) = 5.14$
E	126.22	6	21.036		
T	414.19	11			

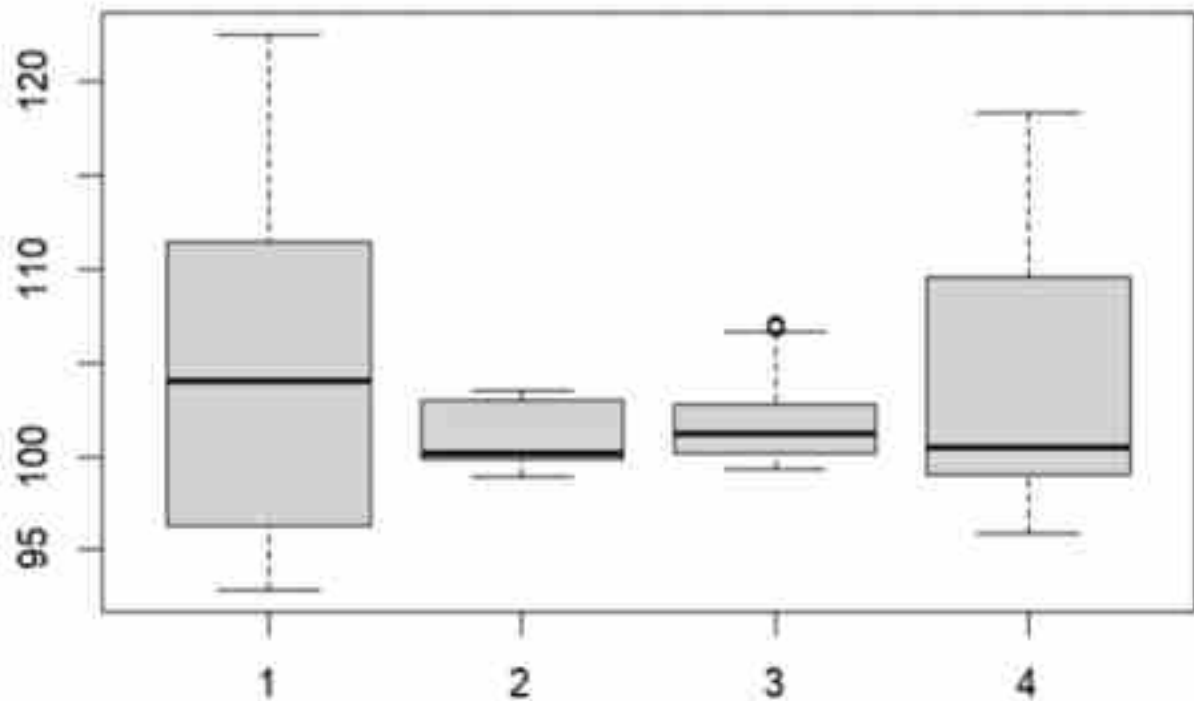


A : $F_0 < F_\alpha \rightarrow H_0$ 채택 \Rightarrow A인자 간의 수준 차이가 없다.

B : $F_0 > F_\alpha \rightarrow H_0$ 기각 \Rightarrow B인자 간의 수준 차이가 있다.

A1 ~ A4 품목별 박스 플롯

*2018.01~2023.10 시점 (각 70개 데이터)



A1 식료품 및 비주류음료

A2 주류 및 담배

A3 오락 및 문화

A4 음식 및 숙박품목

A에 대한 95% 신뢰구간

$$\widehat{\mu}_A = \overline{x_{i.}} \pm t_{\frac{\alpha}{2}}(\phi_E) \cdot \sqrt{\frac{V_E}{m}}$$

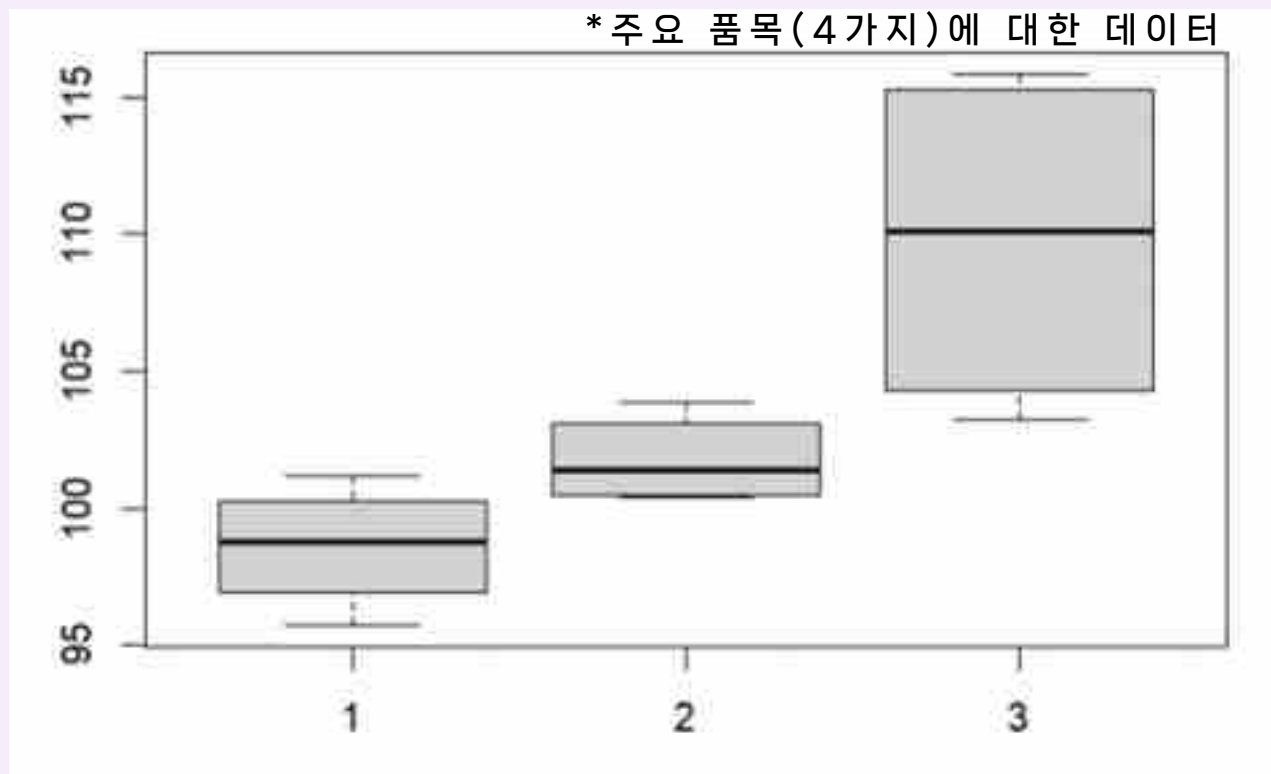
$$\begin{aligned} A_1 &: (105.14 \pm 2.447 \times 2.648) \\ &= (105.14 \pm 6.47) \\ &= (98.67, 111.61) \end{aligned}$$

$$\begin{aligned} A_2 &= (101 \pm 6.47) \\ &= (94.55, 107.47) \end{aligned}$$

$$\begin{aligned} A_3 &= (102.32 \pm 6.47) \\ &= (95.85, 108.79) \end{aligned}$$

$$A_3 = (98.59, 111/53)$$

B1 ~ B4 시기별 박스 플롯



B1 코로나 이전 (2018.01~2019.12)

B2 코로나 진행 (2020.01~2022.04)

B3 코로나 이후 (2022.05~2023.10)

B에 대한 95% 신뢰구간

$$\widehat{\mu}_B = \overline{x}_{\cdot j} \pm t_{\frac{\alpha}{2}}(\phi_E) \cdot \sqrt{\frac{V_E}{l}}$$

$$B_1 : (98.6 \pm t_{0.025}(6) \cdot \sqrt{\frac{21.036}{4}})$$

$$(98.6 \pm t_{0.025}(6) \cdot \sqrt{\frac{21.036}{4}})$$

$$= (98.6 \pm 5.61)$$

$$= (92.99, 104.21)$$

$$B_2 = (96.14, 107.36)$$

$$B_3 = (104.18, 115.37)$$

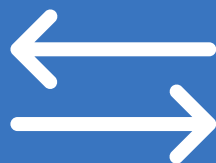
R을 이용한 연구 데이터 분석

```
> # 데이터 프레임 생성
> data <- data.frame(
+   A = factor(rep(c("A1", "A2", "A3", "A4"), each = 3)),
+   B = factor(rep(c("B1", "B2", "B3"), times = 4)),
+   Value = c(95.74, 103.88, 115.82, 99.37, 100.43, 103.2, 101.14, 100.47, 105.35, 98.17, 102.24, 114.79)
+ )
>
> # 이원배치법 모델 적합
> model <- aov(Value ~ A * B, data = data)
>
> # 분산분석 결과 출력
> summary(model)
```

	Df	Sum Sq	Mean Sq
A	3	38.26	12.75
B	2	266.12	133.06
A:B	6	109.81	18.30

결론

A인자 간 차이가 없고,
B인자 간 차이가 있다.



B3(코로나 이후)기간에
물가 변화가 가장 컸다.



코로나19가 실제로
물가 변화에 영향을 미쳤으며,



A1B3 조합(식료품 및 비주류 음료)이
코로나 이후 가장 크게 변했다.