

〈권이태〉

Problem. 1. 한국이는 인구수가 충분히 많은 A시의 소득수준을 조사하였다.

(a) 한국이는 A시에서 3000명의 사람들의 소득을 조사하여, 월 소득이 300만원 이상인 사람들의 수를 조사하였다. 그 결과 월 소득이 300만원 이상인 사람들이 1560명, 그렇지 않은 사람들이 1440명이었다고 하자. 유의수준 0.05에서 월 소득의 중앙값이 300만원이라는 데에 대한 카이제곱 적합성 검정을 수행하여라. 단, $\chi_{0.05}^2(1) = 3.838$, $\chi_{0.05}^2(2) = 5.994$ 임을 이용하여도 좋다.

(b) 한국이는 추가 분석을 위하여 20명에 대한 인터뷰를 통해 만 원 단위의 월 소득을 확인하였다. 그 결과는 아래와 같다. 단위는 만 원이다.

290, 193, 280, 273, 291, 284, 260, 261, 281, 225

311, 1678, 500, 301, 321, 306, 324, 333, 302, 303

데이터 특성에 맞는 적절한 비모수적 검정을 통하여 중앙값이 300만원인지 유의수준 0.05에서 근사적으로 검정하시오.

(c) 중위소득이 300만원인지에 대해 검정하고자 한다. (b)에서 얻은 데이터를 사용할 때, 부호순위검정과 t 검정을 사용하기 어려운 이유, 혹은 그래서 안 되는 이유를 논하시오.

(d) 데이터에 t 검정을 사용해도 되는지 확인하기 위하여 20개의 월 소득 자료를 이용하여 정규분포 가정을 확인하려 한다. **두 개 이상**의 정규분포 검정 방법과 그 검정통계량, (근사적) 기각역을 소개하여라. 실제 데이터로부터 관측된 검정통계량의 값을 계산할 필요는 없다.

Problem. 2. 한국이는 실업급여 제도가 실업자들의 구직의사에 어떤 영향을 미치는지 확인해보고자 한다.

- (a) 314명의 실업자들에 대하여 실업급여를 제공하였다. 그 결과 지급 전과 후에 모두 구직의사가 있는 사람이 101명, 지급 전에는 구직의사가 없었으나 지급 후 구직의사가 생긴 사람이 121명, 지급 전에는 구직의사가 있었으나 지급 후 구직의사가 없어진 사람이 59명, 전후에 모두 구직의사가 없는 사람이 33명이었다. 이에 대한 분할표를 작성하고, 한국이에게 필요한 적절한 귀무가설을 설정한 뒤 유의수준 $\alpha = 0.05$ 에서 이를 (근사적) 검정하여라. $\chi_{0.05}^2(1) = 3.838$ 임을 이용하여도 좋다.

- (b) 한국이는 실업급여 액수에 따라서 구직의사에 미치는 영향이 달라질 것으로 보고 있다. 따라서 난괴법을 통하여 이를 검정해 보고자 한다. A_1, A_2, A_3 는 각각 실업급여가 0원, 100만원, 200만원 지급된 그룹이고, B_1, B_2, B_3 는 사회경제적 위치가 비슷한 실업자들을 묶어둔 블록이다. 표 안의 값은 10점 만점의 구직의사 점수를 의미한다.

	A_1	A_2	A_3
B_1	7	9	10
B_2	3	6	8
B_3	3	4	9

실업급여 액수에 따른 구직의사 정도에 차이가 있는지 유의수준 $\alpha = 0.05$ 에서 검정하여라. $F_{0.05}(2, 4) = 6.944$ 임을 이용하여도 좋다.

- (c) 실업급여 액수가 올라감에 따라 구직의사가 높아지는지를 검정해보고 싶다. 이를 검정하기 위한 비모수적 검정방법을 소개하고, (b)에서 검정통계량을 구하여라.

Problem. 3. 한국이는 이변량 자료에서 해당 자료의 평균이 0인지를 검정하려 한다.

(a) 한국이가 가진 이변량 자료는 $X_i = (X_{i1}, X_{i2})^T \sim_{i.i.d.} N(\mu, I_2)$ 로부터 비롯되었다고 한다. 한국이는 $H_0 : \mu = (0, 0)^T$ 을 검정하려 한다. 따라서 아래의 과정을 통하여 유의수준 $\alpha = 0.0025$ 에서 검정을 진행하였다.

- 먼저, n 개의 랜덤포본 X_1, \dots, X_n 을 구한다.
- X_{i1} 들의 표본평균 \bar{X}_1 , X_{i2} 들의 표본평균 \bar{X}_2 를 구한다.
- 만약 $|\sqrt{n}\bar{X}_1| \geq z_{0.00125} = 3.02$ 이거나 $|\sqrt{n}\bar{X}_2| \geq 3.02$ 이라면, 귀무가설을 기각한다.

한국이가 이러한 방식으로 검정을 진행하면 안 되는 이유를 논하시오.

(b) 한국이는 마지막 단계만

- 만약 $\min\{|\sqrt{n}\bar{X}_1|, |\sqrt{n}\bar{X}_2|\} \geq k$ 이라면, 귀무가설을 기각한다.

로 바꾸었다. 이때 한국이의 검정이 가지는 제1종의 오류를 정확히 0.0025로 하기 위해서는 k 가 어떤 값이 되어야 하는가?

(c) 한국이가 가진 자료가 $X_i = (X_{i1}, X_{i2})^T \sim_{i.i.d.} N(\mu, \Sigma)$ 로부터 비롯되었다고 하자. $\Sigma \succ 0$ 이다. \bar{X} 가 X_i 의 표본평균이라 할 때, $n(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)$ 는 어떤 분포를 따르는가?

(d) (c)의 결과를 이용하여 한국이가 (a)에서 검정하려 한 귀무가설을 유의수준 $\alpha = 0.05$ 에서 검정하는 방법을 설명하여라. $\chi_{0.05}^2(2) = 5.994$ 임을 이용하여도 좋다.

Problem. 4. 다음은 균형 불완비블록계획에 따라 실험한 결과이다.

블록	처리							계
	1	2	3	4	5	6	7	
1	2.10	2.67		2.91				7.68
2		1.14	3.00		3.10			7.24
3			2.92	3.14		2.99		9.05
4				3.13	2.63		2.75	8.51
5	1.85				2.84	3.13		7.82
6		1.82				3.01	2.99	7.82
7	3.54		3.78				2.86	10.18
계	7.49	5.63	9.70	9.18	8.57	9.13	8.60	58.30

(a) 위 데이터에서 데이터 구조식과 제약조건을 써라. 또한 해당 자료가 균형 불완비블록인 4가지 이유를 밝혀라.

(b) 분산분석표를 작성하여라.

요인	S	ϕ	V	F_0
블록(미수정)		-		-
처리(수정)				
오차				-
계			-	-

(c) 모든 처리가 유의하다고 가정하고, 대비

$$\frac{1}{2}[\mu(A_1) + \mu(A_2)] - \mu(A_4)$$

의 95퍼센트 신뢰구간을 구하여라. 단 $t_{0.025}(6) = 2.447, t_{0.025}(8) = 2.306$ 임을 이용하여도 된다.