

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (15점)

다음 각 문제에 대해 답하라. (각 3점)

- (1) 식약청에서 중국에서 수입된 농산물의 농약 잔류검사를 위해 다음과 같은 귀무가설 (null hypothesis)을 세웠다.

H_0 : 농약잔류치는 허용치 이하이다.

중국산 농산물의 잔류치에 대한 자료를 얻은 다음 검정한 결과 유의수준 (level of significance) 0.05에서 귀무가설을 기각하지 않았다. 만약 같은 자료를 이용하여 유의수준 0.01과 0.1에서 검정을 시행한다면 각각 어떤 결과가 나올 것인가?

- (2) 어떤 정책에 대한 지지도를 조사하여 95% 신뢰구간을 구했더니 (35%, 45%)였다. 이 신뢰구간이 의미하는 바는 무엇인가?
- (3) 대학수학능력시험의 결과는 원점수, 변환표준점수, 백분위수 등으로 수험생 개개인에게 주어진다. 어떤 학생의 언어와 외국어 모두 90점인데 변환표준점수는 각각 125점과 145점으로 나타났다. 표준변환점수가 계산되는 과정을 설명하고 이 처럼 원점수는 같은데 변환표준점수가 다르게 나타나는 이유를 설명하라.
- (4) 확률변수 Z_1, Z_2, Z_3, Z_4 가 서로 독립이고 표준정규분포인 $N(0, 1)$ 을 따른다고 가정하자. 다음 각 변수들의 분포를 명시하라.

$$(a) \frac{Z_1^2 + Z_2^2}{Z_3^2 + Z_4^2} \quad (b) \frac{Z_1}{|Z_2|} \quad (c) \frac{Z_1 + Z_2}{\sqrt{Z_3^2 + Z_4^2}}$$

- (5) 확률변수 T 가 자유도 n 인 t -분포를 따를 때에 $t_{0.025}(n)$ 는 $P[T > t_{0.025}(n)] = 0.025$ 을 만족하는 실수값이다. n 이 변함에 따라 $t_{0.025}(n)$ 이 다음과 같이 변한다. n 이 매우 커지면 ($n \rightarrow \infty$) $t_{0.025}(n)$ 은 어떠한 값으로 수렴하는가?

n	$t_{0.025}(n)$
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
10	2.228
20	2.086
:	:
:	:

제 2 문 (10점)

회귀모형

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

을 적합시킨 결과가 다음과 같다. (각 2점)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1422.80	711.40	111.21	<.0001
Error	10	63.97	6.40		
Corrected Total	12	1486.77			

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-65.02	15.07	-4.32	0.0015
X1	1	2.37	0.17	13.86	<.0001
X2	1	0.43	0.07	5.77	0.0002

- (1) 전체 관측치 수 n 은 얼마인가?
- (2) 반응변수(Y)의 분산에 대한 추정치는 얼마인가?
- (3) X_1 의 회귀계수 추정치인 2.37의 의미는 무엇인가?
- (4) 각 설명변수에 대응되는 모수를 각각 $\beta_0, \beta_1, \beta_2$ 라고 표시할 때 $H_0: \beta_1 = \beta_2 = 0$ 의 검정을 유의수준 5%에서 실시하라.
- (5) 첫 번째 관측치가 (120, 70, 50)으로 주어졌다. 이 관측치에 대한 잔차(residual)를 계산하라. 단 자료는 $(Y_i, X_{1i}, X_{2i}), i=1, 2, \dots, n$ 으로 주어져 있다.

제 3 문 (13점)

매번 시행에서 성공의 확률이 $p(0 < p < 1)$ 일 때 첫 번째 성공을 얻기 위해 필요한 시행횟수를 X 라고 하자. 단 각 시행은 서로 독립이다. 즉, i 번째 시행결과가 j 번째 시행 결과에 영향을 미치지 않는다. (단, $i \neq j$)

- (1) 확률변수 X 의 확률분포함수 (probability distribution function)를 구하라. (3점)
- (2) (1)의 결과를 이용하여 확률변수 X 의 기댓값 (expected value)을 구하라. (4점)
- (3) 한국인의 모든 부부가 자녀를 딸 선호사상에 의해서 낳는다고 가정하자. 딸 선호사상이란 첫 아이가 딸이면 더 이상 낳지 않고 첫 아이가 아들이면 둘째를 낳아서 딸이면 그만 낳고 아들이면 또 낳는다. 즉, 딸 선호사상이란 첫 딸을 낳을 때까지 계속 낳고 첫 딸을 낳으면 더 이상 낳지 않는 것이다. 단, 아들과 딸을 낳을 확률은 각각 0.5로 동일하며 임신 중에 초음파 검사 등을 통해 태아성감별을 하여 강제유산 시키는 일은 없다고 가정한다. 이렇게 딸 선호사상으로 모든 부부가 자녀를 낳을 경우 많은 시간이 지난 후 한국인의 성비는 어떻게 되겠는가? (2)의 결과를 이용하여 답하라. (6점)

제 4 문 (12점)

어떤 회사에서 입사시험을 통해 신입직원을 채용하고 있다. 입사시험 성적 우수자가 입사 후 근무성적이 우수한지를 알아보려고 2008년 입사자들을 대상으로 입사시험 성적(A,B,C,D)이 1년 후 근무성적(0-100점)에 어떠한 영향을 미치는가를 조사하고자 한다. 이에 적절한 통계적 분석방법을 제시하라.