

통계학 (강좌) 기말고사 (16:00~18:00)

※ 소수점 셋째자리까지 쓰세요.

1. (총 10점, 각 2점) 다음 명제에 대하여 맞으면 O, 틀리면 X로 답하시오.

- (1) 확률변수 e_i ($i=1, \dots, n$)가 정규모집단의 랜덤포본일때 확률변수 $\sum_{i=1}^n e_i$ 은 정규분포를 따른다.
- (2) $Cov(Y, X)=0$ 또는 $Cor(Y, X)=0$ 이면, Y와 X사이에 아무 관계가 없다고 결론 지을수 있다.
- (3) 동일한 자료와 동일한 귀무가설 하에서 유의확률은 대립가설의 형태에 따라 달라진다.
- (4) 회귀분석의 잔차 제곱합(SSE)은 오차분산에 대한 불편 추정량이다.
- (5) 모평균에 대한 95% 신뢰구간을 구했을 때, 신뢰구간이 모평균 μ_0 를 포함한다면 유의수준 5% 의 양측 검정에서 귀무가설 $H_0: \mu = \mu_0$ 를 기각할 수 없다.

2. (총 17점) 다음의 절편이 없는 단순 선형 회귀모형에 대하여 다음을 답하여라.

$$Y_i = \beta x_i + e_i, \quad e_i \stackrel{iid}{\sim} (0, \sigma^2), \quad i=1, \dots, n.$$

(1) (4점) 최소제곱법에 의한 모수 β 에 대한 추정량(LSE) $\hat{\beta}$ 를 구하시오.

(2) (4점) 최소제곱 추정량 $\hat{\beta}$ 에 대하여 $E(\hat{\beta}), Var(\hat{\beta})$ 를 구하시오.

(3) (9점) 최소제곱법에 의해 얻어진 잔차 $\hat{e}_i = Y_i - \hat{Y}_i$ 에 대하여, $Var(\hat{e}_i) = \left(1 - \frac{x_i^2}{\sum_j x_j^2}\right) \sigma^2$ 임을 보이시오.

3. (총 10점) n 개의 자료 $(x_1, y_1), \dots, (x_n, y_n)$ 을 이용하여 최소제곱회귀직선 $\hat{y} = \hat{\alpha}_1 + \hat{\beta}_1 x$ 를 구하였다. 이들 n 개 자료의 표본평균을 (\bar{x}_n, \bar{y}_n) 이라 하자. 이들 n 개의 자료에 추가로 관측된 $n+1$ 번째 자료 (x_{n+1}, y_{n+1}) 를 포함하여 새롭게 최소제곱회귀직선을 추정하여 $\hat{y} = \hat{\alpha}_2 + \hat{\beta}_2 x$ 를 구하였다. 다음 물음에 답하시오.

(1) (5점) $n+1$ 번째 자료의 설명변수 x_{n+1} 이 n 번째 자료까지의 표본평균 \bar{x}_n 와

일치($x_{n+1} = \bar{x}_n$)할 때 $\hat{\beta}_1 = \hat{\beta}_2$ 임을 보이시오.

(2) (5점) $n+1$ 번째 자료의 설명변수 x_{n+1} 이 n 번째 자료까지의 표본평균 \bar{x}_n 와

일치($x_{n+1} = \bar{x}_n$)하고 $n+1$ 번째 반응변수의 관측값 y_{n+1} 이 n 번째 자료까지의 표본평균 \bar{y}_n 와

일치($y_{n+1} = \bar{y}_n$)할 때 $\hat{\alpha}_1 = \hat{\alpha}_2$ 임을 보이시오.

4. (총 10점) 식품회사가 두 가지 종류의 새로운 통조림을 개발하였다. 두 가지 통조림에 대한 선호도가 지역적으로 차이가 있는지를 알아보기 위하여 A지역에서 160명, B지역에서 140명을 랜덤하게 추출하여 선호도를 조사한 결과 다음과 같다.

	참치 통조림	고등어 통조림
A 지역	95	65
B 지역	81	59

(1) (5점) 두 지역에서 참치통조림에 대한 선호도가 다르다고 할 수 있는지 두 모집단의 비율 검정(Z -검정 통계량 이용)을 이용하여 유의수준 5%에서 검정하여라.

(2) (5점) 두 지역에서 참치통조림에 대한 선호도가 다르다고 할 수 있는지 카이제곱 통계량을 이용하여 검정하고, (1)번 결과와 비교하여라.

5. (총 10점) 전구생산업체인 A회사는 기존 제품에 비해 평균 수명이 2배 넘게 늘어난 새로운 전구를 개발하였다고 광고하고 있다. 이 회사의 주장이 타당한지 검증하기 위해 각각 11개씩의 전구를 랜덤추출한 후 수명을 측정하여 아래 표와 같은 결과를 얻었다.

번호	1	2	3	4	5	6	7	8	9	10	11	평균
기존전구의 수명(X_i)	3	5	6	4	4	4	3	3	5	4	3	4
새전구의 수명(Y_i)	11	12	12	9	7	9	11	13	9	7	10	10

X_i 와 Y_i 들이 각각 서로 독립이고 평균이 μ_1, μ_2 , 분산이 $\sigma^2, 4\sigma^2$ 인 정규분포를 따른다고 가정할 때 가설 $H_0: \mu_2 = 2\mu_1$ vs $H_1: \mu_2 > 2\mu_1$ 을 검정하려 한다. 아래 물음에 답하시오.

(1) (5점) σ^2 이 알려져 있다고 가정할 때 이 가설은 $\bar{Y} - 2\bar{X}$ 의 분포를 이용하여 검정할 수 있다. 귀무가설 하에서 $\bar{Y} - 2\bar{X}$ 의 분포를 구하시오. (풀이과정을 명시하고 μ_1, μ_2, σ^2 을 이용하여 나타내시오.)

(2) (5점) $\sigma^2 = 1.1$ 이라는 사실이 알려져 있는 경우, 유의수준 $\alpha = 0.05$ 에서 가설을 검정하시오.

6. (총 28점) A 마트에서는 빵을 판매하고 있다. 빵을 전시하는 면적과 전시된 선반의 높이가 빵의 매출에 영향을 줄 수 있을 것이라 판단하였다. 전시면적(대,소)과 선반높이(상,중,하)를 달리 하며 총 6가지의 경우에서, 하루 동안 판매된 빵의 개수를 두 번씩 반복하여 관측하였다.

선반높이 \ 전시면적	대		소	
	상	중	상	중
상	30	28	31	35
중	57	51	56	60
하	36	32	35	29

(1) (3점) 이 자료에 대한 적절한 분산분석 모형을 제시하고 그에따른 가정사항을 기술하시오.

(2) (15점) 분산분석표를 작성하고 (10점), (1)의 모형에서 교호작용의 유의성을 유의수준 5%에서 검정하여라.

(3) (10점) (2)의 검정 결과에 따라 설정하게 되는 새로운 모형을 나타내고 (2점), 그 모형에서의 새로운 분산분석표를 작성하시오 (5점). 작성된 분산분석표를 이용하여 적절한 가설을 제시하고 유의수준 5%에서 검정하여라 (3점).

7. (총 25점) A업체는 기능공들의 작업 속도를 향상시키기 위해 새로운 연수 프로그램을 개발하였다. 숙련도가 유사한 8명의 기능공을 랜덤추출하여 연수를 받도록 한 후 작업시간을 측정하여 다음과 같은 결과를 얻었다.

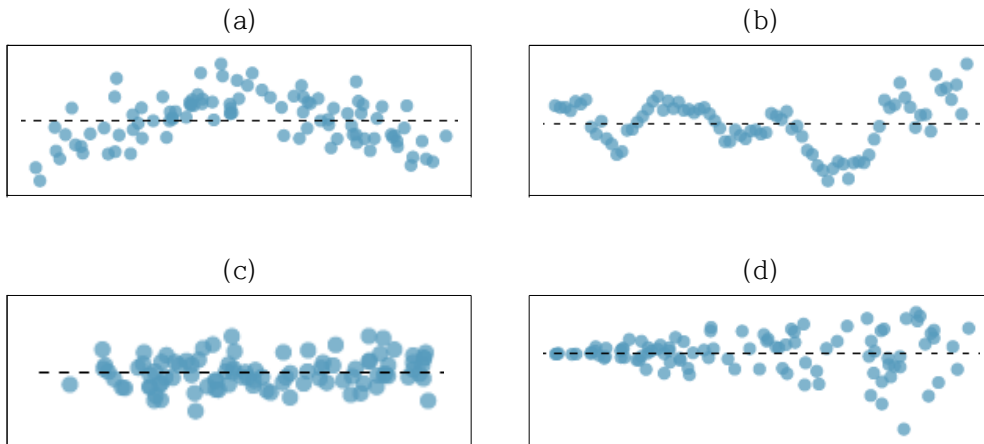
기능공	A	B	C	D	E	F	G	H	평균
작업시간(분)	9	10	6	10	8	9	15	13	10

- (1) (5점) 연수를 받은 기능공들의 작업시간의 기댓값(모평균 μ)을 추정하려 한다. 적절한 가정을 기술하고 μ 에 대한 95% 신뢰구간을 구하시오.
- (2) (10점) 연수프로그램 개발자는 기능공들의 연수 프로그램 이수 시간이 아래 표와 같이 차이가 있으며 연수 프로그램 이수시간과 작업시간 사이에는 유의한 선형관계가 있다고 주장하고 있다. 연수프로그램 개발자의 주장을 수식으로 표현하고 적절한 가정을 서술한 후 이 주장이 타당한지 유의수준 5%에서 F 분포를 이용하여 검정하시오.

기능공	A	B	C	D	E	F	G	H	평균
연수시간(x_i , 시간)	13	9	20	6	16	11	1	4	10
작업시간(Y_i , 분)	9	10	6	10	8	9	15	13	10

(필요한 경우 $\sum x_i^2 = 1080$, $\sum y_i^2 = 856$, $\sum x_i y_i = 681$ 을 이용하시오.)

- (3) (5점) 연수시간이 1시간 늘어나면 작업시간은 β 분 변한다고 할 때 β 의 95% 신뢰구간을 구하시오.
- (4) (5점) 10시간 연수를 받은 기능공의 평균작업시간($E[Y|x=10] = \alpha + 10\beta$)이 11분보다 짧다고 할 수 있는지($H_0: \alpha + 10\beta = 11$ vs $H_1: \alpha + 10\beta < 11$) 유의수준 5%에서 검정하시오.
8. (총 10점) 다음은 단순선형회귀모형을 적합시켜 구한 잔차의 산점도이다. 그림을 이용하여 아래 물음에 답하시오.



- (1) (5점) 단순 선형 회귀모형을 적합하기 위해 필요한 오차항의 세 가지 가정을 기술하시오.
- (2) (5점) (a)~(d)의 잔차도에 대하여 (1)에서 기술한 가정들을 만족한다고 할 수 있는지 답하고 그 이유를 간략하게 서술하시오.

9. (총 20점) 휘발유 종류별로 연비에 차이가 있는지 확인하기 위해 동일한 차종의 차량 18대를 휘발유 종류별로 각각 6대씩 완전 랜덤화 계획에 따라 배치하여 아래 표와 같은 연비 측정값의 통계량을 얻었다. 다음 물음에 답하시오.

	휘발유1	휘발유2	휘발유3	
표본평균	14	13	15	전체평균 : 14
제곱합($\sum x_i^2$)	1180	1024	1360	총제곱합 3564

(문제 (1)~(2)) 먼저 휘발유1과 휘발유2의 연비 차이가 있는지 확인해 보고자 한다.

(1) (5점) 휘발유1과 휘발유2를 사용한 차들의 연비에 대한 검정을 실시하기에 앞서 분산이 동일하다고 말할 수 있는지($H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \text{not } H_0$) 검정하고자 한다. 측정된 연비가 정규분포를 따른다고 가정할 때 이를 유의수준 5%에서 검정하시오.

(2) (5점) 휘발유1과 휘발유2를 사용한 차들의 연비에 차이가 있는지($H_0 : \mu_1 = \mu_2$ vs $H_1 : \text{not } H_0$) 유의수준 5%에서 검정하시오. (1)에서의 검정 결과에 따라 필요한 경우 다음과 같은 t 분포의 근사자유도를 사용하시오.

$$df = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{1}{n_1 - 1}(S_1^2/n_1)^2 + \frac{1}{n_2 - 1}(S_2^2/n_2)^2}$$

(문제 (3)~(4)) 이번에는 세 종류의 휘발유에 대해 연비 차이가 있는지 확인해 보고자 한다.

(3) (3점) 주어진 자료에 대한 일원배치법 모형과 필요한 가정을 함께 제시하시오.

(4) (7점) 분산분석표를 작성하고(5점) 휘발유 종류별로 연비 차이가 있는지 ($H_0 : \mu_1 = \mu_2 = \mu_3$ vs $H_1 : \text{not } H_0$) 유의수준 5%에서 검정하시오.(2점)