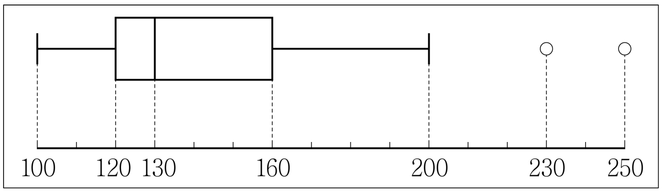
통계학개론

문 1. 다음 상자그림에 대한 설명으로 옳은 것은?



① 자료의 표준편차는 150이다.

② 자료의 중앙값은 130이다.

③ 자료의 최빈값이 중앙값보다 크다.

④ 자료의 사분위수범위는 100이다.

문 2. n개의 자료 $(x_i,\ y_i)$ $(i=1,2,...,n)$의 표본상관계수는 0.2이다. $v_i=x_i+1$, $w_i=2y_i-1$이라고 할 때, 변환된 n개의 자료 $(v_i,\ w_i)$ $(i=1,2,...,n)$의 표본상관계수는?

① 0.2

② 0.3

③ 0.4

④ 0.5

문 3. 어느 회사에서 차량 내비게이션을 연령대가 30대부터 50대까지인 고객에게 한 달 동안 사용하게 한 후 구매 의사를 알아보기 위해 240명을 임의추출하여 다음과 같은 분할표를 작성하였다.

연령대 / 구매 의사	30대	40대	50대	계
있음	40	60	40	140
없음	20	40	40	100
계	60	100	80	240

연령대와 구매 의사는 서로 독립이라는 귀무가설을 검정하기 위한 카이제곱통계량의 $p-$값(유의확률)이 0.128일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. 귀무가설이 참일 때 30대이며 구매 의사가 있는 고객의 수의 기대도수는 35이다.
ㄴ. 귀무가설이 참일 때 카이제곱통계량의 자유도는 2이다.
ㄷ. 유의수준 5%에서 귀무가설을 기각한다.

① ㄴ

② ㄱ, ㄴ

③ ㄱ, ㄷ

④ ㄱ, ㄴ, ㄷ

문 4. 단순선형회귀모형 $y_i=\alpha+\beta x_i+\epsilon_i\ (i=1,2,...,n)$에서 최소제곱법에 의하여 추정된 회귀계수를 $\hat{\alpha}$와 $\hat{\beta}$, y_i의 예측값을 $\hat{y}_i$, 잔차를 $e_i=y_i-\hat{y}_i$, 총제곱합(SST)을 $\sum_{i=1}^{n}(y_i-\bar{y})^2$, 회귀제곱합(SSR)을 $\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$, 잔차제곱합(SSE)을 $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$이라고 할 때, 이에 대한 설명으로 옳지 않은 것은? (단, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이다)

① SSE는 σ^2의 불편추정량(unbiased estimator)이다.

② 결정계수(R^2)는 $\frac{\text{SSR}}{\text{SST}}$이다.

③ $\sum_{i=1}^{n}e_i=0$이다.

④ $\sum_{i=1}^{n}x_ie_i=0$이다.

문 5. 어느 대학교의 4개 단과대학인 인문대, 사회대, 자연대, 공대에서 각각 50명씩 임의추출하여 남학생과 여학생의 수를 조사한 분할표가 다음과 같다.

단과대학 / 성별	인문대	사회대	자연대	공대	계
남학생	20	25	25	30	100
여학생	30	25	25	20	100
계	50	50	50	50	200

4개 단과대학의 남학생 비율이 모두 같다는 귀무가설을 검정하기 위한 카이제곱통계량의 값과 유의수준 5%에서 검정 결과를 옳게 짝 지은 것은? (단, $\chi^2_{\alpha}(k)$는 자유도가 k인 카이제곱분포의 제$100\times(1-\alpha)$ 백분위수를 나타낼 때 $\chi^2_{0.05}(3)=7.81$, $\chi^2_{0.05}(4)=9.49$이다)

	카이제곱통계량의 값	검정 결과
①	4	귀무가설을 기각한다.
②	4	귀무가설을 기각하지 못한다.
③	8	귀무가설을 기각한다.
④	8	귀무가설을 기각하지 못한다.

문 6. 확률변수 X에 대하여 $E[X(X-1)]=3$, $E[X(X+1)]=5$일 때, X의 분산은?

① 3

② 4

③ 5

④ 6

문 7. 갑, 을 두 사람이 가위바위보 게임을 10회 시행하여 갑이 이기는 횟수를 확률변수 X라 할 때, 이에 대한 설명으로 옳은 것은? (단, 갑, 을은 각자 임의로 가위, 바위, 보 중에서 하나를 내고, 비기는 경우도 1회 시행으로 간주한다)

① X의 확률분포는 대칭이다.

② 확률변수 $Y=10-X$는 10회의 시행에서 갑이 지는 횟수이다.

③ $Var(X)=\frac{5}{2}$이다.

④ $E(3X+5)=15$이다.

문 8. 단순선형회귀모형 $y_i=\alpha+\beta x_i+\epsilon_i\ (i=1,2,...,n)$에 최소제곱법을 적용하여 얻은 분산분석표의 일부가 다음과 같을 때, 이에 대한 설명으로 옳지 않은 것은? (단, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이다)

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
회귀	10	(㉠)	10	5	0.017
잔차	36	18	(㉡)		
계	46				

① ㉠과 ㉡의 값은 각각 1과 2이다.

② $F-$값은 회귀평균제곱(MSR)을 잔차평균제곱(MSE)으로 나눈 값이다.

③ 전체 자료의 수 n은 19이다.

④ 유의수준 5 %에서 회귀모형이 유의하다.

문 9. 정규분포 $N(\mu,\sigma^2)$을 따르는 모집단에서 n개의 표본을 임의추출하여 구한 μ의 95 % 신뢰구간이 $(-0.1, 0.7)$일 때, 이에 대한 설명으로 옳지 않은 것은?

① 가설 $H_0:\mu=0$ 대 $H_1:\mu\neq 0$을 검정할 때 유의수준 5 %에서 귀무가설을 기각하지 못한다.

② 표본평균은 0.3이다.

③ μ가 구간 $(-0.1, 0.7)$에 속하지 않을 확률은 0.05이다.

④ μ의 90 % 신뢰구간의 길이는 0.8보다 작다.

문 10. 어느 회사 신입사원의 입사 시험성적은 평균이 μ인 정규분포를 따른다고 한다. 신입사원 중에서 n명을 임의추출하여 구한 시험성적의 표본평균은 60보다 크고, 가설 $H_0:\mu=60$ 대 $H_1:\mu\neq 60$에 대한 $p-$값은 0.1이었다. 동일한 표본에 대해 가설 $H_0:\mu=60$ 대 $H_1:\mu>60$에 대한 $p-$값은 a이고, 가설 $H_0:\mu=60$ 대 $H_1:\mu<60$에 대한 $p-$값은 b일 때, $b-a$의 값은?

① 0.05

② 0.1

③ 0.9

④ 0.95

문 11. 50개의 관측값을 가장 작은 값부터 가장 큰 값까지 크기순으로 나열하여 $x_{(1)}, x_{(2)}, ..., x_{(50)}$으로 나타낼 때, 이에 대한 설명으로 옳지 않은 것은?

① 제20 백분위수는 $x_{(20)}$이다.

② 제3 사분위수는 $x_{(38)}$이다.

③ 중앙값은 $\frac{x_{(25)}+x_{(26)}}{2}$이다.

④ 사분위수범위는 $(x_{(38)}-x_{(13)})$이다.

문 12. 10개의 자료 $(x_1, y_1), (x_2, y_2), ..., (x_{10}, y_{10})$에 대하여 단순선형회귀모형 $y_i=\alpha+\beta x_i+\epsilon_i\ (i=1,2,...,10)$를 적합하고자 한다.

$\overline{x}=20,\ \overline{y}=80,\ \sum_{i=1}^{10}(x_i-\overline{x})^2=20,\ \sum_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})=60$

일 때, 최소제곱법을 이용하여 추정한 회귀계수 $\hat{\alpha}$와 $\hat{\beta}$의 값은? (단, $\overline{x}, \overline{y}$는 각각 x, y의 표본평균이고, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이다)

① $\hat{\alpha}=30,\ \hat{\beta}=2$

② $\hat{\alpha}=30,\ \hat{\beta}=3$

③ $\hat{\alpha}=20,\ \hat{\beta}=2$

④ $\hat{\alpha}=20,\ \hat{\beta}=3$

문 13. 인자의 수준수가 k이고 각 수준에서의 반복수가 n인 일원배치법에서 얻은 분산분석표의 일부가 다음과 같을 때, 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
처리	60		20	(㉠)	<0.001
오차	28	28			
계	88				

① 인자의 수준수 k는 4이다.

② 반복수 n은 8이다.

③ ㉠의 값은 20이다.

④ 인자 수준에 따라 모평균이 같은지를 유의수준 5 %에서 검정할 때 모평균은 모두 같다고 할 수 있다.

문 14. 전체 고등학생이 치른 수학 시험점수는 평균이 60점, 표준편차가 12점인 정규분포를 따른다고 한다. 전체 고등학생 중에서 임의로 뽑은 한 명의 수학 시험점수를 X라 하고, 임의로 뽑은 36명의 수학 시험점수의 표본평균을 Y라 할 때, 이에 대한 설명으로 옳지 않은 것은?

① $E(X)=E(Y)$

② $Var(X)=2\times Var(Y)$

③ $P(X>60)=P(Y>60)$

④ $P(X<72)=P(Y<62)$

문 15. 정규분포 $N(\mu, 2^2)$을 따르는 모집단에서 임의추출한 9개 표본의 표본평균이 12일 때, 가설 $H_0 : \mu = 10$ 대 $H_1 : \mu \neq 10$에 대한 $p-$값과 같은 것은? (단, Z는 표준정규분포를 따르는 확률변수이다)

① $2 \times P(Z \geq 3)$
② $P(Z \geq 3)$
③ $P(Z \leq 3)$
④ $2 \times P(Z \leq 3) - 1$

문 16. 어느 공장에서 생산되는 제품의 불량률 p에 대한 가설 $H_0 : p = 0.04$ 대 $H_1 : p > 0.04$를 검정하려고 한다. 이 공장에서 임의추출한 384개 제품 중 불량품의 개수를 X라고 하자. 기각역으로 $R : \frac{X}{384} > 0.06$을 사용할 때, 제1종 오류를 범할 확률과 같은 것은? (단, Z는 표준정규분포를 따르는 확률변수이다)

① $P(Z > 1.5)$
② $P(Z > 2)$
③ $P(Z > 2.5)$
④ $P(Z > 3)$

문 17. 1부터 13까지의 자연수 중 하나와 4개의 무늬 ♣, ♠, ♥, ◆ 중 하나를 조합하여 각각 적은 같은 크기의 52장의 카드가 있다. 이 52장의 카드에서 임의로 한 장의 카드를 뽑을 때, 이 카드에 ♣ 무늬가 적힌 사건을 A, 숫자 1이 적힌 사건을 B라고 하자. 이에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. $P(A) = \frac{1}{4}$
ㄴ. $P(B) = \frac{1}{13}$
ㄷ. 사건 A와 B는 서로 독립이다.
ㄹ. $P(A^C \cup B) = \frac{9}{13}$ (단, A^C은 A의 여사건이다)

① ㄱ, ㄹ
② ㄴ, ㄷ
③ ㄱ, ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ, ㄹ

문 18. 성공의 확률이 p인 베르누이 시행을 독립적으로 100회 반복할 때 성공 횟수를 확률변수 X라 하자. 이에 대한 설명으로 옳지 않은 것은? (단, $0 < p < 1$)

① $p = \frac{1}{2}$이면 $P(X \leq 49) = P(X \geq 51)$이다.
② X의 분산은 $p = \frac{1}{2}$일 때 최댓값을 가진다.
③ $p < \frac{1}{2}$이면 $P(X \leq 49) < P(X \geq 51)$이다.
④ $p > \frac{1}{2}$이면 $P(X = 49) < P(X = 51)$이다.

문 19. 분산 σ^2이 알려진 정규분포 $N(\mu, \sigma^2)$을 따르는 모집단에서 임의추출한 n개 표본을 이용하여 유의수준 α에서 가설 $H_0 : \mu = \mu_0$ 대 $H_1 : \mu \neq \mu_0$을 검정하려고 한다. 검정통계량이 $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, $0 < \alpha < 1$, μ_0은 상수이고 $\overline{X}$는 표본평균이다)

ㄱ. $|Z| \geq z_\alpha$이면 H_0을 기각한다. (단, z_α는 표준정규분포의 제$100 \times (1-\alpha)$ 백분위수이다)
ㄴ. μ_0이 μ에 대한 $100 \times (1-\alpha)\,\%$ 신뢰구간에 속하지 않으면 $p-$값이 α보다 작다.
ㄷ. $Z^2 \geq \chi^2_\alpha(1)$이면 H_0을 기각한다. (단, $\chi^2_\alpha(k)$는 자유도가 k인 카이제곱분포의 제$100 \times (1-\alpha)$ 백분위수이다)

① ㄴ
② ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

문 20. 한 요인의 3개 수준 A, B, C에서 10회씩 반복 측정한 자료에 대해 요인의 수준에 따라 반응변수 Y의 평균에 차이가 있는지 알아보기 위하여 다음 선형회귀모형을 적합하고자 한다.

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i \ (i = 1, 2, \ldots, 30)$$

위 모형에서 i번째 자료의 요인의 수준이 A이면 $(x_{1i}, x_{2i}) = (0, 0)$, 요인의 수준이 B이면 $(x_{1i}, x_{2i}) = (1, 0)$, 요인의 수준이 C이면 $(x_{1i}, x_{2i}) = (0, 1)$이다. 모수 β_1, β_2의 가설검정 결과에 대한 해석으로 옳은 것은? (단, ϵ_i는 $N(0, \sigma^2)$을 따르고 서로 독립이다)

① 가설 $H_0 : \beta_1 = 0$ 대 $H_1 : \beta_1 \neq 0$에서 H_0을 기각하지 못한다면 수준 A와 수준 B의 반응변수의 평균은 유의한 차이가 있다.
② 가설 $H_0 : \beta_1 = 0$ 대 $H_1 : \beta_1 \neq 0$에서 H_0을 기각한다면 수준 B와 수준 C의 반응변수의 평균은 유의한 차이가 있다.
③ 가설 $H_0 : \beta_2 = 0$ 대 $H_1 : \beta_2 \neq 0$에서 H_0을 기각하지 못한다면 수준 B와 수준 C의 반응변수의 평균은 유의한 차이가 없다.
④ 가설 $H_0 : \beta_2 = 0$ 대 $H_1 : \beta_2 \neq 0$에서 H_0을 기각한다면 수준 A와 수준 C의 반응변수의 평균은 유의한 차이가 있다.