통계학개론

문 1. 다음은 18명의 지능지수(IQ)를 측정한 자료를 줄기-잎 그림으로 정리한 것이다. 이 자료의 중앙값은?

7	8
8	59
9	2899
10	14458
11	689
12	34
13	8

① 102 ② 103
③ 104 ④ 105

문 2. 음료를 판매하는 회사에서 나온 어느 제품의 한 개당 용량은 평균이 100 ml, 표준편차가 10 ml인 정규분포를 따른다고 할 때, 이 제품에서 임의로 추출한 25개의 표본평균이 97 ml 이상 102 ml 이하일 확률은? (단, 아래의 표는 Z가 표준정규분포를 따르는 확률변수일 때 $P(Z \le z)$의 값에 상응하는 z의 값을 나타낸 것이다)

$P(Z \le z)$	z
0.8413	1.0
0.9332	1.5
0.9772	2.0

① 0.6826
② 0.7745
③ 0.8185
④ 0.9104

문 3. 평균이 μ, 분산이 σ^2인 모집단에서 추출한 임의표본(random sample)의 표본평균에 대한 설명으로 옳은 것만을 모두 고른 것은? (단, $0 < \sigma^2 < \infty$이다)

> ㄱ. 모집단이 정규분포를 따를 때 표본평균의 분포는 정규분포이다.
> ㄴ. 표본의 크기가 커질수록 표본평균의 분산은 커진다.
> ㄷ. 표본평균은 모평균의 불편추정량이다.

① ㄱ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

문 4. 다음은 어느 지역에서 80명을 임의로 추출하여 출생한 계절을 조사한 결과이다.

출생계절	봄	여름	가을	겨울
관측도수	28	12	16	24

이 지역의 계절별 출생률이 같다는 귀무가설에 대한 검정에서 카이제곱 검정통계량의 값과 유의수준 5%에서의 검정결과에 대한 설명으로 옳은 것은? (단, $\chi^2_{\alpha}(k)$는 자유도가 k인 카이제곱분포의 $(1-\alpha)\times 100$ 번째 백분위수를 나타내고, $\chi^2_{0.05}(3)=7.8147$, $\chi^2_{0.05}(4)=9.4877$이다)

① 검정통계량의 값이 8로 계절별 출생률이 같다고 할 수 있다.
② 검정통계량의 값이 8로 계절별 출생률이 같다고 할 수 없다.
③ 검정통계량의 값이 9로 계절별 출생률이 같다고 할 수 있다.
④ 검정통계량의 값이 9로 계절별 출생률이 같다고 할 수 없다.

문 5. X와 Y가 확률변수이고 a와 b가 상수일 때, 이에 대한 설명으로 옳지 않은 것은?

① $E(aX+b)=aE(X)+b$
② $Var(aX+b)=a^2 Var(X)$
③ $Cov(aX, bY)=ab\times Cov(X, Y)$
④ X와 Y가 서로 독립이면 $Var(X-Y)=Var(X)-Var(Y)$ 이다.

문 6. 다음은 회사 A, B, C에서 생산하는 직물의 인장강도를 비교하기 위하여 각 회사별로 제품 4개씩을 임의로 추출하여 실험한 결과이다.

(단위 : kg/cm^2)

회사	A	B	C
인장강도	55	52	49
	54	50	51
	52	51	52
	55	51	52
표본평균	54	51	51
표본분산	2	2/3	2

이 결과를 이용하여 일원배치 분산분석표를 작성할 때, ㉠과 ㉡의 값은?

요인	제곱합	자유도	평균제곱	F 값
처리	㉠			
잔차	㉡			
계				

	㉠	㉡
①	6	10
②	6	14
③	24	10
④	24	14

문 7. 표본상관계수 r에 대한 설명으로 옳은 것만을 모두 고른 것은?

> ㄱ. r가 -1이면 산점도에서 모든 관측값은 일직선 위에 있다.
> ㄴ. r가 -0.5이면 단순회귀분석에서 결정계수는 0.25이다.
> ㄷ. r가 양수이면 단순회귀분석에서 기울기의 추정값도 양수이다.

① ㄱ, ㄴ　② ㄱ, ㄷ
③ ㄴ, ㄷ　④ ㄱ, ㄴ, ㄷ

문 8. 어느 공장에서 생산되는 제품의 불량률이 5%라고 한다. 이 공장에서 생산되는 제품을 임의로 추출하여 추출된 순서대로 조사할 때 10번째 제품이 첫 번째 불량품일 확률은?

① 0.05×0.95^9　② 0.05×0.95^{10}
③ $\binom{10}{1}\times0.05^9\times0.95$　④ $\binom{10}{1}\times0.05\times0.95^9$

문 9. 두 변수 X와 Y에 대한 10개의 관측쌍 $(x_1,y_1),(x_2,y_2),\cdots,(x_{10},y_{10})$에 대하여 x_i와 y_i를 다음과 같이 표준화하였다.

$$z_{1i}=\frac{x_i-\overline{x}}{s_x},\ z_{2i}=\frac{y_i-\overline{y}}{s_y},\ i=1,2,\cdots,10.$$

$\sum_{i=1}^{10}z_{1i}z_{2i}=0.9$일 때, X와 Y의 표본상관계수는?

(단, $\overline{x}=\sum_{i=1}^{10}x_i/10$, $\overline{y}=\sum_{i=1}^{10}y_i/10$, $s_x=\sqrt{\sum_{i=1}^{10}(x_i-\overline{x})^2/9}$, $s_y=\sqrt{\sum_{i=1}^{10}(y_i-\overline{y})^2/9}$ 이다)

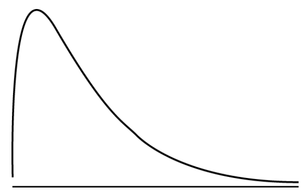
① 0.1　② -0.1
③ 0.9　④ -0.9

문 10. 단순회귀모형 $Y_i=\beta_0+\beta_1X_i+\epsilon_i$, $i=1,2,\cdots,n$에서 최소제곱법으로 구한 추정회귀식이 $\widehat{Y}_i=\hat{\beta}_0+\hat{\beta}_1X_i$이고, 잔차가 $e_i=Y_i-\widehat{Y}_i$일 때, 이에 대한 설명으로 옳은 것만을 모두 고른 것은? (단, $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, $\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$이다)

> ㄱ. $\overline{Y}=\hat{\beta}_0+\hat{\beta}_1\overline{X}$
> ㄴ. $\sum_{i=1}^{n}e_i=0$
> ㄷ. $\sum_{i=1}^{n}\widehat{Y}_i e_i=0$

① ㄱ, ㄴ　② ㄱ, ㄷ
③ ㄴ, ㄷ　④ ㄱ, ㄴ, ㄷ

문 11. 그림과 같이 분포가 오른쪽으로 긴 꼬리를 가지는 확률변수에 대해 최빈값(mode), 평균(mean), 중앙값(median)을 작은 것부터 순서대로 바르게 나열한 것은?



① 최빈값 < 평균 < 중앙값
② 중앙값 < 평균 < 최빈값
③ 최빈값 < 중앙값 < 평균
④ 평균 < 중앙값 < 최빈값

문 12. 작년 입사 지원자 중 60%가 남자이고 40%가 여자인 어느 회사에서 남자 지원자 중 8%, 여자 지원자 중 6%가 합격하였을 때 이 회사의 작년 입사 지원자 중 합격한 사람의 비율은?

① 7.2%　② 7.4%
③ 7.6%　④ 7.8%

문 13. 두 이산확률변수 X와 Y의 결합확률분포표가 다음과 같을 때, $2X-Y$의 기댓값은?

X \ Y	-1	0	1
1	0.1	0.1	0.2
2	0.1	0.2	0.3

① 2.6　② 2.7
③ 2.8　④ 2.9

문 14. 다음은 어느 정책에 대하여 세 도시 A, B, C의 만족도를 조사한 결과이다.

도시 \ 만족도	만족	보통	불만족	합계
A	34	39	12	85
B	29	31	26	86
C	33	29	19	81
합계	96	99	57	252

만약 세 도시 A, B, C에서 이 정책에 대한 만족도의 분포가 동일하다면 B 도시의 '만족' 셀의 기대도수는?

① $\frac{96\times86}{252}$　② $\frac{29}{86}\times\frac{29}{96}\times252$
③ $\frac{96}{3}\times\frac{86}{3}\times\frac{1}{29}$　④ 29

문 15. 인자 A의 처리수준 A_1, A_2, A_3에 대하여 A_1에서 5회, A_2에서 10회, A_3에서 15회를 랜덤하게 실험하였다. 이 실험을 통하여 얻은 자료의 일원배치 분산분석에서 처리제곱합의 자유도(㉠)와 잔차제곱합의 자유도(㉡)를 바르게 연결한 것은?

	㉠	㉡
①	2	26
②	2	27
③	3	26
④	3	27

문 16. 가설검정에서 검정력(power)에 대한 설명으로 옳은 것은?

① 참인 귀무가설을 기각할 확률

② 참인 귀무가설을 기각하지 않을 확률

③ 거짓인 귀무가설을 기각할 확률

④ 거짓인 귀무가설을 기각하지 않을 확률

문 17. 두 확률변수 X와 Y의 상관계수 ρ에 대한 설명으로 항상 옳은 것만을 모두 고른 것은?

ㄱ. X와 Y가 서로 독립이면 $\rho = 0$이다.

ㄴ. $\rho = 1$이면 $Y = X$이다.

ㄷ. $X+1$과 $\dfrac{Y+2}{2}$의 상관계수는 ρ이다.

① ㄱ

② ㄱ, ㄷ

③ ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ

문 18. 어느 지역의 남녀 출생률이 같은지를 검정하기 위하여 이 지역에서 태어난 100명의 신생아를 임의로 추출하여 조사하였더니 이 중 남아가 57명이었다. 이 지역의 남녀 출생률에 대한 설명으로 옳은 것은? (단, Z가 표준정규분포를 따르는 확률변수일 때, $P(Z \geq 1.96) = 0.025$, $P(Z \geq 1.645) = 0.05$, $P(Z \geq 1.282) = 0.1$로 가정한다)

① 유의수준 5%에서 남녀 출생률이 같다고 할 수 없다.

② 유의수준 5%에서 남아의 출생률이 여아의 출생률보다 더 크다고 할 수 있다.

③ 유의수준 10%에서 남아의 출생률이 여아의 출생률보다 더 크다고 할 수 있다.

④ 유의수준 10%에서 여아의 출생률이 남아의 출생률보다 더 크다고 할 수 있다.

문 19. 다음은 어느 식이요법이 몸무게에 미치는 영향을 알아보기 위하여 임의로 추출한 10명의 식이요법 전후 몸무게를 측정하여 얻은 결과이다.

(단위 : kg)

실험대상	1	2	3	4	5	6	7	8	9	10	평균	표준편차
식이요법 전	68	76	74	71	71	72	75	83	75	74	73.9	4.01
식이요법 후	67	77	74	74	69	70	71	77	71	74	72.4	3.34
차이	1	-1	0	-3	2	2	4	6	4	0	1.5	2.68

식이요법 전과 후의 몸무게 평균을 각각 μ_1, μ_2라 할 때, 귀무가설 $H_0 : \mu_1 = \mu_2$ 대 대립가설 $H_1 : \mu_1 \neq \mu_2$에 대한 검정의 유의확률($p-$값)은?

(단, $t_1 = \dfrac{73.9 - 72.4}{\sqrt{\dfrac{9 \times 4.01^2 + 9 \times 3.34^2}{18}} \times \sqrt{\dfrac{1}{10} + \dfrac{1}{10}}}$,

$t_2 = \dfrac{1.5}{2.68/\sqrt{10}}$이며 T_1은 자유도가 18인 t 분포를 따르는 확률변수이며 T_2는 자유도가 9인 t 분포를 따르는 확률변수이다)

① $P(T_1 > t_1)$

② $P(T_2 > t_2)$

③ $2P(T_1 > |t_1|)$

④ $2P(T_2 > |t_2|)$

문 20. $\sum_{i=1}^{n}(X_i - \overline{X})^2 = 80$, $\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = 360$인 두 변수 X와 Y에 대하여 단순회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, 2, \cdots, n$에서 추정된 회귀직선이 $\hat{Y}_i = 0.8 + 2.0 \times X_i$일 때, 잔차제곱합 $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$은? (단, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$이다)

① 40

② 160

③ 200

④ 280