1. 표준편차가 4인 정규분포로부터 나온 확률표본 16개의 표본평균이 10일 때 모평균 μ에 대한 95% 신뢰구간은?
① $10 \pm 1.65 \times 4$ ② $10 \pm 1.96 \times 4$
③ 10 ± 1.65 ④ 10 ± 1.96

2. 확률변수 Y의 표준편차가 2일 때 확률변수 $Z = 3Y$의 분산은?
① 36 ② 18 ③ 6 ④ 12

3. 성공확률이 p인 베르누이 시행을 독립적으로 반복 시행하여 k번 성공할 때까지의 총 시행횟수를 확률변수 X라 하자. 이때 확률변수 X는 어떤 분포를 따르는가?
① 이항분포 ② 초기하분포
③ 기하분포 ④ 음이항분포

4. 다음 중 확률표본추출방법이 아닌 것은?
① 가용표본추출(Available Sampling)
② 단순무작위추출(Simple Random Sampling)
③ 집락(군집)추출(Cluster Sampling)
④ 층화추출(Stratified Sampling)

5. 프로야구 경기에서 연장전을 할 확률이 0.2라고 할 때 10번의 경기에서 연장전을 치르는 경기의 횟수 X는 이항분포 $Bin(10,\ 0.2)$를 따른다. 이때 적어도 한 경기 이상이 연장전으로 이어질 확률은?
① $P(X=1) = \binom{10}{1}(0.2)^1(0.8)^9$
② $1 - P(X=1) = 1 - \binom{10}{1}(0.2)^1(0.8)^9$
③ $1 - P(X=0) = 1 - \binom{10}{0}(0.2)^0(0.8)^{10}$
④ $P(X=0) = \binom{10}{0}(0.2)^0(0.8)^{10}$

6. 상자그림(Box Plot)으로부터 알 수 없는 통계량은?
① 최솟값(Minimum), 최댓값(Maximum)
② 제1사분위수, 제3사분위수
③ 중앙값(Median)
④ 평균(Mean)

7. 서로 독립인 두 확률변수 X, Y가 각각 $N(0,1)$, $N(1,3)$을 따를 때, 확률 $P(X+Y \geq 4.92)$는 얼마인가?
(단, $P(|Z| \geq 1.96) = 0.05$, $P(|Z| \geq 1.645) = 0.1$)
① 0.025 ② 0.05
③ 0.075 ④ 0.1

8. 확률변수 X가 평균 $\mu = 60$과 분산 $\sigma^2 = 16$을 가질 때 체비셰프 부등식에 의한 확률 $P(54 < X < 66)$의 하한은?
① $\frac{1}{4}$ ② $\frac{4}{9}$
③ $\frac{5}{9}$ ④ $\frac{2}{5}$

9. 단순회귀분석에서 독립변수 X와 종속변수 Y의 관계를 파악하기 위해 수집한 자료 10개에 대해 다음과 같은 요약 값을 얻었다. 최소제곱법에 의해 추정된 회귀선은?
$\sum_{i=1}^{10} X_i = 40,\ \sum_{i=1}^{10} X_i^2 = 175,\ \sum_{i=1}^{10} Y_i = 50,\ \sum_{i=1}^{10} Y_i^2 = 215,\ \sum_{i=1}^{10} X_i Y_i = 179$
① $\hat{Y} = 10.6 - 1.4X$
② $\hat{Y} = 10.6 - 2.33X$
③ $\hat{Y} = 8.6 - 1.4X$
④ $\hat{Y} = 8.6 - 2.33X$

10. 다음은 자동차 부품을 생산하는 어느 공장의 현황이다.

	생산량	불량률
기계 1	30%	1%
기계 2	40%	3%
기계 3	30%	2%

이 공장에서 생산되는 제품 가운데 임의로 뽑힌 1개의 제품이 불량품일 때, 기계 1에서 생산되었을 확률은?
① $\frac{1}{7}$ ② $\frac{2}{7}$
③ $\frac{3}{7}$ ④ $\frac{4}{7}$

11. $X_1, X_2, \cdots, X_{12}$가 균일분포 $U(0,2)$의 확률표본이라 할 때 확률변수 $Y = \sum_{i=1}^{12} X_i$의 분포는 중심극한정리에 의해 근사적으로 정규분포 $N(\mu, \sigma^2)$을 따른다. 이때 Y의 평균과 분산은 각각 얼마인가?

	평균	분산		평균	분산
①	12	4	②	12	6
③	6	4	④	6	6

12. $X_1, X_2, \cdots, X_{10}$은 정규분포 $N(\mu_1, \sigma^2)$의 확률표본이고 $Y_1, Y_2, \cdots, Y_{15}$은 정규분포 $N(\mu_2, \sigma^2)$의 확률표본이다. X와 Y는 서로 독립이며 σ^2은 모르는 상수이다. S_1^2과 S_2^2이 각각 X와 Y의 표본분산일 때 공통분산 σ^2의 추정량인 통합표본분산(Pooled Sample Variance) S_p^2를 잘 나타낸 것은?
① $\frac{1}{23}\left[9S_1^2 + 14S_2^2\right]$ ② $\frac{1}{2}\left[S_1^2 + S_2^2\right]$
③ $\frac{1}{23}\left[14S_1^2 + 9S_2^2\right]$ ④ $\frac{1}{25}\left[15S_1^2 + 10S_2^2\right]$

13. 표본의 크기가 n인 어느 자료에 단순회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,\ \epsilon_i \sim N(0,\sigma^2)$을 적용하여 모형의 유의성 검정을 하였다. 다음 설명 중 <u>틀린</u> 것은?

① 모형의 유의성 검정을 위한 귀무가설은 $H_0 : \beta_1 = 0$ 이다.
② Y_i와 ϵ_i의 분산은 같다.
③ 분산분석 검정을 위한 검정 통계량은 귀무가설 하에서 $F(1, n-1)$분포를 따른다.
④ 최소제곱법에 의해 얻은 $\hat{\beta}_1$은 β_1의 불편추정량이다.

14. 확률변수 X의 기댓값 $E(e^{tX})$를 X의 적률생성함수라 할 때, 성공확률이 p인 베르누이 분포의 적률생성함수는? (단, $p+q=1$)

① $p+qe^t$ ② $q+pe^t$
③ $1+e^{pt}$ ④ $e^{pt}+e^{qt}$

15. 표준오차에 대한 설명 중 <u>틀린</u> 것은?

① 표본평균의 분포가 가지는 표준편차, 즉 표본평균의 확률분포가 갖는 흩어짐의 정도를 의미한다.
② 표본의 크기가 클수록 표준오차는 커진다.
③ 일반적으로 표준오차는 모집단의 표준편차보다 작다.
④ 표본평균의 표준편차를 표준오차라 한다.

16. 확률변수 X의 누적분포함수 $F(x)$가 다음과 같이 주어져 있다. 이때 X의 기댓값은?

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x^2, & 0 \le x < 1 \\ \frac{1}{2}, & 1 \le x < 2 \\ 1, & 2 \le x \end{cases}$$

① $\frac{2}{5}$ ② $\frac{2}{4}$
③ $\frac{4}{3}$ ④ $\frac{5}{4}$

17. 분포에 대한 설명 중 <u>틀린</u> 것은?

① 두 확률변수 X, Y가 서로 독립이며 각각 이항분포 $Bin(n,p_1)$, $Bin(m,p_2)$를 따를 때 $Z = X+Y$는 $p_1 = p_2$가 성립하면 이항분포를 따른다.
② 두 확률변수 X, Y가 서로 독립이며 각각 지수분포 $\exp(\theta_1)$, $\exp(\theta_2)$를 따를 때 $Z = X+Y$는 $\theta_1 = \theta_2$가 성립하면 지수분포를 따른다.
③ 두 확률변수 X, Y가 서로 독립이며 카이제곱분포를 따를 때 $Z = X+Y$는 카이제곱분포를 따른다.
④ 두 확률변수 X, Y가 서로 독립이며 표준정규분포를 따를 때 $U = X+Y$, $V = X-Y$로 정의하면 두 확률변수 U, V는 같은 분포를 갖는다.

18. 공평한 동전 4개를 동시에 던졌을 때, 적어도 1개의 앞면이 나올 확률은?

① $\frac{1}{2}$ ② $\frac{1}{4}$
③ $\frac{15}{16}$ ④ $\frac{1}{16}$

19. 베이즈 추론의 세 가지 기본요소에 해당되지 <u>않는</u> 것은?

① 사전분포(Prior Distribution)
② 가능도 함수(Likelihood Function)
③ 표본분포(Sample Distribution)
④ 사후분포(Posterior Distribution)

20. 어느 나라 국민의 2%가 특정 질병에 걸려있다고 한다. 이 병의 검진방법에 의하면 감염자의 98%가 양성 반응을 보이며, 비감염자의 경우도 3%가 거짓양성반응을 보인다고 한다. 한 사람이 검진 결과 양성을 보였다면, 이 사람이 실제로 병에 걸렸을 확률은?

① 0.060 ② 0.196
③ 0.400 ④ 0.490

21. 확률변수 X의 적률생성함수가 다음과 같이 주어졌다.

$$M(t) = E(e^{tX}) = [pe^t + 1 - p]^n,\ 0 < p < 1$$

주어진 적률생성함수를 한 번 미분한 다음 $t = 0$을 대입하면 X의 평균, 즉 $E(X) = M'(0)$임이 알려져 있다. 이때 X의 평균은?

① p ② np
③ $np(1-p)$ ④ $p(1-p)$

22. 성별에 따라 색상에 대한 선호도 차이가 있는지를 알아보기 위해 어느 학교 남녀학생 각각 500명씩을 임의로 추출하여 노랑, 빨강, 파랑 세 가지 색상에 대한 선호도를 조사한 결과가 다음과 같다. 만일, 성별에 따라 선호하는 색상에 차이가 없다면 빨강을 선호하는 여학생 수에 대한 기대도수(Expected Frequency)는?

	노랑	빨강	파랑	합계
남학생	170	150	180	500
여학생	170	200	130	500
합 계	340	350	310	1,000

① 200 ② 175
③ 350 ④ 150

23. 자동차 부품을 생산하는 각 회사의 제품품질이 균일한지 알아보기 위해 네 회사로부터 부품 3개씩을 임의로 추출하여 부품의 강도를 측정한 결과 다음과 같은 분산분석표를 얻었다. a의 값은?

요인	제곱합	자유도	F 비
처리	a		4
잔차			
계	70		

① 25 ② $\frac{23}{1120}$

③ $\frac{3}{140}$ ④ 42

24. $X_1, X_2, \cdots, X_9$는 감마분포 $Gam(4,3)$의 확률표본이고, $\overline{X}$는 표본평균이다. 중심극한정리를 이용하여 구한 $P(8.71 \le \overline{X} \le 15.92)$의 근사확률은?

(단, $P(|Z| \ge 0.65) = 0.5156$, $P(|Z| \ge 0.55) = 0.5824$, $P(|Z| \ge 1.96) = 0.05$, $P(|Z| \ge 1.645) = 0.1$)

① 0.451 ② 0.549

③ 0.85 ④ 0.925

25. $X_1, X_2, \cdots, X_n$이 정규분포 $N(\mu, \sigma^2)$의 확률표본일 때, 확률변수 $Y = (n-1)S^2/\sigma^2$은 자유도 $(n-1)$인 카이제곱 $\chi^2(n-1)$ 분포를 따른다. 다음 기댓값은 알려졌다고 하자.

$$E\left[\frac{1}{S}\right] = \sqrt{\frac{n-1}{2}} \cdot \frac{\Gamma\left(\frac{n}{2}-1\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{1}{\sigma}$$

$n=5$일 때 $\frac{C}{S}$가 $\frac{1}{\sigma}$의 불편추정량이면 C의 값은?

(단, S^2은 표본분산, C는 상수, $\Gamma(1/2) = \sqrt{\pi}$)

① $\sqrt{\pi}$ ② $\sqrt{\frac{\pi}{2}}$

③ $\frac{1}{\sqrt{\pi}}$ ④ $\sqrt{\frac{2}{\pi}}$

26. 어느 도시 성인들의 키(단위: cm)와 몸무게(단위: kg)의 관계를 알아보기 위해 이 도시의 성인 100명을 임의로 뽑아 키를 독립변수, 몸무게를 종속변수로 갖는 단순회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$를 적합한 결과, 회귀식 $\hat{Y} = 0.6X - 32$과 결정계수 $R^2 = 0.75$을 얻었다. 다음 설명 중 <u>틀린</u> 것은?

① 오차인 ϵ_i와 종속변수 Y_i는 동일한 분산을 갖는다.

② 키 X가 $\overline{X}$일 때 몸무게 Y는 $\overline{Y}$이다.

③ 키가 10cm 증가하면 평균 몸무게는 6kg 증가한다.

④ 적합된 회귀식에 의해 키가 몸무게를 설명하는 정도는 75%이므로 키와 몸무게의 피어슨 상관계수는 0.75이다.

27. X_1, X_2, X_3, X_4는 균일분포 $U(0,1)$의 확률표본이고, $X_{(1)} \le X_{(2)} \le X_{(3)} \le X_{(4)}$는 X_1, X_2, X_3, X_4의 순서통계량이라 하자. $E[X_{(2)}] = a$, $P(X_{(4)} < 1/2) = b$라 하면 $a+b$는?

① $\frac{9}{16}$ ② $\frac{11}{35}$

③ $\frac{25}{62}$ ④ $\frac{37}{80}$

28. 상관계수의 성질에 대한 설명 중 <u>틀린</u> 것은?

① 공분산과 상관계수의 부호는 서로 같다.

② 상관계수는 -1과 +1 사이의 값을 갖는다.

③ 두 변수 간 관계의 성질과 강도를 알려준다.

④ 적합된 회귀선 기울기가 양(+)이면 상관계수는 음(-)의 값을 갖는다.

29. 공평한 육면체 주사위를 던져 나오는 수를 X라 하고 조건부 분포 $Y|X$는 성공확률이 0.2인 이항분포 $Bin(X, 0.2)$를 따른다고 하자. 이때 확률변수 Y의 평균 $E(Y)$는?

① 0.3 ② 0.5

③ 0.7 ④ 0.9

30. 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,\ i = 1,2,\cdots,n$ 에 대한 설명 중 <u>틀린</u> 것은?

① 오차항 ϵ_i는 관측될 수 없는 값이기 때문에 오차항의 추정량인 잔차를 이용하여 등분산성, 독립성, 정규성에 대한 타당성을 조사한다.

② 오차항 ϵ_i는 서로 상관을 가지고, 평균이 0, 분산이 σ^2 인 동일한 분포를 따른다.

③ β_0는 회귀직선의 절편이다.

④ β_1은 회귀직선의 기울기이다.

31. 분산분석의 성질에 대한 설명 중 <u>틀린</u> 것은?

① 각 집단의 관측된 데이터 수는 반드시 같아야 한다.

② 3개 이상 집단의 평균을 비교하는 분석방법이다.

③ 각 집단의 분포가 정규분포이고, 개별 관측값은 독립이라는 가정이 필요하다.

④ 총변동을 집단 간의 변동과 집단 내의 변동으로 분해하여 집단 내의 변동보다 집단 간의 변동이 더 큰지를 분석한다.

32. 눈덩이표본추출방법(Snowball Sampling)이 필요한 경우로 가장 적합한 조사대상끼리 짝지어진 것은?

① 도서지역 주민, 교도소 수감자

② 마약중독자, 불법체류자

③ 불법체류자, 국민연금 수령자

④ 기초노령연금 수급자, 문맹자

33. 다음과 같은 확률함수를 갖는 분포에서 얻은 두 확률표본을 X_1, X_2라 하자. 이때 확률 $P(X_1 + X_2 \le 1)$는?

$$f(x) = \begin{cases} x, & 0 \le x \le 1, \\ 2 - x, & 1 < x \le 2. \end{cases}$$

① $\frac{1}{8}$ ② $\frac{1}{11}$

③ $\frac{1}{17}$ ④ $\frac{1}{24}$

34. 다음 중 변동계수에 대한 설명 중 <u>틀린</u> 것은?

① 변동계수는 평균을 표준편차로 나눈 값이다.

② 관측값의 산포 정도를 수치화하여 상대적으로 비교할 때 이용한다.

③ 평균의 차이가 큰 두 집단의 산포 정도를 비교할 때 이용한다.

④ 관측값의 단위가 서로 다른 두 집단의 산포 정도를 비교할 때 이용한다.

35. 회귀분석(Regression Analysis)과 상관분석(Correlation Analysis)에 대한 설명 중 <u>틀린</u> 것은?

① 회귀분석이란 독립변수의 특정한 값에 따른 종속변수의 값을 예측하는 기법이다.

② 상관분석이란 두 변수 사이의 밀접성(선형관계)의 강도(Strength)와 방향(Direction)을 나타내는 수치를 구하는 기법이다.

③ 상관분석만으로도 두 변수 사이에 관계가 있는지는 물론, 관계의 원인과 결과인 인과관계까지를 손쉽게 규명할 수 있다.

④ 회귀분석이나 상관분석을 할 때, 두 변수 사이의 관계를 개략적으로 알아보기 위해 산포도(Scatter Plot)를 그려서 활용한다.

36. 김 대리가 받는 이메일의 20%에는 '행운'이란 단어가 포함되어 있다. '행운'이란 단어가 포함된 이메일의 60%는 광고이고, '행운'이란 단어가 포함되지 않은 이메일의 30%도 광고이다. 김 대리가 받은 한 이메일이 광고일 때, 이 이메일에 '행운'이란 단어가 포함될 확률은?

① $\frac{1}{5}$ ② $\frac{3}{5}$

③ $\frac{24}{100}$ ④ $\frac{1}{3}$

37. 어느 회사에서 현재의 직위가 회사가 마련한 정년연장안에 대한 의견에 영향을 미치는지 알아보기 위해 전체 직원 1,000명을 대상으로 설문조사를 실시하였다. 조사결과의 일부가 표에 주어졌다. 현재의 직위와 정년연장안에 대한 의견이 서로 독립이라고 가정할 때, a에 들어갈 기대도수(Expected Frequency)는 얼마인가?

	정년연장 찬성	정년연장 반대	합계
대리 이하	a		700
과장 이상			300
합 계	350	650	1,000

① 455 ② 245
③ 105 ④ 195

38. $X_1, X_2, \cdots, X_n$이 정규분포 $N(\mu, \sigma^2)$의 확률표본일 때 모평균 μ에 대한 95% 신뢰구간을 구하고자 한다. 모집단의 표준편차 σ를 알고 있는 경우 $\sqrt{n}(\overline{X}-\mu)/\sigma$는 정규분포를 따르며, 모집단의 표준편차 σ를 모르는 경우 $\sqrt{n}(\overline{X}-\mu)/S$는 자유도 $(n-1)$인 $t(n-1)$분포를 따르고 표본크기 $n=5$인 경우 $E(S)=0.94\sigma$가 된다. σ를 모르는 경우와 비교하여 σ를 아는 경우 정보가 많아 신뢰구간의 크기가 감소할 것으로 예상된다. 이때 신뢰구간의 크기가 평균적으로 얼마나 감소하는가?(단, $z_{0.025}=1.96$, $t_{0.025}(4)=2.78$)

① 약 13% ② 약 20%
③ 약 25% ④ 약 33%

39. 모평균 μ에 대한 신뢰구간의 크기를 1/4로 줄이기 위해서는 원래의 표본 크기(n)를 어떻게 변화시켜야 하는가?

① 4배 증가 ② 4배 감소
③ 4^2배 증가 ④ 4^2배 감소

40. 다음 A, B 두 회사는 조사를 실시하여 조사결과를 향후 정책수립에 반영하고자 한다. 다음과 같은 조사 목적에 적합한 귀무가설(H_0)과 대립가설(H_1)은?

	현황과 목표
A 회사	현재 월평균 냉장고 판매량은 500대인데, 새로운 판매보너스 지급정책이 월평균 판매량을 증가시킬 수 있는지 조사하고자 함.
B 회사	생산제품의 용기에 담긴 내용물의 중량이 750g에 미달하거나 초과하는 경우 내용물을 담는 기계를 재조정하려고 조사하고자 함.

		H_0	H_1
①	A 회사	$\mu>500$	$\mu\le 500$
	B 회사	$\mu\neq 750$	$\mu=750$
②	A 회사	$\mu=500$	$\mu>500$
	B 회사	$\mu=750$	$\mu\neq 750$
③	A 회사	$\mu=500$	$\mu\neq 500$
	B 회사	$\mu>750$	$\mu\le 750$
④	A 회사	$\mu\neq 500$	$\mu=500$
	B 회사	$\mu\le 750$	$\mu>750$