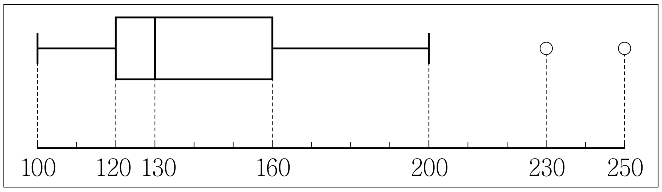
통계학개론

☞ 급하게 해설을 만들다보니 오류가 있을 수 있습니다. 언제든지 문의주세요. ^^

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.

[정답] ②
[출제포인트] 산포도 : 상자(-수염) 그림
[해설]
① (×) 위의 상자그림에 있는 정보로는 표준편차를 구할 수 없다.
② (○) 상자 안에 종으로 그린 선이 중앙값을 나타내므로 자료의 중앙값은 130이다.
③ (×) 그림의 형태를 볼 때 왼쪽으로 치우친 분포를 나타내므로 최빈값<중앙값<평균 순의 대푯값을 갖는다.
④ (×) 자료의 사분위수 범위는 40(= $Q_3 - Q_1$=160-120)이 된다.

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

[정답] ①
[출제포인트] 표본상관계수 계산
[해설]
상관계수의 특징에 의해 임의의 상수 a,b,c,d에 대하여 $X,\ Y$의 상관계수는, $aX+b$와 $cY+d$의 상관계수와 $ac>0$일 때는 동일하며, $ac<0$일 때는 부호만 달라진다. 따라서 $X,\ Y$의 표본상관계수가 0.2이고, $1\times 2>0$이므로 $X+1, 2Y-1$의 표본상관계수는 0.2로 동일하다.

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.

구매 의사 \ 연령대	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

[정답] ②
[출제포인트] 카이제곱 독립성검정
[해설]
ㄱ. (○) 기대도수 $E_{11} = \frac{O_{i.} \times O_{.j}}{n} = \frac{140 \times 60}{240} = 35$이다.
ㄴ. (○) 카이제곱 독립성검정의 검정통계량에 대한 자유도 $(p-1)(q-1) = (2-1)(3-1) = 2$이다.
ㄷ. (×) 카이제곱통계량 χ^2의 유의확률($p-$값)이 유의수준 5%(=0.05)보다 큰 값을 갖게 되므로, 귀무가설은 채택된다, 즉 기각되지 못한다. 여기서 유의확률(p-값)이란 귀무가설(H_0)이 참일 때 표본에서 얻어진 결과가 귀무가설을 기각하게 하는 확률 이다. p-값이 기각치 a보다 작거나 같으면 H_0을 기각하고 p-값이 a보다 크면 H_0을 기각하지 못하게 된다.

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 $\sum_{i=1}^{n}(y_i - \overline{y})^2$, 회귀제곱합(SSR)을 $\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$, 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 $\dfrac{\text{SSR}}{\text{SST}}$이다.

③ $\sum_{i=1}^{n} e_i = 0$이다.

④ $\sum_{i=1}^{n} x_i e_i = 0$이다.

[정답] ①
[출제포인트] 단순회귀모형 : 오차항의 성질
[해설]

① (×) 잔차항의 분산 $Var(\epsilon_i^2) = \sigma^2$에 대한 불편추정량은 $MSE = \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$ 이다.

② (○) 단순회귀분석의 분산분석(총변동의 분해)에 따라 결정계수 $R^2 = \dfrac{SSR}{SST} = 1 - \dfrac{SSE}{SST}$이 된다.

③, ④ (○) 단순회귀분석의 기본 가정이 성립할 때 주어진 잔차의 합과 가중합은 $\sum_{i=1}^{n} e_i = 0$와 $\sum_{i=1}^{n} x_i e_i = 0$라는 성질을 만족한다.

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

성별 \ 단과대학	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.

[정답] ②
[출제포인트] 카이제곱 동질성검정
[해설]
범주형 자료가 주어졌을 때 여러 모집단들이 주어진 특성에 대해 서로 동일한 분포를 갖는지에 대해 검정하는 것을 카이제곱 동질성 검정이라고 한다. 이 때 카이제곱 검정통계량에 대한 자유도는 $(p-1)(q-1) = (2-1)(4-1) = 3$이며, 카이제곱 검정통계량 $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{4}\dfrac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2(3)$을 따른다.

각 셀의 기대도수는 $E_{ij} = E_{11} = \dfrac{O_{i.} \times O_{.j}}{n} = \dfrac{100 \times 50}{200} = 25$이므로,

$$\chi^2 = \frac{(20-25)^2}{25} + \frac{(25-25)^2}{25} + \cdots + \frac{(25-25)^2}{25} + \frac{(25-20)^2}{25} = \frac{5^2 + 5^2 + 5^2 + 5^2}{25} = 4$$

유의수준 5% 하에서의 기각치는 $\chi^2_{0.05}(3) = 7.81$이므로, 주어진 검정통계량은 4이기에 기각치보다 작은 값을 나타내므로 귀무가설을 채택한다. 따라서 유의수준 5% 하에서 "4개 단과대학의 남학생 비율이 모두 같다"는 귀무가설은 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

[정답] ①

[출제포인트] 확률변수의 기댓값과 분산

[해설]

확률변수의 기댓값의 성질에 따라

$E[X(X-1)]=E(X^2-X)=E(X^2)-E(X)=3$

$E[X(X+1)]=E(X^2+X)=E(X^2)+E(X)=5$이 된다.

$E[X(X+1)]-E[X(X-1)]=2E(X)=5-3=2$,

$\therefore E(X)=1, E(X^2)=4\,(=3+E(X))$

확률변수 X의 분산 $Var(X)=E(X^2)-[E(X)]^2=4-1^2=3$이 된다.

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 $Var(X)=\dfrac{5}{2}$이다.

④ $E(3X+5)=15$이다.

[정답] ④

[출제포인트] 확률분포 : 이항분포

[해설]

확률변수 $X\sim B(\frac{10}{3},\frac{1}{3})$인 이항분포를 따르며($n=10, p=\frac{1}{3}, q=1-p=\frac{2}{3}$),

① (×) 이항분포는 $np=\frac{10}{3}<5$이므로 정규분포로 근사하지 않는다.

② (○) 확률변수 $Y=10-X$는 독립 시행을 따르며 X의 선형변환일 뿐이다.

③ (×) 확률변수 X의 분산 $Var(X)=npq=10\times\frac{1}{3}\times\frac{2}{3}=\frac{10}{3}$이다.

④ (×) 기댓값의 성질에 따라 $E(3X+5)=3E(X)+5=3\times\frac{10}{3}+5=15$이 된다.

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.

[정답] ③

[출제포인트] 단순회귀모형과 분산분석

[해설]

① (○) 단순회귀모형이므로 회귀변수의 자유도(㉠)는 1이 되고,
평균제곱오차(㉡) $MSE=\frac{SSE}{n-2}=\frac{36}{18}=2$이 된다.

② (○) 검정통계량 $F=\frac{MSR}{MSE}$이다.

③ (×) 잔차항의 자유도는 $n-2$가 되므로, $\therefore n=18+2=20$

④ (○) 유의확률($p-$값)이 유의수준 5%(=0.05)보다 작으므로 귀무가설을 기각하게 된다. 따라서 유의수준 5% 하에서 위의 회귀모형은 통계적으로 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.

[정답] ③

[출제포인트] 신뢰구간과 가설검정

[해설]

추정량 μ의 분포가 정규분포를 따른다고 할 때.

μ의 95% 신뢰구간은 $\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}, \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)$이 된다.

① (○) 귀무가설 하에서 주어진 신뢰구간이 μ의 95% 신뢰구간 내 포함되는 영역과 포함되지 않는 영역을 모두 갖고 있으므로 귀무가설을 기각하지 못한다.

② (○) 정규분포는 대칭적이기 때문에 주어진 신뢰구간을 이용하여
표본평균 $\overline{X}=0.3\,(\because 2\overline{X}=0.7+(-0.1))$이 된다.

③ (×) 신뢰수준이 95%로 주어졌을 때 신뢰구간의 정의에 따라 추정량 μ가 포함되었을 확률이 95%이라는 것을 의미하며, 이 구간에 속하지 않을 확률은 분명하지 않다.

④ (○) 신뢰수준이 줄어들수록(커질수록) 신뢰구간의 길이는 작아진다(커진다).

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

[정답] ③
[출제포인트] 가설검정 : 단측검정과 양측검정
[해설]
양측검정의 $p-$값

$$p-value = P(|\overline{X}| > |\overline{x}|\,|\,\mu=60) = P\left(|Z| > \left|\frac{\overline{x}-60}{\sigma/\sqrt{n}}\right|\right)$$
$$= P(|Z| > z_{\overline{x}}) = P(Z < -z_{\overline{x}}) + P(Z > z_{\overline{x}}) = a + (1-b) = 0.1$$
$\therefore b-a = 1-0.1 = 0.9$

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

① 제20 백분위수는 $x_{(20)}$이다.
② 제3 사분위수는 $x_{(38)}$이다.
③ 중앙값은 $\dfrac{x_{(25)}+x_{(26)}}{2}$이다.
④ 사분위수범위는 $(x_{(38)}-x_{(13)})$이다.

[정답] ①
[출제포인트] 대푯값 : 중앙값, 사분위수, 백분위수
[해설]
① (×) 백분위수는 주어진 자료를 크기 순으로 나열했을 때 백분율로 나타낸 특정 위치의 값을 나타내는데, 전체 자료 50개 중 제20 백분위수는 10번째 자료가 되므로 $x_{(20)}$이 된다.
② (○) 제3 사분위수 Q_3는 주어진 자료에서 세 번째 사분위점을 의미하므로, Q_3의 위치는

$$[d(M)] + \frac{[d(M)]+1}{2} = \left[\frac{50+1}{2}\right] + \frac{\left[\frac{50+1}{2}\right]+1}{2} = 25+13 = 38\text{이며,}$$
$Q_3 = x_{(38)}$이 된다.
③ (○) 중앙값 M은 주어진 자료에서 2등분이 되는 중앙에 위치하는 값을 의미하므로, M의 위치는 $d(M) = \dfrac{n+1}{2} = \dfrac{50+1}{2} = 25.5$이며, $M = \dfrac{x_{(25)}+x_{(26)}}{2}$가 된다.
④ (○) 자료의 사분위수 범위

$$IQR = Q_3 - Q_1 = x_{(38)} - x_{(13)} \ \left(\because Q_1\text{의 위치는 } \frac{[d(M)]+1}{2} = 13\right)\text{이 된다.}$$

문 12. 10개의 자료 $(x_1, y_1), (x_2, y_2), \ldots, (x_{10}, y_{10})$에 대하여 단순선형회귀모형 $y_i = \alpha + \beta x_i + \epsilon_i \ (i = 1, 2, \ldots, 10)$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 $\hat{\alpha} = 20,\ \hat{\beta} = 3$

[정답] ④
[출제포인트] 단순회귀모형 : 최소제곱법
[해설]
단순선형회귀모형에서의 추정 회귀계수는 각각

$$\hat{\beta} = \frac{\sum_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{10}(x_i-\overline{x})^2} = \frac{60}{20} = 3$$
$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} = 80 - 3\times 20 = 20$이 된다.

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.

[정답] ④
[출제포인트] 일원배치 분산분석 : 반복수가 있는 경우
[해설]
위의 반복수(n)가 있는 일원배치 분산분석표를 구하면 다음과 같다.

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
처리	60	3 (=4-1)	20	㉠=20 (=20/1)	< 0.001
오차	28	28 (=32-4)	1 (=28/28)		
계	88	31 (=32-1)			

① (○) 처리의 자유도가 3이므로, 인자의 수준수 $k=3+1=4$이 된다.
② (○) 오차의 자유도 $N-4=28$, $N=32=n\times k=4n$이므로, 반복수 $n=32/4=8$이 된다.
③ (○) $F=MST/MSE=20/1=20$이 된다.
④ (×) 유의확률이 0.001보다 작은 값으로 유의수준 5%보다 작은 값을 갖게 되므로 귀무가설은 기각되게 되며, 이 때 “인자 수준에 따른 모평균은 다르다”가 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 $P(X<72)=P(Y<62)$

[정답] ②
[출제포인트] 정규분포, 표본평균과 분산의 성질
[해설]
① (○) 동일한 분포를 가지는 확률변수 X, Y의 기댓값은 동일하다.
$\therefore E(X)=E(Y)$
② (○) 또한, 동일한 분포를 가지는 확률변수 X, Y의 기댓값은 동일하다.
$\therefore Var(X)=Var(Y)$
③ (○) $P(X>60)=P(Z>\dfrac{60-60}{12/\sqrt{1}})=P(Z>0)$
$=P(Z>\dfrac{60-60}{12/\sqrt{36}})=P(Y>60)$
④ (×) $P(X<72)=P(Z>\dfrac{72-60}{12/\sqrt{1}}=1)=P(Z>1)$
$=P(Z>\dfrac{62-60}{12/\sqrt{36}})=P(Y>62)$

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

[정답] ①
[출제포인트] 유의확률(p-값)
[해설]
유의확률($p-$값)은

$$p-value=P(|\overline{X}|>|12|\mid\mu=10)=P(|Z|>\frac{12-10}{2/\sqrt{9}})$$
$$=P(|Z|>3)=P(Z<-3)+P(Z<3)$$
$$=2\times P(Z<-3)=2\times P(Z\geq 3)$$

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 $R: \dfrac{X}{384} > 0.06$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 $P(Z > 3)$

[정답] ②
[출제포인트] 제1종 오류를 범할 확률 (단측 검정)
[해설]
제1종 오류는 귀무가설이 참일 때 대립가설을 채택하게 되는 오류를 의미하며,
불량률 $\hat{p} = \dfrac{X}{384}$이라 할 때 제1종 오류를 범할 확률은
$P(p > 0.06 | p = 0.04)$ 이 된다. 모비율이 알려져 있고 n이 충분히 큰 경우라면,
$P(p > 0.06 | p = 0.04)$
$= P(Z > \dfrac{0.06 - 0.04}{\sqrt{(0.04 \times 0.96)/384}}) = P(Z > 2)$

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. $P(A) = \dfrac{1}{4}$

ㄴ. $P(B) = \dfrac{1}{13}$

ㄷ. 사건 A와 B는 서로 독립이다.

ㄹ. $P(A^C \cup B) = \dfrac{9}{13}$ (단, A^C은 A의 여사건이다)

① ㄱ, ㄹ

② ㄴ, ㄷ

③ ㄱ, ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ, ㄹ

[정답] ③
[출제포인트] 사건과 확률
[해설]
ㄱ. (○) $P(A) = \dfrac{n(A)}{N_A} = \dfrac{{}_1C_1}{{}_4C_1} = \dfrac{1}{4}$ 이다.
ㄴ. (○) $P(B) = \dfrac{n(B)}{N_B} = \dfrac{{}_1C_1}{{}_{13}C_1} = \dfrac{1}{13}$ 이다.
ㄷ. (○) 사건 A, B는 임의로 추출할 때 발생하게 되어 각각 영향을 미치지 않게 되므로 서로 독립이다.
ㄹ. (×) $P(A^C \cup B) = P(A^C) + P(B) - P(A^C \cap B)$
$= (1 - P(A)) + P(B) + (1 - P(A)) \times P(B)$
$= (1 - \dfrac{1}{4}) + \dfrac{1}{13} + (1 - \dfrac{1}{4}) \times \dfrac{1}{13} \neq \dfrac{9}{13}$

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 $p = \dfrac{1}{2}$이면 $P(X \le 49) = P(X \ge 51)$이다.

② X의 분산은 $p = \dfrac{1}{2}$일 때 최댓값을 가진다.

③ $p < \dfrac{1}{2}$이면 $P(X \le 49) < P(X \ge 51)$이다.

④ $p > \dfrac{1}{2}$이면 $P(X = 49) < P(X = 51)$이다.

[정답] ③
[출제포인트] 베르누이 시행과 이항분포 근사
[해설]
확률변수(성공횟수) X가 $X \sim B(100, p)$ 인 이항분포를 따르고, n이 충분히 클 때, $Z = \dfrac{X - np}{\sqrt{npq}}$(단, $q = 1 - p$)는 근사적으로 표준정규분포를 따르게 된다.
① (○) $X \sim B(100, \dfrac{1}{2})$인 경우
$P(X \le 49) = P(Z \le \dfrac{49 - 100 \times 1/2}{\sqrt{100 \times 1/2 \times 1/2}}) \fallingdotseq P(Z \le -0.2)$
$= P(Z \ge 0.2) = P(Z \ge \dfrac{51 - 100 \times 1/2}{\sqrt{100 \times 1/2 \times 1/2}}) = P(X \ge 51)$
② (○) $Var(X) = npq$이므로 $n = 100$으로 주어졌을 때 pq를 가장 큰 경우에 $Var(X)$는 최대가 된다. 따라서 $p = \dfrac{1}{2}$일 때 $Var(X)$는 최대값을 가진다.
③ (×) $p < \dfrac{1}{2}$, 예를 들어 $X \sim B(100, \dfrac{1}{4})$이면
$P(X \le 49) = P(Z \le \dfrac{49 - 100 \times 1/4}{\sqrt{100 \times 1/4 \times 3/4}}) \fallingdotseq P(Z \le \dfrac{48}{5\sqrt{3}})$
$P(X \ge 51) = P(Z \ge \dfrac{51 - 100 \times 1/4}{\sqrt{100 \times 1/4 \times 3/4}}) = P(Z \ge \dfrac{52}{5\sqrt{3}})$
$\therefore P(X \le 49) > P(X \ge 51)$
④ (○) $p > \dfrac{1}{2}$, 예를 들어 $X \sim B(100, \dfrac{3}{4})$이면
$P(X = 49) = P(Z = \dfrac{49 - 100 \times 3/4}{\sqrt{100 \times 3/4 \times 1/4}}) \fallingdotseq P(Z = -\dfrac{52}{5\sqrt{3}})$
$P(X = 51) = P(Z \ge \dfrac{51 - 100 \times 3/4}{\sqrt{100 \times 3/4 \times 1/4}}) = P(Z = -\dfrac{46}{5\sqrt{3}})$
$\therefore P(X = 49) < P(X = 51)$

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 $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$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. $Z^2 \geq \chi_\alpha^2(1)$이면 H_0을 기각한다. (단, $\chi_\alpha^2(k)$는 자유도가 k인 카이제곱분포의 제$100 \times (1-\alpha)$ 백분위수이다)

① ㄴ

② ㄷ

③ ㄴ, ㄷ

④ ㄱ, ㄴ, ㄷ

[정답] ③

[출제포인트] 가설검정 : Z검정, 카이제곱검정

[해설]

ㄱ. (×) 양측검정인 경우 $|Z| \geq |z_\alpha|$인 경우 귀무가설을 기각하게 된다.

ㄴ. (○) 신뢰구간이란 신뢰수준이 주어졌을 때 추정하고자 하는 모수가 포함될 확률을 의미하며, 주로 95% 수준을 주로 사용한다. 따라서 μ_0가 신뢰구간에 속하지 않으면 귀무가설을 기각하게 되며 $p-$값이 α보다 작게 된다.

ㄷ. (○) 카이제곱 검정의 경우 양측검정일 때 $Z^2 = \chi^2 \geq \chi_\alpha^2(1)$인 경우 귀무가설을 기각하게 된다.

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.

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

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.

[정답] ④

[출제포인트] 다중회귀분석: 추정계수의 유의성 검정

[해설]

귀무가설 $H_0 : \beta_2 = 0$, 대립가설 $H_1 : \beta_2 \neq 0$이라 할 때, H_0을 기각한다면 x_{2i}의 추정계수가 통계적으로 유의하다는 의미를 갖게 되므로 $\triangle x_{2i}$ 변동을 설명할 수 있는 수준 A와 수준 C의 반응변수의 평균은 유의한 차이가 있다고 할 수 있다.