확률및통계 3차 과제물

제출일 : 12월 4일 수업 전까지
파일명 : 확률및통계(3차과제_학번_이름) ~ 한글, 워드, pdf, 파일 중 하나로 정리하여 LMS 과제란에 업로드하세요.

[문제 1]

두 확률변수 X와 Y의 결합확률분포가 다음과 같을 때, 다음 물음에 답하라.

$$f(x, y) = \begin{cases} \dfrac{6}{5}(x^2 + y), & 0 < x < 1, 0 < y < 1 \\ 0 \quad , & \text{그 외에서} \end{cases}$$

(1) 확률변수 X의 주변확률밀도함수를 구하라.

(2) 확률 $P(0 < X < 0.5, 0 < Y < 0.5)$을 구하라.

(3) 두 확률변수의 상관계수를 구하라.

(4) 두 확률변수는 독립인가? 그 이유는?

solution)

(1) $f(x) = \int_y f(x, y)\,dy = \int_0^1 \frac{6}{5}(x^2 + y)\,dy$

$$= \frac{6}{5}\left[x^2 y + \frac{1}{2}y^2\right]_0^1 = \frac{6}{5}x^2 + \frac{3}{5}, \quad 0 < x < 1$$

(2) $P(0 < X < 0.5, 0 < Y < 0.5) = \int_0^{0.5}\int_0^{0.5} \frac{6}{5}(x^2 + y)\,dx\,dy$

$$= \frac{6}{5}\left[\frac{1}{24}y + \frac{1}{4}y^2\right]_0^{0.5}$$

(3) $f(y) = \int_x f(x,y)\,dx = \int_0^1 \frac{6}{5}(x^2 + y)\,dx = \frac{6}{5}\left[\frac{1}{3}x^3 + xy\right]_0^1 = \frac{6}{5}y + \frac{2}{5}, \quad 0 < y < 1$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x\left(\frac{6}{5}x^2 + \frac{3}{5}\right)dx = \frac{6}{5}\left[\frac{1}{4}x^4 + \frac{1}{4}x^2\right]_0^1 = \frac{3}{5}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y)\,dy = \int_0^1 y\left(\frac{6}{5}y + \frac{2}{5}\right)dy = \frac{6}{5}\left[\frac{1}{3}y^3 + \frac{1}{6}y\right]_0^1 = \frac{3}{5}$$

$$E(X^2)=\int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2\left(\frac{6}{5}x^2+\frac{3}{5}\right)dx = \frac{6}{5}\left[\frac{1}{5}x^5+\frac{1}{6}x^3\right]_0^1 = \frac{11}{25}$$

$$E(Y^2)=\int_{-\infty}^{\infty} y^2 f(y)\,dy = \int_0^1 y^2\left(\frac{6}{5}y+\frac{2}{5}\right)dy = \frac{6}{5}\left[\frac{1}{4}y^4+\frac{1}{9}y^3\right]_0^1 = \frac{13}{30}$$

$$Var(X)=E(X^2)-[E(X)]^2=\frac{2}{25}\,,\quad Var(Y)=E(Y^2)-[E(Y)]^2=\frac{11}{150}$$

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f(x,y)\,dx\,dy = \int_0^1\int_0^1 xy\frac{6}{5}(x^2+y)\,dx\,dy=\frac{7}{20}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=-0.01$$

$$\rho(X,Y)=Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_x\sigma_y}=-0.1306$$

(4) $f(x)f(y)=\left(\frac{6}{5}x^2+\frac{3}{5}\right)\left(\frac{6}{5}y+\frac{2}{5}\right)\neq f(x,y)=\frac{6}{5}(x^2+y)$이므로 두 확률변수 X, Y는 서로 독립이 아니다.

[문제 2]

빨간색 공(R) 3개, 파란색 공(B) 4개, 흰색 공(W)이 5개 들어 있는 항아리에서 임의로 두 개의 공을 선택한다. 다음 물음에 답하라.

(1) 확률변수 X를 뽑힌 공 가운데 빨간색 공의 수, 확률변수 Y를 뽑힌 공 가운데 흰색 공의 수라고 할 때, 확률변수 X와 Y의 결합확률분포를 구하라.

(2) (1)의 결과를 이용하여, 확률변수 X와 확률변수 Y의 주변확률분포를 구하라.

(3) E(X), E(Y), E(XY)를 구하라.

(4) 확률변수 X, Y가 독립인가? 아니라면 공분산을 구하라.

solution)

(1)

X \ Y	0	1	2
0	$\frac{12}{132}$	$\frac{40}{132}$	$\frac{20}{132}$
1	$\frac{24}{132}$	$\frac{30}{132}$	0
2	$\frac{6}{132}$	0	0

(2)

X	0	1	2
$P(X=x)$	$\frac{72}{132}$	$\frac{54}{132}$	$\frac{6}{132}$

Y	0	1	2
$P(Y=y)$	$\frac{42}{132}$	$\frac{70}{132}$	$\frac{20}{132}$

(4) $E(X)=(1\times\frac{54}{132})+(2\times\frac{6}{132})=\frac{1}{2}$, $E(Y)=(1\times\frac{70}{132})+(2\times\frac{20}{132})=\frac{5}{6}$

(5) $P(X=0)P(Y=0)=\frac{21}{121}\neq P(X=0,\ Y=0)=\frac{12}{132}$ 이므로 두 확률변수 X, Y 는 서로 독립이 아니다.

[문제 3]

정규분포 $N(30,\ 4)$를 따르는 모집단으로부터 크기 36인 확률표본을 추출하였다. 이때의 표본평균이 $\overline{X}$라고 할 때, 다음에 답하라.

(1) $P(\overline{X}\geq 30)$의 값은?

(2) $P(\overline{X}\geq a)=0.05$를 만족하는 a의 값은?

(3) $P(-b\leq\overline{X}-30\leq b)=0.95$를 만족하는 b의 값은?

solution)

(1) $P(\overline{X}\geq 30)=P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\geq\frac{30-30}{2/\sqrt{36}}\right)$

$=P(Z\geq 0)=0.5$

(2) $P(\overline{X}\geq a)=P\left(Z\geq\frac{a-30}{0.33}\right)=0.05$, $\frac{a-30}{0.33}=1.645$

$a=30+0.548=30.548$

(3) $P(-b\leq\overline{X}-30\leq b)=0.95=P\left(\frac{-b}{0.33}\leq\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\leq\frac{b}{0.33}\right)$

$=P\left(|Z|\leq\frac{b}{0.33}\right)=0.95$

$\frac{b}{0.33}=1.96$, $a=0.6468$

[문제 4]

성인 남자의 고혈압 환자 비율에 대한 조사를 하고자 한다. 신뢰수준 95%하에서 오차한계는 0.02를 넘지 않기 위해서, 다음 각 경우에 대해 얼마나 많은 표본을 추출해야 하는지를 결정하라.

(1) 실제 고혈압 보유 비율을 알 수 없는 경우

(2) 예년의 조사에 따라 고혈압 환자 비율이 0.07이라고 판단되는 경우

(3) (1)과 (2)의 결과에 따라 어떤 결론을 내릴 수 있는가?

solution)

✎해답 (1) $n \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{d}\right)^2 = \frac{1}{4}\left(\frac{1.96}{0.02}\right)^2 = 2401$

(2) $n \geq p^*(1-p^*)\left(\frac{z_{\alpha/2}}{d}\right)^2 = 0.07 \times 0.93\left(\frac{1.96}{0.02}\right)^2 = 625.22$

즉, 필요한 표본의 크기는 약 626명이다.

(3) 따라서, 아무 정보가 없을 때 더 많은 표본이 필요한 것을 알 수 있다.

[문제 5]

어떤 TV 방송국에서 특집 프로그램을 방영하고 있는데 시청률이 30%라고 한다. 전국에서 900명을 임의로 추출하여 이 특집 프로그램을 시청하고 있는 비율을 추정하고자 한다.

(1) 표본자료의 비율로 전체 시청률을 추정하고자 할 때 적용되는 표본분포를 구하시오.

(2) 표본비율이 32% 이상일 확률을 구하시오. (비율의 표본분포, 이항분포의 정규 근사방법, 연속성 수정을 각각 이용하여 구하여라)

solution)

(1) n=900으로 대표본 CLT, np와 nq>5로 $\hat{p} \sim$ 정규분포를 따른다.

표본비율의 표본분포는 $\hat{p} \sim N(p\,,\, \frac{p(1-p)}{n}) = N(0.3\,,\, 0.015)$

(2) $P(\hat{p} \geq 0.32) = P(Z \geq \frac{0.32-0.30}{0.015}) = P(Z \geq 1.33) = 0.092$

[문제 6]

베르누이 분포로부터 200개의 표본을 추출하여, 성공 71, 실패 129를 얻었다.

(1) 성공의 확률이 30% 보다 큰가를 알아보고자 한다. 가설을 설정하라.

(2) 영가설하에서의 검정통계량 값을 계산하라.

(3) 유의수준 5%에서 검정하고 결론을 설명하라.

(4) 유의확률(p-value)을 구하여 검정하고 결론을 설명하라.

(5) 성공의 확률이 30%와 다른가를 알아보고자 한다면, 가설을 설정하고 신뢰구간을 구하여 검정하라.

solution)

해답 (1) $H_0 : p = 0.3, \quad H_1 : p > 0.3$

(2) $T_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0.355 - 0.3}{\sqrt{\dfrac{0.3(1-0.3)}{200}}} = 1.697$

(3) $C_{0.05} = \{ T_0 > z_{0.05} = 1.645 \} \Rightarrow T_0 > z_{0.05}$이면 H_0 를 기각

$T_0 \not\in C_{0.05}$가 되어 영가설을 채택한다. 즉, 유의수준 $\alpha = 0.05$ 일 때 모비율이 0.3이다.

(4) P-value < 0.05 -> 귀무가설 기각

(5) CI 판정 : (L, U)$\not\in$ 0.3 -> 귀무가설 기각