과제 1

제출 기한 : 10월 18일 금요일 14시 00분

1. 두 확률변수 X, Y 의 결합확률분포가 다음과 같다.

Y \ X	1	2	3
1	0.01	0.02	0.25
2	0.02	0.03	0.2
3	0.02	0.1	0.05
4	0.15	0.1	0.05

(1) $E(X), E(Y)$ 를 구하시오.

$E(X) = 2.35, \quad E(Y) = 2.49$

(2) $Corr(X, Y)$를 구하시오

$Cov(X, Y) = E(XY) - E(X)E(Y) = \ = -0.5815$

$V(X) = E(X^2) - E(X)^2 = 0.6275, \ V(Y) = 1.4099$

$Corr(X, Y) = \ -0.62$

(3) 두 확률변수 X, Y 는 서로 독립인가? 그렇다면 혹은 그렇지 않다면 그 이유는 무엇인가?

상관계수가 -0.62 이므로 음의 선형관계. 따라서 두 확률변수는 독립이 아니다.

2. 다음과 같은 분포를 갖는 무한 모집단을 생각해보자.

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

(1) 주어진 모집단에서 선택한 랜덤 표본을 $(X_1, X_2, \ldots, X_n)$라고 하자. X_i 의 기댓값과 분산을 구하시오.

X_i의 분포는 모집단 분포에 대해 iid.

$E(X_i) = 1, \ V(X_i) = \frac{1}{6}$

(2) (5점) 주어진 모집단에서 크기 $n = 40$ 의 랜덤 표본을 선택하여 $Y = \sum_{i=1}^{40} X_i$ 로 정의하였다. 이 때, $P(Y < k) = 0.9$ 를 만족하는 k를 구하시오. 단, Y 의 분포를 사용하여 답하시오.

(풀이) 중심극한정리에 의해, $(n > 40)$, $Y \dot{\sim} N(40, \ 40/6)$

$P(Y < k) = P(Z < \frac{k-40}{\sqrt{40/6}}) = 0.9, \quad \frac{k-40}{\sqrt{40/6}} = z_{0.1} = 1.28$

$\therefore k = 43.3024$

3. {3,6,9,12,15} 의 다섯 개의 원소로 구성된 크기 $N = 5$의 모집단이 있다. 단순 랜덤 비복원 추출을 사용하여 크기 $n = 3$의 랜덤 표본을 선택하는 경우를 생각해보자.

(1) 표본 평균과 표본 중앙값의 확률 분포표를 각각 구하시오.

$\overline{X}$	6	7	8	9	10	11	12
$P(\overline{X})$	0.1	0.1	0.2	0.2	0.2	0.1	0.1

X_{med}	6	9	12
$P(X_{med})$	0.3	0.4	0.3

(2) 표본 평균과 표본 중앙값은 모두 모평균 $\mu=9$에 대해 비편향 추정량임을 보이시오.

$E(\overline{X})=6*0.1+7*0.1+...+12*0.1=9$

$E(X_{med})=6*0.3+9*0.4+12*0.3=9$

4. 다음은 지난 5년동안 12월 26일과 27일에 고객들에 의해 반품되어 돌아온 제품의 수(X_i)를 조사하여 그 분포를 나타낸 표이다. 올해의 반품 제품수도 동일한 분포를 계속해서 따른다고 한다.

x	1	2	3	4	5	6
$p(x)$	0.25	0.28	0.20	0.17	0.08	0.02

올해 45명의 고객을 랜덤하게 선택하여 조사하였을 때, 평균 반품 제품 개수가 2.9개를 초과할 확률을 근사적으로 구하시오.

(풀이) 주어진 분포표를 사용하여 계산하면, $\mu=2.61,\ \sigma^2=1.34^2$ 이다.
45명의 고객들이 반품한 평균 제품수를 $\overline{X}$ 라고 하면 중심극한 정리에 의해 $\overline{X}\dot{\sim}N(2.61,\ 0.2^2)$를 따른다.
따라서 구하는 확률은 $P(\overline{X}>2.9)=P\left(Z>\frac{2.9-2.61}{0.2}\right)=P(Z>1.45)=0.0735$

5. 한 사회학자는 최근들어 가구원수가 지속적으로 감소하고 있으며, 도시의 평균 가구원수는 2.5명 미만이라고 주장한다. 이를 위해 도시 지역에서 150 가구를 무작위로 추출하여 가구원수를 조사하였고, 그 결과 평균 2.3명, 표준편차 1.5명인 것으로 나타났다. 이 조사결과는 사회학자의 주장을 뒷받침한다고 볼 수 있는가? 유의수준 1%에서 이를 <u>유의확률을 사용하여</u> 검정하시오. 검정 결과의 의미는 무엇인가?

(풀이)
1. 가설 : $H_0: \mu=2.5 \qquad H_1: \mu<2.5$
2. 유의수준 : 0.01
3. 검정통계량 및 관측치 : <u>n>30이므로 CLT에 의해 표본평균은 정규분포를 따른다.</u>

$$t=\frac{\overline{X}-2.5}{s/\sqrt{n}}=-1.633\sim t(149)$$

4. 유의확률 및 검정 결과 : 자유도 30 이상의 t 분포이므로 정규분포로 근사하여 유의확률을 구할 수 있다.
$P(\ T_{(149)}<-1.633)\approx P(\ Z<-1.633)=0.0516$ 이므로 유의확률이 유의수준보다 크기 때문에 귀무가설

을 기각할 수 없다.
5. 검정 결과의 의미 : 가구원수가 지속적으로 감소하고 있다고 볼 만한 충분한 근거가 없다.

6. 제약 회사에서 생산하는 멀티비타민 영양제 안에는 한 알당 비타민 B6가 평균 50mg 포함되어 있고, 통상적으로 1개월마다 1mg의 영양소 감손이 발생하므로 3개월이 지나도 영양제 한 알당 비타민 B6는 47mg이 포함되어 있다고 한다. 소비자 그룹은 제약회사의 주장보다 비타민 감손이 더 많이 발생할 것이고 그 결과, 3개월 후에는 그보다 더 적은 양의 비타민 B6가 영양제에 포함되어 있다고 생각한다. 이에 따라 생산된 지 3개월이 지난 영양제 20알을 무작위로 선택하여 비타민 B6의 용량을 측정하였고, 그 결과 평균 46.94, 표준편차 0.15로 나타났다. 조사된 결과는 소비자 그룹의 주장을 뒷받침 하는가? 유의수준 5%에서 이를 검정하시오. 단, 모집단은 정규분포를 따른다고 가정하자.

(풀이)
1. 가설 : $H_0 : \mu = 47 \qquad H_1 : \mu < 47$
2. 유의수준 : 0.05
3. 검정통계량 및 관측치 : 모집단의 정규분포 가정에 의해 표본평균은 정규분포를 따른다.

$$T = \frac{\overline{X} - 47}{s/\sqrt{n}} = -1.789 \sim t(19)$$

4. 기각역 및 검정 결과 : 기각역이 $T < -t_{0.05}(19) = -1.729$ 이므로 검정통계량의 관측치는 기각역에 속한다. 따라서 귀무가설을 기각할 수 있다.
5. 검정 결과의 의미 : 유의수준 5%에서 비타민 B6의 용량은 제약사의 주장보다 더 적게 포함되어 있다고 볼 수 있는 충분한 통계적 근거가 있다.

7. 자동차 회사에 부품을 납품하는 공장에서는 부품의 품질 확인을 위해 제품의 회전력을 정기적으로 측정하는데, 이 회전력은 평균 125, 표준편차 9.6의 정규분포를 따른다고 알려져 있다. 15개의 무작위로 선택된 부품에 대해 품질 측정을 한 결과, 회전력의 평균이 127.6 으로 나타났면, 부품의 평균 회전력이 달라졌다고 할 수 있는가? 유의수준 5%에서 유의확률을 이용하여 검정하여라. 이 검정 결과의 의미는 무엇인가?
(단, 유의확률의 계산은 분포표로 구한 대략적인 유의확률을 사용하여도 무방합니다.)
(풀이) μ = 제품의 회전력

1. 가설 : $H_0 : \mu = 125 \qquad H_1 : \mu \neq 125$
2. 유의수준 : 0.05
3. 검정통계량 및 관측치 : 모집단의 정규분포 가정에 의해 표본 평균은 정규분포를 따른다.

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{127.6 - 125}{9.6/\sqrt{15}} = 1.049 \sim N(0,1)$$

4. 유의확률 및 검정 결과 : 유의확률이 $P(|Z| > 1.049) = 0.294$ 이므로 주어진 유의수준 0.05보다 크다. 따라서 귀무가설을 기각할 수 없다.
5. 검정 결과의 의미 : 유의수준 5%에서 부품의 평균회전력이 달라졌다고 볼 수 있는 충분한 통계적 근거가 없다.

8. 우리나라 성인들의 일일 TV 시청시간은 평균이 2.98 시간, 표준편차가 2.66 시간인 정규분포를 따른다고 한다. 한 연구자는 대학생들의 경우 인터넷과 게임 등에 시간을 더 소비하기 때문에 TV 시청시간은 이보다 더 적을 것이라고 생각하였다. 이를 위해 75명의 대학생들을 무작위로 선택하여 TV 시청시간을 조사한 결과, 평균 TV 시청시간은 2.5 시간으로 조사되었다. 실제 대학생들의 TV 시청시간의 평균이

2시간이라고 할 때, 유의수준 5%에서의 검정력을 구하시오.

(풀이) 유의수준 5%에서 아래꼬리 검정의 기각역의 임계치는 $-z_{0.05} = -1.645$ 이므로, 귀무가설하에서 표본평균이

$2.98 - 1.645 \times (2.66/\sqrt{75}) = 2.475$ 보다 작은 경우에 귀무가설을 기각하게 된다.

따라서 실제 대학생들의 TV 시청시간의 평균이 2시간이라고 할 때, 검정력은

$$P(\overline{X} < 2.475 \mid \mu = 2) = P\left(\frac{\overline{X} - 2}{2.66/\sqrt{75}} < \frac{2.475 - 2}{2.66/\sqrt{75}}\right) = P(Z < 1.55) = 0.9394$$

9. 과거 자료에 따르면 A 지역에 내린 비의 평균 산성도가 5.2(pH)라고 한다. 최근 들어 이 수치에 변화가 나타났는지를 알아보기 위해 12개의 산성도 조사 자료를 랜덤하게 수집하였고, 그 결과 표본평균이 5.667, 표준편차가 0.921로 나타났다. 이 조사 결과를 통해 최근 이 지역에 내리는 비의 산성도 수치가 변화하였다고 할 수 있는가? 유의수준 5%에서 이를 검정하시오. 검정 결과의 의미는 무엇인가? 단, 모집단은 정규분포를 따른다고 가정하자.

(풀이)

1. 가설 : $H_0 : \mu = 5.2 \qquad H_1 : \mu \neq 5.2$
2. 유의수준 : 0.05
3. 검정통계량 및 관측치 : 모집단은 정규분포를 따르므로 표본평균은 정규분포를 따른다.

$$t = \frac{\overline{X} - 5.2}{s/\sqrt{n}} = 1.76 \sim t(11)$$

4. 기각역 및 검정 결과 : 기각역이 $|T| > t_{0.025}(11) = 2.2$ 이므로 검정통계량의 관측치는 기각역에 속하지 않는다. 따라서 귀무가설을 기각할 수 없다.
5. 결론의 의미 : 유의수준 5%에서 수치가 변화하였다고 할 수 있는 충분한 증거가 없다.

10. 확률변수 $X \sim N(4, 2^2)$에 대해 $P(X^2 - 8X < 2)$ 의 값을 구하시오.

$$\frac{(X-4)^2}{4} \sim \chi^2(1), \quad X^2 - 8X = (X-4)^2 - 16$$

$$P(X^2 - 8X < 2) = P((X-4)^2 - 16 < 2) = P\left(\frac{(X-4)^2}{4} < \frac{18}{4}\right)$$

$$= P(\chi^2(1) < 4.5) = P(-2.12 < Z < 2.12) = 0.966$$