

통계학입문(002), 과제1

소속학부/과	소프트웨어학부-컴퓨터과학과				학 번			2016173			
이 름	이우진				제출일			2020.05.04.(월)			
점 수	1번	2번	3번	4번	5번	6번	7번	8번	9번	10번	소계
											/50

- 과제1은 7쪽에 걸쳐 10문제로 구성되어 있고, 총점은 50점입니다.
- 과제1의 제출기한은 5월 4일 월요일 23시 55분까지입니다. 정해진 기간 안에 문제풀이를 모두 끝내고 과제를 제출할 수 있도록 유의하시기 바랍니다.
- 과제를 제출할 때에는 다음과 같이 “스노우보드 과제 제출하기” 메뉴를 이용하시기 바랍니다.
 - 스노우보드 ▶ 통계학입문(002) 강의실 ▶ 과제 제목 “과제1입니다.” 클릭 ▶ “과제 제출하기” 버튼 클릭 ▶ 과제풀이파일(PDF, XLSX) 업로드하여 제출
 - 과제를 제출한 후에도 제출기한 종료 전까지 수정 및 재제출이 가능하며, 과제를 수정한 후에는 최초 제출시간 기록은 사라지고 수정한 일시로 제출 일자가 변경됩니다.
- 2번부터 6번까지는 계산기를 이용하고, 9번과 10번은 엑셀 추가기능 KESS를 이용하여 풀이합니다.
- 1번부터 8번까지는 다음 두 방법 중 개인별로 편한 방법을 선택하여 풀이합니다.
 - 태블릿 등에서 <과제1 PDF파일>에 직접 수기로 풀이하고, 파일명을 “소속_학번_이름” 으로 저장한 PDF파일을 제출합니다. (파일명 예: 통계학과_2012345_김숙명)
 - <과제1 PDF파일>의 한 쪽이 A4용지 한 면에 인쇄되도록 출력하고, 인쇄용지에 직접 수기로 풀이합니다. 풀이 완료 후 A4 용지를 스캔하여 파일명을 “소속_학번_이름” 으로 저장한 PDF파일을 제출합니다.
- 9번과 10번은 <통계학입문 과제1 자료 엑셀파일>에 새로운 sheet를 생성하여 결과를 분석합니다. 분석 완료 후 파일명을 “소속_학번_이름” 으로 저장한 엑셀파일을 제출합니다. (파일명 예: 통계학과_2012345_김숙명)
- 계산 결과는 소수점 이하 넷째자리에서 반올림하여, 셋째자리까지만 작성하십시오. (예: 9.8765 → 9.877)
- 모호하거나 다수의 답을 제출하지 않도록 하고, 풀이과정 없이 답만 제출하는 것은 인정되지 않으므로 모든 계산과정을 제시하시기 바랍니다.
- 만약 불명확한 점이 있으면 Q&A 게시판을 통해 저에게 질문하여 주시기 바랍니다.
- 상기 규정들을 모두 준수하시어 과제1을 풀이하고 제출하십시오.

1. (주교재 연습문제 4.3번; 5점)

주사위를 던지고 그 결과에 따라 동전을 던지는 실험을 생각하자. 즉 주사위가 홀수가 나오면 동전을 2번 던져 결과를 관측하고 짝수가 나오면 동전을 3번 던져 결과를 관측한다. 표본공간을 나열하라.

$(1, H, H), (1, H, T), (1, T, H), (1, T, T),$
 $(3, H, H), (3, H, T), (3, T, H), (3, T, T),$
 $(5, H, H), (5, H, T), (5, T, H), (5, T, T),$
 $(2, H, H, H), (2, H, H, T), (2, H, T, H), (2, H, T, T),$
 $(2, T, H, H), (2, T, H, T), (2, T, T, H), (2, T, T, T),$
 $(4, H, H, H), (4, H, H, T), (4, H, T, H), (4, H, T, T),$
 $(4, T, H, H), (4, T, H, T), (4, T, T, H), (4, T, T, T),$
 $(6, H, H, H), (6, H, H, T), (6, H, T, H), (6, H, T, T),$
 $(6, T, H, H), (6, T, H, T), (6, T, T, H), (6, T, T, T)$

2. (주교재 연습문제 4.7번; 5점)

숙대에서 세종문화회관까지 택시를 타고 갈 때 걸리는 시간이 X 평균이 30분, 표준편차가 4분인 정규분포라 가정하자. 택시를 타고 시계를 보았더니 약속시간까지 34분이 남았다면, 약속시간에 늦을 가능성은 얼마인가?

$$X \sim N(30, 4^2)$$

이를 표준화시키면

$$Z = \frac{X - \mu}{\sigma}, \quad Z \sim N(0, 1^2)$$

약속시간에 늦을 가능성은 가는데 걸리는 시간 X 가 34분을 초과할 가능성과 같으므로

$$P(X > 34) = P\left(Z > \frac{34 - 30}{4}\right)$$

$$= P(Z > 1)$$

$$= 0.1586 \dots$$

$$\therefore 0.159$$

3. (주교재 연습문제 3.3번; 5점)

$n=18$

다음 자료는 어린이보호구역(30km/h 이내)에서 속도위반으로 단속된 차량의 속도를
줄가-잎 그림으로 정리한 것이다.

4		68	2
5		0012235578	10
6		0113	4
7		7	1
8		0	1

(1) 단속된 차량의 속도에 대한 평균과 표준편차를 계산하라. (2점)

$$\text{평균} = \frac{1}{18} \sum_{i=1}^{18} x_i = \frac{1}{18} \cdot 1029 = 57.1666 \dots \quad \therefore 57.167$$

$$\begin{aligned} \text{표준편차} &= \sqrt{\frac{1}{17} \left\{ \sum_{i=1}^{18} x_i^2 - \frac{1}{18} \left(\sum_{i=1}^{18} x_i \right)^2 \right\}} \\ &= \sqrt{\frac{1}{17} (60241 - 58824.5)} \\ &= 9.1281 \dots \quad \therefore 9.128 \end{aligned}$$

(2) 최고속도로 단속된 차량이 이상점인지를 확인하라. (3점)

*상자누점그림그리기

① 사분위누점범위구하기

$$\text{사분위누점치값 } KQ_1 = 17 \cdot \frac{1}{4} + 1 = 5.25$$

$$KQ_2 = 17 \cdot \frac{2}{4} + 1 = 9.5$$

$$KQ_3 = 17 \cdot \frac{3}{4} + 1 = 12.75$$

$$\text{사분위누 } Q_1 = 51 \times 0.75 + 52 \times 0.25 = 51.25$$

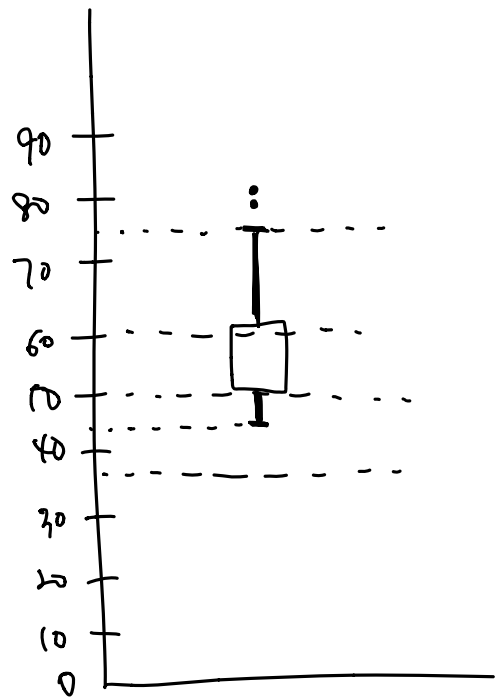
$$Q_2 = 55 \times 0.5 + 55 \times 0.5 = 55$$

$$Q_3 = 60 \times 0.25 + 61 \times 0.75 = 60.75$$

$$\rightarrow IQR = Q_3 - Q_1 = 9.5$$

$$\begin{aligned} \text{상단값 } U &= Q_3 + 1.5 \times IQR \\ &= 75 \end{aligned}$$

$$\begin{aligned} \text{하단값 } L &= Q_1 - 1.5 \times IQR \\ &= 37 \end{aligned}$$



$\therefore 68$ 이상점이다

4. (주교재 연습문제 4.5번; 5점)

한 자동차 보험회사에서는 보험 가입자들을 두 가지 집단(고위험군, 저위험군)으로 나누어 관리하고 있다. 자료에 따르면 보험 가입 후 1년 이내 사고를 일으키는 확률은 고위험군이 0.7, 저위험군이 0.1 이다. 또한 이 회사 보험 가입자들의 10%가 고위험군으로 분류되어 있다. 임의로 추출한 보험 가입자 한명이 1년 내에 사고를 일으키지 않았을 때 이 사람이 저위험군이었을 확률을 구하라.

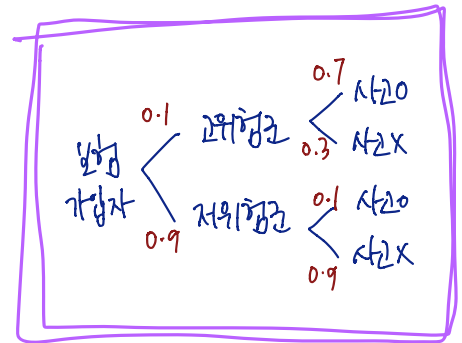
↪ A

↪ B

A가 일어났을 때의 B의 확률 → 조건부 확률이다.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{0.9 \times 0.9}{0.1 \times 0.7 + 0.9 \times 0.9}$$



$$= 0.9642 \dots$$

$$= 0.964$$

5. (주교재 연습문제 4.8번; 5점)

다음은 어느 도시의 임금 근로자들의 부양가족수에 대한 확률분포이다. 이 확률분포의 평균과 표준편차를 구하라.

확률분포

부양가족수	0	1	2	3	4	5
확률	0.10	0.20	0.40	0.20	0.05	0.05

$$E(X) = 0 \times 0.10 + 1 \times 0.20 + 2 \times 0.40 + 3 \times 0.20 + 4 \times 0.05 + 5 \times 0.05$$

$$= 2.05$$

$$\sigma(X) = \sqrt{E(X^2) - E(X)^2}$$

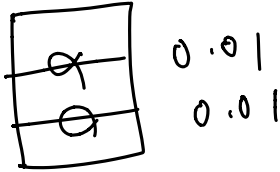
$$= \sqrt{(0 \times 0.10 + 1^2 \times 0.20 + 2^2 \times 0.40 + 3^2 \times 0.20 + 4^2 \times 0.05 + 5^2 \times 0.05) - 2.05^2}$$

$$= 1.2031 \dots$$

$$\therefore 1.203$$

6. (주교재 연습문제 4.6번; 3점)

한 전기제품 내부에 부품 2개가 병렬로 연결되어 있어, 둘 다 고장나지 않는 한 전기제품은 정상적으로 움직인다. 각 부품이 고장나는 사건은 서로 독립적이고 부품이 고장날 확률이 0.01 이라 하면, 이 전기제품이 정상적으로 움직일 확률은 얼마인가?



두 부품이 모두 고장나야 전기제품이 작동하지 않는데 각각 고장날 확률이 독립이다.

독립사건의 교집합의 확률을 곱으로 구하므로
여사건의 확률공식을 이용하면

$$(전체확률) - (두제품이 모두 고장날 확률)$$

$$= 1 - 0.01 \times 0.01$$

$$= 0.9999$$

7. (주교재 연습문제 4.9번; 2점)

→ P103, 189

우리나라 각 도시를 대상으로 커피 전문점수(X)와 이혼건수(Y) 분포로부터의 상관계수가 0.9로 계산되었다고 가정하자. 이로부터 커피전문점수가 늘어나면 이혼건수가 증가(또는 이혼건수가 증가하면 커피전문점수가 증가)한다고 주장하는 것이 논리적으로 타당한지를 밝히고, 만약 타당하지 않다면 그 이유를 간략히 설명해 보라.

타당하지 않다. 상관계수는 X와 Y가 같이 변하는 관계를 측정할 값이고, 문제의 경우 X와 Y는 0.92 양의 상관계수를 가진다. 이는 통계적으로 X와 Y가 같이 증가한다는 방향성과 정도를 나타내는 것이지만, 그 전에 두변수 사이의 인과관계가 우선시되어야 한다. 커피전문점수와 이혼건수는 일반적인 상식선에서 인과관계 등의 직접적 연관이 없다고 판단되므로, 두변수가 상관관계가 있다고 주장하는 것은 논리적으로 타당하지 않다.

8. (주교재 연습문제 2.2번; 5점)

우리나라 통계청 등의 국가통계작성기관에서는 사회·경제적인 변화를 진단하고 합리적인 정책을 수립하는 것을 목적으로 다양한 국가통계를 생산하고 있다. 통계청 사이트를 방문하여, 국가통계작성기관에서는 표본조사를 통해 어떤 국가통계를 생산하고 있는지 살펴보고, 그 중 하나의 통계를 선택해 해당 통계를 생산하기 위해 구체적으로 어떤 방식으로 표본을 추출하여 조사를 하는지 정리해 보라. (상세정리방법은 다음을 참조하시오.)

- (1) 관심 문제 또는 주제(영역)를 선정한다.
- (2) 국가통계포털 사이트(<http://kosis.kr/>)와 통계설명자료 사이트(<http://meta.narastat.kr/>)를 함께 참조하여, 관심 문제/주제와 관련된 표본조사 목록을 살펴본다. (전수조사에 관심이 있다면, 이를 선택해도 무방하다.)
- (3) 표본(전수)조사 목록 중 하나의 조사를 선택하고, 조사항목이나 통계지표들이 관심 문제/주제와 관련된 정보를 제공하는지 확인한다.
- (4) 최종 선택한 표본(전수)조사의 조사명, 작성기관, 조사목적, 모집단, 조사대상, 표본추출방법, 조사방법, 관심 조사항목 또는 결과 등을 간략히 요약하여 정리한다. (국가통계포털 사이트에 제공된 정보가 부족하다면, 관심 표본조사의 작성기관 사이트를 직접 방문하면 더 많은 정보를 수집할 수 있다.)

- 조사명: 공연예술조사 - 공연시설및단체수
- 작성기관: 문화체육관광부
- 조사목적: 객관적이고 신뢰성있는 공연예술분야의 통계산출 및 공연예술시장 현황상태에 대한 체계적 분석으로 예술정책을 위한 기초자료로 활용
- 모집단: 전년도모집단을 기준으로 매년 문화체육관광부, 한국문화예술위원회 등에서 업데이트한 모집단 리스트. 전국의 공연기관, 시설, 단체
- 조사대상: 전국 공연행정기관, 공연시설, 공연단체
- 표본추출방법: 공연행정기관 - X. 전수조사
공연시설
공연단체) 다단계층화계통추출법 (네이만배분+전수조사+지역별비례배분..)
- 조사방법: 면접조사

9. (주교재 연습문제 3.1번; 5점)

별첨된 엑셀자료(sheet명: 9번)는 우리나라 통계청에서 공표한 1980~2018년 인구동태와 관련된 자료를 연도별로 정리한 것이다. 다음 각 변수에 대해 2000년 이전(1980~1999년)과 이후(2000~2018년)의 기술통계값(표본크기, 평균, 표준편차, 중앙값, 최소값, 최대값, 분산, 범위, 사분위수범위, 왜도, 첨도, 사분위수)을 엑셀 추가기능 KESS를 이용하여 계산하고 비교하시오.

※분석변수: 출생아수(명), 사망자수(명), 합계출산율(명), 출생성비(명), 기대수명(계), 기대수명(남), 기대수명(여)

※엑셀 파일명은 “학과_학번_이름”(예: 통계학과_2012345_김숙명)으로, sheet명은 “9번 결과”로 저장하시오.

10. (주교재 연습문제 3.2번; 10점)

별첨된 엑셀자료(sheet명: 10번)는 2018년 12월 31일을 기준으로 수행된 ‘고등교육기관 졸업자 취업통계조사’에서 60개 전공별 졸업현황 자료 일부를 발췌한 것이다. 다음 각 물음에 대하여 엑셀 추가기능 KESS를 이용하여 분석하시오.

※엑셀 파일명은 “학과_학번_이름”(예: 통계학과_2012345_김숙명)으로, sheet명은 “10번 결과”로 저장하시오.

(1) “취업률(남)”과 “취업률(여)” 자료를 비교하기 위해 상자그림을 그리시오. (5점)

(2) 2018년 고등교육기관 졸업자 전체 취업률은 남자 66.7%, 여자 61.8% 라고 한다. 각 전공별 취업률이 전체 취업률 이상인지 미만인지를 구분한 변수인 “기준취업률(남)”과 “기준취업률(여)” 변수를 이용하여 다음 도수분포표를 작성하시오. (5점)

	기준취업률(남)		기준취업률(여)	
	도수	상대도수	도수	상대도수
미만				
이상				
합계				