통계학입문(002), 과제2

소속학부/과	컴퓨터과학과			학 번	2016133			
이 름	이유진			제출일	2020.06.17.(수)			
점 수	1번	2번	3번	4번	5번	6번	7번	8번	소계
									/40

1. 과제2는 8쪽에 걸쳐 8문제로 구성되어 있고, 총점은 40점입니다.
2. 과제2의 제출기한은 6월 17일 수요일 23시 55분까지입니다. 정해진 기간 안에 문제풀이를 모두 끝내고 과제를 제출할 수 있도록 유의하시기 바랍니다.
3. 과제를 제출할 때에는 다음과 같이 "스노우보드 과제 제출하기" 메뉴를 이용하시기 바랍니다.
   - 스노우보드 ▶ 통계학입문(002) 강의실 ▶ 과제 제목 "과제2 입니다." 클릭 ▶ "과제 제출하기" 버튼 클릭 ▶ 과제풀이파일(PDF, XLSX) 업로드하여 제출
   - 과제를 제출한 후에도 제출기한 종료 전까지 수정 및 재제출이 가능합니다. 단, 과제를 수정한 후에는 최초 제출시간 기록은 사라지고 수정한 일시로 제출 일자가 변경됩니다.
4. 1번부터 6번까지는 계산기를 이용하여 풀이하고, 7번과 8번은 엑셀 추가기능 KESS를 이용하여 풀이합니다.
5. 1번부터 6번까지는 다음 두 방법 중 개인별로 편한 방법을 선택하여 작성합니다.
   1) 태블릿 등에서 <과제2 PDF파일>에 직접 수기로 풀이하고, 파일명을 "소속_학번_이름"으로 저장한 PDF파일을 제출합니다. (파일명 예: 통계학과_2012345_김숙명)
   2) <과제2 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.
6. 7번과 8번은 <통계학입문 과제2 자료> 엑셀파일에 새로운 sheet를 생성하여 결과를 분석합니다. 분석 완료 후 파일명을 "소속_학번_이름"으로 저장한 엑셀파일을 제출합니다. (파일명 예: 통계학과_2012345_김숙명)
7. 필요시 계산결과의 소수점 이하 셋째자리에서 반올림하여 둘째자리까지만 제시하십시오. (예: 9.876 → 9.88)
8. 모호하거나 다수의 답을 제출하지 않도록 하고, 풀이과정 없이 답만 제출하는 것은 인정되지 않으므로 모든 계산과정을 제시하시기 바랍니다.
9. 만약 불명확한 점이 있으면 Q&A 게시판을 통해 질문하여 주시기 바랍니다.
10. 상기 규정들을 모두 준수하시어 과제2를 풀이하고 제출하십시오.

$n=10$

$\bar{x}=94.3$

1. (주교재 연습문제 5.2번; 5점)

다음 자료는 어느 도서지역 중학생 10명의 IQ검사 결과이다.

87, 102, 94, 81, 75, 74, 116, 98, 114, 102

(1) 우리나라 중학생 IQ가 평균이 100이고 분산이 $\sigma^2=100$인 정규분포를 따를 때, 이 지역 중학생의 평균 IQ가 우리나라 중학생 평균 IQ보다 낮다고 말할 수 있는지 귀무가설과 대립가설을 세우고 유의수준 $\alpha=0.05$로 검정하라. (4점)

X

(2) (1)번에서 사용된 검정통계치의 유의확률(p-값)을 계산하라. 이 통계치로는 유의수준 얼마까지 귀무가설을 기각할 수 있는가? (1점)

(1) 귀무가설 H_0 : $\mu=100$

대립가설 H_1 : $\mu<100$ → 왼쪽꼬리검정

귀무가설이 참이라 가정하고 표본자료의 발생확률이 낮음을 보인다

$X \sim N(100, 100)$

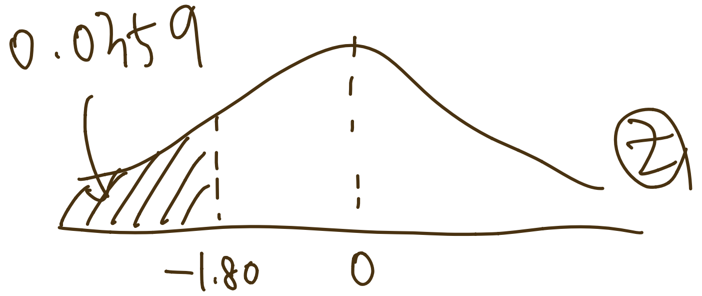
검정통계치 $z=\dfrac{94.3-100}{10/\sqrt{10}}=-1.802\cdots \Rightarrow -1.80$

$z \leq -z_{0.05}=-1.645$ 이므로 귀무가설을 기각한다.

∴ 이지역 중학생의 평균 IQ는 우리나라 중학생 평균 IQ보다 낮다

(2) $P(Z<-1.80)=0.0359$ 이다.

따라서 검정통계치 $z=-1.80$으로 유의수준 0.0359일때까지 귀무가설을 기각할수있다.



X

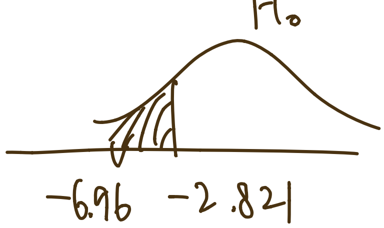
2. (주교재 연습문제 5.3번; 5점) 〈모분산을 모르는 경우〉

기댓값

A질환 유병자 10명의 혈중 칼륨 농도(단위: %)를 측정하여 표본평균 3.4와 표본표준편차 0.5를 얻었다.

(1) 건강한 사람의 혈중 칼륨 농도가 평균 4.5인 정규분포를 따른다면, A질환 유병자의 평균 혈중 칼륨 농도는 유의수준 $\alpha = 0.01$에서 건강한 사람의 평균보다 낮다고 할 수 있는가? (3점)

(2) A질환 유병자의 혈중 칼륨 농도의 모평균에 대한 99% 신뢰구간을 구하라. (2점)



(1) $n=10,\ \bar{x}=3.4,\ S=0.5$

$H_0: \mu = 4.5$

$H_1: \mu < 4.5$ → 왼쪽꼬리검정

검정통계치 $t = \dfrac{\bar{x}-\mu_0}{S/\sqrt{n}} = \dfrac{3.4-4.5}{0.5/\sqrt{10}} = -6.957\cdots \to -6.96$

$t < -t_{0.01}(9) = -2.821$ 이므로 귀무가설 기각 가능

즉 A질환 유병자의 평균 혈중 칼륨 농도는 건강한 사람에 비해 낮다.

(2) 99% 신뢰구간이므로 $\alpha = 0.01$

$$\bar{x} - t_{\alpha/2}(n-1) \times \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}(n-1) \times \frac{S}{\sqrt{n}}$$

$$3.4 - t_{0.005}(9) \times \frac{0.5}{\sqrt{10}} < \mu < 3.4 + t_{0.005}(9) \times \frac{0.5}{\sqrt{10}}$$

$t_{0.005}(9) = 3.25$ 이므로 계산하면

$2.886\cdots < \mu < 3.913\cdots$

정리하면

$2.89 < \mu < 3.91$

3. (주교재 연습문제 5.5번; 5점)

작년 한 해 동안 우리나라 거주 가구 중 20%가 1주일 이상 휴가를 다녀왔다.

어느 여행사가 금년에 조사를 해보니 200가구 중 30가구가 1주일 이상 휴가를 다녀왔다고 응답했다.

(1) 올해 1주일 이상 휴가를 다녀온 가구의 비율이 작년과 달라졌는지 귀무가설과 대립가설을 세우고 유의수준 $\alpha = 0.05$로 검정하라. (4점)

(2) (1)번에서 사용된 검정통계치의 유의확률(p-값)을 구하고, 이 통계치로는 유의수준 얼마까지에서 귀무가설을 기각할 수 있는지 밝혀라. 단, 이 문제의 답은 소수점 넷째자리까지의 값으로 제시하시오. (1점)

(1) $p = 0.2$, $n = 200$

$\hat{p} = \frac{30}{200} = 0.15$

$H_0 : p = 0.2$

$H_1 : p \neq 0.2$ → 양측꼬리검정

검정통계치 $z = \dfrac{0.15 - 0.2}{\sqrt{\dfrac{0.2 \times 0.8}{200}}} = -1.767\cdots = -1.77$

$|z| < z_{0.025} = 1.96$ 이므로 귀무가설을 기각하지 못한다.

(2) 유의확률 $P(|z| \geq 1.77) = P(z \geq 1.77) + P(z \leq -1.77)$

$= 0.0384 \times 2$

$= 0.0768$

4. (주교재 연습문제 5.6번; 5점)

도시 거주자가 농촌 거주자보다 혈중 납 수준(단위: $\mu g/dL$)이 더 높은지 알아보고자 도시 거주자 35명과 농촌 거주자 30명을 임의로 추출하여 다음 표를 얻었다.

거주지	표본크기	표본평균	모표준편차(σ)
X 도시	n_1 35	$\bar{x}$ 14.7	σ_1 7.0
Y 농촌	n_2 30	$\bar{y}$ 9.9	σ_2 4.9

(1) 귀무가설과 대립가설을 세우고 유의수준 $\alpha = 0.01$로 검정하라. (3점)

(2) 도시 거주자와 농촌 거주자의 혈중 납 성분의 모평균 차이에 대한 99% 신뢰구간을 구하라. (2점)

(1) 거주지 도시를 X, 농촌을 Y라고하고 X의 모평균을 μ_1, Y의 모평균을 μ_2 라고하면

$$\begin{cases} H_0 : \mu_1 = \mu_2 \cdots \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 > \mu_2 \cdots \mu_1 - \mu_2 > 0 \rightarrow \text{오른쪽꼬리검정} \end{cases}$$

$$\text{검정통계치 } z = \frac{\bar{x} - \bar{y} - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{14.7 - 9.9 - 0}{\sqrt{\frac{7.0^2}{35} + \frac{4.9^2}{30}}} = 3.235 \cdots \Rightarrow 3.24$$

$z > z_{0.01} = 2.327$ 이므로 귀무가설을 기각한다.

(2) 99% 신뢰구간

$$\bar{x} - \bar{y} - z_{0.005}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x} - \bar{y} + z_{0.005}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$14.7 - 9.9 - 2.575\sqrt{\frac{7.0^2}{35} + \frac{4.9^2}{30}} < \mu_1 - \mu_2 < 14.7 - 9.9 + 2.575\sqrt{\frac{7.0^2}{35} + \frac{4.9^2}{30}}$$

계산하면

$$0.980 \cdots < \mu_1 - \mu_2 < 8.619 \cdots$$

정리하면 $0.98 < \mu_1 - \mu_2 < 8.62$

5. (주교재 연습문제 5.7번; 5점)

벼의 시험재배구역 10구역 중 4곳은 재래종을 6곳은 개량종을 심어서 수확량(단위: kg)을 조사하였다.

재래종과 개량종의 수확량에 대한 모분산이 같음을 가정할 수 있을 때, 다음 문제를 풀이하시오.

X 재래종	30, 34, 29, 27
Y 개량종	23, 25, 28, 29, 30, 27

$n_1=4, \bar{x}=30$

$n_2=6, \bar{y}=27$

(1) 합동표본분산 S_p^2을 구하라. (2점)

(2) 모집단 평균 수확량의 차이의 90% 신뢰구간을 구하라. (3점)

(1) 모분산을 모르지만 등분산 가능이 되는 경우.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$S_1^2 = \frac{1}{n_1-1}\left\{\sum_{i=1}^{n_1} x_i^2 - \frac{1}{n_1}\left(\sum_{i=1}^{n_1} x_i\right)^2\right\} = \frac{26}{3}$$

$$S_2^2 = \frac{1}{n_2-1}\left\{\sum_{i=1}^{n_2} y_i^2 - \frac{1}{n_2}\left(\sum_{i=1}^{n_2} y_i\right)^2\right\} = \frac{34}{5}$$

정리하면

$$S_p^2 = \frac{(4-1)\cdot\frac{26}{3} + (6-1)\cdot\frac{34}{5}}{4+6-2} = \frac{60}{8} = 7.5$$

(2) 90% 신뢰구간이므로 $\alpha=0.1$

$$(\bar{x}-\bar{y}) - t_{\alpha/2}(n_1+n_2-2) \times S_p \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2$$
$$< (\bar{x}-\bar{y}) - t_{\alpha/2}(n_1+n_2-2) \times S_p \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(30-27) - t_{0.05}(8) \times \sqrt{7.5} \times \sqrt{\frac{1}{4}+\frac{1}{6}} < \mu_1-\mu_2$$
$$< (30-27) + t_{0.05}(8) \times \sqrt{7.5}\sqrt{\frac{1}{4}+\frac{1}{6}}$$

$t_{0.05}(8) = 1.86$ 이므로 계산하면

$$-0.288\cdots < \mu_1-\mu_2 < 6.288\ldots$$

정리하면 $-0.29 < \mu_1-\mu_2 < 6.29$

6. (주교재 연습문제 6.1번; 5점)

세 종류의 다른 온도가 식물의 성장에 차이를 미치는지 조사하고자 한다. 관심 있는 품종의 식물의 씨앗을 담은 화분을 랜덤하게 뽑아 세 가지 다른 온도에 배치하고, 한 달이 지난 후 식물의 키를 측정하였다. 여기서 사용된 화분은 15개이고 다음 표는 실험 결과를 보여준다.

	온도1 ($i=1$)	온도2 ($i=2$)	온도3 ($i=3$)
반복 1	10	14	17
반복 2	12	19	19
반복 3	15	21	25
반복 4	14	20	27
반복 5	10	23	28
	$T_1=61$	$T_2=97$	$T_3=116$

$\rightarrow T=274$

(1) 이 실험에 대한 적절한 귀무가설과 대립가설을 기술하라. (1점)

(2) 분산분석표를 작성하고, 유의수준 5%에서 검정하라. (4점)

(1) $H_0 : \mu_1 = \mu_2 = \mu_3$ (세가지 다른온도에 배치된 식물의 성장에 차이는없다)

H_1 : 최소 하나의 모평균 μ_i는 다른 모평균과 같지 않다.

(2)

변동요인	자유도	제곱합	평균제곱	검정통계량 F^*
처리	2	312.13	156.07	11.50
오차	12	162.80	13.57	
전체	14	474.93		

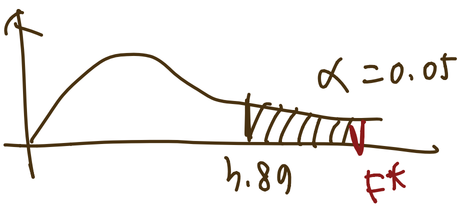
$SST = 5480 - \frac{274^2}{15} = 474.933\cdots = 474.93$

$SSTR = \frac{61^2+97^2+116^2}{5} - \frac{274^2}{15} = 312.133\cdots = 312.13$

평균제곱합 = 제곱합 / 자유도

$F^* = 156.07/13.57 = 11.5$

$F^* \sim F(2,12)$ 이고 $F_{0.05}(2,12) = 3.89$ 이다



$F^* > F_{0.05}(2,12)$ 이므로

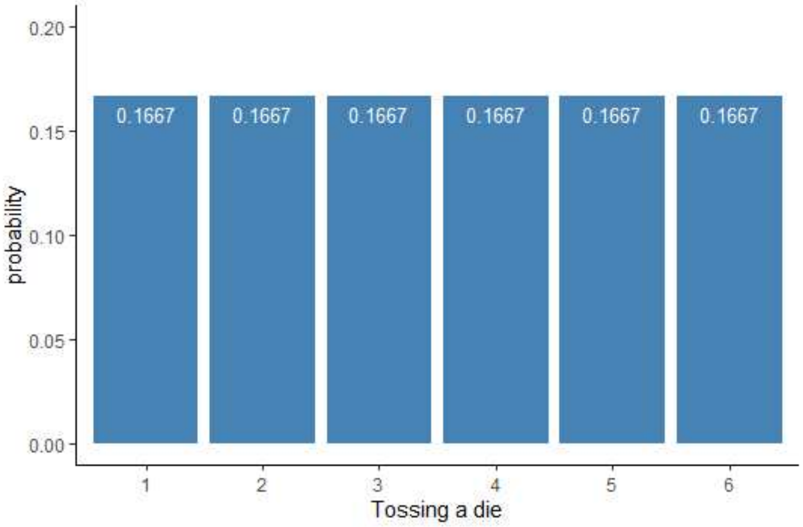
귀무가설 기각 가능

$\therefore$ 온도는 식물의 성장에 영향을 미친다.

7. (예5.1, 그림5.3 참조; 5점)

공정한 정육면체 주사위를 던져 각 눈이 나올 확률이 다음 그림과 같이 동일하다고 가정하자.

엑셀의 추가기능 KESS를 이용하여 한 개의 주사위를 10번, 30번, 100번, 500번 던지는 모의실험을 각각 1만번씩 반복 시행하고, 각 주사위 눈의 표본평균에 대한 히스토그램을 그리시오.



※분석결과는 엑셀 파일명을 “학과_학번_이름”(예: 통계학과_2012345_김숙명)으로, sheet명을 “7번 결과”로 저장하여 제출하시오.

8. (주교재 연습문제 5.8번; 5점)

별첨된 엑셀자료(sheet명: 8번 자료)는 시판중인 어느 해열제의 효과를 알아보기 위해 9명의 고열 환자를 대상으로 해열제 복용 전후의 체온(단위: ℃)을 측정한 결과이다. 이 해열제가 고열 환자에게 1℃를 초과하는 체온강하 효과를 주는지 **엑셀 추가기능** KESS를 **이용하여** 유의수준 $\alpha = 0.05$에서 검정하라.

※분석결과는 엑셀 파일명을 “학과_학번_이름”(예: 통계학과_2012345_김숙명)으로, sheet명을 “8번 결과”로 저장하여 제출하시오.