2021-1학기 경제통계분석 졸업시험(재시) 가이드라인

※ 2021년 1학기 경제통계분석 졸업시험(재시)의 범위는 다음과 같습니다.

(가) 통계학의 기본 개념
(나) 기술통계학
(다) 표본공간과 확률
(라) 확률분포
- 확률분포의 개념
- 확률분포의 기댓값 및 확률분포의 함수의 기댓값 계산 (모평균, 모분산)
- 결합확률분포, 확률변수의 독립, 모공분산
- (이산 및 연속) 균등분포, 베르누이 분포, 이항분포, 정규분포, t-분포

(마) 표본분포
- 표본 평균의 분포
- 표본 분산의 분포

(바) 통계적 추론 (모평균에 대한 추론)
- 통계적 추론의 개념 (추정량, 추정치, 추정량의 성질[불편성, 효율성, 일치성], 1/2종 오류)
- 모평균의 추정
- 모평균에 대한 가설 검정

※ 졸업시험(재시)은 본시험 때와 문항 구성 방식이 달라질 수 있습니다.

※ 예상문제 혹은 그 풀이에 오류를 발견하거나 질문이 있는 응시생은 황진태교수님(jhwang@daegu.ac.kr)께 연락 바랍니다.

경제통계분석 예상문제 모음

Part I [4지선다형 문제]

※ 다음 물음에 가장 알맞은 답을 보기에서 골라 답안지에 기입하시오. (각 5점)

1. 다음 중 통계학의 용어에 대한 설명으로 맞는 것은?

① 자료를 모집단으로부터 생성된 표본의 실현값으로 이해하고, 표본의 관측치를 이용해 모집단의 특성을 알아내려는 통계학 분야를 기술통계학이라 부른다.

② 확률실험으로 나타날 수 있는 모든 결과를 모아 놓은 집합을 모집단이라 부른다.

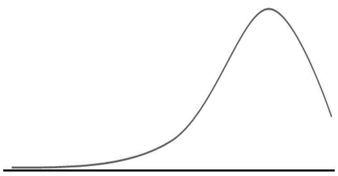
③ 표본공간의 결과에 수치를 대응시킨 것을 확률변수라고 부른다.

④ 두 확률변수가 각각 어떤 값을 가질 확률을 표현한 것을 한계확률분포라 부른다.

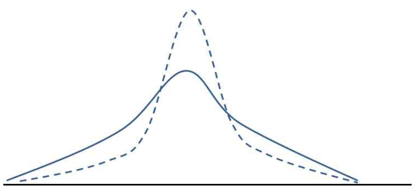
=> 정답: ③

풀이: ①은 추론통계학에 대한 설명, ②는 표본공간에 대한 설명, ④는 결합확률분포에 대한 설명

2. 다음 그림 중 <그림 1>과 <그림 2>는 관측치의 히스토그램을, <그림 3>은 한 자료의 두 변수의 산포도를 개략적으로 그린 것이다. 다음 중 가장 옳지 <u>않은</u> 설명은?



<그림 1>



<그림 2>

<그림 3>

① <그림 1>이 나타내는 자료의 평균은 그 자료의 중위값보다 작을 것이다.

② <그림 2>에서 점선의 히스토그램이 나타내는 자료의 분산은 실선의 히스토그램이 나타내는 자료의 분산보다 클 것이다.

③ <그림 3>이 나타내는 자료에서 두 변수는 약한 선형관계를 가진다.

④ <그림 3>이 나타내는 자료에서 두 변수의 상관계수는 음수일 것이다.

=> 정답: ②

풀이: 평균이 중위값보다 극단적 값의 영향을 많이 받기 때문에 히스토그램의 왼쪽 꼬리가 길면 평균이 중위값보다 작음. 분포가 평균 주위에 몰려 있을수록 분산은 작아짐. <그림 3>의 산포도는 두 변수 사이의 약한 음의 선형관계를 보여줌.

3. 대고대학 경제학과의 경제통계학 중간고사에 출제된 4지선다형 문제는 맞으면 4점, 틀리면 0점, 미기입의 경우 1점이 부여된다고 한다. 담당교수는 정답이 ①, ②, ③, ④일 확률이 모두 1/4라고 하였다. 안공부 학생은 경제통계학을 전혀 공부하지 않았고, 단면이 정사각형인 공정한 연필을 굴려 답을 쓰려고 마음먹고 있다. 다음 설명 중 맞지 않는 것은?

① 공정한 연필을 굴려 답을 썼을 경우 맞을 수 있는 점수의 기댓값은 1이다.

② 공정한 연필을 굴려 답을 썼을 경우 맞을 수 있는 점수의 분산은 3이다.

③ 답을 쓰지 않았을 경우 맞을 수 있는 점수의 기댓값은 1이다.

④ 답을 쓰지 않았을 경우 맞을 수 있는 점수의 분산은 3이다.

=> 정답: ④

풀이: 답을 쓰지 않았을 때 점수의 분산은 0

4. 철수는 동전 2개와 주사위 1개를 동시에 던지는 실험을 하였다. 이 때 표본공간의 원소 개수는 얼마인가? (단, 동전 두 개가 동시에 앞면 또는 뒷면으로 나타난 경우 하나의 결과로 간주한다.)

① 4 ② 12 ③ 24 ④ 48

=> 정답: ③

풀이: $S=\{(H,H,1),(H,T,1),(T,H,1),(T,T,1),(H,H,2),\ldots,(T,T,6)\}$

$n(S)=4\times 6=24$

5. 4번 문제와 관련하여 적어도 동전 앞면이 하나 이상이면서 주사위 눈이 3이하일 사건의 확률은 얼마인가?

① $\frac{1}{2}$ ② $\frac{2}{3}$ ③ $\frac{3}{4}$ ④ $\frac{3}{8}$

=> 정답: ④

풀이: $E=\{(H,H,1),(H,T,1),(T,H,1),(H,H,2),\ldots,(T,H,3)\}$

동전 두 개 모두가 뒷면인 경우를 제외 → $n(E)=12-3=9$

따라서 $\Pr(E)=\frac{n(E)}{n(S)}=\frac{9}{24}=\frac{3}{8}$ 임.

6. 4번 문제와 관련하여 주사위 눈이 5가 나타났다는 가정하에 동전 앞면이 하나만 나타날 사건의 확률은 얼마인가?

① $\frac{1}{2}$ ② $\frac{2}{3}$ ③ $\frac{3}{4}$ ④ $\frac{4}{5}$

=> 정답: ①

풀이: $F=\{(H,H,5),(H,T,5),(T,H,5),(H,H,5)\}$

$E=\{(H,T,1),(T,H,1),(H,T,2),(T,H,2),\ldots,(H,T,6),(T,H,6)\}$

$E\cap F=\{(H,T,5),(T,H,5)\}$

$\Pr(E|F)=\frac{n(E\cap F)/n(S)}{n(F)/n(S)}=\frac{n(E\cap F)}{n(F)}=\frac{2}{4}=\frac{1}{2}$

7. 공정한 동전을 3번 던지는 확률실험에 대하여 다음 중 맞지 <u>않는</u> 것을 고르시오.

① 2번 던졌을 때 모두 앞면이 나왔다면 세 번째에는 뒷면이 나올 확률이 높다.

② 두 번째에 앞면이 나오는 사건과 두 번째에 뒷면이 나오는 사건은 배반이다.

③ 두 번째에 앞면이 나오는 사건과 세 번째에 뒷면이 나오는 사건은 독립이다.

④ 앞면이 나오는 횟수를 X라 하면 X는 확률변수이다.

=> 정답: ①

풀이: 공정한 동전을 반복해 던지는 실험에서 각 시행에서의 사건은 다른 시행에서의 사건과 독립.

8. 다음 확률에 대한 설명 중 맞지 <u>않는</u> 것은?

① 표본공간 전체는 하나의 사건이며 그 사건이 나타날 확률은 1이다.

② 표본공간을 분할하는 사건들은 서로 배반이다.

③ 통계적 확률은 실험을 통한 상대도수로 도출되는 확률이다.

④ 공집합이 아닌 두 사건이 배반이면 두 사건은 독립이다.

=> 정답: ④

풀이: 배반인 두 사건은 그 중 한 사건의 발생 여부가 다른 사건의 발생 확률에 영향을 주므로 독립이 아님.

9. 다음 조건부확률 및 독립사건에 대한 설명 중 맞지 <u>않는</u> 것은?

① $P(A)=P(A|B)$일 때 두 사건 A와 B는 독립이다.

② 두 사건 A와 B가 독립이면 $P(A\cup B)=P(A)+P(B)-P(A)P(B)$이다.

③ 두 사건 A와 B가 독립이면 A와 B^c도 독립이다.

④ 두 사건 A와 B에 대하여 $P(A^c|B)=1-P(A|B^c)$가 성립한다.

=> 정답: ④

풀이: $P(A^c|B)=1-P(A|B)$

10. 사진동아리 “이안”에는 10명의 학생이 회원으로 있는데, 이들 중 3명이 남학생이라 한다. 10명 중 2명을 무작위로 뽑을 때 다음 중 맞는 설명은?

① 순서대로 한 명씩 뽑을 때 첫 번째로 남학생을 뽑는 사건과 두 번째로 남학생을 뽑는 사건은 독립이다.

② 2명이 모두 남학생일 확률은 6/100이다.

③ 한명은 남학생, 한명은 여학생일 확률은 1/2보다 작다.

④ 한명 이상이 남학생일 확률은 1/2보다 작다.

=> 정답: ③

풀이: ①에서 비복원추출이므로 두 사건은 독립이 아님. ②의 확률은 (3/10)*(2/9)=1/15. ③의 확률은 (3/10)*(7/9)+(7/10)*(3/9)= 14/30 < 1/2. ④에서 모두 여학생일 확률이 (7/10)*(6/9)=7/15 이므로, 한명 이상이 남학생일 확률은 8/15>1/2.

11. 다음 중 연속확률변수로 볼 수 있는 것은?

① 공정한 주사위를 두 번 던지는 확률실험에서 두 눈의 합

② 생산품 중 불량품의 갯수

③ 안공부학생이 4지선다형 10문제에서 정답을 제시한 횟수

④ 휘발유의 월평균가격

=> 정답: ④

12. 다음 설명 중 옳은 것은?

① X가 확률변수이면 모집단 평균 $E(X)$도 확률변수이다.

② X가 확률변수이면 표본평균 $\overline{X}$도 확률변수이다.

③ 두 확률변수 X와 Y에 대해 $E(X+Y)=E(X)+E(Y)+2Cov(X,Y)$이다.

④ 확률변수 X가 $n=12$이고 $p=0.1$인 이항분포를 따를 때 X의 기댓값은 120이다.

=> 정답: ②

풀이: ①에서 $E(X)$는 확률변수가 아닌 값.

③에서 $E(X+Y)=E(X)+E(Y)$. 반면 $V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$.

④에서 $E(X)=np=1.2$

13. 확률변수 X는 베르누이분포 $Bernoulli(0.2)$을, 확률변수 Y는 이항분포 $B(8,0.25)$을 따른다고 할 때 새로운 확률변수 $X+Y$의 평균 $E(X+Y)$는 얼마인가?

① 5 ② $\frac{9}{2}$ ③ $\frac{11}{2}$ ④ $\frac{11}{5}$

=> 정답: ④

풀이: $E(X)=\frac{1}{5}$, $E(Y)=8\times\frac{1}{4}=2$, $E(X+Y)=E(X)+E(Y)=\frac{11}{5}$

14. 아래는 일부 수치가 지워진 X와 Y의 결합확률분포표이다. 두 확률변수가 서로 독립일 때 다음 중 옳지 않은 것은?

$Y \diagdown X$	0	1	2	$p(y)$
1	(다)	0.15	(나)	0.3
3	0.07	(가)	0.28	0.7
$p(x)$	0.1	0.5		

① $p(1,1) = P(X=1, Y=1) = 0.15$이다.

② 표본공간에서 $X=1$에 대응되는 사건과 $Y=1$에 대응되는 사건은 서로 독립이다.

③ (가)에 알맞은 확률은 0.35이다.

④ (나)에 알맞은 확률은 0.14이다.

정답: ④ => (나)에 알맞은 확률은 0.12

15. 확률변수 X가 정규분포 $N(9,100)$을 따른다고 할 때 해당 모집단에서 50개의 관측치로 이루어진 하나의 표본 $(X_1,\ X_2,\ ...,\ X_{50})$을 복원추출 방식으로 구성하였다고 하자. 이 때 추정량(estimator) 중 하나로 표본평균 $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_{50}}{50}$을 사용한다면 해당 표본평균의 평균 $E(\overline{X})$는 얼마인가? (힌트: $X_i \sim i.i.d\ N(9,100)$)

① 1　　② 3　　③ 9　　④ 10

=> 정답: ③

풀이: 확률변수 X가 정규분포를 따르므로 $E(X) = 9$임. $E(\overline{X}) = E(X) = 9$

16. 15번 문제와 관련하여 $Var(\overline{X})$는 얼마인가?

① 2　　② $\dfrac{50}{3}$　　③ $\dfrac{100}{3}$　　④ $\dfrac{200}{3}$

=> 정답: ①

풀이: $Var(X) = 100$, $Var(\overline{X}) = \dfrac{Var(X)}{n} = \dfrac{100}{50} = 2$

17. 15번 문제와 관련하여 모집단 분산($Var(X)$)을 추정하기 위해 불편추정량(unbiased estimator)인 표본분산 $S^2 = \dfrac{\sum_{i=1}^{50}(X_i - \overline{X})^2}{50-1}$을 사용한다면 $E(S^2)$은 얼마인가?

① $\dfrac{25}{3}$　　② $\dfrac{50}{3}$　　③ 100　　④ $\dfrac{200}{3}$

=> 정답: ④

풀이: S^2이 $Var(X)$의 불편추정량이라고 하였으므로 $E(S^2) = Var(X) = 100$임.

18. 15번 문제와 관련하여 $Q^2 = \dfrac{\sum_{i=1}^{50}(X_i - \overline{X})^2}{50}$ 을 사용할 때 $E(Q^2)$은 얼마인가?

① $\dfrac{25}{3}$ ② $\dfrac{50}{3}$ ③ 50 ④ 98

=> 정답: ④

풀이: $Q^2 = \dfrac{50-1}{50}S^2$이므로 $E(Q^2) = \dfrac{49}{50}E(S^2) = \dfrac{49}{50} \times 100 = 98$

19. 확률변수 X에 대해 $E(XE(X))$와 동일한 것은?

① $E(X^2)$ ② $(E(X))^2$ ③ X^2 ④ $XE(X)$

=> 정답: ②

풀이: X는 확률변수이나 $E(X)$는 모수(parameters)의 함수에 해당하여 가정상 확률변수가 아님. 따라서 $E(X)$는 일반적인 상수처럼 expectation operator 바깥으로 나올 수 있으므로 $E(XE(X)) = E(X) \times E(X) = (E(X))^2$으로 표현될 수 있음.

20. 표본평균 $\overline{X}$의 표본분포는

① 항상 정규분포를 따른다.

② 항상 이항분포를 따른다.

③ 표본의 크기가 충분히 크면 정규분포를 따른다.

④ 모집단의 분포가 정규분포가 아니라면 정규분포를 따른다.

정답: ③

21. 표본분포에 대한 다음 설명 중 옳은 것은?

① 1개의 표본에서 각 관찰값들이 나타내는 분포를 의미한다.

② 표본평균의 표본분포는 표본의 크기가 작더라도 정규분포를 따른다.

③ 표본분산의 기댓값은 표본의 크기가 작아도 모분산과 같아진다.

④ 표분분산은 확률변수가 아니다.

정답: ③ => 표본분산은 모분산의 불편추정량이다.

22. 통계적 추정에 관한 다음 기술 중 옳지 않은 것은?

① 표본통계량이 모집단의 모수를 추정하기 위해 사용될 때 이 통계량을 추정량이라 한다.

② 구간추정량은 추정량이 가진 정보의 크기를 반영한다는 점에서 점추정량보다 좋다.

③ 표본분산은 모분산의 불편추정량이다.

④ 어떤 추정량이 다른 추정량보다 작은 분산을 가질 때 앞의 추정량이 뒤의 추정량보다 일치성이 뛰어나다고 한다.

정답: ④ => 효율성

23. 아래의 표본 크기 중 다른 조건이 동일하다면 가장 좁은 95%의 신뢰구간을 제시하는 것은?

① 2200　② 200　③ 20　④ 2

정답: ①

24. 신뢰구간의 넓이에 대한 다음 설명 중 옳지 않은 것은?

① 다른 조건이 같다면 표본의 크기가 클수록 신뢰구간이 좁아진다.

② 다른 조건이 같다면 신뢰수준이 높을수록 신뢰구간이 넓어진다.

③ 다른 조건이 같다면 모평균이 클수록 신뢰구간이 넓어진다.

④ 다른 조건이 같다면 모표준편차가 작을수록 신뢰구간은 좁아진다.

정답: ③

25. 표본의 크기를 두 배로 늘리면

① 표본평균의 표준오차(standard error)가 현재 수준의 절반으로 줄어든다.

② 표본평균의 표준오차가 현재 수준의 약 70% 수준으로 줄어든다.

③ 표본평균의 표준오차에는 아무런 영향도 미치지 않는다.

④ 표본평균의 표준오차를 두 배로 늘린다.

정답: ② => 표준오차는 통계량의 표준편차로 표본평균의 표준편차는 $\sigma/\sqrt{n}$임. 따라서 표본의 크기가 2배가 되면 표준오차는 $1/\sqrt{n}$배, 즉 약 70% 수준으로 감소

26. 어떤 확률변수 X의 모평균과 모표준편차가 각각 1와 5이다. 모집단으로부터 표본 크기가 25인 임의표본을 반복적으로 뽑았을 때, 표본평균의 표준편차와 평균은 각각

① 모집단의 분포와 관계없이 5와 1이다.

② 모집단의 분포가 정규분포를 따르는 경우 5와 1이다.

③ 모집단의 분포와 관계없이 1과 1이다.

④ 모집단의 분포가 정규분포를 따르는 경우 1과 1이다.

정답: ③ => 주의: 중심극한정리는 표본의 크기가 클 때 표본분포가 정규분포를 따른다는 정리. 위의 결과는 중심극한정리와 무관하게 성립

27. 대구대학교 경제학과에서는 설문을 통해 학생들의 겨울방학 중 평균수면시간을 추정하려 한다. 모표준편차가 1시간이고 95% 신뢰수준에서 추정오차의 크기를 0.5시간으로 하려면 얼마나 많은 학생에게 설문을 해야 하는가?

① 5명 ② 15명 ③ 50명 ④ 102명

=> 정답: ② 풀이: 추정오차가 $z_{0.025}\cdot\sigma/\sqrt{n}=1.96\cdot 1/\sqrt{n}$이므로 $0.5=1.96/\sqrt{n}$에서

$n=(1.96/0.5)^2\approx 15.37$

28. 신용카드회사가 고객의 평균결제금액을 추정하고자 한다. 모표준편차가 $300이고 99%의 신뢰수준에서 추정오차의 크기를 $75로 하고자 한다면 얼마나 많은 고객의 자료가 필요한가?

① 3382 ② 629 ③ 87 ④ 107

정답: ④ => 추정오차가 $z_{0.005}\dfrac{\sigma}{\sqrt{n}}=2.58\cdot\dfrac{300}{\sqrt{n}}$이고 이 값이 75가 되려면 $n=\left(\dfrac{2.58\cdot 300}{75}\right)^2\approx 107$

29. 어떤 귀무가설이 5%의 유의수준에서 기각된다면, 그 귀무가설은

① 1%의 유의수준에서도 반드시 기각될 것이다.

② 1%의 유의수준에서는 반드시 기각되지 않을 것이다.

③ 1%의 유의수준에서의 가설검정은 불필요하다.

④ 1%의 유의수준 하에서는 기각될지 그렇지 않을지 알 수 없다.

정답: ④ => 5% 유의수준에서 기각되었다는 것은 p-값이 5%보다 작다는 의미. 이 때 p-값이 1%보다 작은지 여부는 알 수 없음.

30. 통계적 가설검정에서 나타날 수 있는 오류에 대한 설명으로 옳지 않은 것은?

① 귀무가설이 옳음에도 이를 기각하는 오류를 제1종 오류라고 한다.

② 검정력이란 귀무가설이 옳지 않을 때 이를 기각할 확률을 의미한다.

③ 통계적 검정에서는 제2종 오류의 확률을 유의수준이라 한다.

④ 일반적으로 제1종 오류의 확률을 줄이면 제2종 오류의 확률이 늘어난다.

정답: ③ => 유의수준은 제1종 오류의 확률. 통상 통계적 가설검정에서는 유의수준을 정해놓고 검정력이 가장 큰 검정통계량을 사용한다.

31. 어떤 통계적 검정에서 귀무가설이 5% 유의수준에서 기각되지 않았다. 이에 대한 해석으로 가장 적절한 것은?

① 귀무가설이 맞을 확률보다 대립가설이 맞을 확률이 더 크다.

② 제2종 오류의 확률이 5%보다 작다.

③ 제1종 오류의 확률을 5%로 할 때 귀무가설이 틀리다는 충분한 근거가 없다.

④ 유의수준을 10%로 해도 유의수준은 기각되지 않을 것이다.

정답: ③

풀이: 제1종 오류는 유의수준으로 결정됨. 유의수준 5%에서 기각되지 않더라도 유의수준 10%에서는 기각될 수 있음.

32. 모집단이 정규분포를 따를 때 모평균에 대한 검정통계량에 대한 설명으로 옳지 않은 것은?

① 모표준편차를 모르면 student-t분포를 이용한다.

② 귀무가설이 $\mu=10$, 대립가설이 $\mu \neq 10$이면 양측검정을 사용한다.

③ 귀무가설이 $\mu=10$, 대립가설이 $\mu > 10$일 때 기각역은 임계치보다 작은 영역이다.

④ 귀무가설이 $\mu=10$, 대립가설이 $\mu < 10$일 때 데이터로부터 계산한 표본평균치가 $\bar{x}$이면 p-값은 $P(\overline{X} < \bar{x})$이다.

정답: ③

풀이: ③에서 기각역은 임계치보다 큰 영역이 됨.

33. 경상대학교 학생식당에서는 새롭게 개발한 메뉴인 카레 돈까스의 수요를 추정하려 한다. 식당 담당자는 하루 매출이 30개보다 크다는 확실한 증거가 있을 때에만 카레 돈까스를 메뉴에 포함시킬 계획이다. 이에 담당자는 경제통계학을 듣고 있는 이백조양에게 설문조사와 가설검정을 의뢰하였다. 이백조양은 경제통계학 시간에 배운 대로 모평균을 μ라 할 때, 귀무가설을 $\mu=30$, 대립가설을 $\mu > 30$이라 하고 가설을 검증하였다. 또한 30일 동안 경상대 식당 사용자 모두에게 설문을 하였더니, 카레 돈까스가 메뉴에 있다면 하루 평균 34.5명이 카레 돈까스를 먹을 것이라 답하였다. 이를 통해 p-값을 계산해 보았더니 3%가 나왔다. 다음 중 맞지 않는 설명은?

① 귀무가설이 기각될 때 카레 돈까스가 메뉴에 포함될 것이다.

② 표본의 크기는 30이다.

③ p-값은 모평균이 34.5일 때 표본평균이 30보다 클 확률이다.

④ 5%의 유의수준으로 가설을 검증했다면 귀무가설은 기각될 것이다.

정답: ③

풀이: p-값은 모평균이 30일 때 표본평균이 34.5보다 클 확률

34. 모표준편차가 알려져 있지 않은 정규분포를 따르는 모집단에서 표본을 16개 뽑아 표본평균의 값과 표본표준편차의 값을 계산하였더니 각각 52와 4가 나왔다. 귀무가설을 H_0: $\mu=50$으로, 대립가설을 H_1: $\mu < 50$로 통계적 검정을 하려 할 때 어떤 유의수준에서 귀무가설을 기각하겠는가?

① 1%　　② 5%　　③ 10%　　④ 어떤 유의수준으로도 기각되지 않는다.

정답: ④

풀이: 표본평균값이 50보다 작은 임계치보다 더 작게 나와야 기각하겠지만, 실제 표본평균값이 50보다 크게 나왔으므로, 어떤 유의수준으로도 기각되지 않음.

Part II [풀이 문제]

※ 다음 질문에 대하여 올바른 답을 답안지에 기입하시오.

1. 어떤 표본공간에서 서로 배반인 사건 A, B, C에 대하여 사건 A에 포함되는 결과가 나오면 1의 값을, 사건 B에 포함되는 결과가 나오면 2의 값을, 사건 C에 포함되는 결과가 나오면 3의 값을 주는 확률변수 X를 생각해보자. 단, $P(A)=0.3$, $P(B^c)=0.6$, $P(A \cup B \cup C)=1$이라고 한다.

   (a) 다음의 확률분포표를 완성하시오.

x	1	2	3	계
$p(x)=P(X=x)$	0.3	0.4	0.3	1

   (b) 확률변수 X의 기댓값 $E[X]$와 분산 $V(X)$를 구하시오.

   $E[X]=2$, $V(X)=0.6$

2. 표본공간에서 사건 A, B에 대하여 다음과 같은 확률이 알려져 있다.

$$P(A) = 0.4,\ P(B^c) = 0.4,\ P(B|A) = 0.5$$

(a) 다음 빈칸을 채워 결합확률표를 완성하라. 여기에서 한계확률(Marginal Probability)는 주변확률이라고도 한다.

	B	B^c	A의 한계확률
A	0.2	0.2	0.4
A^c	0.4	0.2	0.6
B의 한계확률	0.6	0.4	

(b) 사건 A에 포함되는 결과가 발생하면 5의 값을 주고, 그렇지 않은 결과가 발생하면 0의 값을 주는 확률변수를 X라 하자. 또한 사건 B에 포함되는 결과가 발생하면 20의 값을 주고, 그렇지 않은 결과가 발생하면 10의 값을 주는 확률변수를 Y라 하자. 다음 X와 Y의 결합확률분포표를 완성하라.

Y＼X	0	5
10	0.2	0.2
20	0.4	0.2

(c) (b)의 결합확률분포표를 이용하여 X와 Y의 한계확률분포표(주변확률분포표)를 작성하라.

x	0	5	계
$p(x) = P(X = x)$	0.6	0.4	1

y	10	20	계
$p(y) = P(X = y)$	0.4	0.6	1

(d) X와 Y의 평균($E[X]$, $E[Y]$) 및 분산($V(X)$, $V(Y)$), X와 Y의 공분산($COV(X, Y)$)을 구하여 답안의 표에 값을 기입하라.

$E[X]$	$V(X)$	$E[Y]$	$V(Y)$	$COV(X, Y)$
2	6	16	24	-2

(e) 확률변수 Z와 확률변수 X는 $Z = 3X + 2$의 관계에 있다고 한다. 확률변수 Z의 평균과 분산 및 확률변수 X와 확률변수 Z의 공분산을 구하여 답안의 표에 기입하라.

$E[Z]$	$V(Z)$	$COV(X, Z)$
8	54	18

=> 풀이: $E(Z) = 3E(X) + 2$, $V(Z) = 3^2 V(X)$, $COV(X, Z) = 3COV(X, X) = 3V(X)$

3. 0, 5, 10으로 표시된 구슬이 동일한 비율로 구성된 모집단에서 2개의 구슬을 복원으로 추출한다고 하자. 이 때 모집단 평균($E(X)$)을 추정하기 위해 표본평균을 사용한다고 하자.

(a) 표본평균($\overline{X}$)이 7.5일 확률은 얼마인가?

=> 풀이:

$i=1,2$에 대해 $X_i \in \{0,5,10\}$이므로 두 개의 관측치로 구성된 표본의 경우는 $\Omega=\{(0,0),(0,5),(0,10),(5,0),(5,5),(5,10),(10,0),(10,5),(10,10)\}$임.

이 때 $\overline{X}=\dfrac{X_1+X_2}{2}$이므로 $\Pr(\overline{X}=0)=1/9$, $\Pr(\overline{X}=2.5)=2/9$, $\Pr(\overline{X}=5)=3/9$, $\Pr(\overline{X}=7.5)=2/9$, $\Pr(\overline{X}=10)=1/9$임. 이를 확률분포표로 표현하면,

$\overline{X}$	0	2.5	5	7.5	10	계
$\Pr(\overline{X})$	1/9	2/9	3/9	2/9	1/9	1

(b) 표본평균($\overline{X}$)의 평균과 분산은 얼마인가?

=> 풀이:

먼저 a)의 확률분포표를 통해 $\overline{X}$의 평균과 분산을 구할 수 있음.

또는 아래와 같이 모집단의 분포를 이용할 수 있음.

$$E(X)=0\times\frac{1}{3}+5\times\frac{1}{3}+10\times\frac{1}{3}=5$$

$$Var(X)=(0-5)^2\times\frac{1}{3}+(5-5)^2\times\frac{1}{3}+(10-5)^2\times\frac{1}{3}=\frac{50}{3}$$

따라서 $E(\overline{X})=E(X)=5$, $Var(\overline{X})=\dfrac{Var(X)}{n}=\dfrac{50}{3}\div 2=\dfrac{25}{3}$임.

4. 철수가 모집단 평균 $E(X)$(즉, 관련 모수)를 추정하기 위하여 모집단으로부터 표본(X_1, X_2, ..., X_n)을 독립적으로 복원추출한 후 그 표본을 이용하여 다음과 같은 세 가지 추정량(estimators)을 고려하고 있다.

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}\ ;\ \dot{X} = \frac{X_1 + X_2 + \cdots + X_n}{n-1}\ ;\ \widetilde{X} = \frac{X_1 + X_n}{2}$$

(a) 세 추정량 각각에 대해 불편추정량 여부를 판단하라.

=> 풀이:

불편추정량(unbiased estimator)은 추정량($\widehat{E(X)}$)의 평균이 모집단 평균($E(X)$과 같아짐을 의미함. 즉, $E(\widehat{E(X)}) = E(X)$가 성립함을 의미함.

$E(\overline{X}) = E(X)\ \rightarrow\ \overline{X}$는 불편추정량

$E(\dot{X}) = \dfrac{n}{n-1}E(X)\ \rightarrow\ \dot{X}$는 편의추정량, 또는 불편추정량이 아님.

$E(\widetilde{X}) = E(X)\ \rightarrow\ \widetilde{X}$는 불편추정량

(b) a)에서의 불편추정량에 대해 효율성 여부를 판단하라.

=> 풀이:

$Var(\overline{X}) = \dfrac{Var(X)}{n}$ vs. $Var(\widetilde{X}) = \dfrac{Var(X)}{2}$

$\overline{X}$가 $\widetilde{X}$보다 분산이 작아 효율적임.

(c) 상기 세 가지 추정량 중 모집단 평균 $E(X)$에 대해 어떤 추정량이 통계학적으로 가장 좋은가? 그리고 그 의미가 무엇인지 간략히 기술하라.

=> 풀이:

표본평균 $\overline{X}$가 불편추정량에다 효율성의 특성을 가지고 있어 $E(X)$에 대한 가장 좋은 추정량임. 즉, 표본을 통해 관측치 값을 모두 더한 뒤 그 관측치 수로 나누어 계산하는 추정방식이 모집단의 참값(true value)에 해당하는 모집단 평균 $E(X)$ 또는 관련 모수를 추정하는데 확률적으로 가장 좋은 방식이라는 의미임.

5. 콜로라도에서 잡히는 송어의 길이는 정규분포를 따르고 평균이 12.5인치이고 표준편차는 1.2인치이다. 다음 질문에 답하라.

(a) 임의로 잡힌 송어의 길이가 11인치에서 13인치 사이일 확률은?

=> 풀이:

송어의 길이를 X라 하면 $X \sim N(12.5, 1.2^2)$이다. $Z = \dfrac{X-12.5}{1.2}$인 정규화를 이용하면

$$\begin{aligned} P(11 < X < 13) &= P(-1.25 < Z < 0.42) = P(Z < 0.42) - (1 - P(Z < 1.25)) \\ &= 0.6628 - (1 - 0.8944) = 0.5572 \end{aligned}$$

정답: 0.5572

(b) 잡힌 송어 중 작은 치어는 방류하고 나머지는 집으로 가져갈 수 있다. 집으로 가져갈 수 있는 송어의 비율이 80%가 되도록 송어 길이의 기준을 정한다면, 가져갈 수 있는 송어의 최소 길이는 얼마인가?

=> 풀이:

$P(X > x_0) = 0.8$인 x_0를 구하는 문제. 즉 $P(X < x_0) = 0.2$인 x_0를 구하면 된다. 위 a)와 같은 정규화를 이용하면

$$P(X < x_0) = P\left(Z < z_0 = \frac{x_0 - 12.5}{1.2}\right) = 0.2$$

이어야 한다. 그런데 z_0가 음수이므로 $P(Z < z_0) = 1 - P(Z < -z_0)$를 이용해야 한다. $P(Z < -z_0) = 0.8$인 $-z_0$를 엑셀이나 표준정규분포표에서 찾으면 $z_0 \approx -0.84$임을 알 수 있다. 따라서 $x_0 = 12.5 + 1.2 \cdot (-0.84) = 11.492$

정답: 11.492

(c) 각 낚시꾼은 4마리의 송어를 잡아 집으로 가져갈 수 있다. 이 때 집으로 가져가는 송어의 길이의 평균이 13인치보다 클 확률은?

=> 풀이:

4마리의 평균을 $\overline{X}$라 하면 이는 표본의 크기가 4인 표본평균으로 모집단이 정규분포를 따를 때 $\overline{X} \sim N\left(12.5, (1.2/\sqrt{4})^2\right) = N(12.5, 0.6^2)$이 된다. $Z = \dfrac{\overline{X} - 12.5}{0.6}$의 정규화를 하여 계산하면

$$P(\overline{X} > 13) = P\left(Z > \frac{13 - 12.5}{0.6}\right) = P(Z > 0.83) = 1 - P(Z \le 0.83) = 1 - 0.7967 = 0.2033$$

정답: 0.2033

6. 하양농장에서 생산되는 유기농 달걀의 무게는 정규분포를 따르고, 평균은 120g이며 표준편차는 20g인 것으로 알려져 있다. 다음 물음에 답하시오.

(a) 달걀을 임의로 하나 뽑았을 때 그 달걀의 무게가 100g에서 140g 사이일 확률은?

정답: 0.6826 혹은 68.26%

풀이: 유기농 달걀 1개의 무게를 X라 하면 $X \sim N(120, 20^2)$이므로 정규화를 위해 $Z = (X - 120)/20$으로 두면

$$\begin{aligned} P(100 < X < 140) &= P\left(\frac{100-120}{20} < Z < \frac{140-120}{20}\right) \\ &= P(-1 < Z < 1) = 2P(0 < Z < 1) \\ &= 2(P(Z < 1) - 0.5) = 2(0.8413 - 0.5) \\ &= 0.6826 \end{aligned}$$

(b) 하양농장에서는 출하되는 달걀의 품질 관리를 위해 일정 무게 이상의 달걀만을 판매하고, 그 나머지는 농장 사람들이 나누어 먹는다고 한다. 출하되는 달걀의 비율이 전체 생산되는 달걀의 90%가 되도록 하려면 몇g 이상의 달걀을 출하하면 되겠는가?

정답: 94.4

풀이: $P(X > \overline{x}) = 0.9$인 $\overline{x}$를 찾으면 된다. 위 (a)의 정규화를 이용하면

$$P\left(Z > \frac{\overline{x} - 120}{20}\right) = 0.9$$

인데 표준정규분포표에서 $P(Z < 1.28) = 0.9$이므로 $P(Z > -1.28) = 0.9$임을 알 수 있다. 따라서 $(\overline{x} - 120)/20 = -1.28$이어야 하고 $\overline{x} = 120 - 1.28 \cdot 20 = 94.4$이다.

(c) 하양농장은 달걀을 4개들이 꾸러미로 포장하여 판매한다. (b)를 무시하고 모든 달걀을 판매한다고 할 때의 4개들이 달걀 꾸러미의 무게가 460g 이하일 확률은? (힌트: 달걀 꾸러미의 무게가 460g 이하이면 꾸러미의 달걀 1개당 평균무게는 115g 이하이어야 한다.)

정답: 0.3085 혹은 30.85%

풀이: 4개의 표본을 뽑았을 때 그 평균이 115 이하일 확률을 구해야 한다. 표본의 크기가 4일 때 표본평균 $\overline{X}$는 $\overline{X} \sim N(120, (20/\sqrt{4})^2) = N(120, 10^2)$의 분포를 따르므로 $Z = (\overline{X} - 120)/10$의 정규화를 이용하면

$$P(\overline{X} < 115) = P(Z < -0.5) = 1 - P(Z < 0.5) = 1 - 0.6915 = 0.3085$$

7. 모집단으로부터 복원으로 추출하여 표본의 크기가 25개인 표본을 계산한 표본평균($\overline{X}$)이 2.0이라고 하자. 이 때 해당 모집단의 확률변수에 대해 $X \sim N(\mu, 16)$이라고 가정한다.

(a) 표본평균($\overline{X}$)을 이용하여 귀무가설 Ho: $\mu = \mu_0 = 0$에 대해 검정통계량을 구한 후 5% 유의수준에서 검정하라. <부록>에 있는 표를 이용하시오.

풀이: 귀무가설 Ho하에서 검정통계량 $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{2-0}{4/5} = 2.5$임. 이 때 해당 검정통계량 Z_0의 경우 표준정규분포를 따르는 값이며, 그 값이 유의수준 5%의 임계치(critical value)인 1.96보다 크므로 귀무가설을 기각함.

(b) 귀무가설 Ho: $\mu \le 0$, 대립가설 Ha: $\mu > 0$라고 할 때 p-값은 얼마인가? <부록>에 있는 표를 이용하시오.

풀이:

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{2-0}{4/5} = 2.5$$

$\Rightarrow$ p-값은 $\Pr(Z > 2.5) = 1 - \Pr(Z \le 2.5) = 1 - 0.9938 = 0.0062$임.

8. 확률변수 X가 정규분포 $N(3,9)$을 따른다고 하자. 해당 모집단에서 10개의 관측치로 이루어진 표본(X_1, X_2, ..., X_{10})을 구성하였다고 하자. 이러한 표본을 이용하여 새로운 추정량 $L^2 = \dfrac{\sum_{i=1}^{10}(X_i - \overline{X})^2}{10} + \overline{X}^{\,2}$을 고려한다고 할 때 해당 추정량의 평균 $E(L^2)$을 구하라.

풀이:

$E(X) = 3,\ Var(X) = 9$

$E(\overline{X}) = E(X) = 3,\ Var(\overline{X}) = \dfrac{Var(X)}{10} = \dfrac{9}{10}$

$E(\overline{X}^2) = Var(\overline{X}) + \left(E(\overline{X})\right)^2 = \dfrac{9}{10} + 3^2 = \dfrac{99}{10}$

$E(L^2) = \dfrac{10-1}{10}E(S^2) + E(\overline{X}^2) = \dfrac{10-1}{10}Var(X) + E(\overline{X}^2) = \dfrac{9}{10} \times 9 + \dfrac{99}{10} = \dfrac{180}{10} = 18$

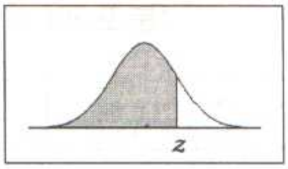
Part III [개념 설명 문제]

※ 경제통계학에서 배운 다음 개념 중 두 개를 골라 간략히 설명하시오. (각 10점)

조건부 확률
확률변수
독립 사건
두 확률변수의 독립
확률밀도함수(Probability Density Function)
대수의 법칙(Law of Large Numbers)
중심극한정리(Central Limit Theorem)
추정량(Estimator)
불편추정량(Unbiased Estimator)
일치추정량(Consistent Estimator)
추정량의 효율성(Efficiency of an Estimator)
신뢰구간
귀무가설과 대립가설
1종 오류와 2종 오류
검정력(Power)
유의수준
단측검정과 양측검정

<부 록>

3. 확률분포표



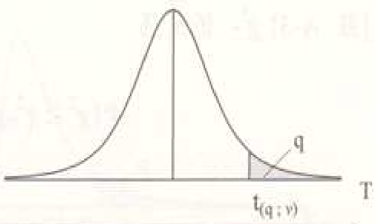
표준정규분포표

$$\left(\Pr(Z \le z) = \Phi(z),\ Z \sim N(0, 1)\right)$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

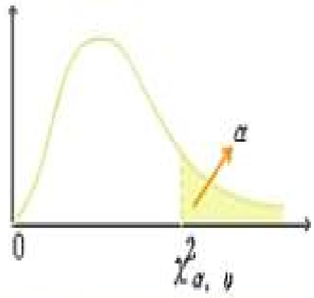
[표 A-2] t-분포표

$$P\{T \geq t_{(q\,;\,\nu)}\} = q$$



자유도	꼬리확률 q									
ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.792	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

〈부표-3〉 χ^2 분포표



ν	α=,995	α=,99	α=,975	α=,95	α=,05	α=,025	α=,01	α=,005	ν
1	,3333930	,000157	,000982	,00393	3,841	5,024	6,635	7,879	1
2	,0100	,0201	,0506	,103	5,991	7,378	9,210	10,597	2
3	,0717	,115	,216	,352	7,815	9,348	11,345	12,838	3
4	,207	,297	,484	,711	9,488	11,143	13,277	14,860	4
5	,412	,554	,831	1,145	11,070	12,832	15,086	16,750	5
6	,676	,872	1,237	1,635	13,592	14,449	16,812	18,548	6
7	,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278	7
8	1,344	1,646	2,180	2,733	15,507	17,535	20,090	21,955	8
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589	9
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188	10
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757	11
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300	12
13	3,565	4,107	5,009	5,892	22,362	24,736	27,688	29,819	13
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319	14
15	4,601	5,229	6,262	7,261	24,996	27,488	30,578	32,801	15
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267	16
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718	17
18	6,265	7,015	8,231	9,390	28,869	31,526	34,805	37,516	18
19	6,844	7,633	8,907	10,117	30,114	32,852	36,191	38,582	19
20	7,434	8,260	9,591	10,851	31,410	34,170	37,566	39,997	20

c) 대구관광당국에 문의해본 결과 일반적인 스키장 이용객의 스키장 방문 횟수의 표준편차는 2라고 한다. 유의수준 5%로 가설검정을 수행할 때 p-값을 구하라.

풀이: 모표준편차가 2임이 알려져 있으므로 표본평균의 분포로 정규분포를 사용한다. 즉, 귀무가설 하에서 표본평균의 분포는 다음과 같다.

$$\frac{\overline{X}-4}{2/\sqrt{64}} \sim N(0,1^2)$$

이 때, p-값은 표본평균이 4.6이상일 확률이므로

$$\begin{aligned} p-\text{값} &= P(\overline{X} \geq 4.6) = P(Z \geq \frac{4.6-4}{2/\sqrt{64}}) = P(Z \geq 2.4) = 1-P(Z < 2.4) \\ &= 1-0.9928 = 0.0072 \end{aligned}$$

(이 값이 유의수준인 5%보다 작으므로 귀무가설은 기각해야 한다.)

정답: 0.0072

7. 용남대학교 학생들의 음주행태를 살펴보는 연구가 진행 중이다. 용남대학교 학생들이 1주일에 마시는 소주량의 표준편차는 0.64병이라고 알려져 있다. 100명의 학생에게 설문조사를 해 보았더니 주량의 표본평균이 2.16병이라는 결과를 얻었다. 표본의 크기 100은 중심극한정리를 적용하기에 충분하다고 하자.

(a) 주어진 자료를 이용하여 용남대학교 학생들의 주당 소주소비량의 평균을 95% 신뢰수준으로 추정하여라. (신뢰구간의 하한과 상한을 제시하면 된다.)

풀이: $2.16 \pm 1.96\frac{0.64}{\sqrt{100}} = 2.16 \pm 0.12544.$

정답: [2.03456, 2.28544]

(b) 음주행태를 연구 중인 경제학과 이백수 교수는 용남대학교 학생의 주당 소주소비량이 2병이라고 주장하는 반면, 김백조 교수는 2병이 아니라고 주장하고 있다. 이백수교수는 자신의 주장을 귀무가설로, 김백조 교수의 주장을 대립가설로 하여 유의수준 5%의 가설검정을 수행하려 한다. 용남대학교 학생들의 주당 소주소비량을 μ라 하고 다음의 빈 간을 채워라. (임계치가 여러 개 있으면 모두 적어라.)

귀무가설: $\mu = 2$ 대립가설: $\mu \neq 2$

임계치: 1.875456, 2.12544

귀무가설의 기각 여부: 기각한다.

(c) 위 가설검정에서 p-값을 구하여라. 또한 귀무가설이 1%의 유의수준과 10%의 유의수준에서 기각되는지 여부를 밝혀라.

정답: p-값: 0.0124

1% 유의수준에서 기각하지 않는다.

10% 유의수준에서 기각한다.

8. 다음은 정규분포를 따르는 모집단에서 16개의 표본을 추출하여 관측값을 기록한 엑셀 시트의 일부분이다. 여기에서 x는 관측값을, $\overline{x}$는 관측값의 평균(즉 표본평균값)을 의미한다.

관측값 번호	x	$x-\overline{x}$	$(x-\overline{x})^2$
1			
2			
중간 생략			
15			
16			
합계	240	0	2160

표본의 크기가 n일 때 관측값을 $x_1, x_2, \cdots, x_n$이라 하면 표본평균($\overline{x}$)과 표본분산(s^2)은 다음과 같은 식으로 계산됨을 기억하자.

$$\overline{x} = \sum_{i=1}^{n} x_i, \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

또한 모표준편차는 알려져 있지 않다. 다음 물음에 답하여라.

(a) 표본평균값과 표본표준편차값은 각각 얼마인가?

=> 정답: $\overline{x} = 240/16 = 15$, $s = \sqrt{2160/(16-1)} = 12$

(b) 모평균 μ에 대하여 귀무가설과 대립가설을 다음과 같이 두고 10%의 유의수준에서 가설검정을 하려 한다.

$$H_0: \mu = 10, \; H_1: \mu > 10$$

① 이 때 검정통계량

$$\frac{\overline{X} - 10}{12/\sqrt{16}}$$

은 어떤 분포를 따르는가? (필요하면 자유도를 밝혀야 한다.)

=> 정답: 자유도가 15인 t-분포

② 문제지 뒤의 t-분포표를 참조하여 표본평균 $\overline{X}$가 어떤 조건일 때 귀무가설을 기각해야 하는지를 적으시오. (즉, $\overline{X}$를 포함하는 부등식을 제시하면 된다.)

=> 정답: $\overline{X} > 14.023$

③ (a)에서 구한 값을 이용하여 귀무가설 기각 여부를 밝혀라.

=> 정답: 기각한다.

(c) 이번에는 모평균 μ에 대하여 귀무가설과 대립가설을 다음과 같이 두고 10%의 유의수준에서 가설검정을 하려 한다.

$$H_0: \mu = 10, \; H_1: \mu \neq 10$$

① 문제지 뒤의 t-분포표를 참조하여 표본평균 $\overline{X}$가 어떤 조건일 때 귀무가설을 기각해야 하는지를 적으시오. (즉, $\overline{X}$를 포함하는 부등식을 제시하면 된다.)

=> 정답: $\overline{X} > 15.256$, $\overline{X} < 4.741$

② (a)에서 구한 값을 이용하여 귀무가설 기각 여부를 밝혀라.

=> 정답: 기각하지 않는다.