

[제31회, 1987년]

제1문 소비수준은 소득수준에 의하여 결정된다는 설이 있는 바, 이를 음미할 수 있는 방법을 제시하라. (50점)

제2문의 ① F-분포는 어떠한 것이며, 이의 이용에 대하여 설명하라. (20점)

제2문의 ② 평균자승오차와 분산의 차이점을 논하라. (15점)

제2문의 ③ 어느 회사에서 생산되는 제품의 지름을 조사하기 위하여 임의로 100개를 뽑아 조사하였더니 평균이 45cm, 표준편차가 2cm 었다. 이 회사가 생산한 제품의 지름에 대한 모평균을 95%의 신뢰수준에서 추정하라. (15점)

[제32회, 1988년]

제1문 χ^2 의 검정법을 설명하라.

제2문의 ① 결정계수

제2문의 ② 추첨에 있어서 첫 번째 뽑는 사람과 두 번째 뽑는 사람 중에서 누가 유리한가?

제2문의 ③ 통계적 품질관리법에 있어서 X-관리도를 설명하라.

[제33회, 1989년]

제1문 단순확률(임의)추출법과 층화추출법을 비교, 설명하라.

제2문의 ① 비모수적 추론과 Wilcoxon의 순위합검정

제2문의 ② 이항분포, Poisson분포 및 정규분포와 이들의 관계

[제34회, 1990년]

제1문 통계적 가설검정의 기본개념을 논하라.

제2문의 ① 체비셰프의 부등식

제2문의 ② 시계열에서 추세치를 추정하는 방법

[제35회, 1991년]

제1문

단순회귀계수와 상관계수와의 관계를 비교, 설명하라.

제2문의 ① Bayes공식을 설명하고 간단히 실례를 들어라.

제2문의 ② 정규분포(모분포)의 모평균추정에 따른 표본크기의 결정 방법

[제36회, 1992년]

제1문

단순선형회귀모형 $Y_i = \alpha + \beta X_i + \epsilon_i (i = 1, \dots, n)$ 에서 필요한 가정들을 열거하고 그 의미를 설명하여라. 또한 잔차분석의 내용과 그 필요성을 논하여라. (50점)

제2문의 ① 가설검정에서 유의확률 p값(p-value) (25점)

제2문의 ② 층화랜덤 추출방법의 내용과 장점 (25점)

[제37회, 1993년]

제1문

확률밀도함수 $f(x; Q)$ 인 모집단으로부터 추출된 랜덤포본 X_1, X_2, \dots, X_n 을 이용하여 얻은 모수 Q에 대한 추정치에 대해서 편의와 분산을 정의하고 통계적 의미를 설명한 다음 그 평균제곱오차와 편의와 분산의 관계를 논하라.(50점)

제2문의 ①

결합확률밀도함수의 분포가 다음과 같을 때 다음을 계산하라.(25점)

X \ Y	0	1	2
0	0.10	0.20	0.15
1	0.05	0.30	0.20

(1) $Pr(X = Y)$

(2) $E(3X^2 + 2) \quad Var(-2X - 2)$

(3) $(X + Y)$ 의 분포

제2문의 ②

독립된 표본과 짝지은 표본에 의한 추론에 있어서 양자를 비교하

라.(25점)

[제38회, 1994년]

제1문

표본조사방법에 있어 단순임의추출, 층화추출, 계통추출, 2단집락추출 방법을 설명하고, 이들 각 방법을 사용할 때 주의점과 전제조건에 관하여 논하시오. (50점)

제2문의 ①

표본통계량(sample statistic)과 그 표본분포(sampling distribution)에 관하여 설명하시오.(30점)

제2문의 ②

두 확률변수 X와 Y가 서로 독립이면 그 공분산은 0이 됨을 증명하고, 그 역은 성립하지 않음을 예를 들어 설명하시오.(20점)

[제39회, 1995년]

제1문

다중선형회귀모형 $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$ 이고 서로 독립.

변수선택방법과 다중공선성이 발생언 경우 그 발생요인과 회귀모형에 미치는 영향을 설명하고,대처방안에 대해 논하시오.(50점)

제2문의 ①

평균 μ 와 분산 σ^2 이 미지인 정규모집단 $N(\mu, \sigma^2)$ 으로부터 추출한 확률표본 X_1, X_2, X_n 에 대하여 모분산 σ^2 을 점추정하는 방법들을 비교, 설명하시오.(25점)

제2문의 ②

1936년 미국대통령 선거시 민주당 루즈벨트후보와 공화당 랜드후보에 대한 지지도 여론조사에서 전화번호부, 클럽회원명부를 표본추출해 조사한 결과 공화당 랜드후보가 57%, 민주당 루즈벨트후보의 43%보다 지지도에 있어서 앞섰다. 그러나 선거결과 득표율에서는 민주당 루즈벨트후보가 63%,공화당 랜드후보가 37%로서 민주당 루즈벨트후보가 당선되었는데, 그러한 오차가 발생한 원인과 그 대처방안을 설명하시오. (25점)

[제40회, 1996년]

제1문

중학생 평균신장이 164cm보다 커졌는지 알아보기 위해서 500명을 대상으로 조사, 분석한 결과가 다음과 같다.

$n = 500$	$\bar{x} = 169.9$
$St\ de = 5.0xx$	
$St\ error = a = 244$	
$v = 25xx$	
Q_1	75%
Q_2	100%
.	
.	
0	

(Computer 출력)

(1) 키 175cm인 학생은 전체에서 상위 몇 %에 속한다고 할 수 있는가?

(2) 모평균 신뢰구간을 추정하라.

(3) 4분위수범위(자료에서 Q_1, Q_3 주어져 있다)

(4) 목적에 따라 가설을 설정하고 검정을 수행하라(과정은 표시하지 구체적 수치는 쓰지 않아도 됨). (총 50점)

제2문

다음의 점선도의 의미를 설명하라. (20점)

(그림은 잔차산점도로서 이분산성을 나타내도록 x값이 커짐에 따라 잔차가 위 아래로 커지도록 그려져있음)

제3문

다음과 같은 경우 출신지역과 전공선택간의 독립성 검정방법을 상세하게 설명하라. (15점)

	자연계	사회계	인문계
농촌			
도시			

제4문

두 확률변수 X , Y 와의 관계에 있어 "독립적(independent)" 이다는 것과 "상관관계에 있지 않다(uncorrelated)" "되어 있다"는 것 간의 관계를 설명하라. (15점)

[제41회, 1997년]

제1문

어떤 모집단에서 모평균 μ 를 추정하기 위하여 랜덤포본 X_1, X_2, \dots, X_n 을 추출하고 $\hat{\theta}$ 를 추정량으로 사용한다고 하자. 확률분포와 표본분포를 정의하고 이들 사이의 관계를 설명하시오. 특히 $\hat{\theta} = \bar{X}$ (표본평균)인 경우에 대하여 표본분포와 추론과정(신뢰구간)을 구체적으로 밝히시오.

제2문

인구 10만명인 어떤 도시에 거주하는 20세 이상의 성인중 신문을 읽는 사람의 비율을 알아보기 위하여 400명을 랜덤 추출하여 조사한 결과 360명이 신문을 읽는다고 대답하였다. 도시 전체의 신문을 읽는 성인의 비율에 대한 95% 신뢰구간을 구하시오. ($Z_{0.025} = 1.96$)

제3문

치약상품인 A, B, C, D의 시장 점유율이 각각 30%, 60%, 8%, 2%라고 알려져 있다. 600명을 랜덤추출하여 조사한 결과가 다음과 같다.

상표	A	B	C	D
선호도(단위 : 명)	192	342	44	22

이 자료로부터 기존에 알려진 시장 점유율이 옳지 않다고 결론 내릴 수 있는가? 가설을 쓰고 유의수준 5%로 검정하시오.
($\chi^2(3, 0.05) = 7.815$)

[제42회, 1998년]

제1문

제현불가(50점)

(작은 문제 여러개로 나뉘어져 있었고 약품에 대한 치료효과를 대응 비교하는 문제였던 것으로 기억됨. 비교집단에 약대신 물을 투약하는 이유를 물어보았던 것으로 기억함)

제2문

확률변수 X 가 이항분포(n, p)를 따르는 경우에 n 이 크고(대략 50 이상), p 가 작으면(대략 0.1 이하) 휴대용 계산기를 이용하여 X 에 대한 확률을 구하기가 쉽지 않다. 이때 포아송 분포를 이용하여 근사적으로 X 에 대한 확률을 계산할 수 있는데, 이 근사법을 증명과 함께 논하시오. (20점)

제3문

시계열(time series)자료를 정의하고 일반적인 자료와의 차이를 비교, 설명하시오. 또한 정상(stationary) 시계열과 비정상(nonstationary) 시계열을 비교, 구분하고 선형과정(linear process)과의 관련성을 논하시오. (15점)

제4문

분산 안정화(variance stabilizing) 변환에 관해 논하시오. (15점)

[제43회, 1999년]

제1문

다음 물음에 답하시오. (50점)

(1) 표집방법으로 무작위추출, 층화추출, 군락추출, 계통(계열)추출 방법을 설명하라.

(2) 층화추출, 군락추출, 계통(계열)추출 방법을 무작위추출과 비교하여 그 장단점을 논하시오.

(3) 아래 자료에 주어진 정보를 이용하여 무작위추출, 층화추출, 군락추출, 계통(계열)추출을 실시했을 때 그 구체적인 예를 적시하시오.

한 대학교의 대학생 10,000명을 대상으로 여론조사를 실시하고자 한다. 각 학생은 고유번호(학번)가 부여되어 있다.

1학년 : 4,000명, 2학년 : 3,000명, 3학년 : 2,000명, 4학년 : 1,000명이다.

모두 10개 학과이고 각 학과는 1,000명으로 구성된다.

교무처 학생부에는 전체 학생목록을 학년, 학과 구별없이 가나다 이 름순으로 기재되어 있다.

제2문

객관적 척도로서 평균, 중앙값, 절사 평균이 있다. 다음 물음에 답하라. (30점)

(1) outlier가 포함되어 있는 경우, 위 세 가지의 척도로서의 적절성을 설명하라.

(2) 다음 자료를 기초로 위 세 척도중 하나를 선택하여 척도로서의 적절성을 보아라.

한 마을에 농가 20가구가 있다. 각 가구의 소득 분포는 다음과 같다.

18은 2가구, 19는 6가구, 20은 4가구, 21은 1가구, 22는 3가구, 23은 3가구, 430은 1가구(단위 1,000원)

제3문

무작위화의 원리를 설명하고, 기계 3대의 생산성을 측정함에 있어서 각 기계당 4명씩 작업자가 배치된 경우에 Randomization을 구성하라. (20점)

[제44회, 2000년]

제1문

반응(또는 종속)변수 Y 를 설명하기 위하여 하나의 설명(또는 독립)변수 X 를 사용하는 단순선형회귀모형을 생각한다. (총 50점)

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon$$

(1) 다음의 자료를 이용하여 위의 모형에 적합시키기 위한 회귀계수 추정량 β_1 의 의미에 대하여 구체적으로 약속하시오. (10점)

(단, 자료를 이용하여 구체적으로 계산할 필요는 없음)

X : 운송회사들의 운송량 Y :안전과 사고대비를 위한 보험료 액 수

X(단위:천톤)	10	16	12	18	17	17	9	19	17	11
Y(단위:백만원)	13	18	14	18	23	21	14	25	23	14

(2) 추정된 회귀직선의 적합도를 나타내는 척도로 결정계수를 사용한다. 결정계수를 정의하고 그 의미에 대해 논하시오. (15점)

(단, 자료를 이용하여 구체적으로 계산할 필요는 없음)

(3) 적합된(fitted) 결과를 이용하여 오차항(ϵ_i)에 대한 가정의 타당성을 검토하는 방법에 대해 논하시오. (25점)

제2문

A 감기약은 부작용으로 혈압이 낮아지는 것으로 알려졌다. 이를 알아보기 위해 10명의 감기환자를 표본추출하여 감기약 복용 1시간전과 복용 2시간 후에 혈압을 조사한 결과가 다음과 같다.

환자번호	1	2	3	4	5	6	7	8	9	10
복용 전	90	56	49	64	65	88	62	91	74	93
복용 후	72	55	56	58	62	79	55	72	73	74

위의 자료를 이용하여 실제로 혈압이 낮아지는 지를 알아보기 위한 통계적 분석방법(필요한 가정, 가설, 검정방법 등)에 대하여 논하시오. (25점)

제3문

어느 시청에서 주민들을 대상으로 행정서비스에 대한 주민들의 의견을 수렴하기 위해 표본조사를 실시하고자 한다. (총 25점)

(1) 어느 특정한 날 시청민원실에 들른 주민의 10%에 해당하는 주민을 표본으로 추출하는 방법에 대해 설명하시오. (10점)

(2) 표본으로 선정된 주민들을 대상으로 다음을 질문하였다.

다음 중 가장 시급하게 보완되어야 할 공공시설은 무엇입니까?

①도서관 ②수영장 ③보육시설 ④공원 ⑤기타

응답자 중 남녀 성별에 따라 응답에 차이가 있는지 알아보기 위한 검정방법에 대해 설명하시오. (15점)

[제45회, 2001년]

제1문

다음은 고등학생 1인당 사교육비 지출액에 대한 자료이다.(단위 : 만 원/월)

35	21	31	14	8	17	2	57	8	15	9	26	10	13	9	46	12
17	50	13	5	15	7	3	3	17	13	31	22	18	52	1	10	14
11	50	10	12	23	30	16	35	16	15	37	27	20	0	12	57	61

1) 자료의 분포적 특성을 잘 보여 줄 수 있는 그림을 하나 그리고 그림에 나타나는 분포적 특징을 기술하시오. (10점)

2) 이 자료의 평균은 20.9이고 중위수는 15이다. 이와 같이 차이가 나는 이유에 대하여 설명하시오. (10점)

3) 평균과 중위수 중 어느 것이 대표값으로 적당하다고 생각하는가? 이유를 들어 설명하시오. (10점)

제2문

어느 제약회사의 연구팀이 개발한 혈중콜레스테롤 수치를 낮춰 주는 신약 A가 기존의 약 B보다 더 효과가 있는 지를 알아보려 한다. 연구 목적에 맞는 귀무가설과 대립가설을 설정하고, 이 가설을 검정하기 위한 자료수집방법과 분석방법을 설명하시오. (20점)

제3문

정규분포를 따르는 모집단에서 추출한 확률표본으로부터 얻은 통계량의 분포와 관련하여 다음 물음에 답하시오. (총 30점)

1) t-분포, 카이제곱(χ^2) 분포, F-분포에 대하여 설명하시오. (10점)

2) 통계적 추론에서 t-분포, 카이제곱(χ^2) 분포, F-분포가 활용되는 예를 제시하시오. 제시한 예에서 사용된 통계량의 분포와 자유도를 제시하고, 그 이유를 1)의 내용을 중심으로 설명하시오. (20점)

제4문

자동차 판매대리점을 운영하고 있는 김씨는 중요한 결정을 해야 할 시점에 놓여 있다. 1월에서 3월까지의 자동차 판매에 대한 예측을 하여 새로이 차를 발주하여야 할 것인 지를 결정해야 한다. 현재의 자동차 재고는 360대이다. 만약 1월에서 3월까지의 자동차 판매 대수 합계가 현재의 자동차 재고를 넘을 확률이 60%보다 크다면 신규 발주를 하려고 한다. 이 자동차 판매 대리점의 1월에서 3월까지의 매월 판매대수는 과거의 실적으로부터 평균 100대, 표준편차 20대인 정규분포를 따른다고 한다. 다음의 각 물음에 답하시오. (총 20점)

1) 1월에서 3월까지의 자동차 판매 대수를 X_1, X_2, X_3 라 할 때, 제 1사분기의 판매대수 $Y = X_1 + X_2 + X_3$ 의 확률분포를 구하시오. (10점)

2) 신규발주를 할 필요가 있는 지 판단하시오. 단, Z가 표준정규분포를 따르는 확률변수일 때, $P(-0.5 < Z < 0.5) = 0.38$, $P(-1 < Z < 1) = 0.68$, $P(-2.0 < Z < 2.0) = 0.95$ 이다. (10점)

통 계 학

2002년 제46회 행정고등고시 제2차시험

응시번호 :

성명 :

제 1 문. 어느 회사 직원들의 TOEFL 시험성적분포가 정규분포 $N(\mu, 8100)$ 이라고 가정하자. 이 회사 직원들 중에서 $n=36$ 명을 랜덤추출하여 시행한 TOEFL성적을 가지고 귀무가설 $H_0: \mu = 530$ 과 대립가설 $H_1: \mu < 530$ 에 대한 검정을 기각역 $C = \{x: \bar{x} \leq 510.77\}$ 으로 실시하고자 한다. (단, \bar{x} 는 시험결과의 평균점수를 나타낸다.) 표준정규확률 변수의 누적확률함수 $\Phi(\cdot)$ 를 사용하여 다음 물음에 답하시오. (총 50점)

- 1) 이 검정의 유의수준 α 를 구하고, 만약 표본의 크기 n 이 유동적이라면 이것의 변화가 α 값에 미치는 영향을 서술하시오. (15점)
- 2) 이 검정의 검정력함수 $K(\mu)$ 를 구하고, $K(\mu)$ 의 그래프의 개형을 그리시오. (15점)
- 3) 시험결과 수험자의 평균점수가 507.33일 때, 이 검정의 p 값(p -value)를 구하시오. (10점)
- 4) 대립가설을 $H_1: \mu \neq 530$ 으로 바꾼 후, 유의수준 0.05하에서 우도비검정을 할 경우 기각역을 구하시오. (10점)

제 2 문. 어떤 안전에 대한 여론조사의 결과 다음과 같은 자료를 얻었다. 이와 같은 범주형 자료의 분석에는 동질성 검정과 독립성 검정의 두 가지가 있다. 다음 물음에 답하시오. (다만, 이 자료를 이용하여 구체적으로 계산할 필요는 없음) (총 30점)

		의 건		합
		찬성	반대	
나이	30세 미만	225	75	300
	30세 이상	174	125	299
합		399	200	599

- 1) 동질성 검정과 독립성 검정을 위한 각각의 가설을 제시하고, 어떤 차이가 있는지 설명하십시오. (5점)
- 2) 동질성 검정과 독립성 검정을 위한 각각의 적절한 자료수집 방법에 대하여 설명하십시오. (10점)
- 3) 두 검정 방법에서 사용되는 추정기대도수의 계산방법과 검정통계량의 자유도는 결과적으로 동일한 형태를 갖게 되지만 이들에 대한 유도과정을 보면 본질적으로 차이가 있다. 어떤 차이가 있는지 설명하십시오. (15점)

제 3 문. 두 사람 A와 B가 최대 4회의 게임을 하는데 매 게임에서 A가 이길 확률이 θ , B가 이길 확률이 $1 - \theta$ 이다. 만약 B가 2승하기 전에 A가 3승하면 A가 최종 승자가 되며, 반대로 A가 3승하기 전에 B가 2승하면 B가 최종 승자가 된다. A가 최종 승자가 될 확률을 구하십시오. (구하는 과정도 쓰시오.) (20점)

확인 : 행정자치부 고시과장

2003년도 제19회 입법고등고시 제2차시험

【 통계 학 】 - 재경직

【문1】

어느 제약회사에서는 새로 개발된 진통제의 효과를 알기 위하여 환자 n 명에게 양을 달리하여 투여한 후, 지속시간을 기록하였다. $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 이 진통제의 양과 지속시간을 기록한 자료이고, 두 변수 사이의 선형관계를 나타내는 통계모형으로 $y_i = \beta x_i + \varepsilon_i$, $i = 1, 2, \dots, n$ (여기서 ε_i 는 오차항으로 서로 독립이고 평균이 0이며 분산이 σ^2 이다.)을 선택했을 때,

1. 오차제곱합 $\sum_{i=1}^n \varepsilon_i^2$ 을 최소화하는 β 의 최소제곱추정량이 $b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$ 로 주어짐을 보여라.
2. b 의 의미를 투여량과 지속시간 두 단어를 넣어 설명하라.
3. b 의 기대값을 구하라.
4. b 의 분산이 $\frac{\sigma^2}{\sum_{i=1}^n x_i^2}$ 로 주어짐을 보여라.

【문2】

자유로를 통과하는 차량의 시속 V 의 분포는 정규분포를 따른다고 가정하자. 시속 80km가 넘는 차량이 전체의 86%, 96km 이하는 70%라는 사실을 알았다. 다음 질문에 답하라.

단, 표준정규분포를 따르는 변수 Z 에 대하여 $P(Z \leq 0.52) = 0.70$, $P(Z \leq 1.08) = 0.86$, $P(Z \leq 1.28) = 0.90$, $P(Z \leq 1.64) = 0.95$, $P(Z \leq 1.9) = 0.97$, $P(Z \leq 1.96) = 0.975$, 이다.

1. 자유로를 통과하는 차량의 시속 V 의 평균(μ)과 표준편차를 구하라.
2. 어느 상수 c 에 대하여 모든 차량속도의 90%가 $\mu \pm c$ 사이에 위치한다면 c 의 값은 얼마인가?
3. 대략 몇 %의 운전자가 시속 110km 이상으로 운전하고 있는가?

【문3】

A회사와 B회사는 합병이 논의되고 있다. 논의되기 이전의 A회사의 주식가격은 4900원, B회사의 주식가격은 2000원이었다고 한다. 합병이 논의되고 있는 현재의 주식가격은 시너지효과에 의해 A회사는 5400원, B회사는 3000원에 거래되고 있다고 한다. 합병이 성사될 확률은 80%이고 합병이 되면 A회사의 주식은 5500원으로 B회사의 주식은 4000원으로 될 것으로 예측된다면, 어떤 회사의 주식을 살 것인지를 기대치의 개념에서 설명하라. 단, 합병이 실패하면 현재의 주식가격은 합병논의 이전의 주식가격으로 환원된다고 가정한다.

통 계 학

2003년 시행 제47회 행정고등고시 제2차시험

응시번호 :

성명 :

제 1 문. 정규분포 $N(\mu_1, \sigma^2)$ 을 따르는 모집단과 정규분포 $N(\mu_2, \sigma^2)$ 을 따르는 모집단에서 크기가 n_1, n_2 인 표본을 추출하였다. 여기서 μ_1, μ_2, σ^2 은 미지의 모수이다. 각 표본의 표본평균을 \bar{X}_1, \bar{X}_2 , 표본분산을 S_1^2, S_2^2 이라 하자. 두 모집단의 평균을 비교하고자 한다. (총 30점)

- 1) 적당한 가설을 설정하시오. (5점)
- 2) 가설을 검정하기 위한 검정통계량을 유도하시오. (10점)
- 3) 유의수준 5%에서 2)의 검정통계량을 이용하여 가설검정을 수행하는 과정을 설명하시오. (15점)

제 2 문. 어느 닷컴 기업에서 등록회원을 대상으로 이용자 만족도를 전화면접으로 조사하고자 한다. 총 회원 100만명 중 1,000명을 표본으로 추출하고자 한다. 다음을 설명하시오. (총 30점)

- 1) 단순임의추출(simple random sampling) 방법 (6점)
- 2) 회원 가입순서에 따라 계통추출(systematic sampling) 방법 (7점)
- 3) 층화임의추출(stratified random sampling) 방법 (7점)
(이때, 회원의 성(gender)과 나이를 층화변수로 할 것)
- 4) 각 방법의 장점 (10점)

제 3 문. 한 여론조사 발표에서 1,000명을 조사한 결과 “찬성이 52%이며 95% 신뢰 수준하에서 최대허용오차는 $\pm 3.1\%$ 이다.”라고 발표하였다. (총 20점)

- 1) 최대허용오차가 $\pm 3.1\%$ 라는 주장의 의미와 이 값의 계산과정을 설명하시오. (10점)
- 2) 이 발표에 의할 때 모집단의 과반수가 찬성한다고 주장할 수 있는가에 대한 의견을 제시하고 그 이유를 설명하시오. (10점)

제 4 문. 어느 인터넷 쇼핑 기업에서 기존회원의 활성화 캠페인으로서 회원에게

A : 1,000원의 사이버 머니를 제공하는 방법

B : 서비스 안내책자를 우송하는 방법

C : 무(無)조치

를 비교하고자 한다. 그리고 캠페인의 효과는 캠페인 후 한 달간의 인터넷 쇼핑 금액으로 측정한다고 하자. (총 20점)

- 1) 두 방법의 효과를 비교하는 실험을 구체적으로 설계하시오. (10점)
- 2) 분석방법을 제시하시오. (10점)

확인 : 행정자치부 고시과장



응시번호 :

성 명 :

【제 1 문】 (50점)

어느 도시는 하루에 발생하는 교통사고 건수를 매일 모니터링 한다. 그리고 사고 건수가 90건에서 110건 사이이면 정상적인 날로, 그렇지 않으면 비정상적인 날로 판단한다고 한다. 실제 그 도시에서 매일매일 발생하는 교통사고 건수는 독립적으로 평균이 100이고 표준편차가 5인 정규분포를 따른다고 한다.

- ① 몇 일만에 최초로 비정상적인 날이 발생할 것으로 기대하는가?(과정을 제시할 것)
- ② 그 도시의 교통사고 건수의 분포가 표준편차는 같지만 평균이 105로 증가하였다면 몇 일만에 최초로 비정상적인 날이 발생할 것으로 기대하는가?(과정을 제시할 것)
단, 표준정규분포를 따르는 변수 Z 에 대하여 $P(Z \geq 1) = 0.16$, $P(Z \geq 2) = 0.02$, $P(Z \geq 3) = 0.001$ 로 가정한다.
- ③ 사고 건수의 평균이 100인 경우에는 평균적으로 150일 만에 최초로 비정상적인 날이 발생하고, 사고 건수의 평균이 110인 경우에는 평균적으로 5일 만에 최초로 비정상적인 날이 발생하도록 모니터링 기준을 결정할 수 있는지의 여부를 판단하여라 (근거를 제시할 것). 그리고 그러한 기준을 결정할 수 있다면, 그 기준을 구하는 과정을 설명하라.

【제 2 문】 (50점)

1. 다음 상황을 고려하자.(20점)

귀하가 지방자치단체의 장이라고 가정하자. 지방자치단체에 대한 중앙정부의 지원금 할당과 관련하여 최근 중앙정부로부터 온 공문의 내용을 간단히 요약하면 다음과 같다. “여러분이 관할하는 지방자치단체 소속 주민의 연간 소득을 대표할 수 있는 수치를 보고하라. 이 수치를 보고 중앙정부의 지원 액수를 결정하고자 하며, 소득수준이 타 지역에 비하여 떨어지는 지방자치단체에 대해서는 지원금을 확대할 예정이다.”

귀하의 대답과 함께 그러한 결론을 이끌어낸 논리를 피력하라.

2. 어느 보험회사에서는 사람들을 두 부류로 나눌 수 있다고 믿고 있다. 즉, 사고성향이 있는 사람과 그렇지 않은 사람으로 나눌 수 있다. 보험회사의 과거 자료에 의하면 사고성향이 있는 사람의 경우 주어진 1년 내에 사고를 일으킬 확률은 0.4이고, 반면에 사고성향이 없는 사람이 사고를 일으킬 확률은 0.2이다.(30점)

- ① 만약 모집단의 30%가 사고성향이 있다고 하면, 어떤 새로운 계약자가 계약한지 1년 내에 사고를 일으킬 확률은?
- ② 새로운 계약자가 계약 1년 내에 사고를 낸다고 가정하자. 그러면 그가 사고성향을 갖고 있을 확률은?
- ③ 어떤 계약자가 계약 첫 해에 사고를 냈다는 조건하에서, 이 계약자가 계약 다음 해에도 사고를 낼 확률은?

통 계 학

2004년 시행 제48회 행정고등고시(행정·공안직) 제2차시험

응시번호 :

성명 :

제 1 문. 한 사업가가 선택할 수 있는 대안이 A_1, A_2 두 가지가 있고, 사업을 시작한 후 예상할 수 있는 경제상황이 S_1, S_2 두 가지가 있으며, 각 경우의 이익금은 아래의 표와 같이 주어진다.

대안	불확실한 상황	
	S_1	S_2
A_1	100	75
A_2	50	120

과거의 경험으로 보아 경제상황 S_1, S_2 가 나타날 확률은 각각 0.4, 0.6으로 예상된다. (40점)

- 1) 최대 기대이익을 기준으로 어느 대안이 선택되나?
- 2) 위의 사업가는 더욱 확실하게 경제전망을 파악하기 위하여 한 경제연구소에 상담을 의뢰하였다. 이 경제연구소의 과거 실적에 의하면 경제상황이 S_1 인 경우는 90 %, 경제상황이 S_2 인 경우는 80 %의 예측 정확도를 보였다. 경제연구소의 전망이 S_1 으로 나왔을 때 경제상황 S_1, S_2 의 사후확률을 계산하고 이를 이용하여 최대 기대이익을 기준으로 하는 대안을 선택하시오.
- 3) 경제연구소의 전망이 S_2 로 나왔을 때 경제상황 S_1, S_2 의 사후확률을 계산하고 이를 이용하여 최대 기대이익을 기준으로 하는 대안을 선택하시오.
- 4) 이러한 상담을 통하여 사업가가 얻을 수 있는 이득에 대해 논하시오.

제 2 문. 모비율 θ 에 대한 추정량 U 를 생각해 보자. 추정량 U 는 표본을 추출하는 대신에 동전을 던져서 앞면이 나오면 $\theta = \frac{1}{3}$ 이라고 추정하고 뒷면이 나오면 $\theta = \frac{2}{3}$ 라고 추정한다. 이 추정량 U 에 대한 평균제곱오차(mean square error)를 구하고, 평균제곱오차의 그래프를 그리시오. (10점)

제 3 문. 가구의 전력소비량 Y 에 대하여 통계적 모형을 설정하고자 한다. (25점)

- 1) Y 에 영향을 주는 변수로 가구원수 $N(=1, 2, 3, \dots)$, 가구원 총소득 I (단위: 만원/월), 가장의 직업 $J(= \text{자영업, 사무직, 노무직, 기타})$ 을 고려한 선형회귀 모형을 구체적으로 제시하여라. 모두 몇 개의 설명 변수가 필요한가?
- 2) 변수 J 가 Y 에 미치는 영향을 통계적으로 유의한가를 검정하는 방법을 제시하시오.
- 3) 변수 Y 와 I 의 분포가 큰 값 쪽으로 긴 꼬리를 갖는다면 어떤 변환을 취하는 것이 좋은가?

제 4 문. 어느 세무서에서는 소득세 신고서를 감사하기 위하여 표본을 추출하기로 하였다.

전체 신고서들을 총소득이 비슷한 세 개의 집단으로 분류하고, 집단의 크기에 비례하도록 총표본 50개를 각 집단에 배정하였다. 그런 다음 각 집단에서 배정된 크기만큼의 신고서를 랜덤하게 추출하여 실시한 결과가 다음과 같았다. 물음에 답하시오. (25점)

집 단	집단의 크기	표본의 크기	표본 내에 있는 부실신고서의 수
1. 2,000이하	600	(a)	3
2. 2,000 ~ 4,000	300	(b)	3
3. 4,000이상	100	(c)	2
합 계	1,000	50	8

- 1) 이러한 표본추출법을 무엇이라고 하는가? 또, 이러한 추출법은 단순랜덤추출법에 비하여 어떤 장점이 있는가?
- 2) 위 표의 (a), (b), (c)에 들어갈 숫자를 구하고, 각 집단에서의 부실 신고율 p_1, p_2, p_3 를 각각 추정하시오.
- 3) 집단 3의 부실 신고율의 추정량에 대한 표준오차를 구하시오.
- 4) 총 부실신고서의 수 X 를 추정하시오.
- 5) 전체 집단의 부실 신고율 p 를 추정하시오.

통 계 학

제 1 문 (40점)

제품조립을 위한 직업훈련프로그램으로서 두 가지 다른 방법(A, B)이 제안되었다. 직업훈련기관은 두 방법 중에서 효과적인 방법을 채택하기 위하여 실험을 하고자 한다. 실험에 소요되는 비용의 제약 때문에 22명의 훈련생만을 대상으로 실험할 수 있으며, 일정기간 훈련을 실시한 후 각 훈련생이 제품을 조립하는데 걸리는 시간을 측정하여 의사결정에 활용하고자 한다.

- (1) 실험을 어떻게 실행해야 하는지 계획을 수립하라.
- (2) 실험에서 방법A에 의해 훈련받은 훈련생의 수와 방법B에 의해 훈련받은 훈련생의 수가 같고, 측정된 자료로부터 다음과 같이 요약된 값들을 얻었다고 가정한다.

[방법A] 훈련생들의 소요시간 합계 = 55, 훈련생들의 소요시간 제곱의 합계 = 285.

[방법B] 훈련생들의 소요시간 합계 = 69, 훈련생들의 소요시간 제곱의 합계 = 451.

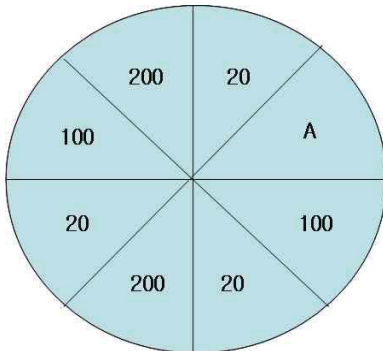
훈련생이 제품을 조립하는데 걸리는 시간이 정규분포를 따른다고 가정하고 의사결정을 위한 통계적 가설을 설정하라.

- (3) 통계적인 방법에 바탕을 둔 의사결정을 하라. (단, 유의수준은 5%를 적용하고, 다음 값들을 참고할 것. $F(a,b)$ 는 자유도 a,b 인 F -분포, $t(c)$ 는 자유도 c 인 t -분포를 의미함)

$$\begin{aligned}
 P(F(10,10) \leq 2.98) &= 0.95, & P(F(10,10) \leq 3.72) &= 0.975, \\
 P(F(11,11) \leq 2.82) &= 0.95, & P(F(11,11) \leq 3.47) &= 0.975, \\
 P(F(20,20) \leq 2.12) &= 0.95, & P(F(20,20) \leq 2.46) &= 0.975, \\
 P(F(22,22) \leq 2.05) &= 0.95, & P(F(22,22) \leq 2.36) &= 0.975, \\
 P(t(10) \leq 1.812) &= 0.95, & P(t(10) \leq 2.228) &= 0.975, \\
 P(t(11) \leq 1.796) &= 0.95, & P(t(11) \leq 2.201) &= 0.975, \\
 P(t(20) \leq 1.725) &= 0.95, & P(t(20) \leq 2.086) &= 0.975, \\
 P(t(22) \leq 1.171) &= 0.95, & P(t(22) \leq 2.074) &= 0.975.
 \end{aligned}$$

제 2 문 (60점)

1. 어느 대학의 입시홍보브로슈어에는 다음과 같은 글이 적혀 있었다. “본 대학의 졸업생들은 졸업 10년 후 평균연봉이 1억원으로 조사되었다.” 동창회 명부에 있는 졸업생들에게 설문지를 보내 응답 설문지를 통해 평균치를 계산한 것이다. 이 브로슈어를 보고 이 대학에 지원하려는 학생은 만약 이 대학에 입학만 하면 앞으로 큰 부자가 될 수 있을 것으로 생각하였다. 이 학생의 생각에 대하여 통계학적인 관점에서 문제점을 지적하시오.(20점)
2. 초등학교 정문앞 문구점에서 아래 그림과 같이 8등분한 원판의 각 면에 숫자가 적혀 있다. 초등학교생이 100원을 걸고 돌아가는 원판에 화살을 던져 맞춘 부분에 있는 숫자만큼 돈을 주기로 할 때 이 문구점이 손해를 보지 않기 위해 A부분에 넣을 수 있는 최대값을 구하는 문제이다.(20점)



- 가. 확률분포표를 만들고, A부분에 넣을 수 있는 최대값을 구하라.
- 나. ‘가’에서 구한 최대값을 이용하여, 문구점 주인이 주는 금액의 분산을 구하라.

3. p -value(p 값)에 대한 정의를 기술하고 통계분석 package의 출현과 더불어 더욱 중요해지는 이유를 설명하라.(20점)

통 계 학

2005년 시행 제49회 행정고등고시(행정공안직) 제2차시험

응시번호 :

성명 :

제 1 문. 종속변수 Y 의 변동을 설명하기 위하여 p 개의 독립변수를 고려한 다중선형 회귀모형이 다음과 같다고 가정하자. (총 20점)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + e_i, \quad i = 1, \dots, n,$$

단, e_i 는 서로 독립이며 $N(0, \sigma^2)$ 이다.

- 1) 위 모형에서 독립변수의 관측값들 사이에 강한 선형관계가 존재할 때 발생하는 회귀분석의 문제점과 그 해결책에 대해 논하시오. (7점)
- 2) 위 모형이 $\beta_2 = \cdots = \beta_p = 0$ 인 단순회귀 모형일 경우 종속변수 Y 와 X_1 의 상관 계수 ρ 와 회귀계수 β_1 간에 관계에 대해 설명하시오. (5점)
- 3) 위 모형에서 종속변수의 관측값(Y_i)이 0과 1의 값만 가질 경우 발생하는 회귀분석 모형의 가정에 대한 문제점과 그 해결책에 대해 논하시오. (8점)

제 2 문. 수식을 사용하지 않고 중심극한정리(central limit theorem)의 의미와 그 중요성에 대하여 기술하시오. (10점)

제 3 문. 두 종류의 자동차 타이어 A와 B의 마모도에 차이가 있는지를 조사하기로 하였다. 먼저, 16대의 택시를 임의(random)로 선택하고, A와 B 각각 32개의 타이어를 16대의 자동차에 임의로 장착하여 3만 km를 주행한 후에 각 타이어의 마모도를 밀리미터(mm) 단위로 측정하기로 한다. (총 10점)

- 1) 택시에 따라 주로 운행하는 도로의 상태가 다르고, 따라서 타이어의 마모도에 차이가 날 수 있다. 각 종류별 32개 타이어를 16대의 자동차에 어떻게 장착하는 것이 마모도의 차이를 정확히 검정할 수 있는가를 설명하시오. (5점)
- 2) 실제, 택시 1대 당 종류별로 2개씩의 타이어를 장착하여 주행하고, 각 택시는 타이어 종류별로 마모도의 평균값만을 보고하였다고 한다. 두 종류의 타이어 마모도에 차이가 있는지를 검정할 수 있는 방법을 제시하시오. (가설과 필요한 가정, 그리고 검정 방법을 제시할 것) (5점)

제 4 문. 앞면에 1부터 18까지의 숫자가 적힌 18장의 카드와 '*'의 표시가 된 2장의 카드가 있고 뒷면은 어느 카드나 같다. 어느 사람이 초지각능력을 가졌다고 주장하고 있어 2장의 '*' 표시 카드를 비복원(非復元)으로 뽑는 실험으로 초능력을 테스트하기로 하였다. 실험 결과, 1장의 '*' 표시 카드와 1장의 숫자 카드가 뽑혔다고 하자. (총 10점)

- 1) 이 실험연구의 귀무가설(영가설)과 대립가설(대안가설)을 기술하시오. (5점)
- 2) p -값(유의확률)을 구하여라. 유의수준 10%에서 검증(검정) 결과를 제시하시오. (5점)

중앙인사위원회 인재채용과장



중앙인사위원회
CIVIL SERVICE COMMISSION

통 계 학



통 계 학

▶ 응시번호 :

▶ 성 명 :

1. 세 개의 투자자산 A, B, C의 수익률 X_1, X_2, X_3 은 다음의 관계를 가지고 있다고 한다.

$$X_1 = Z + \varepsilon_1, \quad X_2 = 2Z + \varepsilon_2, \quad X_3 = 3Z + \varepsilon_3.$$

단, 여기에서 $Z, \varepsilon_1, \varepsilon_2, \varepsilon_3$ 는 서로간에 독립이며, Z 는 시장수익률을 나타내며 평균은 0.1%, 표준편차는 5%이라고 한다. 그리고 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 는 모두 평균이 0이며 표준편차는 각각 $\sqrt{75}\%$, $\sqrt{125}\%$, $\sqrt{175}\%$ 라고 한다. A, B, C 세 개의 자산에 대한 기대수익률이 0.22%인 포트폴리오(portfolio, 분산투자) 중 최소분산을 갖는 포트폴리오를 구하시오. (30점)

2. 시계열 데이터 X_1, \dots, X_n 이 주어졌을 때 1차 자기 상관 계수는

$$\hat{\rho}_1 = \frac{\sum_{i=1}^{n-1} (X_i - \bar{X})(X_{i+1} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

으로 정의된다. 이 값은 미래의 값 X_{n+1} 을 예측하는데 중요한 역할을 하는 것으로 알려져 있다.

주가의 움직임을 예측하는데 있어서 5일 평균이 30일 평균을 상향 돌파하는 경우, 추가 상승이 기대된다는 말을 많이 하는데 이를 Golden Cross 가 발생한다고 말한다. 이러한 골든 크로스 현상을 위의 1차 자기상관계수를 통해 설명하시오. (30점)

3. 어떤 조합의 임원들은 그들이 관리하는 근로자들이 회사로부터 받는 봉급의 적정수준에 관심을 가지고 있다. 그래서 시간당 평균 임금을 평가하기 위해 회사로부터 n명을 골라, 적정수준의 여부를 조사하기로 결정하였다. 만일 회사의 봉급의 격차가 시간 당 10(천)원의 범위(최대값-최소값)를 가지는 것으로 알려져 있다면, 시간당 평균 임금 μ 를 0.6(천)원의 허용오차를 가지는 95%신뢰구간으로 추정하기 위한, 샘플의 크기 n을 결정하시오. (단위=천원) (20점)

4. 서로 다른 세 종류의 감귤나무에 대한 4가지 다른 살충제의 효과를 검증하고자 하는 실험이 수행되었다. 세 종류의 나무는 과수원에서 임의로 추출되었고 4종류의 살충제는 각 종류의 나무에 랜덤한 순서로 할당되었다. 이 실험의 반복수가 5번일 때 교호작용을 검증하기 위한 가설을 세우고 이에 대한 검증 절차를 설명하시오. (20점)

통 계 학

2006년 시행 행정고등고시(행정·공안직) 제2차시험

응시번호 :

성명 :

제 1 문. 어떤 제품의 품질변수 X 가 평균이 μ , 분산이 4인 정규분포를 따르고, 제품의 규격은 $8 \leq X \leq 12$ 이다. X 가 규격 내에 속하면 A원의 이익이, $X < 8$ 이면 B원의 손실이, $X > 12$ 이면 C원의 손실이 발생한다. 다음 물음에 답하시오. (총 20점)

<참고>

- 아래의 표준정규확률변수 Z 의 근사적 성질을 이용하시오
 $P[-1 \leq Z \leq 1] = 0.68$, $P[-2 \leq Z \leq 2] = 0.95$, $P[-3 \leq Z \leq 3] = 0.99$
- 표준정규누적확률함수의 기호 $\Phi(\cdot)$ 을 사용하시오

- 1) μ 가 10인 경우와 12인 경우에 대하여, 규격 상한을 초과할 확률과 규격 내에 속할 확률을 각각 구하고, μ 가 증가할 때 각 확률의 변화를 설명하시오. (5점)
- 2) 현재 μ 가 10이라고 알고 있으며, 연속적으로 제조될 100개의 제품 중에서 90개 이상이 규격 내에 속하면 변수의 분산이 감소된 것으로 판단한다. 실제 분산이 4인데도 분산이 감소된 것으로 판단할 확률을 구하시오. (5점)
- 3) μ 가 8인 경우에 최초로 규격 상한을 초과하는 제품은 몇 번째일 것으로 기대하는가? (5점)
- 4) 이 제품의 제조 과정에서 변수 X 의 평균 μ 를 조정할 수 있다고 한다. 기대이익을 최대로 하는 μ 의 값을 결정할 수 있는지를 설명하시오. (5점)

제 2 문. 어떤 건전지의 수명 X 에 대한 확률밀도함수가 다음과 같다.

$$f(x) = \alpha e^{-\alpha x}, \quad 0 < x < \infty$$

서로 독립인 두 건전지의 수명을 각각 X_1 , X_2 라 하고, $Y_1 = X_1 + X_2$ 라 할 때 다음 물음에 답하시오. (총 15점)

- 1) Y_1 의 확률밀도함수를 구하시오. (7점)
- 2) Y_1 과 X_1 사이의 상관계수를 구하시오. (8점)

제 3 문. 정책 효과를 평가하기 위하여 n 명으로 패널을 구성하여 정책시행 이전과 이후의 수행성과에 관한 측정을 하여 다음과 같은 자료를 산출하였다. 다음 각 경우에 대하여 시행전후의 차이가 통계적으로 유의한가를 검증하는 방법을 제시하시오. (총 15점)

패널번호	시행전	시행후
1	y_{11}	y_{12}
2	y_{21}	y_{22}
\vdots	\vdots	\vdots
n	y_{n1}	y_{n2}

관측값 y_{ij} ($i = 1, \dots, n; j = 1, 2$) 가,

- 1) 연속형 측정값일 때 (7점)
- 2) 명목형 범주 c_1, \dots, c_k 중 1개를 취할 때 (8점)

중앙인사위원회 출제관리과장

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (40점)

100명의 학생이 두 개의 선택과목 중 한 과목을 선택하여 시험을 친 결과가 다음과 같다.

구 분	인 원	평 균	분 산
선택과목 1	40	80	100
선택과목 2	60	70	25

- 1) 과목 선택을 무시하고 전체 학생의 평균과 분산을 계산하라
- 2) ‘선택과목 1’의 답안지 중 1장이 잘못 채점되어 실제로 10점인데 90점으로 계산되었다. 실제 ‘선택과목 1’의 평균과 분산을 계산하라.
- 3) ‘선택과목 1’의 난이도가 ‘선택과목 2’ 보다 낮은 것으로 보인다. 이 경우 ‘선택과목 2’를 선택한 학생들이 불이익을 받지 않도록 난이도를 조정하는 방안을 제시하라.

제 2 문 (60점)

1. 우리나라 국민들을 대상으로 정부에서 입안한 새로운 교육정책에 대해 찬성하는 사람들의 비율 p 를 파악하고자 한다. 이와 관련해 두개의 조사기관에서 각각 독립적으로 랜덤하게 표본을 추출하여 여론조사가 수행되었다. 조사기관 ‘갑’에서는 800명을 랜덤하게 추출하여 조사한 결과 찬성률 $\hat{p}_1 = 480/800 = 0.6$ 을 얻었고, 다른 조사기관 ‘을’에서는 랜덤하게 추출된 200명을 조사하여 찬성률 $\hat{p}_2 = 80/200 = 0.4$ 를 얻었다. 정책 담당자는 두 조사기관에서 얻은 결과를 종합하여 좀 더 신뢰할 수 있는 찬성률 p 를 추정하기 위해 다음과 같은 두 가지 추정방법을 고려하고자 한다.(40점)

① $\hat{p} = (0.6 + 0.4)/2 = 0.5$

② $\tilde{p} = (480 + 80)/(800 + 200) = 0.56$

- 1) 두 가지 추정방법에 대해 각각 편향(bias)이 발생하는지 밝혀라.
 - 2) 두 가지 추정방법에 대해 추정량의 분산을 구하라.
 - 3) 제시된 \hat{p} 과 \tilde{p} 는 \hat{p}_1 과 \hat{p}_2 의 가중평균 형식의 추정량이다 이런 가중평균 추정량 중 가장 효율적인 추정량을 구하는 방법에 대해 설명하라.
2. 여러 개의 변수가 주어진 경우 이들 사이의 관계를 알아보고자 한다.(20점)
- 1) 편상관계수(partial correlation coefficient)를 정의하고 그 의미를 설명하라.
 - 2) 회귀분석에서 편상관계수가 어떻게 활용되는지 설명하라.

통 계 학

2007년 시행 행정고등고시(행정직) 제2차시험

응시번호 :

성명 :

제 1 문. 어느 지역 특정작물의 평균수확량을 추정하고자 세 그루의 작물을 임의로 추출하여 얻은 수확량을 각각 X_1, X_2, X_3 이라 하자. 각 수확량은 평균이 μ 이고 분산이 σ^2 인 모집단에서의 임의표본(random sample)이라고 할 때, μ 의 추정량으로서 두 추정량 $T_1 = \frac{X_1 + X_2 + X_3}{3}$ 과 $T_2 = \frac{Y + X_3}{2}$ 을 고려하기로 하자.

(단, $Y = \frac{X_1 + X_2}{2}$) (총 20점)

- 1) 두 추정량의 특성을 추정량이 갖추어야 할 성질을 중심으로 기술하시오. (5점)
- 2) 두 추정량 중 하나를 선택한다면 어떤 추정량을 선택하겠는가? 그 이유를 설명하시오. (5점)
- 3) X_1, X_2, X_3 중에서 X_1 과 X_2 는 서로 이웃한 작물의 수확량이어서 상관관계가 있다고 가정하자. X_1 과 X_2 의 상관계수를 ρ 라 할 때, ρ 의 값에 따른 적절한 추정량을 선택하고 그 이유를 설명하시오. (10점)

제 2 문. 주머니에 5개의 공이 들어 있는데, 그 중에 θ 개가 빨간색 공이고 나머지는 파란색 공이다. 가설 $H_0: \theta=2$ 대 $H_1: \theta=4$ 를 검정하기 위하여 2개의 공을 비복원 추출하여 2개 모두 빨간색 공이면 귀무가설 H_0 을 기각하고 그 이외의 경우에는 귀무가설을 기각하지 않는다. (총 10점)

- 1) 이 검정방법에 대한 제1종의 오류를 범할 확률을 구하시오. (5점)
- 2) 이 검정방법에 대한 제2종의 오류를 범할 확률을 구하시오. (5점)

제 3 문. 어떤 회사의 세탁기는 3개의 공장에서 생산된다. 생산된 전체 세탁기 중에서 20%는 A공장에서, 30%는 B공장에서, 그리고 나머지 50%는 C공장에서 생산되었다고 한다. A공장에서 생산된 세탁기의 불량률은 5%, B공장에서 생산된 세탁기의 불량률은 2%, 그리고 C공장에서 생산된 세탁기의 불량률은 2%라고 한다. (총 10점)

- 1) 이 회사에서 생산되는 세탁기의 불량률을 구하시오. (5점)
- 2) 임의로 추출한 2개의 세탁기가 모두 불량품일 때, 두 세탁기 모두 C공장에서 생산되었을 확률을 구하시오. (5점)

제 4 문. 다음은 두 확률변수 X 와 Y 의 결합확률분포이다. (총 10점)

$$P(X=0, Y=0) = p_{00}, P(X=0, Y=1) = p_{01},$$

$$P(X=1, Y=0) = p_{10}, P(X=1, Y=1) = p_{11},$$

여기서 $p_{00} + p_{01} + p_{10} + p_{11} = 1$ 이다.

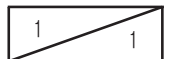
- 1) X 와 Y 가 독립이기 위한 조건을 구하시오. (5점)
- 2) X 와 Y 의 상관계수가 0인 조건이 X 와 Y 가 서로 독립이기 위한 필요충분 조건임을 설명하시오. (5점)

중앙인사위원회 출제관리과장



중앙인사위원회
CIVIL SERVICE COMMISSION

통 계 학



통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (40점)

3가지 공무원 연수방법의 효과를 서로 비교하기 위하여 총 30명의 연수 참가자를 대상으로 3개월간 연수를 실시한 후 연수 이전과 이후 두 차례 평가를 실시하여 연수 효과를 측정하기로 하였다.

- 1) 참가자의 나이가 연수 결과에 영향을 줄 수 있다고 할 때 적절한 실험계획법과 통계모형을 제시하시오.
- 2) 위의 실험결과를 이용하여 3가지 연수방법에 따른 차이가 나는지를 검정하는 방법을 구체적으로 서술하시오.

제 2 문 (20점)

주머니 속에 2개의 검정 공이 들어 있다. 이 때 주사위를 한 번 던져서 나온 눈의 수만큼 주머니에 흰 공을 추가로 넣어 두었다.

이 주머니에서 공을 한개 무작위로 뽑는 실험을 하였다.

- 1) 뽑힌 공이 검정 공일 확률은 얼마인가 ?
- 2) 뽑힌 공이 검정 공이라고 가정할 때, 주사위의 눈이 1이었을 확률은 얼마인가?

제 3 문 (20점)

정부에서 입안한 환경관련 정책에 대한 찬성률에 있어서 도시와 농촌 간에 차이가 있는지 알아보기로 하였다. 임의로 뽑힌 도시에 사는 성인 100명 중 70명이 정책에 찬성하였고, 또한 농촌에 사는 성인 100명 중 50명이 찬성하였다.

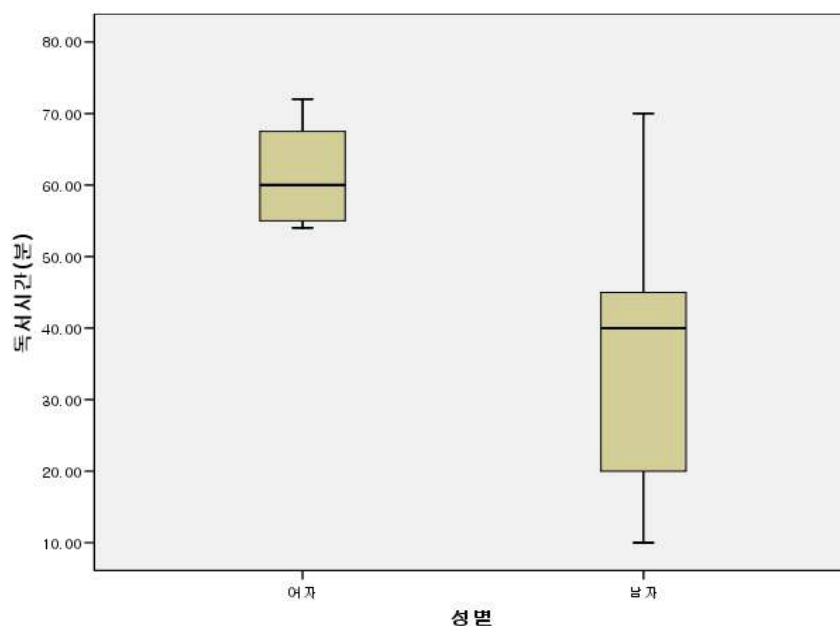
- 1) 도시의 찬성률이 농촌의 찬성률 보다 높은지 유의수준 5% 하에서 검정하시오.
- 2) 두 지역 간 찬성률에 있어 차이가 있다면 그 차이가 얼마나 되는지 신뢰수준 95% 하에서 추정하시오.
- 3) 도시와 농촌간의 인구비율이 7:3 이라고 할 때, 이 정책에 대한 전체 국민의 찬성률을 추정하는 방법을 논하시오.

※ 표준정규분포 확률변수 Z 에 대하여 $P(Z > 1.96) = 0.025, P(Z > 1.645) = 0.05$ 이다.

제 4 문 (20점)

다음 그림은 여자 11명과 남자 14명의 주당 독서시간(분) 자료에 대한 상자그림이다.

남녀 간 평균독서시간이 차이가 있는지 알아보려 한다. 적절한 통계분석 방법을 서술 하시오.



통 계 학

2008년 시행 행정고등고시(행정직) 제2차시험

응시번호 :

성명 :

제 1 문. 신, 구형 휴대폰 단말기의 선호도를 알아보기 위해 랜덤 추출한 36명을 2그룹으로 나누어서 각각의 단말기에 대한 선호도를 50점 만점의 점수로 평가하도록 하였다. 신형 단말기와 구형 단말기의 선호도 사이에 차이가 있는지에 대한 검정을 수행하고자 한다. 자료로부터 다음과 같은 t검정 결과를 얻었다. (총 10점)
(단, 원 자료에서 각 점수는 정규분포를 따르는 것으로 가정한다)

	n	평균	표본분산
구형 단말기	18	32	38
신형 단말기	18	39	34
T-검정 : $ T = 3.50$			
분산의 합동추정량(Pooled Variance) : 36			

- 1) 모분산이 동일하다는 가정 하에서는 모분산에 대한 합동추정량을 사용한다. 모분산이 동일한가를 위의 자료를 이용하여 검정하는 방법을 설명하시오. (5점)
- 2) 모분산이 동일하다는 가정 하에서, 신형 단말기의 선호도와 구형 단말기의 선호도 사이에 차이가 있는지에 대한 가설을 세우고, 검정통계량의 값(= 3.50)과 p값(유의 확률)을 구하는 과정을 구체적으로 설명하시오. (5점)

제 2 문. 서로 독립인 확률변수 X_1, X_2, X_3 의 분포는 각각 $X_i \sim N(\mu_i, \sigma^2), i = 1, 2, 3$ 이라고 한다. 각 모집단으로부터 각각 크기가 1인 표본을 추출한 결과 $x_1 = 1, x_2 = 2, x_3 = 3$ 을 얻었다. (총 15점)

- 1) $\mu_1 = \mu_2 = \mu_3$ 로 알려져 있는 경우에, σ^2 의 95 % 신뢰구간을 구하시오. (5점)
(단, $V \sim \chi^2(2)$ 일 때, $P(V \leq 0.05) = 0.025, P(V \leq 7.38) = 0.975$)
- 2) $\sigma^2 = 6$ 으로 알려져 있는 경우에, $\mu_1 + 2\mu_2 - \mu_3$ 의 95 % 신뢰구간을 구하되 그 과정을 구체적으로 설명하시오. (5점)
(단, $Z \sim N(0,1)$ 일 때, $P(Z \geq 1.96) = 0.025$)
- 3) $\sigma^2 = 6$ 으로 알려져 있는 경우에, 귀무가설 $H_0: \mu_1 + 2\mu_2 - \mu_3 = 8$ 과 대립가설 $H_1: \mu_1 + 2\mu_2 - \mu_3 \neq 8$ 에 대해 유의수준 5 %로 검정하기 위한 기각역(critical region)을 구하고, 이에 따라 귀무가설을 검정하시오. (5점)

제 3 문. X_1, X_2, \dots, X_n 이 균일분포(uniform distribution) $U(0, \theta)$ 에서 추출한 확률표본이고, 그 중 최대값을 T 라고 하자. (총 10점)

- 1) $Y = \frac{T}{\theta}$ 의 누적분포함수(cumulative distribution function)가 다음과 같음을 보여라. (5점)

$$F_Y(y) = P(Y \leq y) = \begin{cases} 0, & y \leq 0 \\ y^n, & 0 < y \leq 1 \\ 1, & y > 1 \end{cases}$$

- 2) $P(y < \frac{T}{\theta} \leq 1) = 0.9$ 가 되는 y 를 구하고, 이를 이용하여 θ 의 90 % 신뢰구간이 $[T, aT)$ 가 되도록 상수 a 를 구하여라. (5점)

제 4 문. 어떤 암의 치료를 위해 새로 개발된 항암제 NEW와 기존의 항암제 OLD의 임상 효과를 실증적으로 비교하고자 한다. 50명의 암환자에게는 항암제 NEW를 투약하고 다른 50명의 암환자에게는 항암제 OLD를 투약하여, 1년간 치료한 후 그 결과가 <표 1>과 같이 나타났다고 하자.

<표 1> 총괄자료

	생 존	사 망	합 계	생존율
항암제 NEW	30	20	50	0.6
항암제 OLD	20	30	50	0.4
합 계	50	50	100	

<표 1>에 의하면 항암제 NEW의 생존율이 항암제 OLD의 생존율보다 더 높은 것으로 나타나고 있다. 그런데 한 통계전문가가 좀 더 엄밀하게 이 자료를 탐사한 결과 위 100명의 암환자를 투약직전의 임상적 진단내용에 따라 경증과 중증 환자로 분류하여 <표 2>의 자료를 구성할 수 있었다.

<표 2> 세부자료

가) 경증환자

	생 존	사 망	합 계	생존율
항암제 NEW	28	12	40	0.7
항암제 OLD	8	2	10	0.8
합 계	36	14	50	

나) 중증환자

	생 존	사 망	합 계	생존율
항암제 NEW	2	8	10	0.2
항암제 OLD	12	28	40	0.3
합 계	14	36	50	

<표 2>의 세부자료에 의하면 경증환자와 중증환자 모두에서 <표 1>과는 달리 항암제 OLD의 생존율이 더 높은 것으로 나타나고 있다. (총 15점)

- 1) 총괄자료에서의 항암제 NEW와 OLD 각각의 생존율을 세부자료에서의 경증환자 생존율과 중증환자 생존율을 이용하여 표현하시오. (5점)
- 2) 총괄자료와 세부자료로부터의 생존율이 상반되어 보이는 현상에 대한 원인을 1)의 결과를 이용하여 설명하시오. (5점)
- 3) 해당 암에 걸린 새로운 환자가 내원했을 때, 어떤 항암제를 사용하는 것이 바람직한지를 판단하고 그 근거를 구체적으로 설명하시오. (5점)

행정안전부 시험출제과장

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (15점)

다음 각 문제에 대해 답하라. (각 3점)

- (1) 식약청에서 중국에서 수입된 농산물의 농약 잔류검사를 위해 다음과 같은 귀무가설 (null hypothesis)을 세웠다.

H_0 : 농약잔류치는 허용치 이하이다.

중국산 농산물의 잔류치에 대한 자료를 얻은 다음 검정한 결과 유의수준 (level of significance) 0.05에서 귀무가설을 기각하지 않았다. 만약 같은 자료를 이용하여 유의수준 0.01과 0.1에서 검정을 시행한다면 각각 어떤 결과가 나올 것인가?

- (2) 어떤 정책에 대한 지지도를 조사하여 95% 신뢰구간을 구했더니 (35%, 45%)였다. 이 신뢰구간이 의미하는 바는 무엇인가?
- (3) 대학수학능력시험의 결과는 원점수, 변환표준점수, 백분위수 등으로 수험생 개개인에게 주어진다. 어떤 학생의 언어와 외국어 모두 90점인데 변환표준점수는 각각 125점과 145점으로 나타났다. 표준변환점수가 계산되는 과정을 설명하고 이 처럼 원점수는 같은데 변환표준점수가 다르게 나타나는 이유를 설명하라.
- (4) 확률변수 Z_1, Z_2, Z_3, Z_4 가 서로 독립이고 표준정규분포인 $N(0, 1)$ 을 따른다고 가정하자. 다음 각 변수들의 분포를 명시하라.

$$(a) \frac{Z_1^2 + Z_2^2}{Z_3^2 + Z_4^2} \quad (b) \frac{Z_1}{|Z_2|} \quad (c) \frac{Z_1 + Z_2}{\sqrt{Z_3^2 + Z_4^2}}$$

- (5) 확률변수 T 가 자유도 n 인 t -분포를 따를 때에 $t_{0.025}(n)$ 는 $P[T > t_{0.025}(n)] = 0.025$ 을 만족하는 실수값이다. n 이 변함에 따라 $t_{0.025}(n)$ 이 다음과 같이 변한다. n 이 매우 커지면 ($n \rightarrow \infty$) $t_{0.025}(n)$ 은 어떠한 값으로 수렴하는가?

n	$t_{0.025}(n)$
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
10	2.228
20	2.086
:	:
:	:

제 2 문 (10점)

회귀모형

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

을 적합시킨 결과가 다음과 같다. (각 2점)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1422.80	711.40	111.21	<.0001
Error	10	63.97	6.40		
Corrected Total	12	1486.77			

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-65.02	15.07	-4.32	0.0015
X1	1	2.37	0.17	13.86	<.0001
X2	1	0.43	0.07	5.77	0.0002

- (1) 전체 관측치 수 n 은 얼마인가?
- (2) 반응변수(Y)의 분산에 대한 추정치는 얼마인가?
- (3) X_1 의 회귀계수 추정치인 2.37의 의미는 무엇인가?
- (4) 각 설명변수에 대응되는 모수를 각각 $\beta_0, \beta_1, \beta_2$ 라고 표시할 때 $H_0: \beta_1 = \beta_2 = 0$ 의 검정을 유의수준 5%에서 실시하라.
- (5) 첫 번째 관측치가 (120, 70, 50)으로 주어졌다. 이 관측치에 대한 잔차(residual)를 계산하라. 단 자료는 $(Y_i, X_{1i}, X_{2i}), i=1, 2, \dots, n$ 으로 주어져 있다.

제 3 문 (13점)

매번 시행에서 성공의 확률이 $p(0 < p < 1)$ 일 때 첫 번째 성공을 얻기 위해 필요한 시행횟수를 X 라고 하자. 단 각 시행은 서로 독립이다. 즉, i 번째 시행결과가 j 번째 시행 결과에 영향을 미치지 않는다. (단, $i \neq j$)

- (1) 확률변수 X 의 확률분포함수 (probability distribution function)를 구하라. (3점)
- (2) (1)의 결과를 이용하여 확률변수 X 의 기댓값 (expected value)을 구하라. (4점)
- (3) 한국인의 모든 부부가 자녀를 딸 선호사상에 의해서 낳는다고 가정하자. 딸 선호사상이란 첫 아이가 딸이면 더 이상 낳지 않고 첫 아이가 아들이면 둘째를 낳아서 딸이면 그만 낳고 아들이면 또 낳는다. 즉, 딸 선호사상이란 첫 딸을 낳을 때까지 계속 낳고 첫 딸을 낳으면 더 이상 낳지 않는 것이다. 단, 아들과 딸을 낳을 확률은 각각 0.5로 동일하며 임신 중에 초음파 검사 등을 통해 태아 성감별을 하여 강제유산 시키는 일은 없다고 가정한다. 이렇게 딸 선호사상으로 모든 부부가 자녀를 낳을 경우 많은 시간이 지난 후 한국인의 성비는 어떻게 되겠는가? (2)의 결과를 이용하여 답하라. (6점)

제 4 문 (12점)

어떤 회사에서 입사시험을 통해 신입직원을 채용하고 있다. 입사시험 성적 우수자가 입사 후 근무성적이 우수한지를 알아보려고 2008년 입사자들을 대상으로 입사시험 성적(A,B,C,D)이 1년 후 근무성적(0-100점)에 어떠한 영향을 미치는가를 조사하고자 한다. 이에 적절한 통계적 분석방법을 제시하라.

통 계 학

2009년 시행 행정고등고시(행정직) 제2차시험

응시번호 :

성명 :

제 1 문. 확률변수 X 의 확률밀도함수가 다음과 같다.

(총 13점)

$$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right), x > 0, \lambda > 0$$

- 1) 확률변수 X 의 제 90 백분위수(percentile)를 구하시오. (4점)
- 2) 위 확률밀도함수를 갖는 분포로부터 크기 n 인 표본을 뽑아서 그것을 X_1, X_2, \dots, X_n 이라고 할 때, 모수 λ 의 최우추정량(maximum likelihood estimator)을 구하시오. (5점)
- 3) 2)에서 구한 최우추정량이 불편성(unbiasedness)을 만족하는지 여부를 밝히시오. (4점)

제 2 문. 통계조사에서 표본추출방법은 일반적으로 확률표본추출법(probability sampling)과 비확률표본추출법(nonprobability sampling)으로 구분된다.

(총 10점)

- 1) 확률표본추출법과 비확률표본추출법을 비교하여 설명하시오. (6점)
- 2) 할당추출법(quota sampling)에 대해 설명하고, 이 방법을 사용하여 조사하는 경우 어떤 문제가 발생할 수 있는지 기술하시오. (4점)

제 3 문. 다음은 어느 회사의 같은 부서에 근무하는 7명의 직원들에 대하여 시행한 직무능력 평가 결과와 그 직원들의 대학교 평균평점 자료이다.

직무능력 평가점수(Y)	80.3	85.7	83.5	92.9	78.1	87.2	90.4
대학교 평균평점(X)	3.4	3.9	3.3	4.3	3.0	3.4	3.9

대학교 평균평점과 직무능력 평가점수 사이에 어떤 연관이 있는지 알아보기 위해 단순선형회귀 모형을 적합하여 다음과 같은 결과를 얻었다.

(총 15점)

Analysis of Variance					
Source	DF	Sum of Square	Mean Squares	F Value	Pr > F
Model	1	129.2	129.2	16.86	0.0003
Error	5	38.3	7.7		
Corrected Total	6	167.5			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	48.1	9.2	5.25	0.0033
X	1	10.4	2.5	4.11	0.0003

- 1) 회귀모형의 유의성을 검정하고자 한다. 귀무가설과 대립가설을 설정하고 유의수준 5%에서 검정하시오. (5점)
- 2) 위의 결과물을 이용하여 회귀모형에 대한 결정계수(R^2)를 소수점 둘째자리까지 구하고 그 의미를 설명하시오. (3점)
- 3) 적합된 회귀식을 기술하고 그 의미를 설명하시오. (5점)
- 4) 어느 직원의 대학교 평균평점이 4.0일 때, 이 직원의 직무능력평가점수를 예측하시오. (2점)

제 4 문. 다음 물음에 답하시오.

(총 12점)

- 1) $P(X=1)=p$, $P(X=0)=1-p$ ($0 < p < 1$)일 때, 확률변수 X 의 분산을 구하시오. (3점)
- 2) U 는 0과 1사이에서 균일분포(uniform distribution)를 갖는 확률변수이고, $P(X=1)=U$, $P(X=0)=1-U$ 라고 한다.
 - ① 확률변수 U 의 기댓값과 분산을 구하시오. (3점)
 - ② 조건부 기댓값의 성질을 이용하여 확률변수 X 의 분산을 구하고, 1)에서 구한 값과 비교하시오. (3점)
 - ③ $X=0$ 일 때, U 의 조건부 확률밀도함수를 구하시오. (3점)

행정안전부 시험출제과장



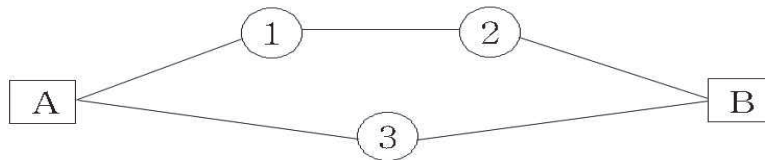
통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (20점)

A지역에서 B지역까지 천연가스 수송관을 아래와 같이 설계하고 주요지점에 압력을 조절하는 장치 ①, ②, ③을 설치하려고 한다. 압력조절 장치에 이상이 있으면 그 부분에서 가스공급이 중단되도록 만들고 각 압력조절 장치는 별개의 업체에서 독립적으로 제작되었으며 수명은 모두 평균이 5년인 지수분포를 따른다고 하자. 압력조절장치를 제외한 다른 부분에서는 문제가 발생하지 않는다고 가정한다. 압력조절 장치를 5년 전 같은 시기에 설치하고 지금까지 보수 및 수리를 하지 않았다. 단, $\exp(-1) = 0.37$, $\exp(-2) = 0.14$, $\exp(-3) = 0.05$ 이고, 평균이 c 인 지수 분포의 확률밀도함수는 $f(x) = \frac{1}{c} e^{-x/c}$, $x > 0$ 이다.



- (1) 압력조절장치 ③이 설치된 이후 10년 이상 정상적으로 작동할 확률을 구하여라. (3점)
- (2) 현재 A지역에서 B지역까지 가스가 공급될 확률을 구하여라. (7점)
- (3) 현재 A지역에서 B지역까지 가스가 공급되고 있을 때, ①과 ②가 정상적으로 작동하고 있을 확률을 구하여라. (6점)
- (4) 현재 ①, ②, ③이 모두 정상적으로 작동하고 있다고 할 때, 5년 후에도 A지역에서 B지역까지 가스가 공급될 확률을 구하고 그 이유를 기술하여라. (4점)

제 2 문 (10점)

어떤 의회는 100명의 의원으로 이루어져 있고 의원 중 45%가 A당, 30%가 B당, 25%가 C당 소속이라고 하자. 본회의에 제출된 법안에 대해 의원 과반수이상의 출석과 출석의원 중 과반수이상의 찬성이 있으면 법률이 통과된다고 하자. 이번에 심의된 법률에 대해 A당 의원의 80%, B당 의원의 20%, C당 의원의 40%가 찬성하고 있다고 할 때 아래의 물음에 답하여라.

- (1) 무작위로 선정된 한 의원이 법안을 찬성하고 있다면, 이 의원이 C당 소속 의원일 확률을 구하여라. (4점)
- (2) 의원 중 3명을 비복원으로 무작위 선정하여 법안에 대해 찬성여부를 알아보았을 때 2명이 찬성할 확률을 구하여라. (6점)

제 3 문 (8점)

어떤 일기예보의 정확도를 확인하기 위해 “비율 확률 50%”로 비가 예보된 날들을 조사해 보았더니 100일 중 67일은 비가 온 것으로 기록되어 있다. 이 자료에 근거하여 “비율 확률 50%”라는 일기예보의 정확도에 관해 어떻게 말할 수 있겠는가? 그 이유를 설명하라.

제 4 문 (12점)

다음은 세 기관(A, B, C)에서 랜덤하게 선정한 연수자 9명의 연간 결석 회수를 분석한 결과이다.

ANOVA Table					
Source	DF	Sum of Squares	Mean Square	F-Value	Pr > F
Model	2	150.0	75.0	5.0	0.0527
Error	6	90.0	15.0		
Corrected Total	8	240.0			

- (1) 세 기관에 속한 연수자들의 평균결석회수가 같지 않다는 주장을 입증하고 싶다. 적절한 귀무가설과 대립가설을 서술하라. (4점)
- (2) 유의수준이란 무엇인지 간단히 설명하고, 세 기관에 속한 연수자들의 평균 결석회수가 다르다는 주장이 채택되기에 유리한 유의수준은 1%, 5%, 10% 중 어느 것인가 선택하라. (4점)
- (3) 위에 주어진 표를 참고하여 (2)에서 선택한 유의수준 하에서 검정한 결과를 서술하라. (4점)

통 계 학

2010년 시행 행정고등고시(행정직) 제2차시험

응시번호 :

성명 :

제 1 문. 두 확률변수 X 와 Y 의 결합확률밀도함수(joint probability density function)는 다음과 같다. (총 12점)

$$f(x, y) = \begin{cases} 3x, & 0 < y < x < 1 \\ 0, & \text{그 외} \end{cases}$$

- 1) X 와 Y 의 주변확률밀도함수(marginal probability density function)를 각각 구하시오. (5점)
- 2) $Y = \frac{1}{2}$ 로 주어졌을 때 X 의 조건부확률밀도함수(conditional probability density function)를 구하시오. (4점)
- 3) $P\left(X \leq \frac{3}{4} \mid Y = \frac{1}{2}\right)$ 을 구하시오. (3점)

제 2 문. A도시에서 한 시간 동안 발생하는 교통사고 건수 X 는 평균이 λ 인 포아송분포를 따른다고 한다. (총 15점)

- 1) 교통사고 발생시점을 T_1, T_2, \dots 와 같이 기록했다고 하자. $T_0 = 0$ 을 자료수집 시작 시점이라고 하면, T_i 는 i 번째 사고 발생시점이다. T_1 의 확률분포를 구하시오. (단, 시점기록 단위는 시간이다. 예를 들어, 1시간 30분만에 사건이 발생했으면 1.5로 기록한다) (5점)
- 2) B도시에서 한 시간 동안에 발생하는 교통사고 건수 Y 는 평균이 θ 인 포아송분포를 따르며, X 와 Y 는 서로 독립이라고 하자.
 - (1) $X+Y$ 의 확률분포를 제시하고, 그 이유를 설명하시오. (5점)
 - (2) $X+Y=3$ 일 때 X 의 조건부확률분포를 제시하고, 그 이유를 설명하시오. (5점)

제 3 문. 상자 안에 숫자 1, 2, 3이 각각 적혀있는 세 장의 카드가 있다. 이 상자에서 두 장의 카드를 임의로 추출하되, 추출방법에 따라 다음과 같이 정의된 두 통계량 T 와 S 를 고려하기로 한다. (총 10점)

$$T = \frac{X_1 + X_2}{2}, \text{ 여기서 } X_1 \text{은 첫 번째 추출된 카드의 숫자, } X_2 \text{는 첫 번째}$$

추출된 카드를 복원시킨 후 두 번째로 추출된 카드의 숫자이다.

$$S = \frac{Y_1 + Y_2}{2}, \text{ 여기서 } Y_1 \text{은 첫 번째 추출된 카드의 숫자, } Y_2 \text{는 첫 번째 추}$$

출된 카드를 복원시키지 않고 두 번째로 추출된 카드의 숫자이다.

- 1) Y_1 과 Y_2 의 공분산(covariance)을 구하시오. (5점)
- 2) T 와 S 의 분산을 각각 구하시오. (5점)

제 4 문. 다음 자료는 서울 강남의 특정 지역에서 오후 1시에서 2시 사이에 발생된 교통사고 건수를 50일 간 관측한 자료이다. 이 자료를 이용하여 이 시간대의 교통사고 건수가 포아송분포를 따른다는 가설에 대한 적합도검정(goodness-of-fit test)을 하려고 한다. (총 13점)

5	5	3	2	5	4	3	3	7	6
7	4	3	6	4	4	5	3	5	3
7	3	2	8	6	7	4	1	9	8
6	4	3	11	9	6	7	4	5	4
4	8	9	3	9	7	7	9	3	10

(합계 : 270)

- 1) 위 자료를 6개의 범주 $A_1 = \{0, 1, 2, 3\}$, $A_2 = \{4\}$, $A_3 = \{5\}$, $A_4 = \{6\}$, $A_5 = \{7\}$, $A_6 = \{8 \text{ 이상}\}$ 으로 분류하여 유의수준 5%로 적합도검정을 하는 절차를 상세히 서술하시오. (7점)
- 2) 위 검정에서 귀무가설이 채택되었을 때, 이 지역에서 오후 1시에서 1시 10분 사이에 교통사고가 1건 발생될 확률을 추정하시오. (6점)

행정안전부 시험출제과장

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (11점)

양탄자를 생산하는 공정과정에서 과거의 기록을 볼 때 생산기계의 성능이 정상으로 가동되는 경우가 90% 이다. 생산기계의 성능이 정상일 때는 불량품의 비율이 10% 인데 기계의 성능이 비정상일 때는 불량품의 비율이 30%가 된다. 관리자는 생산되는 양탄자의 불량 여부를 관찰하다가 기계의 성능이 의심스러워지면 가동을 멈추고 기계의 이상 여부를 검사하고자 한다.

어느 날 12개의 생산품을 조사한 결과 불량품(F) 와 양품(S)이 다음과 같은 차례로 얻어졌다. 즉, 첫 번째 생산된 양탄자는 양품(S), 두 번째 생산된 양탄자는 불량품(F), 세 번째 생산된 양탄자는 양품(S),... 등이다.

S F S F S F S S S S F S

- (1) 첫 번째 관측치를 얻은 후 첫 번째 관측치가 주어진 조건 하에서 기계가 정상일 조건부 확률을 구하라. 단, 이하 모든 계산을 소숫점 두자리까지만 계산한다. (3점)
- (2) 두 번째 관측치를 얻은 후 첫 번째와 두 번째 관측치가 주어진 조건 하에서 기계가 정상일 조건부 확률을 구하는데 (1)번의 결과를 이용하라. (4점)
- (3) 매 관측치 이후에 그 시점까지의 자료가 주어진 조건하에서 기계가 정상일 조건부 확률을 계산하여 이 확률이 70% 이하이면 기계가동을 멈추고 기계를 검사하였다면 몇 번째 생산품 이후에 기계를 멈추어야 했는가? (계산과정을 보이시오.) (4점)

제 2 문 (12점)

어느 직장에서 교육 프로그램 A와 B를 비교하기 위하여 직원 100명을 임의로 선택하여 직원 50명을 교육 프로그램 A를 수강하도록 하였고 나머지 50명은 교육 프로그램 B를 수강하도록 하였다. 6개월 뒤에 100명에 대한 근무성적을 측정한 후 교육 프로그램 A를 수강한 직원의 평균 근무성적(μ_1)과 교육 프로그램 B를 수강한 직원의 평균 근무성적(μ_2)의 차 $\mu_1 - \mu_2$ 에 대하여 합동표본분산(sample pooled variance)과 표준정규분포를 이용한 90% 신뢰구간은 (0.21, 6.79)이다.

- (1) $\mu_1 - \mu_2$ 에 대한 95% 신뢰구간을 계산하라. 표는 Z 가 표준정규분포를 따르는 확률변수일 때 z 값에 따른 $P(Z \leq z)$ 의 값을 나타낸 것이다. (4점)

z	$P(Z \leq z)$
-1.282	0.100
-1.645	0.050
-1.960	0.025
-2.326	0.010

- (2) H_0 : 두 교육 프로그램에 대한 근무성적의 차이가 없다 대 H_1 : 두 교육 프로그램에 대한 근무성적의 차이가 있다 에 대한 검정을 하고자 한다. 이를 위한 가정을 쓰고 유의수준 5%에서 실시하라. (4점)
- (3) (2)의 검정에서 귀무가설을 기각할 수 있는 최소의 유의수준의 값을 구하고 (2)의 결과와 비교하여 설명하라. (뒤의 표를 이용하시오.) (4점)

제 3 문 (15점)

다음은 종속변수 Y 와 독립변수 X_1, X_2, X_3 의 관계를 분석한 결과로 상관계수 및 다중 선형회귀모형의 결과이다.

Correlations

	Y	X1	X2	X3
Y	1.00	0.81	0.83	0.36
X1	0.81	1.00	0.96	0.28
X2	0.83	0.96	1.00	0.41
X3	0.36	0.28	0.41	1.00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	VIF
(Intercept)	2.4530	1.4912	1.645	0.109	
X1	0.4760	0.6586	0.723	0.475	13.7
X2	1.1822	0.7355	1.607	0.117	15.2
X3	0.2386	0.4514	0.529	0.600	1.4

Residual standard error: 2.044 on 36 degrees of freedom

Multiple R-Squared: 0.6975, Adjusted R-squared: 0.6723

F-statistic: 27.67 on 3 and 36 DF, p-value: 1.853×10^{-9}

- (1) 다중선형회귀모형 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ 에서 오차항 ε_i 에 대한 가정들을 기술하라. (3점)
- (2) (1)의 모형에서 회귀모형의 유의성을 검정하려 할 때 귀무가설과 대립가설을 기술하고 유의수준 1%에서 검정하라. (3점)
- (3) 독립변수 X_1, X_2, X_3 각각이 종속변수 Y 에 유의한 영향을 미치는 지를 유의수준 5%에서 검정하라. (2)의 결과와 비교하여 (3)의 검정결과에 대한 이유를 설명하라. (4점)
- (4) (3)의 이유에 대한 해결방안을 설명하라. (5점)

제 4 문 (12점)

어느 지역에서 새 정책에 대한 찬성과 반대를 조사하기 위하여 이 지역 주민 1210명을 임의로 뽑아 주민의 성별(남, 녀)와 학력(대졸 이상, 대졸 미만)과 새 정책에 대한 찬반을 조사하였다. 먼저 학력과 새 정책의 찬반에 대한 분할표는 표 1과 같다.

	찬성	반대	합
대졸 이상	240	420	660
대졸 미만	200	350	550
합	440	770	1210

표 1

또한, 남자와 여자 각각에 대한 학력과 새 정책의 찬반에 대한 분할표는 아래와 같다.

남자

	찬성	반대	합
대졸 이상	135	415	550
대졸 미만	5	45	50
합	140	460	600

여자

	찬성	반대	합
대졸 이상	105	5	110
대졸 미만	195	305	500
합	300	310	610

표 2

- (1) 표 1에서 학력이 새 정책의 찬반에 영향을 준다고 할 수 있는지를 서술하시오. (3점)
- (2) 표 2에서 남자에 대한 분할표에서의 독립성 검정을 위한 카이제곱통계량의 값이 5.42이고 이에 상응하는 p-값은 0.0199이다. 또한, 표 2의 여자에 대한 분할표에서 독립성 검정을 위한 카이제곱통계량의 값은 115.0이고 이에 상응하는 p-값은 0.001보다 작다. 조사 대상 중 남자에 대하여 학력이 새 정책의 찬반에 영향을 준다고 할 수 있는가? 또한, 여자에 대해서는 어떻게 설명할 수 있는가? 이에 대하여 서술하시오. (4점)
- (3) (1)과 (2)의 결과를 종합적으로 분석하시오. (5점)

통 계 학

2011년 시행 5급(행정) 공채 제2차시험

응시번호 :

성명 :

제 1 문. 어떤 대규모 입사시험에서 수험자가 주어진 과제를 해결하는데 걸리는 시간은 평균이 5분인 지수분포를 따른다고 한다. 지수분포의 확률밀도함수(probability density function)는 다음과 같다. 주어진 <표>의 지수함수 값을 이용하여 다음 물음에 답하시오. (총 15점)

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x > 0, \lambda > 0.$$

<표> 지수함수 값

x	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
e^{-x}	0.607	0.368	0.223	0.135	0.082	0.050	0.030	0.018	0.011	0.007

- 임의로 선택된 한 수험자가 주어진 과제를 5분 안에 해결할 확률을 구하시오. (3점)
- 수험자들을 A, B 두 개의 그룹으로 나눈 후 A 그룹에서 5명을, B 그룹에서 7명을 무작위로 선택하여 과제를 해결하도록 하였다. 선택된 12명 중에서 과제를 5분 안에 해결한 사람이 5명이라고 할 때, 이 중 2명이 A 그룹의 수험자일 조건부확률을 구하시오. (6점)
- 무작위로 선택된 100명의 수험자 중 과제 완료시간이 15분 이상인 수험자가 2명 이상일 확률을 근사적으로 구하시오. (6점)

제 2 문. 확률변수 X 가 시행횟수가 6, 성공확률이 $p(0 < p < 1)$ 인 이항분포를 따른다고 할 때, 다음과 같은 가설을 검정하고자 한다. 이 가설검정에서 기각역의 형태는 ' $X \leq c$ '이다. (단, c 는 0이상의 정수) (총 10점)

$$H_0: p = \frac{1}{2} \quad \text{대} \quad H_1: p < \frac{1}{2}$$

- X 의 관측값으로 1을 얻었을 때, 이 가설검정의 유의확률(p값)을 구하시오. (4점)
- 제1종 오류를 범할 확률을 α , $p = \frac{1}{3}$ 에서 제2종 오류를 범할 확률을 β 라고 할 때, $(2\alpha + \beta)$ 의 값을 최소로 하는 기각역의 상수 c 값을 구하시오. (6점)

제 3 문. A, B 두 운동화 회사에서 생산한 운동화 바닥의 평균마모정도를 비교하기 위하여 8명을 임의로 뽑은 후, 임의로 선택된 한쪽 발에는 A 회사의 운동화를, 다른 한쪽 발에는 B 회사의 운동화를 신게 하고 일정 시간이 흐른 뒤에 운동화 바닥의 마모정도를 조사하여 다음과 같은 결과를 얻었다. 이 때, 다른 요인이 개입되지 않도록 운동화 디자인은 같게 생산하였고, 운동화 바닥의 마모정도는 정규분포를 따른다고 가정한다. (총 10점)

회사	1	2	3	4	5	6	7	8
A	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8
B	14.0	8.8	11.2	14.2	11.8	6.4	9.8	11.3

- 1) 두 회사에서 생산한 운동화 바닥의 마모정도가 다른지를 검정하기 위한 적절한 가설을 제시하시오. (2점)
- 2) 1)의 가설을 검정하기 위한 검정통계량의 값으로 -2.68 을 얻었을 때, 이를 얻기 위한 계산 절차를 설명하시오. (단, 자세한 계산은 생략 가능) (5점)
- 3) 2)에서 계산된 검정통계량에 대응되는 유의확률(p값)이 0.0316 이라고 할 때, 1)의 가설을 유의수준 5% 에서 검정하시오. (3점)

제 4 문. 다음의 단순선형회귀모형을 고려하기로 한다.

(총 15점)

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \text{는 서로 독립이고 평균 } 0, \text{ 분산 } \sigma^2 \text{인 동일한 분포를 따르며, } i = 1, 2, \dots, n.$$

- 1) 회귀계수 α 와 β 에 대한 최소제곱추정량(least squares estimator)을 구하시오. (5점)
- 2) 1)에서 구한 β 의 최소제곱추정량이 β 의 불편추정량(unbiased estimator)임을 보이시오. (5점)
- 3) \hat{y}_i 를 y_i 의 적합값(fitted value), 잔차(residual) e_i 를 $(y_i - \hat{y}_i)$ 라 할 때, $\sum_{i=1}^n \hat{y}_i e_i = 0$ 이 성립함을 보이시오. (5점)

행정안전부 시험출제과장

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (10점)

지난 수년간 100개 공기업의 평균매출액은 평균 9천억원, 표준편차 3천억원, 그리고 평균비용은 평균 7천억원, 표준편차 4천억원으로 알려져 있다. 평균이익을 평균매출액에서 평균비용을 차감한 것으로 정의할 때, 손실이 발생하지 않을 확률을 구할 수 있으면 그 근거를 밝히고 계산하라. (단, 평균이익과 평균매출액은 서로 독립이라 가정한다.)

제 2 문 (10점)

어떤 자료에 단지 두개의 관찰점 (x_1, y_1) 과 (x_2, y_2) 가 있다고 가정한다. 구할 수 있는 표본상관계수 ((Pearson's) sample correlation coefficient)에 대하여 논의하여라.

제 3 문 (15점)

두개의 확률변수 X, Y 가 아래의 결합확률밀도함수를 가질 때 각 물음에 답하여라.

$$f(x, y) = \begin{cases} 24xy & 0 < x < 1, 0 < y \leq 1 - x \\ 0 & \text{otherwise} \end{cases}$$

- (a) 주변확률밀도함수 $h(x) = (\quad)$ (5점)
- (b) 조건부확률밀도함수 $g(y|x) = (\quad)$ (5점)
- (c) $P(Y < X | X = \frac{1}{3}) = (\quad)$ (5점)

제 4 문 (15점)

새로운 복지정책에 대한 전체 국민들의 지지율을 알아보기 위하여 한 여론조사업체가 전화번호부에 기재된 모든 유선전화번호 중 400개를 무작위로 추출하여 이 복지정책에 대한 지지율을 조사하였다. 전체 응답률이 25%라고 할 때 다음에 답하시오.

- (a) 이 정책에 대한 여론조사 지지율이 59%였다면, 이 결과로부터 유의수준 5%하에서 통계적으로 과반수이상의 국민들이 이 정책에 지지한다고 볼 수 있는지에 대하여 논하시오. (7점)
- (b) 이 정책에 대한 여론조사 지지율이 80%였다면, 전체 국민의 정책에 대한 지지율의 95%신뢰구간을 구하시오. (8점)

NORMAL DISTRIBUTION TABLE

Entries represent the area under the standardized normal distribution from $-\infty$ to z , $Pr(Z \leq z)$

The value of z to the first decimal is given in the left column. The second decimal place is given in the top row

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
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

통 계 학

2012년 시행 5급(행정) 공채 제2차시험

응시번호 :

성명 :

제 1 문. 어느 전구를 생산하는 회사에서 전구의 수명을 조사하였더니 평균 3,800시간 이고 표준편차 150시간이었다. 그 후 평균 수명을 증가시키기 위해 생산공정을 새롭게 바꾸었고, 바뀐 공정을 통해 생산한 전구들 중 25개를 임의로 추출하여 구한 수명의 표본평균이 3,875시간이었다. 이것을 토대로 회사 측에서는 새로운 생산공정을 통해 전구의 평균 수명이 증가했다고 광고하였다. 새로 생산된 전구의 수명은 평균 μ 인 정규분포를 따르고, 공정이 바뀌어도 표준편차는 변하지 않았다고 할 때, 회사측의 광고가 옳은 것인지를 검정하고자 한다. (단, 확률변수 Z 는 $N(0, 1)$ 을 따르며, $z_{0.025} = 1.96$, $z_{0.05} = 1.645$ 이다) (총 12점)

- 1) 가설을 세우고, 유의수준 $\alpha = 0.05$ 에서 기각역을 설정하여 검정하시오. (3점)
- 2) 1)에서 유의확률(p -값)을 $P(Z > a)$ 의 형태로 나타내고, 유의확률에 근거한 검정방법을 설명하시오. (3점)
- 3) 유의수준 $\alpha = 0.05$ 에서, $\mu = 3,900$ 일 때 제2종의 오류를 범할 확률을 $P(Z < a)$ 의 형태로 나타내시오. (3점)
- 4) 유의수준 $\alpha = 0.05$ 에서, $\mu = 3,850$ 일 때 검정력이 0.95가 되기 위해서는 몇 개의 전구를 조사하여야 하는지 계산하시오. (3점)

제 2 문. 다음의 각 물음에 대해 <예시>와 같이 답하시오. 즉, 적용할 수 있는 적절한 검정방법을 <보기>에서 선택하여, 선택의 이유를 간략히 설명하고, 필요한 가정이 있으면 기술하시오. 이 과정에서 관심의 대상이 되는 모수를 정의 하고, 검정하고자 하는 귀무가설과 대립가설을 기술하시오. (총 12점)

<예 시>

(물음) 전국 20개 하천에서 상류가 하류보다 BOD가 낮은지 검정하고자 한다.

[정답예시] ③ 대응비교(쌍체비교): 자료가 상류와 하류로 짝지어져 있으므로 대응비교를 실시한다. 짝을 이룬 관측자료가 20개로 작으므로 모집단의 정규분포 가정이 필요하다.

모수: 상류의 평균 BOD를 μ_1 , 하류의 평균 BOD를 μ_2

가설: $H_0: \mu_1 - \mu_2 = 0$ 대 $H_1: \mu_1 - \mu_2 < 0$

<보 기>

- | | |
|------------------------|---------------|
| ① 회귀분석 | ② 일원배치 분산분석 |
| ③ 대응비교(쌍체비교) | ④ 모비율에 대한 검정 |
| ⑤ 두 모집단의 모비율의 차에 대한 검정 | ⑥ 카이제곱 적합도 검정 |

- 1) 86명의 환자를 대상으로 신약이 기존약보다 1주일 후의 완치율이 높은지 검정하고자 한다. (4점)
- 2) 위의 1)에서 기존약의 1주일 후 완치율이 80%라고 주어져 있는 상태에서 검정하고자 한다. (4점)
- 3) 위의 <예시>의 (물음)에서 상류의 BOD가 높은 하천은 하류도 BOD가 높은 경향이 있는지 검정하고자 한다. (4점)

제 3 문. 총 콜레스테롤(CHOL)과 체질량지수(BMI)가 HDL 콜레스테롤(HDL)에 어떤 연관이 있는 지를 알아보기 위해 다중회귀모형을 적합하여 다음과 같은 결과를 얻었다. (총 13점)

Variable	DF	Parameter Estimate	Standard Error	T value	Prob > T
INTERCEPT	1	-24.990	38.234	-0.654	0.515
BMI	1	2.459	1.651	1.489	0.139
CHOL	1	0.498	0.181	2.753	0.007
BMI*CHOL	1	-0.019	0.008	-2.406	0.018

- 1) 적합된 회귀식을 기술하시오. (2점)
- 2) BMI*CHOL에 대한 회귀계수의 유의성을 검정하고자 한다. 귀무가설과 대립가설을 설정하고, 유의수준 5%에서 검정하시오. (3점)
- 3) 총 콜레스테롤(CHOL)이 129일 때, 체질량지수(BMI)가 HDL 콜레스테롤(HDL)에 미치는 효과에 대하여 설명하시오. (4점)
- 4) BMI*CHOL에 대한 회귀계수의 의미를 설명하시오. (4점)

제 4 문. 다음은 어느 시즌 한국프로농구(KBL) 우승팀이 정규리그에서 치른 총 45게임의 득점 자료를 요약한 것이다. 이 결과를 보고 각 물음에 답하시오. (총 13점)

줄기와 잎 그림

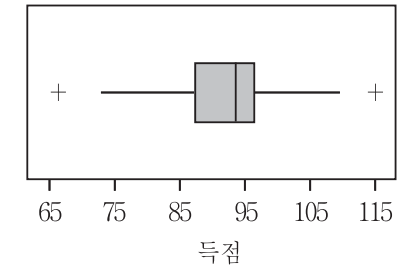
Leaf Unit = 1.0

```

1 | 6 | 6
2 | 7 | 3
2 | 7 |
8 | 8 | 000134
15 | 8 | 5777899
(13) | 9 | 0001123344444
17 | 9 | 55666678
9 | 10 | 00114
4 | 10 | 789
1 | 11 |
1 | 11 | 5

```

상자그림



통계량

n	평균	중앙값	표준편차	표준오차	제1사분위수(Q1)	제3사분위수(Q3)
45	92.29	()	9.29	1.39	()	96.50

- 1) 중앙값(median)을 구하고, 그 의미를 설명하시오. (3점)
- 2) 사분위수 범위(inter-quartile range)를 구하고, 그 의미를 설명하시오. (3점)
- 3) 상자그림으로부터 자료의 치우침(skewness)에 대해 해석하시오. (3점)
- 4) 상자그림에서 +로 표시된 부분의 의미를 밝히고, 그 값이 +로 표시되는 이유를 구체적으로 설명하시오. (4점)

행정안전부 시험출제과장

통 계 학

▶ 응시번호 :

▶ 성 명 :

제 1 문 (10점)

세 회사 A, B, C에 근무하는 근로자의 연령을 조사한 결과 다음의 결과를 얻게 되었다.

	근로자 수	평균	표준편차
회사 A	80	32	7
회사 B	20	44	5
회사 C	60	36	9

위 결과를 이용하여 세 회사를 합한 전체 근로자의 평균과 표준편차를 구하되, 풀이 과정을 자세히 기술하여라.

	근로자 수	평균	표준편차
전체	160		

여기서 표준편차(s)는 편의상 $s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ 으로 정의하기로 한다.

제 2 문 (10점)

동일한 모집단에 대해 두 회사가 독립적으로 여론조사를 한다. 회사 A는 표본의 크기를 1,000명으로 하고 회사 B는 2,000명으로 표본설계를 한다고 하자. 모비율을 추정할 때 두 회사 모두 신뢰수준은 95%로 한다.

가. 각 회사의 최대 오차의 한계는 각각 얼마인가?

나. 두 회사의 여론조사 결과 두 회사간 표본비율의 차이가 0.1 보다 클 최대 확률을 구하는 방법을 서술하여라. 단, 정규근사를 활용한다고 하자.

제 3 문 (15점)

다음은 독립변수 x 와 반응변수 y 에 대한 관측값이다. 이 자료에 대해 단순선형회귀모형을 적합하고자 한다.

x	1	2	3	5	6	8
y	0.1	0.15	0.2	0.3	1.5	3.0

$$\sum x = 25, \sum y = 5.25, \sum xy = 35.5, \sum x^2 = 139$$

- 가. 단순선형회귀모형과 오차항에 대한 가정을 기술하고, 회귀계수를 추정하여라.
나. 피어슨(Pearson) 상관계수와 스피어만(Spearman) 상관계수를 각각 구하여라.
다. 잔차분석을 실시하고, 이를 통해 위 모형에 대한 개선방안을 제시하여라.

제 4 문 (15점)

균일분포 $U(\mu-1, \mu+1)$ 으로부터 추출된 크기가 1인 확률표본을 X 라고 하자.

- 가. 한 개의 표본 X 를 이용하여 다음의 가설에 대한 검정을 실시하고자 한다.

$$H_0 : \mu \leq 10, H_1 : \mu > 10$$

유의수준 5% 에서의 기각역을 구하여라.

- 나. 위 가.에서 구한 기각역을 사용하여 검정을 실시할 때, 검정력 함수를 구하고 이를 그림으로 그려라.

- 다. 위 가.에서 구한 기각역을 사용하여 아래 가설에 대한 검정을 실시하고자 한다.

$$H_0 : \mu = 10, H_1 : \mu = 11$$

제 1종 오류와 제 2종 오류 확률을 각각 구하여라.

통 계 학

2013년 시행 5급(행정) 공채 제2차시험

응시번호 :

성명 :

제 1 문. A시는 기차역 승차권 매표 대기시간에 대한 승객들의 불만을 감소시키기 위하여 번호표 제도를 도입하고자 한다. 이 제도를 시범 실시한 후 25명의 승객을 임의로 뽑아서 대기시간을 조사하였더니 표준편차가 2분이었다. 현행 대기시간이 표준편차가 4분인 정규분포를 따른다고 할 때, 번호표 제도의 도입이 대기시간의 분산을 줄인다고 볼 수 있는지를 검정하고자 한다. 다음 물음에 답하시오. (단, $\chi^2_{0.975}(24) = 12.40$, $\chi^2_{0.95}(24) = 13.85$, $\chi^2_{0.05}(24) = 36.42$, $\chi^2_{0.025}(24) = 39.36$, $\chi^2_{\alpha}(df)$ 는 분포 $\chi^2(df)$ 에서의 $100(1-\alpha)$ 분위수이다)

(총 15점)

- 1) 모분산(σ^2)의 식으로 귀무가설과 대립가설을 세우고, 이 가설을 검정하기 위한 검정통계량과 그 분포를 기술하시오. (6점)
- 2) 1)에서 제안한 검정통계량의 값을 구하여 유의수준 5%에서 가설검정을 하고, 그 결과를 구체적으로 기술하시오. (6점)
- 3) 이 검정문제에서의 제1종 오류의 의미를 구체적으로 기술하시오. (3점)

제 2 문. 어느 자동차 보험회사에서는 보험가입시 자동차 사고의 위험도에 따라 운전자를 위험도 낮음(L), 보통(M), 높음(H)의 세 등급으로 분류한다. 이 자동차 보험회사에 가입한 운전자의 30%가 L등급, 50%가 M등급, 20%가 H등급으로 분류되어 있다. 임의의 한 운전자에 대해서 1년당 사고의 수는 L등급은 평균이 0.01, M등급은 평균이 0.03, H등급은 평균이 0.08인 포아송(Poisson) 분포를 따르며, 운전자들이 자동차 사고에 관련될 가능성은 각각 독립이라고 가정한다. (단, 평균이 m 인 포아송분포의 확률질량함수는 $p(x) = \frac{e^{-m}m^x}{x!}$, $x=0, 1, 2, \dots$ 이고, 지수함수 값은 다음과 같다)

<표> 지수함수 값

x	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
e^{-x}	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.92	0.91	0.90

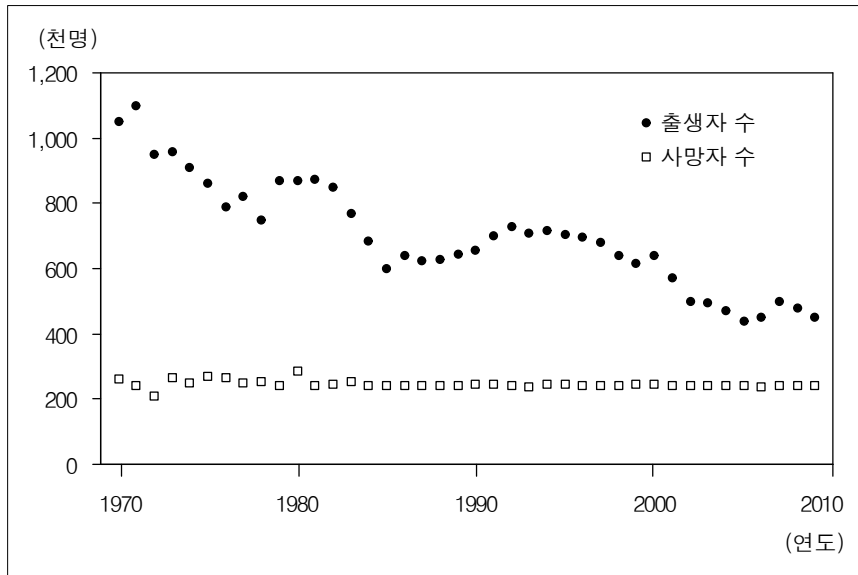
다음 물음에 답하시오.

(총 12점)

- 1) 이 보험회사의 가입자들 중 임의로 뽑은 한 명의 운전자가 1년 동안 어떤 자동차 사고에도 연루되지 않았을 확률을 계산하시오. (4점)
- 2) 이 보험회사의 가입자들 중 임의로 뽑은 두 명의 운전자가 모두 1년 동안 어떤 사고에도 연루되지 않았을 때, 한 사람은 M등급에, 또 다른 한 사람은 L등급에 속할 확률을 계산하시오. (4점)
- 3) 이 보험회사의 가입자들 중 n 명의 표본을 임의로 뽑을 때, 그 표본에 포함된 운전자 중 적어도 한 명이 H등급에 포함될 확률이 최소한 0.90이 되기 위한 가장 작은 n 의 값을 구하시오. (단, $\log_{10}2 = 0.301$ 이다) (4점)



제 3 문. 아래의 그림은 1970년부터 2009년까지 40년 간의 우리나라 출생자 수(천명)와 사망자 수(천명)를 연도별로 표시한 것이다.



연도(t)를 설명(독립)변수라고 하고, 출생자 수와 사망자 수를 반응(종속)변수로 설정하였다. 각각의 반응변수에 대하여 단순선형회귀모형을 적합한 결과가 아래와 같다.

출생자 수 분석결과

- 표본평균 = 699.4, 표본표준편차 = 163.9
- 회귀분석

	Estimate	Std. Error	T-value	Pr(> T)
(Intercept)	26534.7	1706.1989	15.55	< 0.0001
t	-13.0	0.8576	-15.14	< 0.0001

사망자 수 분석결과

- 표본평균 = 245.1, 표본표준편차 = 11.3
- 회귀분석

	Estimate	Std. Error	T-value	Pr(> T)
(Intercept)	658.1	304.8287	2.159	0.0372
t	-0.2	0.1532	-1.355	0.1834

다음 물음에 답하시오.

(총 12점)

- 1) 회귀계수에 대한 검정을 이용하여 연도(t)와 출생자 수의 관계를 기술하고 최종 관계식을 유도하시오. (단, 유의수준은 5%로 한다) (4점)
- 2) 회귀계수에 대한 검정을 이용하여 연도(t)와 사망자 수의 관계를 기술하고 최종 관계식을 유도하시오. (단, 유의수준은 5%로 한다) (4점)
- 3) 인구동태에 영향을 미치는 다른 변인(예. 이민 등)이 없이 순수하게 출생자 수와 사망자 수만 고려할 때, 1)과 2)의 결과를 이용하여 총 인구수가 감소하는 시점을 예측하시오. (4점)

제 4 문. 어느 지방 세무서에서는 관할 지역에 있는 편의점들의 하루 평균 매출액에 대하여 알아보고자 한다. 표준정규분포를 따르는 확률변수 Z 에 대한 $P(Z > 1.96) = 0.025$, $P(Z > 1.645) = 0.05$ 를 이용하여, 다음 물음에 답하시오.

(총 11점)

- 1) 하루 평균 매출액의 95 % 신뢰구간의 길이가 30만원을 넘지 않기 위해서 몇 개의 편의점을 대상으로 조사를 해야 하는지 계산하시오. (단, 과거의 조사로부터 표준편차는 60만원이라고 가정한다) (7점)
- 2) 1)에서 구한 수를 n 이라 할 때, 임의로 뽑은 n 개의 편의점을 대상으로 매출액을 조사한 후 하루 평균 매출액의 95 % 신뢰구간을 구하였더니 (125만원, 155만원)이었다. 이 때 ‘관할 지역에 있는 편의점들의 하루 평균 매출액이 (125만원, 155만원)에 포함될 확률이 95 %다.’라는 주장이 옳은지 그른지 판단하고 그 이유를 기술하시오. (4점)

안전행정부 시험출제과장

