국가기술자격검정 필기시험문제
2008년 기사 제1회 필기시험(문제지 A형)
사회조사분석사 2급

*** 제2과목 :사회통계**

61.다음 표의, x y에 대한 상관계수 설명으로 맞는 것은?

x	1	2	3	4	5
y	5	4	3	2	1

가. 완전 앙의 상관관계 나. 상관이 없다
다. 강한 앙의 상관관계 라. 완전 음의 상관관계

62.상관관계(correlation)에 대한 설명 중 옳은 것은?
가. 두 변수 간에 강한 상관관계가 존재하면 두 변수는 서로 독립적이라고 한다
나. 두 변수 간의 상관관계로부터 인과관계를 도출할 수 있다.
다. 두 변수 간에 상관관계가 없다면 피어슨 상관계수의 값은 0이다.
라. 피어슨 상관계수의 값은 항상 0 이상 1 이하이다.

63.종속변수 Y를 독립변수 x, 과 X2로 설명하는 일반적인
선형회귀모형은?
가. $Y_{ij}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\epsilon_{ij}$
나. $Y=X_1+X_2+\varepsilon$
다. $Y=\alpha+\beta X_1X_2+\epsilon$
라. $Y=\beta_0+\beta_1X_1+\beta_2X_2+\epsilon$

64. 만일 자료에서 모평균 μ에 대한 95% 신뢰구간이 (-0.042, 0.522)로 나왔다면, 유의수준 0.05에서 귀무가설 H_0 : μ=0 대 대립가설 H_1 : μ≠0의 검증 결과는 어떻게 해석할 수 있는가?
가. 신뢰구간과 가설검증은 무관하기 때문에 신뢰구간을 기초로 검증에 대한 어떠한 결론도 내릴 수 없다.
나. 신뢰구간이 0을 포함하기 때문에 귀무가설을 기각할 수 없다.
다. 신뢰구간의 상한이 0.522로 O보다 상당히 크기 때문에 귀무가설을 기각해야 한다.
라. 신뢰구간을 계산할 때 표준정규분포의 임계값을 사용 했는지 또는 t 분포의 임계값을 사용했는지에 따라 해석이 다르다.

65.회귀분석에서 관찰값과 예측값의 차이는?
가. 오차(error) 나. 편차(deviation)
다. 잔차(residual) 라. 거리(distance)

66.자료 A에 대한 단순회귀분석의 결과 $\widehat{y_A}=\widehat{\alpha_A}+\widehat{\beta_A}x$, 자료 B에 대한 단순회귀분석의 결과로 $\widehat{y_B}=\widehat{\alpha_B}+\widehat{\beta_B}x$ 를 얻었다. 그런데 결과적으로 $\widehat{\alpha_A}=\widehat{\alpha_B}$, $\widehat{\beta_A}=\widehat{\beta_B}$ 였다고 하자. 그리고 오차분산의 추정값도 같으나 독립변수 X 의 분산이 자료 A보다 자료 B에서 크다고 할 때 다름 중 맞는 것은?
가. $\mathrm{Var}\widehat{\beta_A}$ 의 추정값 = $\mathrm{Var}\widehat{\beta_B}$ 의 추정값
나. $\mathrm{Var}\widehat{\beta_A}$ 의 추정값 < $\mathrm{Var}\widehat{\beta_B}$ 의 추정값
다. $\mathrm{Var}\widehat{\beta_A}$ 의 추정값 > $\mathrm{Var}\widehat{\beta_B}$ 의 추정값
라. 경우에 따라 다르다

67.A고등학교 1학년 학생 1000명의 성적분포가 평균 80점, 표준편차 20점인 정규분포로 나타났다. 이 경우에 60점 이상 100점 이하의 점수를 얻은 학생은 약 몇 명인가?
[단, $P(Z\leq0.5)=0.68$, $P(Z\leq1.0)=0.84$,
$P(Z\leq1.5)=0.93$, $P(Z\leq2.0)=0.98$]
가. 350 나. 680 다. 790 라. 850

68.“성과 정당지지도 사이에 관계가 있는가?"를 살펴보기 위하여 설문조사를 실시 분석한 결과 Pearson 카이제곱 값이 32.29, 자유도가 2, 유의확률이 0.0000이었다. 이 분석에 근거할 때, 유의수준 0.05에서 ”성과 정당 지지도 사이의 관계”에 대한 결론은?
가. 위에 제시한 통계량으로는 성과 정당지지도 사이의 관계를 알 수 없다.
나. 성과 정당지지도 사이에 유의미한 관계가 있다.
다. 성과 정당지지도 사이에 유의미한 관계가 없다
라. 남성이 여성보다 정당지지도가 높다.

69. “학력별로 월평균소득에 차이가 있는가?"를 파악하기 위해 분산분석을 실시하였다. 아래와 같은 통계량을 구하였을 때, 유의수준 0.05에서 학력과 월평균소득의 관계에 대한 가설검정의 결과는?

제곱합	자유도	F	유의확률
집단간 1,836.25	2	6.078	0.003
집단내 4,184.66	277		
6,020.91	279		

가. 학력별로 월평균소득에 차이가 있다.
나.. 학력별로 월평균소득에 차이가 없다.
다. 위의 결과로는 가설검정을 할 수 없다.
라. 남성 대졸학력자가 여성 대졸학력자 보다 월평균소득이 높다.

70.확률변수 X가 정규분포 N(μ, σ^2) 을 따를 때 Z=$\frac{X-\mu}{\sigma}$는 어떤 분포를 따르는가?

가. N(0, 1)　　나. N(1, 1)

다. N(μ, σ^2)　　라. N(μ, 1)

71.자료에 특이값(이상값, outlier)이 일부 섞여 있는 경우에도 비교적 안정적인 산포의 측도는?

가. 표준편차　　나. 분산

다. 범위　　라. 사분위수 범위

72.어떤 사회정책에 대한 찬성률 θ를 추정하고자 크기 n인 임의표본(확률표본)을 추출하여 자료를 X_1 , ... , X_n 으로 입력하였을 때 θ에 대한 점추정치로 합당한 것은? (단, 찬성이면 0, 반대면 1로 코딩)

가. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i$　　나. $\frac{1}{n}\sum_{i=1}^{n}x_i$

다. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-x_i)$　　라. $\frac{1}{n}\sum_{i=1}^{n}(1-x_i)$

73.독립시행의 수가 n, 성공비율이 θ인 이항 모집단에서 성공 수 x 를 관측하였다. 영 가설과 대립가설을 각각 $H_0:\theta=0.25.25$, $H_1:\theta=0.75$ 라고 할 때, p-값(유의확률)은?

가. $\binom{n}{x}0.25^x0.75^{n-x}$　　나. $\binom{n}{x}0.75^x0.25^{n-x}$

다. $\sum_{i=x}^{n}\binom{n}{i}0.25^i0.75^{n-i}$　　라. $\sum_{i=x}^{n}\binom{n}{i}0.75^i0.25^{n-i}$

74.다음의 분산분석표에서 유의확률(P값) 0.0212 값이 의미하는 것은?

변동 요인	자유도	제곱합	평균 제곱	F값	P값
모형 잔차	1 10	1519.98 759.02	1519.98 75.90	(　)	0.0212
총합	11	2279.00			

가. 5% 수준에서 유의하다

나. 1% 수준에서 유의하다.

다 10% 수준에서 유의하다.

라. 통계적으로 유의하지 않다.

75.귀무가설이 옳음에도 불구하고 대립가설로 잘못 결론을 내리는 오류는?

가. 제1종 오류　　나. 제2종 오류

다. 알파오류　　라. 베타오류

76.가설검증에서 오류를 범할 확률의 최대허용치는?

가. 유의수준　　나. 유의확률(P값)

다. 오차한계　　라. 신뢰수준

77.짝을 이룬 표본(paired sample)의 t검증과 유사한 비모수 통계기법은?

가. Willcoxon 검증

나. Mann-Whitney 검증

다. Kruskal-Walls 검증

라. Kolmogorov-Smirnov 검증

78.직업별로 소비자행동에 어떤 차이가 있는지를 보기 위해 취업주부를 대상으로 전문직, 사무직, 생산직으로 나누어 소비성향을 측정하였다. 이 때 소비행동은 여러 개의 문항을 이용하여 연속변수의 척도를 구성하였다. 직업별로 소비행동의 차이가 있는지를 알아보려면 어떤 통계적 분석을 실시하는 것이 가장 적합한가?

가. 분할표 분석　　나. 회귀분석

다. 상관관계분석　　라. 분산분석

79.다음과 같이 4가지의 회귀분석의 모델이 있다고 할 때, 이에 관한 설명으로 맞는 것은?

모델 1 : 소득 = 성, 연령, 교육 (R^2=0.6)
모델 2 : 소득 = 성, 성장지역, 교육, 직업 (R^2=0.73)
모델 3 : 소득 = 성, 연령, 직업, 근속연수 (R^2=0.75)
모델4 : 소득 = 성, 성장지역, 직업 (R^2=0.5)

가. 모델 3의 R^2 가 가장 높으므로 네 모델 중 가장 좋은 모델이다.

나. 네 모델 중 서로 비교할 수 있는 모델은 모델 2와 모델 4뿐이다.

다. 모델 3은 모델 4보다 R^2 가 높으므로 더 좋은 모델이라고 할 수 있다.

라. 모델 2를 모델 3과 비교하여 어느 모델이 더 좋은지 판별할 수 있다.

80.학생들의 영어교육방법에 따라 영어실력에 차이가 있는지를 비교하기 위해 영어성적이 유사한 학생들을 랜덤하게 3그룹으로 나누어 3가지 교육방법을 실시하였다. 영어교육방법에 차이가 있는가를 비교하기 위한 다음의 설명 중 틀린 것은?

가. 분산분석의 모형은 $Y_{ij}=\mu+\alpha_i+\beta_j+\epsilon_{ij}$, i=1, 2, 3), j=1, 2, ... , n 이다.

나. 오차항 ϵ_{ij}는 서로 독립이고 N(0,σ^2)를 따른다

다. $\alpha_i=\mu_i-\mu$라고 할 때 $\sum\alpha_i=0$ 이다

라. Y_{ij}는 서로 독립이고, N($\mu+\alpha_i,\sigma^2$)를 따른다

81.변량 $x_1, x_2, \ldots, x_n$ 에 대하여
$|x_i - a| + |x_2 - a| + \ldots |x_n - a|$ 을 최소로 하는 대표값은?
가. 산술평균　　나. 중위수(중앙값)
다. 최빈수　　라. 기하평균

82. 다음에 알맞은 검증방법은?

> 도시지역과 시골지역의 가족 수의 평균에 차이가 있는지 알아보기 위해 도시지역과 시골지역 중 각각 몇 개의 지역을 골라 가족 수를 조사하였다.

가. 독립표본 t-검증　　나. 더빈 왓슨검증
다. χ^2-검증　　라. F-검증

83. 다음의 정규분포에 대한 설명 중 옳은 것은 모두 몇 개인가?

> -정규분포 곡선의 모양은 평균과 분산에 의하여만 결정 된다
> -정규분포곡선은 평균에 대하여 좌우대칭이다
> -서로 독립인 두 확률번수 X와 Y가 정규분포를 따르면 (X + Y) 정규분포를 따른다
> -정규분포 곡선은 분산에 대하여 좌우 대칭한다

가.1개　나.2개　다.3개　라.4개

84.확률번수 X는 평균이 2이고 표준편차가 2인 분포 Y = -2X + 10 의 평균과 표준편차는?
가. 6, 4　나. 6, 6　다. 14, 4　라. 14, 6

85.변수 X와 Y의 상관계수가 0.1일 때, 2X와 3Y의 상관계수는?
가. 0.6　나. 0.3　다. 0.2　라. 0.1

86.두 개의 사상 A와 B에 대한 확률의 법칙 중 일반적으로 성립하지 않는 것은?
가. $P(A) = P(A \cap B) + p(A \cap B^c)$
나. $P(A \cup B) = PiA) + P(B) - P(A) \cdot P(B|A)$
다. $P(A \cup A^c) = 1$
라. $P(A \cap B) = P(A) \cdot P(B)$

87.어느 고등학교 1학년 학생 전체 1000명의 IQ를 조사했더니 평균이 105이고 표준편차가 15로 나타났다. 그 학교 1학년 학생 중 IQ가 75에서 135 사이에 있는 학생은 최소한 몇 명인가?
가. 750명　나. 850명　다. 900명　라. 950명

88.모든 조건이 동일한 경우 표본의 수를 네 배로 늘릴 때 표본평균의 신뢰구간은?
가. 4배로 넓어진다　나. 2배로 넓어진다.
다. 1/2로 준다　라. 1/4로 준다

89.인구수가 생산량에 미치는 영향을 분석하기 위하여 인구수(단위, 100명)를 독립변수, 생산량(단위, 1만원)을 종속변수로 설정하여 회귀분석을 실시하였다 분석결과 다음과 같은 회귀방정식을 구하였을 때, 회귀방정식에 대한 설명으로 맞는 것은?

회귀방정식:$Y = 273.50 + 1.47X$,　$R^2 = 0.25$

가. 인구 수가 0이면, 평균생산량도 0이다
나. 회귀선은 곡선형이다
다. 인구 수가 100명 증가할 경우에, 평균생산량은 약 1만4천7백원 증가한다
라. 회귀방정식에 의하여 인구 수가 생산량을 설명하는 정도는 75%이다

90.다음 자료들을 확률변수로 나타냈을 때 이산확률변수가 되는 것은?
가. 어느 중학교 학생들의 몸무게
나. 습도 80%의 대기 중에서 빛의 속도
다. 장마기간 동안 A시 OO구의 강우량
라. 어느 프로야구 선수가 한 시즌 동안 친 홈런의 수

91.이상점 자료에 대한 설명으로 틀린 것은?
가. 상자그림 요약에서 안쪽 울타리를 벗어나는 자료는 이상점 자료이다.
나. 이상점 자료는 반드시 제외하고 분석하는 것이 바람직하다.
다. 절사평균은 산술평균에 비해 이상점 자료에 덜 민감하다.
라. 중위수는 산술평균에 비해 이상점 자료에 덜 민감하다

92.단순회귀분석에서 회귀선이 $\hat{y}$=0.5-2x 와 같이 주어졌을 때 틀린 것은?
가. 반응변수는 y이고 설명변수는 x이다.
나. 설명변수가 한 단위 증가할 때 반응변수는 2단위 감소한다.
다. 반응변수와 설명변수의 상관계수는 0.5이다
라. 설명변수가 0일 때 반응변수의 예측값은 0.5이다.

93.다음 중 상관계수의 의미로 맞는 것은?
가. 두 변수간에 차이가 있는가를 나타내는 측도이다
나. 두 변수간의 분산의 차이가 있는가를 나타내는 측도이다.
다. 두 변수간의 곡선관계를 나타내는 측도이다.
라. 두 변수간의 선형관계를 나타내는 측도이다.

94.단순선형회귀모형 $y=\beta_0+\beta_1 x+\epsilon$에서 오차항 ϵ의 분포가 평균이 0이고 분산이 σ^2인 정규분포를 따른다고 가정하자. 22개의 자료들로부터 회귀식을 추정하고 나서 잔차제곱 합(SSE)을 구하였더니 그 값이 4000이었다. 이 때 분산 σ^2의 불편추정값은?
가. 100 나. 150 다. 200 라. 250

95.곤충학자가 70마리의 모기에게 A 회사의 살충제를 뿌리고 생존시간을 관찰하여 $\bar{x}$=18.3, s=5.2를 얻었다. 생존시간의 모평균에 대한 99% 신뢰구간은?
가. 8.6 < μ < 28.0 나. 17.1 < μ < 19.5
다. 18.1 < μ < 18.5 라. 16.7 < μ < 19.9

96.A 도시에서는 실업률이 5.5%라고 발표하였다. 그러나 관련 민간단체에서는 실업률 5.5%는 너무 낮게 추정된 값이라고 믿고 이에 대해 확인하고자 한다. 노동력인구 중 520 명을 임의 추출하여 39명이 직업이 없음을 알게 되었다. 이 문제에 대한 적합한 검정통계량 값은?
가. -2.58 나. 1.96 다. 2.00 라. 1.75

97.확률변수 X의 기대값이 5이고, 확률변수 Y의 기대값이 10일 때, 확률변수 X+2Y의 기대값은?
가. 10 나. 15 다. 20 라. 25

98.사건 A가 일어날 확률이 0.5, 사건 B가 일어날 확률이 0.6, A 또는 B가 일어날 확률이 0.8일 때, 사건 A와 B가 동시에 일어나는 확률은?
가. 0.3 나. 0.4 다. 0.5 라. 0.6

99.모표준편차가 σ=10으로 알려진 정규모집단에서 n=25개의 표본을 랜덤하게 추출한 결과 표본평균 $\bar{x}$=40이었다. 모평균 μ의 추정값 $\hat{\mu}=\hat{x}$=40의 95% 오차한계는?
[단, Z~N(0,1)일 때 P(Z>1.96)=0.025, P(Z>1.645)=0.05]
가. 3.29 나. 0.658 다. 3.92 라. 0.784

100.가설검정을 할 때 대립가설(H_1)이 사실인 상황에서 귀무가설(H_0)을 기각할 확률은?
가. 검정력 나. 제2종의 오류
다. 유의수준 라. 신뢰수준

라 다 라 나 다 / 다 나 나 가 가
라 라 다 가 가 / 가 가 라 나 가
나 가 다 가 라 / 라 가 다 다 라
나 다 라 다 라 / 다 라 가 다 가