**11차 [통계 기반 데이터 분석] 실습형 평가지**

|  |  |  |  |
| --- | --- | --- | --- |
| **훈련 과정 명** | 데이터베이스기반 빅데이터 분석가 양성과정 | **훈 련 기 간** | 2019-11-26 ~ 2020-04-21 |
| **NCS 직종** | 빅데이터 분석 | **평 가 일 자** | 2020년 02월 21일 |
| **교 과 목** | 빅데이터 환경 분석 및 운영 | **문 항 수** | **06 문항** |
| **능력단위 수준** | 6 수준 | **평 가 시 간** | **35 분** |
| **강 사 명** | 김 동 희 | **평가 총 배점** | **100 점** |
| **훈 련 생 명** |  | **평 가 총 점** | 점 |
| **훈련생 확인** | **▣ 본 실습형 평가지는 개별 온라인으로 제출하며, 출력된 평가지에 대한 확인을 생략합니다.**  **▣ 본 실습형 평가 결과는 평가자 체크리스트의 평가 기준이 됩니다.** | | |

|  |  |
| --- | --- |
| **훈련생**  **유의사항** | ○ 평가자 및 담당자의 지시에 따라 평가에 임하며, 평가 유의사항을 준수하여야 합니다.  ○ 실습형 평가지 파일을 다운로드 받은 즉시 평가지의 과목, 인쇄 상태, 문항수(첫 장에 기재된  문항수와 실제 문항수 확인) 등을 확인합니다.  ○ 평가지에 기재되어있는 평가 시간을 엄수하여 평가를 진행합니다.  ○ 실습형 평가가 진행되는 동안 휴대용 전화기, 훈련 교재 등을 활용할 수 없으며, 적발 시  부정행위로 간주하여 0점 처리합니다.  ○ 평가와 관련된 부정행위를 하는 경우 채점 대상에서 제외합니다.  ○ 지정된 좌석에 앉아서 평가에 임합니다.  ○ PC 모니터는 책상의 정면 중앙에 놓고 평가에 임합니다.  ○ 평가 진행 도중 해당 평가 관련 문의사항이 있을 경우, 먼저 평가자 및 담당자에게 조용히  손을 든 후 담당자의 안내에 따라 문의하도록 합니다.  ○ 평가 도중 평가 시간이 종료될 경우에는 평가 진행을 멈추고, 평가를 종료하도록 합니다.  ○ 평가 시간이 종료되어도 평가를 완성하지 못한 경우에는 평가를 진행한 부분까지 채점을  진행합니다.  ○ 본인의 평가가 종료되어 서술형 평가지와 실습형 평가지를 모두 제출한 후에는 되도록 퇴실  하지 않고, 타인에게 방해가 되는 행동을 하지 않습니다.  ○ 매 평가 종료 후 정답을 확인하고 이상이 있거나 의문사항이 있을 경우 반드시 당일 중으로  평가자 및 담당자에게 문의합니다. |
| **실습형 평가**  **답안 작성**  **유의사항** | ○ 훈련기관의 PC 장비에 설치되어 있는 오피스 프로그램을 사용하여 실습형 평가를 진행합니다.  (기타 개인 장비를 활용하여 평가 진행에 따른 파일 오류 등 정답을 확인하기 어려울 경우  감점 및 0점 처리합니다.)  ○ PC를 켜고 실습형 평가를 시작한 후 엎어놓은 서술형 평가를 다시 진행할 경우 부정행위로  간주합니다.  ○ 훈련생은 평가지 첫 장 및 평가지 파일명 제목의 “홍길동”을 본인의 성명으로 기재하여야  합니다.  ○ 훈련생은 실습형 평가 진행에 사용되는 프로그램(각 평가 별 평가자 안내) 이외의 어떠한  프로그램도 열지 않습니다.  ○ 평가 시간 동안은 채팅방 등을 사용하지 않으며, 사용시 부정행위로 간주합니다.  ○ 답안은 평가지에 작성하고, 문항에서 요구한 가지 수(항 수) 이상을 정답란에 표기한 경우 및  오답이 함께 기재되어 있을 경우 부분 점수 요건에 맞게 채점됩니다.  ○ 정답란에 문항과 관련 없는 불필요한 낙서나 특이한 기록사항 등이 부정의 목적이 있다고  판단될 경우에는 모든 문항이 0점으로 처리됩니다.  ○ 실습형 평가 종료 후 담당자의 메일로 완료된 평가지를 회신합니다. |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **관련**  **능력단위** | 통계기반  데이터분석 | **관련**  **능력단위**  **요소** | 가설  설정하기 | **평가방법** | 실습형 | **난이도** | ★★★☆☆ | **배점** | **10** | **평가**  **시간** | **5분** |
| **문 항** | **1. 금년에 벼품종 A를 조사한 30개 지역의 평균수량 500kg/10a이고, 작년에 조사한 490kg/10a 와  차이가 있는지 알려고 한다. 이때, 양측검정을 위한 통계적 가설을 세우시오.** | | | | | | | | | | |
| **평가 시**  **유의사항** | [출제 의도]  통계적 가설로, 귀무가설(H0)와 대립가설(H1)을 실험목적과 의미에 맞게 설정할 수 있는 역량 판단. | | | | | | | | | | |
| **정 답** | |  |  | | --- | --- | | **득 점** |  |   **① 귀무가설(H0): 금년과 작년의 수량은 차이가 없다**  **② 대립가설(H1): 금년과 작년의 수량은 차이가 있다.** | | | | | | | | | | |
| **부분 점수 정답&배점** | ※ 각 항목별 50% 부분점수 배점. 주요 키워드 기재여부에 따라 탄력적 배점. | | | | | | | | | | |
| **관련**  **능력단위** | 통계기반  데이터분석 | **관련**  **능력단위**  **요소** | 가설  설정하기 | **평가방법** | 실습형 | **난이도** | ★★★☆☆ | **배점** | **10** | **평가**  **시간** | **5분** |
| **문 항** | **2. 기존 연구결과를 통해 알려진 우리나라 성인의 평균키는 170cm 이다. 금년에 새로이 우리나라 성인인의 키를 표본으로 수집하여 나온 평균키는 175cm 이다. 추정된 모평균과 표본에서 산출된 평균키를 통해**  **우리나라 성인의 키가 이전에 비해 더 커졌음을 검정하고자 한다. 이를 위한 통계적 가설을 세우시오.** | | | | | | | | | | |
| **평가 시**  **유의사항** | [출제 의도]  통계적 가설로, 단측검정(우측검정)에 대한 귀무가설(H0)과 대립가설(H1) 설정 역령 판단. | | | | | | | | | | |
| **정 답** | |  |  | | --- | --- | | **득 점** |  |   **① 귀무가설(H0): 기존 알려진 성인의 평균키 170cm와 차이가 없다.**  **② 대립가설(H1): 기존 알려진 성인의 평균키 170cm 보다 크다.** | | | | | | | | | | |
| **부분 점수 정답&배점** | ※ 각 항목별 50% 부분점수 배점. 주요 키워드 기재여부에 따라 탄력적 배점. | | | | | | | | | | |
|  |  | | | | | | | | | | |
| **관련**  **능력단위** | 통계기반  데이터분석 | **관련**  **능력단위**  **요소** | 통계 모델 개발하기 | **평가방법** | 실습형 | **난이도** | ★★★★☆ | **배점** | **10** | **평가**  **시간** | **10분** |
| **문 항** | **3. 연비데이터 셋인 mpg{ggplot2} 을 이용하여, compact 와 suv 두 종류의 모든 차종에 대해, 고속도로**  **연비에 대해, 평균연비를 산출 및 출력하고, 만일 이 두 종류의 차종의 평균연비에 차이가 발생한**  **경우, 두 차종의 평균연비에 대한 차이가 통계적으로 유의미한지 검정하고자 한다.**  **이를 위한 검정을 수행하고 검정결과를 밝히시오. (단, 유의수준은 0.05로 한다)** | | | | | | | | | | |
| **평가 시**  **유의사항** | [출제 의도]  통계모델을 개발하기 위한, 기본통계량 산출역량과 두집단의 평균의 차이에 대한 검정능력 판단. | | | | | | | | | | |
| **정 답** | |  |  | | --- | --- | | **득 점** |  |   **① 두 종류의 차종(compact, suv)에 대한 평균연비 산출코드**  **mpg <- as.data.frame(ggplot2::mpg)**  **library(dplyr)**  **mpg\_diff <- mpg %>%**  **select(class, cty) %>%**  **filter(class %in% c("compact", "suv"))**  **head(mpg\_diff)**  **## class cty**  **## 1 compact 18**  **## 2 compact 21**  **## 3 compact 20**  **## 4 compact 21**  **## 5 compact 16**  **## 6 compact 18**  **table(mpg\_diff$class)**  **##**  **## compact suv**  **## 47 62**  **② 두 차종의 평균연비 차이에 대한 통계적 검증 수행(두 집단의 평균의 차이에 대한 검증):**  **t.test(data = mpg\_diff, cty ~ class, var.equal = T)**  **##**  **## Two Sample t-test**  **##**  **## data: cty by class**  **## t = 11.917, df = 107, p-value < 2.2e-16**  **## alternative hypothesis: true difference in means is not equal to 0**  **## 95 percent confidence interval:**  **## 5.525180 7.730139**  **## sample estimates:**  **## mean in group compact mean in group suv**  **## 20.12766 13.50000**  **③ 검정결과에 대한 해설:**  **유의확률(p-value)의 값이 유의수준 0.05에 비해 매우 작으므로, 귀무가설(평균연비의 차이가 없다)를**  **기각하고, 대립가설(평균연비의 차이가 있다)를 채택하므로, “5% 유의수준에서 두 차종의 연비의 차이가**  **있다” 또는 “두 차종의 평균연비의 차이가 통계적으로 유의미하다”** | | | | | | | | | | |
| **부분 점수 정답&배점** | ※ 각 항목별 33% 부분점수 배점. 주요 키워드 및 동일기능 코드 기재여부에 따라 탄력적 배점. | | | | | | | | | | |
| **관련**  **능력단위** | 통계기반  데이터분석 | **관련**  **능력단위**  **요소** | 통계 모델 개발하기 | **평가방법** | 실습형 | **난이도** | ★★★★☆ | **배점** | **10** | **평가**  **시간** | **10분** |
| **문 항** | **4. economics{ggplot2} 데이터셋을 이용하여, 미국의 실업자수(unemploy)와 개인소비경향(pce) 변수 간**  **에 관계가 있는지 검정하고자 한다. 이때, 상관분석을 수행하여, 상관계수(r)를 도출하고, 두 변수**  **간에 어떤 관계가 있는지 판단하고, 관계가 있는 경우, 그 관계가 통계적으로 유의미한 지 검정하시오.** | | | | | | | | | | |
| **평가 시**  **유의사항** | [출제 의도]  통계적 분석방법으로 상관분석 방법과 상관계수의 해석, 그리고 가설검정 역량 판단. | | | | | | | | | | |
| **정 답** | |  |  | | --- | --- | | **득 점** |  |   **① 두 변수(unemploy, pce)간에 상관분석 수행**  **economics <- as.data.frame(ggplot2::economics)**  **cor.test(economics$unemploy, economics$pce)**  **##**  **## Pearson's product-moment correlation**  **##**  **## data: economics$unemploy and economics$pce**  **## t = 18.605, df = 572, p-value < 2.2e-16**  **## alternative hypothesis: true correlation is not equal to 0**  **## 95 percent confidence interval:**  **## 0.5603164 0.6625460**  **## sample estimates:**  **## cor**  **## 0.6139997**  **② 상관계수의 산출 및 해석:**  **상관계수의 값은 0.6139997로, 두 변수 간에 “뚜렷한 정의 선형관계”가 있음을 보여준다.**  **③ 검정결과에 대한 해설:**  **상관분석결과 유의확률(p-value)이 p-value < 2.2e-16 으로, 유의수준 0.05 보다 매우 작으므로**  **귀무가설을 기각하고 대립가설 채택하여, 두 변수 간의 상관관계가 통계적으로 유의미하다.** | | | | | | | | | | |
| **부분 점수 정답&배점** | ※ 각 항목별 33% 부분점수 배점. 주요 키워드 및 동일기능 코드 기재여부에 따라 탄력적 배점 | | | | | | | | | | |
|  |  | | | | | | | | | | |
| **관련**  **능력단위** | 통계기반  데이터분석 | **관련**  **능력단위**  **요소** | 통계  모델 평가 검증하기 | **평가방법** | 실습형 | **난이도** | ★★★★☆ | **배점** | **5** | **평가**  **시간** | **3분** |
| **문 항** | **5. 통계모형에 대한 평가를 위한 여러 방법 중, 아래와 같은 조건에 부합하는 평가방법을 기재하시오.**  **(조건) 임의의 시계열 데이터가 주어졌을 때 훈련 데이터 세트와 검증 데이터 세트로 구분하고,**  **훈련 데이터 세트를 활용하여 ARIMA 모델을 생성 한다. 이 후, 검증 데이터 세트를**  **이용하여 MAE, MSE, MAPE 등과 같은 예측오차를 통해 ARIMA 모델의 성능을 평가하고자 한다**  **(방법들)**  **(1) 예측오차를 이용한 예측모델 성능평가**  **(2) 교차유효성 검사를 통한 예측모델 성능평가**  **(3) ROC곡선을 통한 분류모델 성능평가**  **(4) 군집모델의 성능평가** | | | | | | | | | | |
| **평가 시**  **유의사항** | [출제 의도]  통계적 모델의 성능평가 방법 중, 적절한 방법의 선택역량 판단. | | | | | | | | | | |
| **정 답** | |  |  | | --- | --- | | **득 점** |  |   **(1) 예측오차를 이용한 예측모델 성능평가** | | | | | | | | | | |
| **부분 점수 정답&배점** | ※ 부분점수 없음. | | | | | | | | | | |
| **관련**  **능력단위** | 통계기반  데이터분석 | **관련**  **능력단위**  **요소** | 통계  모델 평가 검증하기 | **평가방법** | 실습형 | **난이도** | ★★★☆☆ | **배점** | **5** | **평가**  **시간** | **2분** |
| **문 항** | **6. 빅데이터를 이용하여, 예측분석모델을 만들엇다고 가정한다. 이때, 예측분석모델의 성능평가를 위한**  **적절한 검사방법 두가지를 기재하시오.** | | | | | | | | | | |
| **평가 시**  **유의사항** | [출제 의도]  빅데이터 기반 예측분석 모델의 성능평가 검사방법에 대한 이해와 선택역량 판단. | | | | | | | | | | |
| **정 답** | |  |  | | --- | --- | | **득 점** |  |   **① 예측 오차 검사**  **② 교차 유효성 검사** | | | | | | | | | | |
| **부분 점수 정답&배점** | ※ 각 항목별 50% 부분점수 배점. 주요 키워드 기재여부에 따라 탄력적 배점 | | | | | | | | | | |