

2022 빅데이터분석기사[실기] 정오표

2022년도 빅데이터분석기사「실기」 제1판 1쇄 독자를 위한 정오표입니다.



2022 빅데이터분석기사[실기] 정오표(제1판 1쇄 발행)

페이지	위치	수정 전	수정 후
작업형 1-6	上 표	연산자 ! 내용 두 개의 ~ 반환하는 연산자	연산자 ! 내용 반대 논리값을 반환하는 not 연산자
작업형 1-18	학습 point	· 4개의 열과 3개의 행으로 ~ · ~ 행은 세 명의 자료	· 4개의 열과 4개의 행으로 ~ · ~ 행은 네 명의 자료
작업형 1-31	上 표	입력값을 절댓값으로 반환하는 사용자 정의 함수를 ~	두 입력 값의 차를 반환하는 사용자 정의 함수를 ~
작업형 2-5, 6	上 표	header 기본값은 FALSE	header 기본값은 TRUE
작업형 2-18	中 표	~ 6보다 작거나 같으면 "S"	~ 5보다 작거나 같으면 "S"
작업형 2-29	19번 문제	~ 값으로 25번째 백분위 수 ~ 수는 무엇인가?	~값은 무엇인가?
작업형 2-50	中 본문	R의 scale 함수에서 center 값을 최댓값	R의 scale 함수에서 center 값을 최소값
작업형 2-92	中 표	cor TRUE 공분산 행렬 사용 FALSE 상관행렬 사용(기본값)	cor TRUE 상관행렬 사용(기본값) FALSE 공분산 행렬 사용
작업형 2-93	上 표	cor = FALSE, # 상관행렬을 사용	cor = FALSE, # 공분산 행렬을 사용
작업형 2-98	上 中 본문	다중 선형 회귀 분석에서는 하나의 독립변수에 따른 ~ 하나의 독립변수와 ~	다중 선형 회귀 분석에서는 2개 이상의 독립변수에 따른 ~ 2개 이상의 독립변수와 ~
작업형 2-123	上 표	test = valid_x, cl = train_y, k = i	test = valid_x, cl = train_y, k = 1)
작업형 2-124	下 표	Accuracy가 0.75로 감소하였음을~ virginica에 대한 Sensitivity(민감도)가~	Accuracy가 0.77로 감소하였음을~ versicolor에 대한 Sensitivity(민감도)가~
작업형 2-166	2번 데이터 세트 2번 해설	> idx <- sample(x= c("train", "test"), + size=nrow(Boston), + replace=TRUE, prob=c(0.8,0.2)) > ds_train <- Boston[idx=="train",] > ds_test <- Boston[idx=="test",] · 데이터 세트를 8:2로 분할한 후 ds_train 변수에 훈련 데이터 생성하고, ds_test 변수에 평가 데이터 생성한다. > median_train <- median(ds_train\$tax) > org_sd <- sd(ds_train\$tax) > ds_train\$tax[is.na(ds_train\$tax)] <- median_train > trans_sd <- sd(ds_train\$tax)	> dshousing <- read.csv("c:/data/Housing.csv") > nrow_end <- nrow(dshousing) * 0.8 > ds_train <- dshousing[c(1:nrow_end),] · nrow 함수를 이용하여 첫 행부터 80%까지의 행을 추출하여 ds_train에 저장한다. > median_train <- median(ds_train\$total_bedrooms, + na.rm = TRUE) > org_sd <- sd(ds_train\$total_bedrooms, na.rm = TRUE) > ds_train\$total_bedrooms[is.na(ds_train\$total_bedrooms)] <- + median_train > trans_sd <- sd(ds_train\$total_bedrooms)
작업형 2-168	1번 해설	1번 문제의 해설	작업형 3-27페이지 14번 문제 해설로 변경
작업형 2-172 (11.19)	4번 해설	dt_airquality <- airquality[c(1: nrow(airquality*0.9)),] + mean(dt_airquality\$Ozone, na.rm = TRUE) + na.rm = TRUE (삭제)	dt_airquality <- airquality[c(1: (nrow(airquality)*0.9)),] + mean(dt_airquality\$Ozone, na.rm = TRUE)
작업형 2-176 (11.19)	11번 해설	11 다음은 ~ 8월 21일의 Ozone 값을 구하시오. + a <- airquality[airquality\$Month == 8 [1] 44	11 다음은 ~ 8월 20일의 Ozone 값을 구하시오. > a <- airquality[airquality\$Month == 8 Ozone 112 44
작업형 2-177	3번 문제	3 다음은 mtcars ~ 자동차의 백분율(%)을 구하시오	3 다음은 mtcars ~ 자동차의 비율을 구하시오
작업형 2-181	20번 문제	"~상위 3위인 국가(country_abrv)를 선택하고 이 국가들~"	"~상위 3위인 국가(country_abrv)를 하나 선택하고 이 국가의~"
작업형 2-182(11.22)	4번 정답	04. 10.62931	04. 10.36634
작업형 2-182	20번 해설	filter(df_fifa_point >= 1765.05) country_abrv 1 GER 2 ITA 3 SUI GER(독일), ITA(이탈리아), SUI(스위스)이다. c('GER', 'ITA', 'SUI')) %>% 348.098	filter(total_points >= 1765.05) country_abrv 1 GER 2 GER 3 GER GER(독일)이다. c('GER')) %>% 421.5051
작업형 2-182	20번 정답	20. 348.098	20. 421.5051
작업형 2-184 (11.19)	1번 해설 下 1번 해설 우측 下	confusionMatrix(y_v\$Churn, predict_rf) ~ 대략 80.2%의 예측 정확도를 보임 confusionMatrix(y_v\$Churn, predict_svm) Reference Prediction No Yes No 758 68 Yes 145 154 Accuracy : 0.8107 95% CI : (0.7865, 0.8332) No Information Rate : 0.8027 P-Value [Acc > NIR] : 0.2636 Kappa : 0.4715	confusionMatrix(predict_rf, y_v\$Churn) ~ 대략 80.9%의 예측 정확도를 보임 confusionMatrix(predict_svm, y_v\$Churn) Reference Prediction No Yes No 757 69 Yes 161 138 Accuracy : 0.7956 95% CI : (0.7708, 0.8188) No Information Rate : 0.816 P-Value [Acc > NIR] : 0.9633 Kappa : 0.4191
작업형 2-186 (11.19)	2번 해설 2번 정답	md_step <- lm(mpg ~ drat + gear + carb, d_train) rmse <- sqrt(mean(d_train\$mpg - pred)^2) [1] 1.728558	md_step <- lm(mpg ~ drat + wt + gear + carb, d_train) rmse <- sqrt(mean((d_test\$mpg - pred)^2)) [1] 3.674629
작업형 3-3 (11.19)	11번 문제	본인 소유의 주택 가격에서 상위 50개의 데이터에 대하여 최솟값으로 변환한 후 타운별 1인당 범죄율 값이 1 이상인 데이터를 구하시오	상위 50개의 MDEV 값을 상위 50개의 MDEV 값의 최솟값으로 변 환한 후 CRIM이 1을 초과하는 값에 대한 CRIM의 평균을 구하시오
작업형 3-6(11.19)	11번 문제	(~ 1.5 표준편차 이하이거나 이상인 값으로~)	(~ 1.5 표준편차 미만이거나 초과인 값으로~)
작업형 3-9(11.22)	13번 문제	(1 : 20 ~ 40세, 2 : 41 ~ 60세, 3 : 60세 이상)	(1 : 20 ~ 40세, 2 : 41 ~ 59세, 3 : 60세 이상)
작업형 3-13	14번 문제	14. 다음은 iris 데이터 세트이다. 주어진 데이터를 이용하여 Species rpart, svm 예측 모형을 만든 후 높은 Accuracy 값	14. 다음은 기업에서 생성된 주문데이터이다. 80,009건의 데이터에 대하여 정시도착 가능 여부 예측 모델을 만들고, 평가데이터에

		을 가지는 모델의 예측값을 CSV 파일로 제출하시오.	대하여 정시도착 가능 여부 예측 확률을 기록한 CSV를 생성하시오.
작업형 3-14	12번 해설	iris_sample70<-iris[c(1:nrow(iris)*0.7),]	iris_sample70<-iris[c(1: (nrow(iris)*0.7)),]
작업형 3-14(11.22)	12번 정답	0.6648607	0.6632932
작업형 3-14(11.19)	13번 정답	11	28
	11번 정답	15.96071	16.05213
작업형 3-17	11번 해설	sd_age <- Carseats %>% filter(Sales <= 5.39 & Sales >= 9.32) %>%	outlier_upper <- mean(Carseats\$Sales) + 1.5*sd(Carseats\$Sales) outlier_lower <- mean(Carseats\$Sales) - 1.5*sd(Carseats\$Sales) sd_age <- Carseats %>% filter(Sales <= outlier_upper & Sales >= outlier_lower) %>%
작업형 3-19	14번 해설	Accuracy : 0.6133	Accuracy : 0.6
	14번 정답	write.csv(pred_md, ~	write.csv(pred_loan, ~
작업형 3-20	11번 정답	54.68845	43.67211
작업형 3-21 (11.19)	11번 해설	alcgp 0-9g/day 10-19 20-29 30+ 0-39g/day 270 94 47 33 40-79 213 102 77 38 80-119 80 68 22 19 120+ 40 30 19 23 X-squared 54.68845	alcgp 0-9g/day 10-19 20-29 30+ 0-39g/day 261 84 42 28 40-79 179 85 62 29 80-119 61 49 16 12 120+ 24 18 12 13 X-squared 43.67211
작업형 3-22(11.22)	13번 해설	2 : 41~60세,	2 : 41~59세,