2024년 1학기 회귀분석 중간고사

|  |  |  |  |
| --- | --- | --- | --- |
| 대상학과 | 첨단공학부 | 교과목명 | 회귀분석 |
| 대상학년 | 2 |
| 공 통 | **문제1.** 16년간 실업률과 임금지수 자료를 이용하여, 다음 물음에 답하시오. (R을 사용하고, 필요한 R문장도 기술하시오.) (40점, 각 5점)   |  |  |  |  | | --- | --- | --- | --- | | 실업률 | 임금지수 | 실업률 | 임금지수 | | 4.5 | 50 | 2.6 | 91 | | 4.4 | 54 | 2.4 | 100 | | 4.1 | 58 | 2.3 | 108 | | 3.8 | 62 | 2.4 | 117 | | 4.0 | 66 | 2.8 | 125 | | 3.8 | 69 | 2.4 | 132 | | 3.1 | 74 | 2.0 | 141 | | 2.5 | 80 | 2.0 | 150 |   (1) 자료의 산점도를 그리고, 산점도 위에 회귀직선을 그려 넣으시오.   |  | | --- | | **A.**  > x <- c(4.5, 4.4,4.1,3.8, 4.0, 3.8, 3.1, 2.5, 2.6, 2.4, 2.3, 2.4, 2.8, 2.4, 2.0, 2.0)  > y <- c(50, 54, 58, 62, 66, 69, 74, 80, 91, 100, 108, 117, 125, 132, 141, 150)  > plot(x, y)  > abline(lm(y~x)) |   (2) 실업률을 설명변수, 임금지수를 반응변수로 하는 선형회귀분석을 실시하려 한다. 최소제곱법에 의한 직선의 식을 추정하여라.(즉, 기울기와 절편 추정량을 구하여라)   |  | | --- | | **A.**  > lm(y ~ x)  > coef(lm(y~x))  (Intercept) x  194.0419 -33.1501 |   (2) 위 모형의 MSE를 구하여라.   |  | | --- | | **A. 252.01 / 220.5064**  > print(sprintf (" 0.2f", sum (model$residuals^ model$df.residual)) |   (4) 주어진 실업률 에 따른 예측치 을 구하여라.   |  | | --- | | **A.**  > predict(lm(y~x), data.frame(lstat = (c(4.5, 4.4, 4.1, 3.8, 4.0, 3.8, 3.1, 2.5, 2.6, 2.4, 2.3, 2.4, 2.8, 2.4, 2.0, 2.0))), interval = "prediction")  1 44.86641 / 2 48.18142 / 3 58.12645  4 68.07149 / 5 61.44147 / 6 68.07149  7 91.27656 / 8 111.16662 / 9 107.85161  10 114.48163 / 11 117.79664 / 12 114.48163  13 101.22159 / 14 114.48163 / 15 127.74167  16 127.74167 |   (5) 기울기와 절편에 대한 95% 신뢰구간을 구하여라.   |  | | --- | | **A.**  > confint(lm(y~x), level = 0.95)  2.5 % 97.5 %  (Intercept) **162.27420 225.80957**  x **-43.12357 -23.17664** |   (6) 회귀계수가 유의한지 검정하여라.(가설, 통계량, 유의확률 값, 검정의 결론 명시) (10점)   |  | | --- | | **A.**  > summary(lm(y~x))  t통계량 : 13.101 / p-value : 3.01e-09  유의수준 0.01수준에서 귀무가설 기각. 따라서 는 0이 아니다.  t통계량 : -7.129 / p-value : 5.10e-06  유의수준 0.01수준에서 귀무가설 기각. 따라서 는 0이 아니다. |   (7) Residual vs Fitted plot, Q-Q plot을 각각 그리시오.   |  | | --- | | **A.** |   **문제2.** 대학입학 학력고사에서 수학성적(), 국어성적() 그리고 IQ()가 대학생활의 성적에 어떤 영향을 미치는가를 분석하기 위하여 1학년 학생 10명을 임의로 선정하여 조사한 자료가 다음과 같다. (40점, 각 10점)   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | 65 | 84 | 78 | 73 | 52 | 82 | 92 | 89 | 75 | 56 | |  | 39 | 71 | 43 | 62 | 21 | 64 | 74 | 47 | 52 | 38 | |  | 52 | 84 | 64 | 59 | 40 | 59 | 72 | 65 | 54 | 45 | |  | 116 | 131 | 129 | 116 | 121 | 124 | 132 | 125 | 112 | 114 |  1. Scatter Plot Matrix를 그리시오  |  | | --- | | **A.**  > score <- c(65, 84, 78, 73, 52, 82, 92, 89, 75, 56)  > Math <- c(39, 71, 43, 62, 21, 64, 74, 47, 52, 38)  > Korea <- c(52, 84, 64, 59, 40, 59, 72, 65, 54, 45)  > IQ <- c(116, 131, 129, 116, 121, 124, 132, 125, 112, 114)  >pairs(cbind(score, Math, Korea, IQ)) | | | |
|  | 1. Multi regression model 2가지 형태를(①, ②) 가정하고 회귀선을 추정하여라.  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **A.**  >model1 <- lm(score ~ Math + Korea + IQ)  >model2 <- lm(score ~ Math + IQ)  >summary(model1)  >summary(model2)   |  |  |  | | --- | --- | --- | |  | Model1 | Model2 | |  | -5.7244 | -32.9491 | |  | 0.3347 | 0.5114 \* | |  | 0.4092 |  | |  | 0.3232 | 0.6674 | |  1. 각 모형의 회귀계수 유의성을 검정하여라.(가설, 통계량, 유의확률 값, 검정의 결론 명시)  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **A.**  >summary(model1)  >summary(model2)   |  |  |  | | --- | --- | --- | |  | Model1 | Model2 | |  | -5.7244 | -32.9491 | |  | 0.3347 | 0.5114 \* | |  | 0.4092 |  | |  | 0.3232 | 0.6674 |   t통계량 : x0=-0.101, x1=1.123, x2=0.811, x3=0.554  p-value : x0=0.923, x1=0.304, x2=0.448, x3=0.599  유의수준 0.01수준에서 귀무가설 기각하지 못함. 따라서  t통계량 : x0=-0.743, x1=3.002, x3=1.1712  p-value : : x0=0.4818, x1=0.0199, x3=0.1306  유의수준 0.01수준에서 귀무가설 기각. 따라서 , |  1. 각 모형의 값을 구하여라.   **Model 1 77.4% / 74.9%**  **Model 2 0.6608 / 0.6774** | | |