**『온라인 제출용 출석수업대체과제물 표지』**

2019학년도 ( 2 )학기 출석수업대체과제물

교과목명 : 회귀모형

학 번 : 201935-368417

성 명 : 유민아

연 락 처 : 010 – 9646 - 6509

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

○ 과 제 명 :

교재 1장 연습문제(39쪽) : 1번, 7번

교재 2장 연습문제(86쪽) : 2번, 3번

교재 78쪽 분석사례 : Rstudio에서 html(또는 pdf) 요약보고서 만들기

- Rstudio 설치후, library(rmarkdown), library(knitr) - 구글(또는 네이버)에서 rmarkdown 검색)

회귀모형 출석수업 대체과제물

## 교재 1장 연습문제(39쪽) : 1번

## 기계들의 사용연도(age)와 정비비용(cost)간의 관계

(1) 데이터의 산점도

|  |
| --- |
|  |
|  |
|  |

(2)최소제곱법에 의한 회귀직선 적합

|  |
| --- |
|  |
|  |
|  |

(3)추정치의 표준오차

|  |
| --- |
|  |
| 🡪Sy.x = square root of MSE = Residual standard error = 44.73 |

(4) 결정계수와 상관계수

|  |
| --- |
|  |
| 🡪결정계수 = Multiple R-squared = 0.1447  🡪상관계수 = 결정계수의 제곱근 = 0.3803 |

(5) 분산분석표 작성, 회귀직선의 유의여부 검정(유의수준 a=0.05사용)

|  |
| --- |
|  |
|  |
| 🡪 p-값이 0.1797로 유의수준 0.05 기준으로 보았을 때 유의하지 않음. Cost에 대한 age의 설명력이 약함을 알 수 있다. |

(6) 사용년도가 4년인 기계의 평균정비비용 추정

|  |
| --- |
|  |
| 🡪(4)summary를 통하여 y hat =49.967 +6.895 \*age 임을 알 수 있다.  x에 4대입하면 y hat = 77.547, 기계의 평균 정비비용은 약 77,547원 |

(7) 잔차와 잔차의 합

|  |
| --- |
|  |
| 🡪 잔차의 합이 0에 가까운 매우 작은 값임을 알 수 있다. |

(8) 잔차들의 xi에 대한 가중합

(9) 잔차들의 yi에 대한 가중합

|  |
| --- |
|  |
|  |

(10) 두변수 X,Y를 표준화된 변수로 고친 후 회귀직선 적합, 그 회귀계수가 두 번수 X, Y간의 상관계수와 같음을 밝혀라.

|  |
| --- |
|  |
|  |
| z.age의 회귀계수는 0.3804로, (4)에서 구한 상관계수와 같음을 알 수 있다. |
|  |

교재 1장 연습문제(42쪽) : 7번

10년간 판매실적과 연도

(1)회귀방정식과 분산분석표

|  |
| --- |
|  |
|  |

(2)각 xi에서 yi^와 잔차 yi-yi^를 구하고 잔차그림을 그려라. 잔차가 가장 큰 연도는 언제인가?

|  |
| --- |
|  |
|  |
| 10번째 데이터(1992년)의 잔차가 29.981818로 가장 크다. |

# 교재 2장 연습문제(86쪽) : 2번

공장 제품의 강도와 온도와 압력

(1) 회귀모형을 추정하라

|  |
| --- |
|  |
| 회귀모형: Y hat = -554.5267 - 0.1743X1 + 11.8449X2 |

(2)오차분산을 MSE로 추정하고 Var(b0), Var(b1), Var(b2)의 추정치를 구하라

|  |
| --- |
|  |
| * 오차분산 = E(잔차평균제곱MSE) = 469.4 * Var(b) = (X’X)^-1\*MSE = XXTXXI \* 469.4 = XXTXXIV에 대입 * 결과값 순서대로 VAR(B0)=38901.52005, VAR(B1)=0.5831197, VAR(B2)=10.4607703 |

(3) X1=200, X2=59에서 평균 제품의 강도의 추정치 Y HAT은 얼마인가? 이 Y HAT의 분산을 추정하라.

🡪 (1)에서 구한 회귀모형 Y hat = -554.5267 - 0.1743X1 + 11.8449X2 에 값 대입하면,

Y HAT = 109.4624

🡪 Y HAT의 분산 추정값 VAR(Y^) = x’(X’X)-1x\*MSE

(4)추정된 회귀계수 b1,b2의 의미는 무엇인가?

🡪 b1과 b2는 기울기 회기계수로서, 다른 변수가 상수로 고정되어 있을 때, b1인 경우 X1(온도)이 1만큼 증가한다면 종속변수 Y(강도)가 -0.1743만큼 평균적으로 감소하고 `b2인 경우 X2(압력)가 1만큼 증가할 때 Y(강도)가 11.8449만큼 평균적으로 증가함을 의미한다.

(5)분산분석표를 작성하고 a=0.05로 F-검정을 행하라

|  |
| --- |
|  |
| 분산분석표: |
| * F값 7.383에 대한 유의확률 p-값=0.03218 로 유의수준 0.05보다 작으므로 귀무가설 기각, 적합된 중회귀모형이 유의함. |

(6)결정계수 R2을 구하라

|  |
| --- |
|  |

* 결정계수 = Multiple R-squared = 0.747

(7) X1,X2,Y를 모두 표준화시키고 표준화된 중회귀방정식을 구하라

|  |
| --- |
|  |

* Y^\*=-0.05499X1 + 0.88252X2

# 교재 2장 연습문제(87쪽) 3번

# 물소비량, 온도, 작업일수, 작업량

(1)회귀방정식을 구하고 어째서 이 모형이 선택되었는지 논하라.

|  |
| --- |
|  |

* 회귀방정식: Y hat = 2.409213 + 0.069788 XX1 - 0.024767 XX2 + 0.005864 XX3

독립변수 XX1(평균온도), XX2(작업일수), XX3(작업량)에 대해 종속변수 Y(물소비량)를 놓고 회귀모형을 돌렸기 때문에 위와 같은 방정식이 나왔다 하지만 유의수준 0.05를 기준으로 보았을 때 XX2, XX3의 P-값이 각각 0.60060, 0.28979로, 유의수준을 벗어났기 때문에 유의하지 않음을 알 수 있다.

(2) b1, b2, b3의 의미는 무엇인가?

🡪 b1, b2, b3은 기울기 회기계수로서, 다른 변수가 상수로 고정되어 있을 때, b1인 경우 X1(온도)이 1만큼 증가한다면 종속변수 Y(물소비량)가 0.069788만큼 평균적으로 감소하고 b2인 경우 X2(작업일수)가 1만큼 증가할 때 Y(물소비량)가 0.024767만큼 평균적으로 감소하며, b3의 경우 X3(작업량)이 1만큼 증가할 때 Y(물소비량)가 0.005864만큼 증가함을 의미한다.

(3)분산분석표를 작성하고, 결정계수 R2을 구하라

|  |
| --- |
|  |
| 분산분석표: |

* 결정계수 = Multiple R-squared = 0.9202

(4) XX1=20, XX2=27, XX3=60 에서 평균 물소비량을 추정하라.

🡪 Y hat = 2.409213 + 0.069788 XX1 - 0.024767 XX2 + 0.005864 XX3 에 값 대입하면 YY^= 평균 물소비량 추정값 = 3.488104