1. Final을 종속변수, t1~t5를 독립변수로 하는 다중회귀분석을 실시하고 그 결과를 restest라는 객체이름으로 저장한 후, 함수 summary를 이용하여 각종 추정값, 결정계수, 유의성 검정 결과를 설명하시오.

|  |
| --- |
| > restest <- lm(final ~ t1 + t2 + t3 + t4 + t5, data = data)  > summary(restest)  Call:  lm(formula = final ~ t1 + t2 + t3 + t4 + t5, data = data)  Residuals:  Min 1Q Median 3Q Max  -13.250 -2.521 0.091 2.622 8.952  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -8.17480 1.86488 -4.384 1.91e-05 \*\*\*  t1 0.32494 0.03613 8.993 < 2e-16 \*\*\*  t2 -0.05013 0.21918 -0.229 0.819332  t3 0.14506 0.03873 3.746 0.000237 \*\*\*  t4 0.44275 0.22029 2.010 0.045831 \*  t5 0.30088 0.03685 8.166 4.00e-14 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 4.071 on 194 degrees of freedom  Multiple R-squared: 0.8385, Adjusted R-squared: 0.8343  F-statistic: 201.4 on 5 and 194 DF, p-value: < 2.2e-16  추정값: t2는 의미가 없는 변수이고, t4는 의미가 거의 없는 변수이다.  결정계수: 0.8343  유의성 검정 결과: F = 201.4이며, p-value값은 유의하다.  추정값은 Estimate 밑 결과값  결정계수는 맞고 p 값은 0;001 보다 작기 때문에 사용하는게 낫다  T2는 유의하지 않다 \*참고 |
|  |

1. 다중공선성을 확인하고 그 결과를 설명하시오.

|  |
| --- |
| t1 t2 t3 t4 t5  1.567852 57.690405 1.801130 58.275192 1.630464  t1, t3, t5는 다중공선성이 없고 t2, t4는 다중공선성이 있다.  cor(data$t2, data$t4)  [1] 0.9912691  실제 두 변수 사이에 강한 상관관계가 있음을 확인할 수 있다.  10이사이면 의심을 해봐야한다. 다중공선성 의심되는 건 2하고 4 |

1. 다중공선성을 유발하는 시험을 삭제한 후 다중회귀분석을 다시 실시하시오. 그 결과는 restest2라는 객체이름으로 저장하고 함수summary를 이용하여 각종 추정값, 결정계수, 유의성 결과를 설명하시오.

|  |
| --- |
| > restest2 <- lm(final ~ t1 + t3 + t5, data = data)  > summary(restest2)  Call:  lm(formula = final ~ t1 + t3 + t5, data = data)  Residuals:  Min 1Q Median 3Q Max  -18.4195 -3.5556 0.1094 4.0704 10.9038  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -2.01840 2.24921 -0.897 0.371  t1 0.33844 0.04576 7.395 4.02e-12 \*\*\*  t3 0.29011 0.04611 6.292 2.01e-09 \*\*\*  t5 0.41182 0.04492 9.168 < 2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 5.161 on 196 degrees of freedom  Multiple R-squared: 0.7376, Adjusted R-squared: 0.7336  F-statistic: 183.7 on 3 and 196 DF, p-value: < 2.2e-16  추정값: t1, t3, t5 모두 \*\*\* 표시가 있으므로 의미가 있는 변수라고 판단할 수 있다.  결정계수: 0.7336  유의성 검정 결과: F비는 183.7이며, p-value값은 유의하다.  **다중공선성 의심되는거 변수 2개를 다 뺄 필요가없다**  **둘 중에 하나만 제거하는 방법을 사용하여 모델링하기**  **변수 두개는 서로 상관관계가 있어 같이 움직이기 때문에 두 변수 중복됨 그래서 둘 중 하나만 씀으로서 모델링 할 수 있다.**  **T4를 제거했을 때와**  **t2를 제거했을 때의 모델 모두 유의성 나타남**  **비교했을 때 결정계수는 딱히. 변화 없지만 p값은 유의함**  **🡪 변수 하나 제거 한다해도 큰 영향이 나타나지 않는다 한다**  **문제 3번은 두 가지의 경우를 생각할 수 있는 것**  **그치만, 두 모델 비교했을 때 결과의 큰 차이는 없기에 둘 중 하나만 사용**  **원래는 t2를 봤을 때 유의성이 아예 없기 때문에 아예 제거를 해야 한다**  **유의하지 않은 것은 모델에 포함시킬 이유가 없다 제거한다.**  **앞으로 다중공선성 변수 2개 나타나면 하나만 제거하고 나타내면 된다.** |