**2013 회귀분석1 실습 기말고사**

**※ 기본적으로 검정은 유의수준 5%에서 실시하고 귀무가설, 대립가설, 검정결과를 모두 기재하라.**

**※ 을 제시하라는 조건이 있는 문제에 한해 을 같이 기재하시오.**

**학과: 학번: 이름:**

**1. 다음 자료는 자동차 연비(km/L)가 자동차 무게(10kg)에 따라 어떻게 변하는지 알아보고자 10대의 차량을 대상으로 조사한 자료이다. (총 25점)**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 12.2 | 9.7 | 10.3 | 15.1 | 13.3 | 11.6 | 12.3 | 8.9 | 14.2 | 11.4 |
|  | 93 | 121 | 108 | 77 | 98 | 105 | 102 | 119 | 88 | 111 |

(1)위 자료를 분석할 수 있는 회귀모형을 세울 때, 그 모형의 독립변수와 종속변수를 적절하게

정하여 연비와 자동차 무게의 관계를 설명하는 직선식(추정회귀모형)을 구하시오. (5점)

(2) 위의 자료로부터 연비가 자동차 무게의 선형식으로 표현될 수 있는지 판단하시오. (5점)

의 test통계량의 p-value(<0.0001)가 유의수준 0.05보다 작으므로 귀무가설을 기각함. 따라서 연비가 자동차 무게의 선형식으로 표현된다고 말할 수 있다.

(3) 자동차의 무게를 100kg 줄일 때 평균 연비가 1이라고 말할 수 있는지 검정하시오. (5점)

test통계량의 p-value(=0.0684)가 유의수준 0.05보다 크므로 귀무가설을 기각하지 못함. 따라서 자동차의 무게를 100kg 줄일 때 평균 연비가 1이 아니라고 말할 수 없다.

(4) (2)의 모형에서 자동차의 무게가 1톤인 자동차들의 평균연비의 95% 신뢰구간을 구하시오. (5점)

(5) (2)의 모형에서 무게의 기울기에 대한 **90%** 신뢰구간을 구하시오. (5점)

**2. 다음은 다섯 개의 변수 에 대한 10개의 관측치를 정리한 것이다. (총 25점)**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 1 | 1 | 1 | 0 | -1 | -1 | -1 | -1 | 0 |
|  | 1 | 2 | 1 | 2 | 0 | -1 | -2 | -1 | -2 | 0 |
|  | 2 | 3 | 6 | 5 | 1 | -2 | 2 | 1 | 0 | 4 |
|  | 1 | 1 | -1 | -1 | 0 | 1 | 1 | -1 | -1 | 0 |
|  | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 |

(1) 을 설명하기 위해 가 포함되는 선형회귀모형(모형 A)을 쓰고, 모형 A의 추정식을

구하시오(모형 A에서 모형의 가정을 같이 표기할 것). (5점)

(2) 모형 A의 독립변수 행렬을 라고 할 때, 의 대각원소들의 합을 구하시오. (5점)

(3) 모형 A의 변수 의 계수가 1인지에 대해서 유의수준 **0.10** 하에서 검정하라. (5점)

test통계량의 p-value(0.0507)가 유의수준 0.10보다 작으므로 귀무가설을 기각함. 따라서 변수 의

계수가 1이 아니라고 말할 수 있다.

(4) 을 설명하기 위해 가 포함되는 보다 간단한 선형회귀모형(모형 B)을 고려한다고 하자. 모형 A와 모형 B중 어떤 모형이 더 타당한지에 대해 검정하시오(을 밝힐 것). (5점)

test통계량의 p-value(=0.3158)가 유의수준 0.05보다 크므로 귀무가설을 기각하지 못함. 따라서 모

형 B가 더 타당하다고 말할 수 있다.

(5) 모형 B을 이용하여 독립변수가 두 번째 관측치일 때, 즉, 일 때 의 기댓값에 대한 **90%** 신뢰구간을 구하시오. (5점)

**3. 제시된 자료를 이용하여 다음의 질문에 답하시오. (총 25점)**

(1) 모든 설명변수를 사용하여 회귀모형을 세우고, 유의한 회귀계수에 한하여 그 의미를 기술하라.

(5점)

(2) (1)의 모형을 사용하였을 때 오차가 정규성 가정을 만족하는지를 검정하시오. (5점)

Shapiro-Wilk 검정에서 test통계량의 p-value(<0.0001)가 유의수준 0.05 보다 작으므로 귀무가설을 기각함. 따라서 오차가 정규분포를 따르지 않는다고 말할 수 있다.

(3) (1)의 모형에 대해 비 정상적인 관측치나 영향력 있는 관측치를 식별하기 위한 통계량을 계산하여, 높은 지레점으로 판별되는 관측치를 찾으시오. (5점)

높은 지레점 : 22, 25

(4) (3)에서 계산한 통계량을 참고하여 일부 관측치를 제거하면 오차의 정규성 가정이 만족되고 모형 적합 결과도 이전보다 훨씬 좋아진다. Adjusted 가 크게 증가하고, 유의하지 않던 변수도 매우 유의하게 나타나게 된다. 이 때 제거된 관측치를 찾으시오. (10점)

제거된 관측치 : 5, 8, 16, 20

**4. 제시된 자료를 이용하여 수학통계학부 학생을 대상으로 개설된 기초통계 강좌의 기말시험 성적()을 설명하는 변수로 중간시험 성적(), 기말시험 바로 전 문제풀이시간의 참석여부( : 참석이면 1, 불참이면 0), 통계학전공 희망여부( : 통계학 전공을 희망하면 1, 수학전공을 희망하면 0)를 고려하였다. (총 25점)**

(1) 기말시험 성적은 중간시험 성적에 의해 설명되는지 알고자 할 때, 사용해야 할 과검정하고자 하는 귀무가설을 밝히고 검정을 실시하시오. (총 5점)

test통계량의 p-value(<0.001)가 유의수준 0.05보다 작으므로 귀무가설을 기각함. 따라서 기말시험

성적은 중간시험 성적에 의해 설명된다고 말할 수 있다.

(2) 와의 관계가 문제풀이 시간의 참석여부에 따라서 다르게 나타나는지 알고자 한다. 오차의 분포는 같다고 가정할 때, 사용해야 할 과 검정하고자 하는 귀무가설을 밝히고 검정을 실시하시오. (총 5점)

test통계량의 p-value(<0.0045)가 유의수준 0.05보다 작으므로 귀무가설을 기각함. 따라서 와 의

관계가 문제풀이 시간의 참석여부에 따라서 다르게 나타난다고 말할 수 있다.

(3) 가 같은 값을 가질 때 문제풀이시간의 참석여부()나 통계학전공 희망여부()가 기말시험의 성적에 영향을 주는 것으로 알려져 있다. 이 때가 0인 경우와 1인 경우에 에 따른 의 기대값의 차이가 다르다고 말할 수 있는지 유의수준 **0.10** 하에서 검정하고자 한다. 사용해야 할 과 검정하고자 하는 귀무가설을 밝히고 검정을 실시하시오. (15점)

다음의 로부터,

*,*

*,*

이므로, 다음의 검정을 통해 가 0인 경우와 1인 경우에 에 따른 의 기대값의 차이가 다르다고 할 수 있는지 판단할 수 있다.

test통계량의 p-value(=0.06)가 유의수준 0.10보다 작으므로 귀무가설을 기각함. 따라서 가 0인 경우와 1인 경우에 에 따른 의 기대값의 차이가 다르다고 말할 수 있다.