

1.(15점) 단순선형회귀모형의 분석에서 다음의 질문에 답하라. (각각 50자 이내로 설명할 것)

1.1 반응변수인 각각의 y_i 가 정규분포를 따른다고 가정하면 회귀계수의 최소제곱추정량도 정규분포를 따른다고 한다. 그 이유를 설명하라. (5점)

회귀계수의 최소제곱추정량은 y_i 의 선형결합으로 표현 가능
→ y_i 가 정규분포를 따르면 회귀계수의 최소제곱추정량도 정규분포를 따른
(정규분포를 따르는 확률변수의 선형결합은 정규분포)

2

1.2 분산분석표의 F -값 = MSR/MSE 을 이용한 유의수준 0.05의 F -검정에서 귀무가설을 기각하지 못한다는 결과를 얻었다. 그런데 가설 $H_0: \beta_1 = 0$, $H_1: \beta_1 > 0$ 의 유의수준 0.05의 t -검정에서는 귀무가설을 기각한다는 결과를 얻었다. 두 검정의 결과가 같은 것으로 알고 있는데 이런 현상이 발생한 이유는 무엇인가? (5점)

회귀계수는 유의하지만 모형 자체가 통계적으로 유의하지 않은 경우
회귀계수의 유의성은 t 통계량으로 확인
모형의 통계적 유의성은 F 통계량으로 확인

1.3 반응변수가 정규분포를 따르면 SSR/σ^2 와 SSE/σ^2 는 서로 독립이고, 각각 자유도가 1인 카이제곱분포와 자유도가 $n-2$ 인 카이제곱분포를 따른다는 것은 적절하지 못한 설명이다. 그 이유는 무엇인가? (5점)

" $\beta_1=0$ 인 경우에만" SSR/σ^2 이 자유도가 1인 카이제곱분포를 따른다.
이 조건과 관계 없이 SSE/σ^2 는 SSR/σ^2 와 서로 독립된 확률변수이며, 자유도가 $n-2$ 인 카이제곱분포를 따른다.

2.(35점) 기계의 사용기간(X , 단위: 년)과 정비비용(Y , 단위: 만원) 사이의 관계를 알아보기 위해 사용기간이 1년에서 10년인 기계 22대를 대상으로 자료를 얻었다. 자료의 요약 결과가 다음과 같다. 선형회귀모형을 가정하고 최소제곱법에 의하여 회귀계수를 추정한 결과 다음의 결과를 얻었다.

$$\bar{x} = 5, \quad \bar{y} = 70, \quad S_{xx} = 50, \quad S_{xy} = 500, \quad S_{yy} = 6000, \quad \hat{\beta}_0 = 20, \quad \hat{\beta}_1 = 10$$

$$t(.025, 20) = 2.086 \quad t(.025, 21) = 2.080 \quad t(.025, 22) = 2.074$$

$$t(.05, 20) = 1.725 \quad t(.05, 21) = 1.721 \quad t(.05, 22) = 1.717$$

2.1 적절한 회귀모형을 적고 가정(분포 포함)을 정확히 기술하라. (7점) (3/4)

모형:	$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, 22$
가정:	$\epsilon_i \sim N(0, \sigma^2), \quad cov(\epsilon_i, \epsilon_j) = 0, \quad i \neq j$ $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), \quad cov(y_i, y_j) = 0, \quad i \neq j$

2.2 추정회귀식을 적어라. (4점)

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 20 + 10x$
--

2.3 잔차의 제곱합 SSE 를 구하라. (4점)

$SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 6000 - 10(500) = 1000$

2.4 오차의 분산의 추정값을 구하라. (4점)

$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{1000}{20} = 50$

2.5 가설 $H_0: \beta_1 = 0$ 대 $H_1: \beta_1 \neq 0$ 에 대한 F -검정을 하고 결과를 자세히 해석하라.(유의수준 $\alpha = 0.05$) (F-검정통계량의 값과 분포를 밝혀야함. t-검정의 결과만으로 분석하면 0점 처리됨.) (8점)

$SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 6000 - 10(500) = 1000$ $SSR = \hat{\beta}_1 S_{xy} = 10 \times 500 = 5000$ $F^* = \frac{MSR}{MSE} = \frac{5000}{50} = 100 > 4.35 = F(0.05, 1, 18) = t(0.025, 20)^2$ $t^* = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{10}{1} = 10 > 2.086 = t(.025, 20), \quad se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 / S_{xx}} = \sqrt{50/50} = 1$
<p>결론 및 해석:</p> <p>$F^* > F$-검정통계량이므로 귀무가설 H_0를 기각한다. 따라서 $\beta_1 \neq 0$이며, 이 추정회귀식은 타당하므로 위 자료의 설명변수와 반응변수는 선형 관계를 갖는다.</p> <p>$t^* > t$-검정통계량이므로 귀무가설 H_0를 기각한다. 따라서 $\beta_1 \neq 0$이며, 이 추정회귀계수는 타당하다.</p>

2.6 사용기간 5년일 때 정비비용(Y)의 기댓값 $E(Y|X=5)$ 이 60만원이 넘는 지를 알아보고자 한다.

적절한 가설을 세우고 검정하고 결과를 자세히 해석하라. (유의수준 $\alpha = 0.05$) (8점) (2/4/2)

가설:

$$H_0 : E(Y|X=5) = 60 \quad \text{or} \quad \leq 60$$

$$H_1 : E(Y|X=5) > 60$$

$$\hat{E}(Y|X=5) = \hat{\beta}_0 + \hat{\beta}_1 x_0 = 20 + 10 \times 5 = 70$$

$$se(\hat{E}(Y|X=5)) = \sqrt{MSE \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} = \sqrt{50 \times \left(\frac{1}{22} + \frac{(5 - 5)^2}{50} \right)} = \sqrt{2.2727} = 1.5076$$

$$t^* = \frac{\hat{E}(Y|X=5) - 60}{se(\hat{E}(Y|X=5))} = \frac{70 - 60}{1.5076} = 6.633 > 1.725 = t(0.05, 20)$$

$$\text{귀무가설 } H_0 : E(Y|X=5) = 60 \quad \text{or} \quad \leq 60 \quad \text{기각함}$$

결론 및 해석:

사용기간 5년일 때 정비비용의 기댓값 $E(Y|X=5)$ 이 60만원이 넘는다.

b

3.(15점) 설명변수의 20개 값에 대하여 반응변수의 값들을 측정하였는데 마지막 5개의 반응변수의 값에 문제가 있는 것으로 판단된다. 측정과정을 조사한 결과 측정기구의 오류로 인하여 마지막 5개의 반응변수값의 측정에서 표준편차가 나머지 15번의 측정에 2배가 된다는 것이 발견되었다. 이 때 반응변수나 오차의 분산은 모두 같다고 가정하기 어렵고 가중최소제곱법을 사용해야한다.

3.1 가중최소제곱법을 적용하기 위해 20개의 자료값의 가중값을 어떻게 주어야 하는가? (8점)

$$Var(y_i) = \sigma^2, \quad w_i = 1, \quad i = 1, \dots, 15$$

$$Var(y_i) = 4\sigma^2, \quad w_i = 1/4, \quad i = 16, \dots, 20$$

3.2 이 자료의 경우 가중최소제곱법에 의한 추정량이 최소제곱추정량 보다 나은점은 무엇인가? (7점)

BLUE

등분산 가정을 할 수 없는 경우 최소제곱추정량은 최량선형비편향추정량(BLUE)가 아님
이 때 가중최소제곱법으로 가중치를 곱해 구한 추정량은 최량선형비편향추정량이다.

4.(15점) 어떤 공장의 제품의 생산과정에서 공정압력(X , 단위: psi)과 제품의 강도(Y , 단위: kg/cm^2)의 관계를 알아보기 위하여 단순선형회귀모형을 생각하였다.[참고: $F(.05;df_1,df_2)=3.10$]

X	n_i	Y						\bar{y}	\hat{y}	$\sum (y - \bar{y})^2$	$\sum (\bar{y} - \hat{y})^2$
2.5	5	11.6	6.0	3.4	8.4	6.1		7.10	11.07	37.84	78.82
5.0	4	9.7	14.5	9.6	14.2			12.00	12.04	22.14	0.01
10.0	6	16.0	13.0	13.8	17.9	20.3	13.2	15.70	13.98	43.24	17.81
20.0	5	22.2	25.8	23.9	21.5			23.35	17.85	11.05	120.89
40.0	5	23.0	25.5	19.8	21.7	25.9	25.4	23.55	25.60	30.54	25.29
합계	25									144.81	242.82

4.1 모형이 적절하였는지를 검정하라. (11점) (8/3)

적합결여검정

$$\begin{cases} H_0: E(Y) = \beta_0 + \beta_1 X \\ H_1: E(Y) \neq \beta_0 + \beta_1 X \end{cases}$$

$$SS_{PE} = 144.81 \quad df_{PE} = n - m = 25 - 5 = 20$$

$$SS_{LOF} = 242.83 \quad df_{LOF} = m - 2 = 52 - 3$$

$$SSE = 387.63$$

$$F^* = \frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)} = \frac{242.82/3}{144.51/20} = \frac{80.94}{7.24} = 11.18 > 3.10 = F(0.05, 3, 20): \text{귀무가설을 기각}$$

결론 및 해석:

$F^* > F$ -검정통계량이므로 귀무가설을 기각한다.
따라서 위 모형은 적절하지 않다.

4.2 가설검정의 결과 만약 모형이 적절하지 않다면 어떠한 조치가 필요한가? (4점)

다른 모형을 사용해야한다.

교재 기준)

설명변수나 반응변수를 변환시키거나, 새로운 설명변수가 추가된 모형을 자료에 적합시킨다.