12장 회귀분석

12.1

(1) $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$

- $\dfrac{\partial}{\partial \beta_0} Y_i = 1,\ \dfrac{\partial}{\partial \beta_1} Y_i = x_i,\ \dfrac{\partial}{\partial \beta_2} Y_i = x_i^2 \quad \Rightarrow$ 선형

(2) $Y_i = e^{\beta_0 + \beta_1 x_i} \epsilon_i$

- $\dfrac{\partial}{\partial \beta_0} Y_i = e^{\beta_0 + \beta_1 x_i} \epsilon_i,\ \dfrac{\partial}{\partial \beta_1} Y_i = x_i e^{\beta_0 + \beta_1 x_i} \epsilon_i \ \Rightarrow$ 비선형

(3) $Y_i = \dfrac{\beta_0}{1 + \exp(\beta_1 + \beta_2 x_i)} + \epsilon_i$

- $\dfrac{\partial}{\partial \beta_0} Y_i = \dfrac{1}{1 + \exp(\beta_1 + \beta_2 x_i)},\ \dfrac{\partial}{\partial \beta_1} Y_i = \dfrac{-\beta_o \exp(\beta_1 + \beta_2 x_i)}{\{1 + \exp(\beta_1 + \beta_2 x_i)\}^2},\ \dfrac{\partial}{\partial \beta_2} Y_i = \dfrac{-\beta_o x_i \exp(\beta_1 + \beta_2 x_i)}{\{1 + \exp(\beta_1 + \beta_2 x_i)\}^2} \ \Rightarrow$ 비선형

(4) $Y_i = \beta_o + \beta_1 \sin(t x_i) + \beta_2 \cos(t x_i) + \epsilon_i,\ t$는 임의의 상수

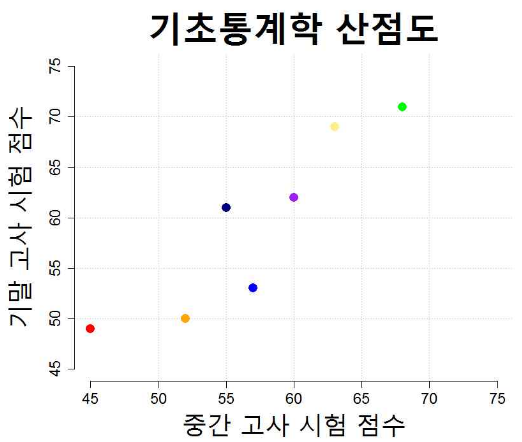
- $\dfrac{\partial}{\partial \beta_0} Y_i = 1,\ \dfrac{\partial}{\partial \beta_1} Y_i = \sin(t x_i),\ \dfrac{\partial}{\partial \beta_2} Y_i = \cos(t x_i) \ \Rightarrow$ 선형

(5) $\log Y_i = \beta_0 + \beta_1 x_i + \log \epsilon_i$

- $\dfrac{\partial}{\partial \beta_0} \log Y_i = 1,\quad \dfrac{\partial}{\partial \beta_1} \log Y_i = x_i \ \Rightarrow$ 선형

12.2

(1)



(2) $\widehat{corr}(X, Y) = 0.9031$

(3) $\hat{\beta}_0 = -1.712,\ \hat{\beta}_1 = 1.067$

(4) ① 65.5379 ② -3.7959 ③ 56.99831 ④ 4.0001

(5) 가설: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

- test statistic = 4.702 > $t_{0.025,5}$ =2.571이므로 유의수준 5%에서 H_0 기각

  ⇒ 5% 유의수준에서 중간시험점수가 기말시험점수에 영향을 준다고 할 수 있음

- p-value = 0.00533 ∴ (유의수준 5% H_0 기각)

(6) $s^2 = \frac{SSE}{n-2} = \frac{87.31535}{5} = 17.46307 \;\Rightarrow\; s = 4.1789$

- $S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{7} = 23196 - \frac{160000}{7} = 338.8571$

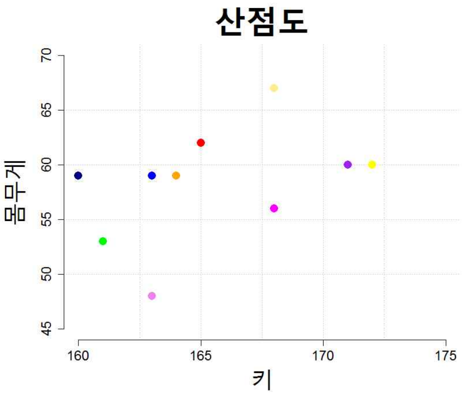
- $\bar{x} = \frac{400}{7} = 57.1429$

⇨ $x = 55$에서의 반응변수 Y(기말점수) 예측값의 표준오차: $s\sqrt{1 + \frac{1}{7} + \frac{(55-\bar{x})^2}{S_{xx}}} = 4.4938$

- 95% 신뢰구간 $56.973 \pm t_{0.025,5} \times 4.4938$=[52.7492, 61.2474]

12.3

(1) 대체로 키와 몸무게 간에는 양의 상관관계를 보인다.



(2) $r_{xy} = 0.7100$

(3) $\hat{y} = -131.7009 + 1.1522x$

(4) 가설: $H_0 : \beta_1 = 1$ vs $H_1 : \beta_1 \neq 1$

- 검정통계량: $t = \frac{1.1522-1}{0.2693} = 0.5652 < t_{0.025,18} = 2.101$이므로 5% 유의수준에서 H_0 기각 못함

⇒ 기울기가 1이라고 할 수 있음

(5) $s^2 = \frac{SSE}{n-2} = \frac{628.059}{18} = 34.8922 \;\Rightarrow\; s = 5.9070$

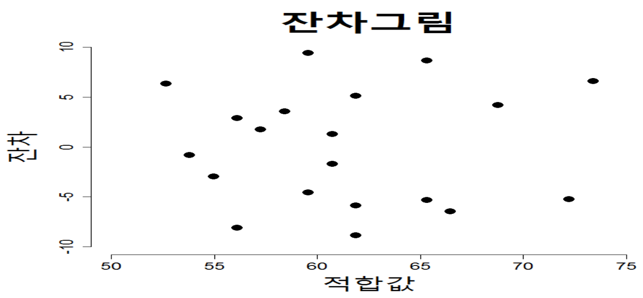
- $S_{xx} = 561941 - \frac{(3351)^2}{20} = 480.95$

- $\bar{x} = 167.55$

⇨ $x = 185$에서 y(몸무게) 예측값의 표준오차: $s\sqrt{1 + \frac{1}{20} + \frac{(185-167.55)^2}{S_{xx}}} = 7.6635$

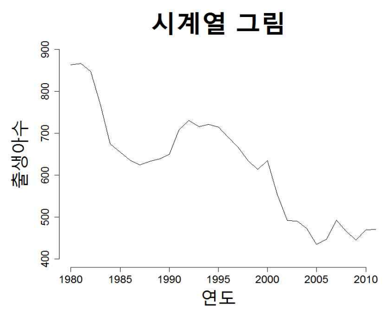
⇒ 95% 신뢰구간: $81.4561 \pm t_{0.025,18} \times 7.6635$ = [65.3553, 97.5569]

(6) 등분산성 가정을 위반하지 않은 것으로 보임

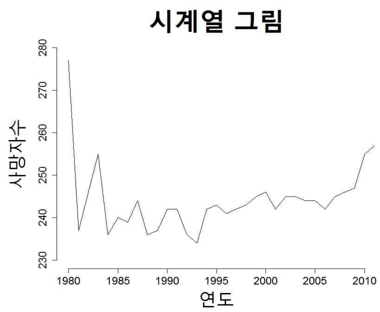


12.4

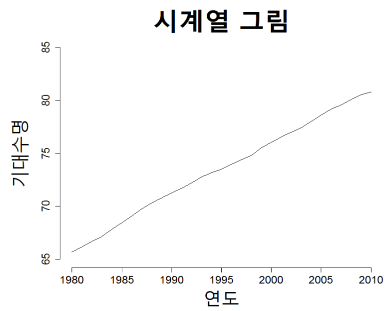
(1)



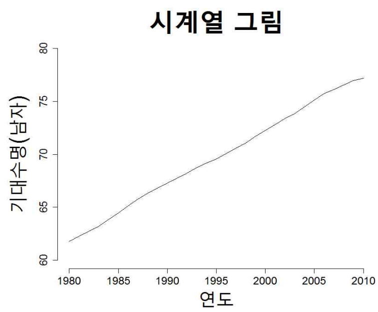
감소추세



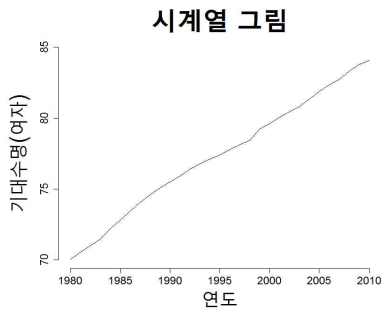
감소하다 증가 추세로 전환



증가추세



증가추세



증가추세

(2) 가설: $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

① 출생아수: $\hat{y}= 24103.3140 - 11.7667x$

- 검정통계량: $|t| = |-9.8591| > 2.042 = t_{0.025,30}$이므로 5% 유의수준에서 H_0 기각

⇒ 5% 유의수준에서 연도에 따라 출생아 수는 차이(감소)가 있다고 할 수 있음

② 사망자수: $\hat{y}= 87.3145 + 0.0786x$

- 검정통계량: $|t| = 0.5065 < 2.042 = t_{0.025,30}$이므로 5% 유의수준에서 H_0 기각 못함

⇒ 5% 유의수준에서 연도에 따라 사망자 수는 차이(감소)가 있다고 할 수 없음

③ 기대수명: $\hat{y}=-933.4614 + 0.5048x$

- 검정통계량: $|t| = 127.4310 > 2.045 = t_{0.025,29}$이므로 5% 유의수준에서 H_0 기각

⇒ 연도에 따라 기대수명은 차이(증가)가 있다고 할 수 있음

④ 기대수명(남자): $\hat{y}=-974.2178 + 0.5233x$

- 검정통계량: $|t| = 156.2161 > 2.045 = t_{0.025,29}$이므로 5% 유의수준에서 H_0 기각

⇒ 연도에 따라 남자기대수명은 차이(증가)가 있다고 할 수 있음

⑤ 기대수명(여자): $\hat{y} = -833.6767 + 0.4567x$

- 검정통계량: $|t| = 83.3794 > 2.045 = t_{0.025,29}$이므로 5% 유의수준에서 H_0 기각

⇒ 연도에 따라 여자기대수명은 차이(증가)가 있다고 할 수 있음

(3)

① 출생아수: 334.58(천 명) ⇨ [192.1728, 476.9872]

② 사망자수: 246.1451(천 명) ⇨ [227.6204, 264.6698]

③ 기대수명: 86.2346(년) ⇨ [85.8247, 86.6445]

④ 기대수명(남자): 82.8482(년) ⇨ [82.5016, 83.1948]

⑤ 기대수명(여자): 88.8573(년) ⇨ [88.2916, 89.4240]

(4) 2028년

(5) 남자의 기대수명 기울기가 큼 ⇨ 가능성 있음 ⇒ 2111년

12.5

(1) $\widehat{s.e}(\hat{\beta}_1) = \dfrac{\sqrt{MSE}}{\sqrt{S_{xx}}} = 0.3$ ⇨ $\sqrt{MSE} = 0.3\sqrt{S_{xx}} = 0.3\dfrac{10}{3} = 1$ ⇒ $MSE = 1$

(2) 가설: $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

- 검정통계량: $|t| = 1.2/0.3 = 4 > t_{0.025,18} = 2.100$이므로 5% 유의수준에서 H_0 기각

⇒ 제품인지도의 제곱이 판매량에 영향을 준다고 할 수 있음

(3) $\widehat{s.e}(\hat{\beta}_0) = \sqrt{MSE}\sqrt{\dfrac{1}{20} + \dfrac{\overline{x}^2}{S_{xx}}} = 1\sqrt{\dfrac{1}{20} + \dfrac{10^2}{100/9}} = 3.0083$

- $\hat{\beta}_0 \pm t_{0.025}(18) \times \widehat{s.e}(\hat{\beta}_0) = 20 \pm 2.100 \times 3.0083 =$[13.6826, 26.3174]

(4) $\hat{y} = 20 + 1.2x^2$ ⇨ $50 = 20 + 1.2x^2$ ⇒ $x = \sqrt{25} = 5$

12.6

(1) $L(\alpha) = \sum_{i=1}^{n}(y_i - \alpha)^2$

- $\dfrac{\partial}{\partial\alpha}L(\alpha) = -\sum_{i=1}^{n}2(y_i - \alpha) = 0$ ⇨ $\hat{\alpha} = \overline{y} = \sum y_i/n$

- $E(\hat{\alpha}) = E(\overline{Y}) = \alpha$

- $Var(\hat{\alpha}) = Var(\overline{Y}) = \dfrac{n\sigma^2}{n^2} = \dfrac{\sigma^2}{n}$

(2) $L(\beta) = \sum_{i=1}^{n}(y_i - \beta x_i)^2$

- $\dfrac{\partial}{\partial\beta}L(\beta) = \sum_{i=1}^{n}2(y_i - \beta x_i)x_i = 0$ ⇨ $\hat{\beta} = \dfrac{\sum x_i y_i}{\sum x_i^2}$

① $E(\hat{\beta}) = \dfrac{\sum E(Y_i)x_i}{\sum x_i^2} = \dfrac{\beta\sum x_i^2}{\sum x_i^2} = \beta$

- $Var(\hat{\beta}) = Var(\sum Y_i x_i / \sum x_i^2) = \dfrac{\sum x_i^2 Var(Y_i)}{(\sum x_i^2)^2} = \dfrac{\sigma^2}{\sum x_i^2}$

⇒ $\hat{\beta} \sim N\left(\beta, \sigma^2 / \sum x_i^2\right)$, $\quad \dfrac{\hat{\beta} - \beta}{\sigma / \sqrt{\sum x_i^2}} \sim N(0,1)$

② σ^2이 알려져 있지 않은 경우, $\widehat{\sigma^2} = MSE = \dfrac{SSE}{n-1}$

- $SSE = \sum (Y_i - \hat{\beta} x_i)^2 = \sum y_i^2 - \dfrac{\left\{\sum x_i y_i\right\}^2}{\sum x_i^2} = \sum y_i^2 - \hat{\beta} \sum x_i y_i$

- 검정통계량: $\dfrac{\hat{\beta}}{\sqrt{MSE} / \sqrt{\sum x_i^2}} \sim t_{n-1}$

12.8

(1) $\hat{y} =$ 11.052+0.065×Time-0.168×Quarter1+0.218×Quarter2+0.161×Quarter3

(2) 2010년 1분기 ⇨ Time=49, Quarter1=1. Quarter2=0, Quarter3=0

⇒ $\hat{y} =$ 11.052+0.065×49-0.168=14.069

(3) 2010년 2분기 ⇨ Time=50, Quarter1=0, Quarter2=1, Quarter3=0

⇒ $\hat{y} =$ 11.052+0.065×50+0.218=14.52

- 2010년 4분기 ⇨ Time=52, Quarter1~3=0

⇒ $\hat{y} =$ 11.052+0.065×52=14.432

- $\left|\hat{y}_2 - \hat{y}_4\right| = 0.088$