10.1절~10.2절
8강 연습문제풀이

14주1강 회귀분석과 상관분석1

10.1절

10.1 산포도와 상관분석

회귀분석과 상관분석

- 두 연속형 변수들 사이의 관계를 분석하는 통계적 분석방법으로 상관분석과 회귀분석이 있다.

① 상관분석

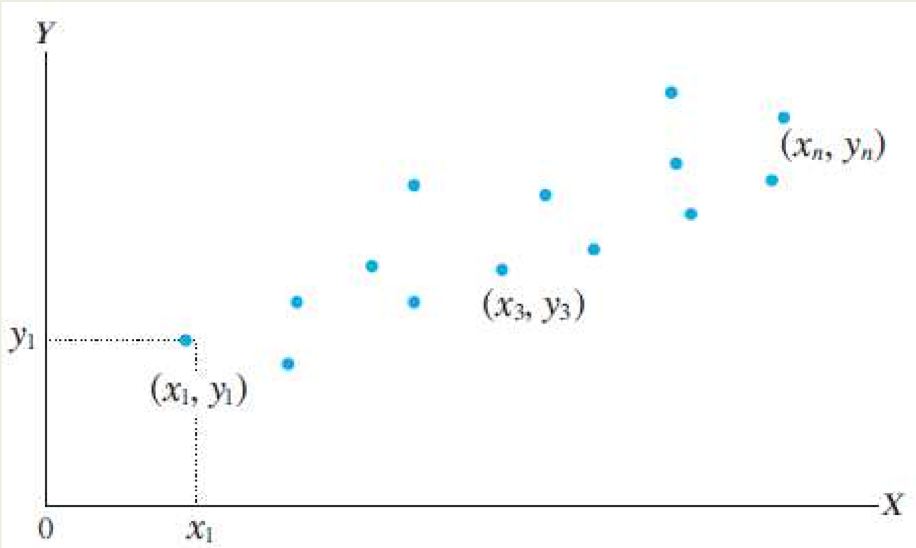
상관계수를 이용하여 두 변수들 사이의 관계가 얼마나 밀접한가를 측정하는 분석방법

② 회귀분석

한 변수를 설명변수(explanatory variable) 라 하고, 다른 변수를 반응변수(response variable) 로 지정한 후 반응변수와 설명변수 사이의 선형식을 구하고 설명변수가 얼마만큼 반응변수를 잘 예측할 수 있는가를 분석하는 통계적 분석방법

산포도

- 서로 연관이 있는 자료가 쌍으로 관측되었을 때, 두 자료 사이의 관계를 분석하고자 하는 경우 가장 기초적인 분석방법은 두 변수 사이의 산포도(scatter plot)를 그리는 것이다.
- 아래의 그림은 각 쌍의 자료 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$을 X와 Y좌표 상에 점으로 표현한 자료이다.



산포도와 상관분석

- 산포도를 이용하여 두 변수 사이의 관계를 분석할 수 있는데, 두 변수 사이의 관계의 밀접도를 측정하는 통계량이 **상관계수(correlation coeffcient)**이다.

- 상관계수 $(r) = \dfrac{s_{XY}}{s_X \cdot s_Y}$

- 여기에서,

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$S_x = \sqrt{S_x{}^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$S_y = \sqrt{S_y{}^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

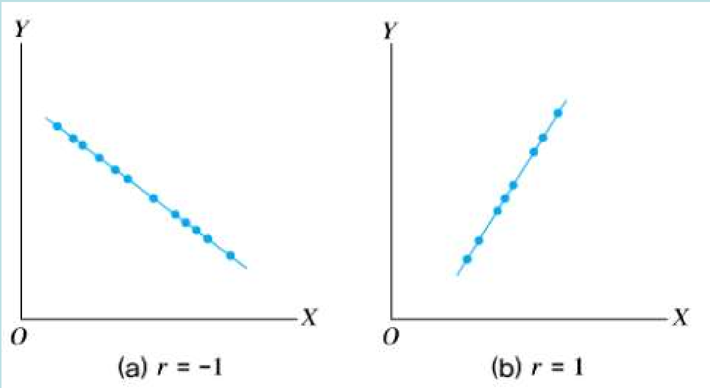
- 상관계수는 $-1 \leq r \leq 1$ 범위의 숫자로 나타나며, 산포도의 중심축에서 흩어진 정도를 측정한 계수이다.

- 상관계수 r 의 부호는 중심축의 방향이 양이면 +, 음이면 –이다.

- 또한 상관계수의 절대값(| r |)의 크기는 산포도가 중심축에 가까이 분포되어 있을수록 1에 가깝다.

산포도와 상관계수

① 완전 선형인 경우



왼쪽 그림과 같이 (a), (b)와 같이 (X, Y)의 관계가 완전선형 $Y = a + bX\ (b \neq 0)$인 경우에는 상관계수가 $r = 1(b > 0)$ 또는 $r = -1(b < 0)$으로 나타난다.

② 어느 정도 상관관계가 있는 경우
산포도의 분포 폭이 중심축으로부터 커질수록 $|r|$ 값의 크기는 0에 가까워진다.

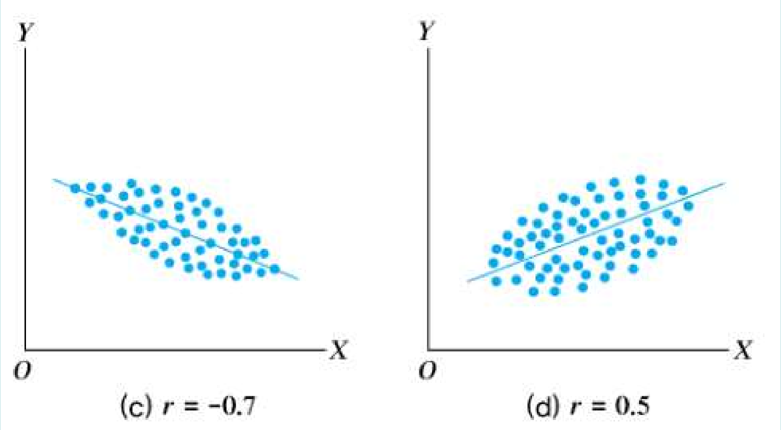


그림 (c) 와 (d) 를 비교할 때
$|r| = 0.5$인 경우가
$|r| = 0.7$인 경우보다 분포의 폭이 더 큰 것을 알 수 있다.

③ 상관관계가 없는 경우

$r = 0$인 경우는 두 변수 사이에 선형관계가 없음을 의미한다.

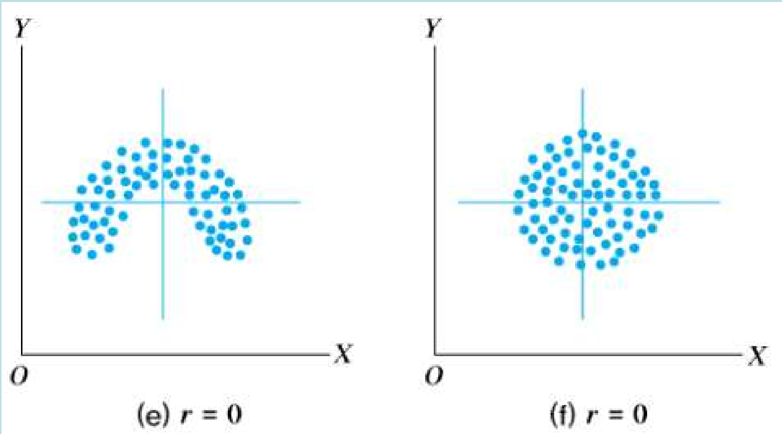


그림 (e), (f)와 같이 산포도가 하나의 특정한 중심축을 그릴 수 없는 경우와 관계식 Y=a+bX에서 b=0인 경우가 여기에 해당된다.

표 10-1 아버지, 어머니, 아들의 키(cm)

가정	아버지(X_1)	어머니(X_2)	아들(Y)
1	170	163	174
2	165	163	173
3	174	166	175
4	180	168	179
5	166	170	172
6	163	158	170
7	172	169	177
8	177	163	183
9	158	166	167
10	176	159	180
11	168	160	175
12	174	155	179
13	166	159	170
14	183	166	187

그림 10-3 아버지 키와 아들의 키 산포도

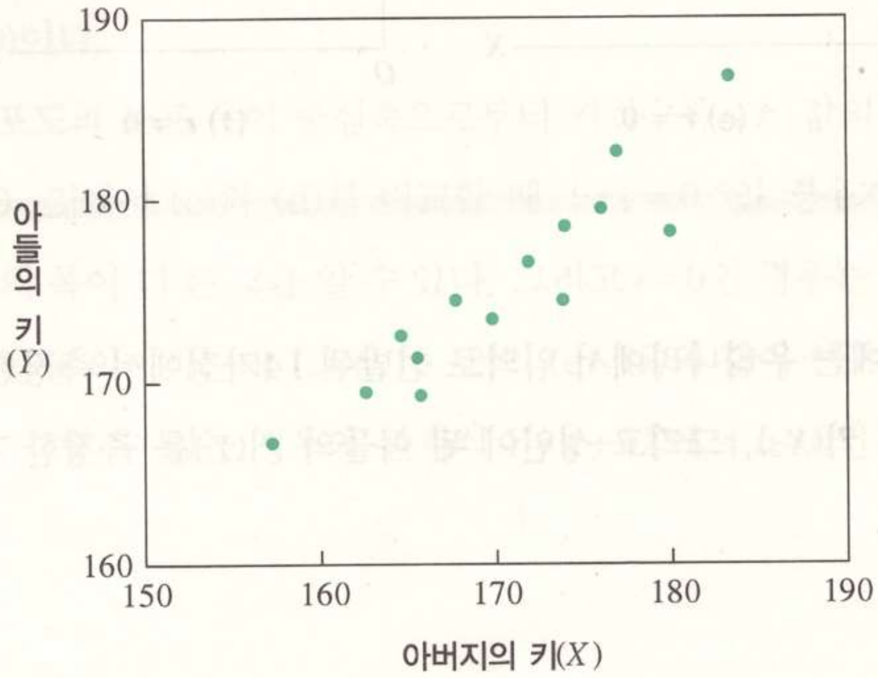
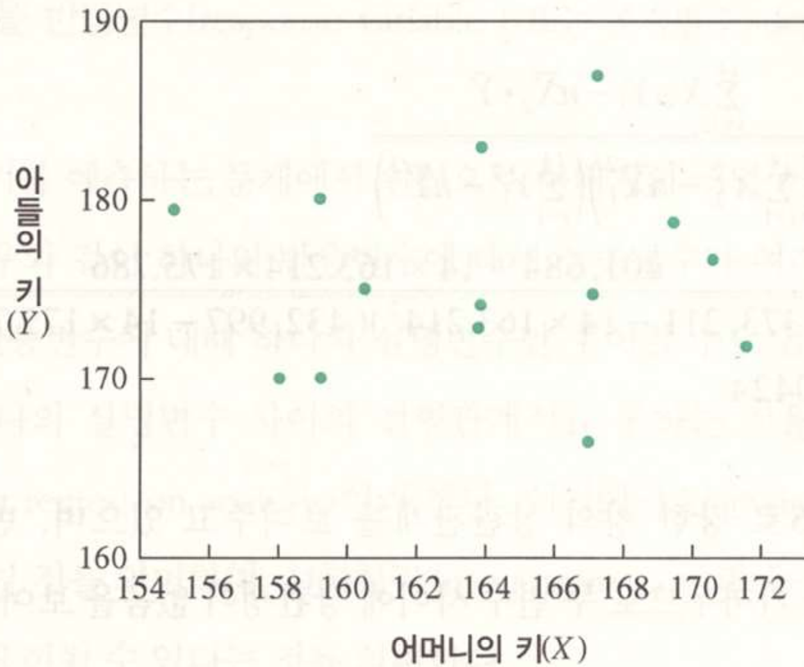


그림 10-4 어머니 키와 아들의 키 산포도



$$r_{x_1 y} = 0.9385$$
$$r_{x_2 y} = 0.0424$$

- 아버지의 키와 아들의 키는 강한 양의 상관관계를 보여주는 반면 어머니의 키와 아들의 키는 상관관계가 거의 없음을 알 수 있다.

10.2절

10.2 단순 선형 회귀분석

단순 선형 회귀분석

- 실제 자료분석에 있어서는 관측값이 쌍으로 주어졌을 때 하나의 변수값을 이용하여 다른 변수값을 예측하고자 하는 경우가 있다. 상관분석에서는 두 변수 사이의 관계에 있어서 원인과 결과의 관계는 설명할 수 없다.

- 선형회귀분석이란, 쌍으로 관찰된 **두 연속형 변수들 사이에 선형관계가 있다고 전제**한 후에 한 변수를 원인으로 하고 다른 변수를 결과로 하여 두 변수 사이의 선형식을 구하는 통계분석방법이다.

- 선형회귀분석에서 **원인의 역할을 하는 변수를 설명변수**(explanatory variable; 독립변수, independent variable)라 하고, **결과를 관측하는 변수를 반응변수**(response variable; 종속변수, dependent variable)라고 한다.

- **단순선형 회귀분석 : 설명변수가 하나이고 하나의 반응변수 사이의 선형 관계식을 구하는 것**

① 단순선형회귀모형의 조건

[조건1] X와 Y 사이의 관계는 선형식으로 표현할 수 있다.

[조건2] 잔차 는 서로 독립이며 평균 0과 분산 σ²을 갖는 정규분포를 따른다.

이를 수식으로 표현하면 다음과 같다.

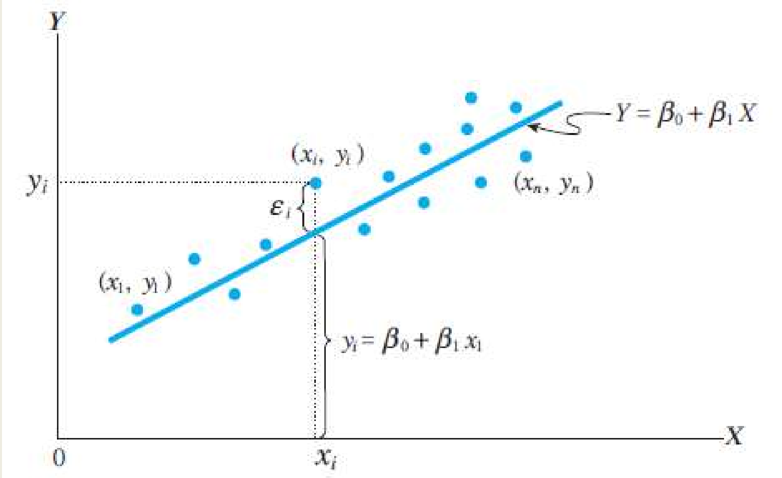
$$COV(\varepsilon_i, \varepsilon_j) = 0, i \neq j \qquad \varepsilon \sim N(0, \sigma^2)$$

② 선형회귀모형

- 단순선형회귀분석에서 설명변수를 X, 반응변수를 Y라고 할 때 회귀모형은 다음과 같이 정의된다.

$$Y = \beta_0 + \beta_1 X + \varepsilon \ , \ \varepsilon \sim N(0, \sigma^2)$$

- Y값은 X의 선형식 $\beta_0 + \beta_1 X$와 잔차 ε의 결합된 형태로 표현되었다. 이를 관측값의 쌍 $(x_1\ y_1), (x_2\ y_2), \dots, (x_n\ y_n)$에 대한 X, Y 산포도를 이용하여 설명하면 다음과 같다.
- X, Y 산포도가 다음과 같이 주어졌을 때 각 Y_i값은 식 $Y_i = \beta_0 + \beta_1 X_i$에 의한 값 $\beta_0 + \beta_1 X_i$와 ε의 합으로 표현됨을 알 수 있다.

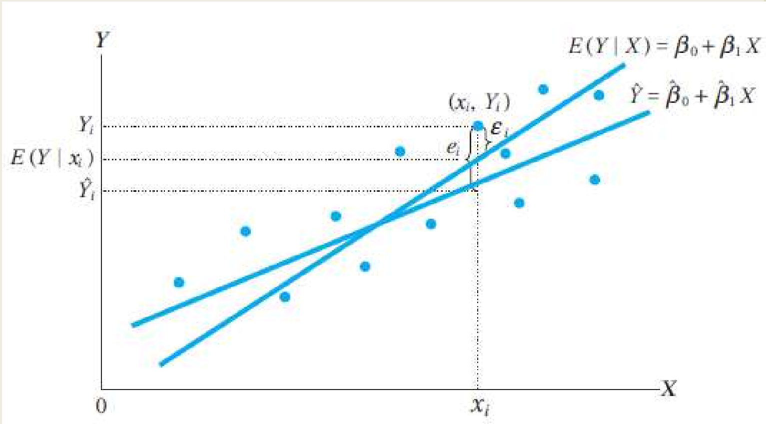


회귀계수의 추정-최소제곱법(least squqre method)

- 회귀분석의 기본은 선형식 $\beta_0 + \beta_1 X$를 구하는 것으로 이는 계수 β_0, β_1을 추정하는 문제이다. 이와 같은 회귀모형에서 계수 β_0, β_1은 대상집단 모두 (모집단)를 대상으로 구하는 계수이다. 그러나 실제 자료분석에서는 대상집단의 일부인 n개의 관측쌍, $(x_1\ y_1), (x_2\ y_2), \dots,\ (x_n\ y_n)$을 이용하여 β_0, β_1의 추정값인 $\hat{\beta}_0$, $\hat{\beta}_1$를 구한다. 추정된 회귀식은 다음과 같다.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

- 회귀모형에서 회귀식은 측정값인 y_i와 회귀식의 값 $\hat{y}_i$의 거리가 가까울 수록 좋다고 할 수 있다. 여기에서 거리는 $e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 X)$와 같이 잔차로 측정되며, 회귀식은 모든 i=1, 2, …, n에 대하여 e_i가 작게 되도록 하는 $\hat{\beta}_0$와 $\hat{\beta}_1$를 구한다. 추정된 회귀식에서 잔차 e는 오차의 추정량이고 모형식과 추정된 회귀식을 그림으로 표현하면 다음과 같다.



$$Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 X)^2$$

① 최소제곱법

- 잔차가 최소가 되도록 다음과 같이 식을 세워 구한다.

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- 회귀계수를 구할 때, 위의 식을 최소화하는 값을 구한다. 이와 같은 원리를 최소제곱법 (least square method)이라고 한다. 추정량을 구하는 과정은 방정식의 극소값을 구하는 방법으로 $Q(\hat{\beta}_0, \hat{\beta}_1)$를 $\hat{\beta}_0$, $\hat{\beta}_1$에 대하여 미분을 취하여 0으로 놓은 후에 $\hat{\beta}_0$와 $\hat{\beta}_1$에 대하여 해를 구한다.

$$\frac{d}{d\hat{\beta}_0}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2=2\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)(-1)=0$$

$$\rightarrow\sum_{i=1}^{n}y_i-n\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{n}x_i=0\rightarrow\bar{y}-\hat{\beta}_0-\hat{\beta}_1\bar{x}=0\rightarrow\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$$

$$\frac{d}{d\hat{\beta}_1}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2=2\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)\cdot(-x_i)=0$$

$$\rightarrow\sum_{i=1}^{n}x_iy_i-\hat{\beta}_0\sum_{i=1}^{n}x_i-\hat{\beta}_1\sum_{i=1}^{n}x_i^2=0\longrightarrow\sum_{i=1}^{n}x_iy_i=\hat{\beta}_0\sum_{i=1}^{n}x_i+\hat{\beta}_1\sum_{i=1}^{n}x_i^2$$

$$\sum_{i=1}^{n} x_i y_i = (\bar{y} - \widehat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i + \widehat{\beta}_1 \sum_{i=1}^{n} {x_i}^2$$

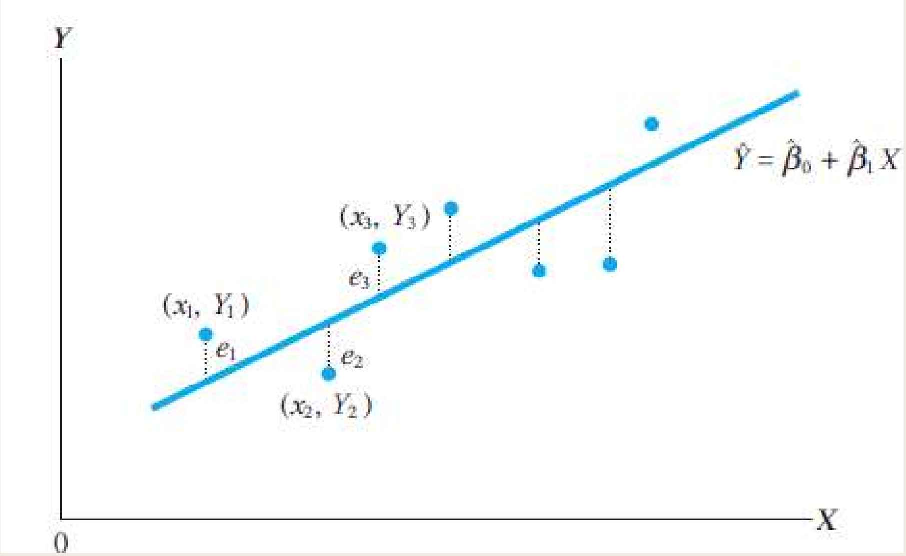
$$\widehat{\beta}_1 (-\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} {x_i}^2) = \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} x_i + n\bar{x}\bar{x}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$



8장 연습문제풀이

5. 미지의 모평균 μ를 갖는 분포로부터 n개의 표본을 추출했다고 할 때, 다음 각 경우에 대한 검정을 실시하라.

(d) $H_0 : \mu = 120, H_1 : \mu \neq 120, n = 18, \bar{X} = 115, S = 13.3, \alpha = 0.05$

풀이) $T = \frac{115-120}{\frac{13.3}{\sqrt{18}}} = \frac{-5}{3.135} = -1.59$

$n = 18$이므로 자유도 17인 t분포를 이용하고, $P_r(|t| > 2.110) = 0.05$이며

$|T| = 1.59 < 2.11$이므로 H_0를 채택한다.

(e) $H_0 : \mu = 48, H_1 : \mu > 48, n = 18, \bar{X} = 52, S = 12.7, \alpha = 0.05$

풀이) $T = \frac{52-}{\frac{12.7}{\sqrt{10}}} = \frac{4}{4.016} = 0.99$

$n = 10$이므로 자유도 9인 t분포를 이용하고, $P_r(t > 1.833) = 0.05$이며

$T = 0.99 < 1.833$이므로 H_0를 채택한다.

6. 한 집단의 49개 관측값으로 부터 다음과 같은 통계량을 구했다고 할 때,

$$\sum_{i=1}^{49} X_i = 50.3, \sum_{i=1}^{49} X_i^2 = 68$$

힌트)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{49} \times 50.3 = 1.0265$$

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] = \frac{1}{48}[68 - 49 \times 1.0265^2]$$

$$= \frac{1}{48}[69 - 51.63] = \frac{1}{48} \times 16.37 = 0.341$$

$$S = \sqrt{0.341} = 0.584$$

(a) H_0 : $\mu = 1.18$에 대한 H_1 : $\mu < 1.18$의 검정을 실시하라. ($\alpha = 0.05$를 이용할 것)

풀이) $T = \frac{1.0265 - 1.18}{\frac{0.584}{\sqrt{49}}} = \frac{-0.1535}{0.0834} = -1.84$

$n = 49$이므로 정규분포를 이용하여 H_1 : $\mu < 1.18$은 왼쪽 단측검정으로

$P_r(Z < -1.645) = 0.05$이다. $T = -1.84 < -1.645$이므로 H_0를 기각한다.

(b) H_0 : $\mu = 1.18$에 대한 H_1 : $\mu \neq 1.18$의 검정을 실시하라. ($\alpha = 0.05$를 이용할 것)

풀이) (a)에서 구한 $T = -1.84$를 이용하며 H_1 : $\mu \neq 1.18$은 양측검정으로

$P_r(|Z| > 1.96) = 0.05$이다. $|T| = 1.84 < 1.96$이므로 H_0를 채택한다.

9. 한 밀가루 제조회사에서는 $5kg$들이 자루에 밀가루를 자동으로 채우는 기계를 도입했다. 이 기계가 실제로 자루의 $5kg$씩 채우는가를 알아보기 위해 이 기계를 이용해 채운 25개 자루의 실제 양을 조사한 결과, 평균과 분산이 각각 $\bar{X} = 5.3kg, S^2 = 1.25kg$이었다. 이 기계가 자루에 밀가루를 $5kg$씩 채운다고 생각할 수 있는가에 대해 검정을 실시하라. ($\alpha = 0.05$)

풀이) $H_0 : \mu = 5, H_1 : \mu \neq 5, \bar{X} = 5.3, S^2 = 1.25, n = 25$이므로

$T = \frac{5.3-5}{\sqrt{\frac{1.25}{25}}} = \frac{0.3}{0.2236} = 1.3417$이다.

자유도 24인 t분포에서 $P_r(|t| > 2.064) = 0.05$이며 $|T| = 1.3417 < 2.064$이므로

H_0를 채택한다. 이 기계가 밀가루를 $5kg$씩 채운다고 할 수 있다.

11. 10명의 학생을 랜덤 추출해 학기말시험 일주일 동안의 공부시간을 알아보고자 한다. 학기말시험에 할애한 시간들이 다음과 같을 때 각 물음에 답하라.

28	**57**	**42**	**35**	**61**	**39**	**55**	**46**	**49**	**38**

분포의 경우 정규분포를 따른다고 가정하자.

(a) 평균과 분산을 구하라.

풀이) $\bar{x} = 45, s^2 = 111.11$

(b) 평균 40시간 이상 공부하는지 유의수준 α=0.05에서 검정을 실시하라.

풀이) $t = 1.5,\ \ H_0$기각할 수 없음

12. 두 모집단의 동일성에 대한 검정$H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$을 실시하기 위해 두 집단으로부터 각각 크기가 n_1과 n_2인 표본을 추출해 다음과 같은 통계량을 구했다. 위의 H_0과 H_1에 대한 검정을 $\alpha = 0.05$하에서 실시하라.

통계량 \ 집단	I	II
n	30	20
$\bar{X}$	23	26
S^2	7.5	6.3

(a) 조건 : $\sigma_1^2 = \sigma_2^2 = \sigma^2$

풀이)

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{29\times7.5+19\times6.3}{30+20-2} = \frac{337.2}{48} = 7.025$$

$$T = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{23-26}{\sqrt{7.025\left(\frac{1}{30}+\frac{1}{20}\right)}} = \frac{-3}{0.765} = -3.922$$

$n_1 + n_2 = 50$이므로 정규분포를 이용하며, $P_r(|Z| > 1.96) = 0.05$이다.

$|T| = 3.922 > 1.96$이므로 H_0를 기각한다.

12. 두 모집단의 동일성에 대한 검정H_0 : $\mu_1 = \mu_2, H_1$: $\mu_1 \neq \mu_2$을 실시하기 위해 두 집단으로부터 각각 크기가 n_1과 n_2인 표본을 추출해 다음과 같은 통계량을 구했다. 위의 H_0과 H_1에 대한 검정을 $\alpha = 0.05$하에서 실시하라.

통계량 \ 집단	I	II
n	50	50
$\bar{X}$	48	53
S^2	17.3	22.8

(b) 두 모분산의 동일성 조건 없음

풀이)

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{48-5}{\sqrt{\frac{17.3}{50} + \frac{22.8}{50}}} = \frac{-5}{0.8955} = -5.583$$

$n_1 + n_2 = 100$이므로 정규분포를 이용하며, $P_r(|Z| > 1.96) = 0.05$이다.

$|T| = 5.583 > 1.96$이므로 H_0를 기각한다.

($\sigma_1^2 = \sigma_2^2$의 조건이 없는 경우에는 이와 다른 방법을 이용하는 경우도 있다.)

12. 두 모집단의 동일성에 대한 검정 $H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$을 실시하기 위해 두 집단으로부터 각각 크기가 n_1과 n_2인 표본을 추출해 다음과 같은 통계량을 구했다. 위의 H_0과 H_1에 대한 검정을 $\alpha = 0.05$하에서 실시하라.

통계량 \ 집단	I	II
n	12	13
$\bar{X}$	17.5	14.3
S^2	9.5	8.7

(c) 조건 : $\sigma_1^2 = \sigma_2^2 = \sigma^2$

풀이)

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{11\times9.5+12\times8.7}{12+13-} = \frac{208.9}{23} = 9.0826$$

$$T = \frac{17.5-14.3}{\sqrt{9.0826\left(\frac{1}{12}+\frac{1}{13}\right)}} = \frac{3.2}{1.2064} = 2.6525$$

$n_1 + n_2 = 25$이므로 자유도 23인 t분포를 이용하며, $P_r(|Z| > 2.069) = 0.05$이다.
$|T| = 2.6525 > 2.069$이므로 H_0를 기각한다.

14. 두 가지 서로 다른 암기훈련 방법을 비교하기 위해 초등학교 1학년생 100명을 임의로 각 50명씩 두 집단으로 나누어 두 가지 서로 다른 암기훈련 방법을 이용해 훈련을 시킨 후 시험을 치른 결과 다음과 같은 통계량을 구했다.

$$\overline{\boldsymbol{X}} = \mathbf{74}, \boldsymbol{S}_{\mathbf{1}} = \mathbf{9}$$
$$\overline{\boldsymbol{X}} = \mathbf{71}, \boldsymbol{S}_{\mathbf{2}} = \mathbf{10}$$

두 가지 훈련방법의 효과에 차이가 있는가를 유의수준 5% 하에서 검정하라.

풀이) $H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$

같은 조건의 학생들을 두 집단으로 나누었기 때문에

두 집단의 모분산이 동일하다고 볼 수 있고 $n_1 = n_2 = 50$이므로

$S_p^2 = \frac{1}{2}(S_1^2 + S_2^2) = \frac{1}{2}(81 + 100) = 90.5$이다.

$$T = \frac{74-71}{\sqrt{90.5\left(\frac{1}{50}+\frac{1}{50}\right)}} = \frac{3}{1.9026} = 1.577$$

$n_1 + n_2 = 100$이므로 정규분포를 이용하며, $P_r(|Z| > 1.96) = 0.05$이다.

$|T| = 1.577 < 1.96$이므로 H_0를 채택한다.

즉, 두 훈련방법의 효과에 차이가 없다고 할 수 있다.

15. 두 가지 서로 다른 체중조절 방법을 비교하기 위해 80명을 랜덤하게 각 40명씩 두 집단으로 나눈 후 두 가지 사로 다른 체중조절 방법을 이용해 얻은 체중조절 효과가 다음과 같이 관측되었다.

두 체중조절 방법의 효과가 동일한가를 유의수준 5% 하에서 검정하라.

통계량 \ 체중조절방법	I	II
n	40	40
$\bar{X}$	$10kg$	$8kg$
S^2	4.3	5.7

풀이) $H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$

동일 조건의 사람들을 두 집단으로 나누었으므로 $\sigma_1^2 = \sigma_2^2$을 전제로 함.

- $n_1 = n_2 = 40$이므로 $S_p^2 = \frac{1}{2}(4.3 + 5.7) = 5$이다.

$$T = \frac{10-8}{\sqrt{5\left(\frac{1}{40}+\frac{1}{40}\right)}} = \frac{2}{0.5} = 4.0$$

$n_1 + n_2 = 80$이므로 정규분포를 이용하며, $|Z| = 4.0 > 1.96$이므로 H_0를 기각한다.

즉, 두 체중조절 방법의 효과가 동일하지 않다고 할 수 있다.

16. 커피자동판매기 업자는 커피자동판매기의 설치장소에 따라 판매량의 차이가 있다고 생각한다. 특정한 두 지역에 설치된 두 대의 자동판매기에서 14일 동안 관측한 평균판매량(잔)과 표준편차가 각각 다음과 같다고 할 때, 두 지역에 설치된 자동판매량의 판매량이 서로 다르다고 할 수 있는지에 대해 유의수준 5% 하에서 검정을 실시하라($\sigma_1^2 = \sigma_2^2$을 가정할 것).

설치 위치 / 통계량	I	II
n	14	14
$\bar{X}$	32.5	28.5
S^2	5	4.5

풀이) $H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$

$n_1 = n_2 = 14$이므로

$S_p^2 = \frac{1}{2}(5.0 + 4.5) = 4.75$이다.

$$T = \frac{32.5-28.5}{\sqrt{4.75\left(\frac{1}{14}+\frac{1}{14}\right)}} = \frac{4}{0.824} = 4.854$$

$n_1 + n_2 = 28$이므로 자유도 26인 t분포로부터 $P_r(|t| > 2.056) = 0.05$이다.

$|T| = 4.854 > 2.056$이므로 H_0를 기각한다.

즉, 두 지역 자동판매기의 판매량이 서로 다르다고 볼 수 있다.

21. 한 다방은 요일별로 판매량이 다르다고 한다. 이 다방에서 판매되는 두가지 음료(커피와 홍차)의 주문량이 서로 다른가를 비교하기 위해 어느 일주일 간의 두 음료에 대한 주문 수량을 조사한 결과가 다음과 같다.

요일	커피	홍차
월	120	91
화	74	43
수	134	169
목	95	112
금	337	238
토	294	221
일	379	325

위의 자료를 이용해 두 음료에 대한 주문량이 서로 다른가에 대해 유의수준 5%하에서 검정을 실시하라.

풀이) 짝진 표본의 모평균에 대한 검정문제로 각 쌍에서 두 값의 차이를 D라 할 때,

H_0 : $\mu_D = 0$, H_1 : $\mu_D \neq 0$이다.

$\bar{d} = 33.4286$, $S_D = 47.504$,

$T = \frac{33.4286}{\frac{47.504}{\sqrt{7}}} = 1.86$이며, T는 자유도가 6인 t분포를 따른다.

자유도 6인 t분포에서 $P_r(|t| > 2.447) = 0.05$이며,

$|T| = 1.86 < 2.447$이므로 H_0를 채택한다.

즉, 두 음료의 주문량이 서로 다르다고 볼 수 없다.

끝~~ ♥♥