보이는 통계학 with 파이썬

연습문제 공개용 정답 이용 안내

- 본 문제 풀이의 저작권은 **손원**과 **한빛아카데미(주)**에 있습니다.
- 이 자료를 무단으로 전제하거나 배포할 경우 저작권법 136조에 의거하여 최고 5년 이하의 징역 또는 5천만원 이하의 벌금에 처할 수 있고 이를 병과(倂科)할 수도 있습니다.

1장 연습문제 공개용 정답

2.

(a) 모집단의 일부 조사단위만 표본으로 추출하여 관찰하기 때문에 발생하는 오차를 표본오차라 한다.

(b) 표본으로 추출된 일부 조사단위가 응답할 수 없을 때 발생하는 오차를 무응답오차라 한다. 포함오차는 모집단에 일부 단위가 누락되거나 필요없는 조사단위가 포함되어 있어 필요한 조사단위가 표본에 포함되지 못하는 경우가 발생한다.

(c) 민감한 주제를 다루는 조사나 기억하기 어려운 내용을 질문할 때 발생하는 오차를 측정오차라 한다.

4.

(a) 층화확률추출법. 설명 생략

(b) 집락추출법. 설명 생략

6.

(a) 참

(b) 거짓 : 실험오차는 조사대상이 되는 개체들의 서로 다른 특성에 의해 발생할 수 있다.

(c) 거짓 : 반복은 동일한 조건 아래에서 여러 개체를 되풀이하여 관찰하는 것을 말한다.

7.

(a) 사료의 효과를 확인하기 위해 각 사료를 50마리의 소에게 먹였다. 이렇게 각 사료를 여러 마리의 소에게 먹임으로써 실험오차를 파악할 수 있으며 추정된 실험오차를 이용하여 실험 결과를 평가할 수 있다.

(b) 문제에서 제시된 것과 같이 소를 랜덤으로 추출해야 할 뿐만 아니라 각 소에게 어떤 사료를 먹일지(실험의 배치)도 랜덤으로 결정해야 한다.

2장 연습문제 공개용 정답

1.

(a) $\bar{x}_{n+1} = \frac{n}{n+1}\bar{x}_n + \frac{1}{n+1}x_{n+1}$

(b) $\bar{x}_{n+k} = \frac{n}{n+k}\bar{x}_n + \frac{k}{n+k}\left(\frac{1}{k}\sum_{i=n+1}^{n+k} x_i\right)$

3.

(a) 15

(b) 25

4.

(a) 171.85

(b) 표본왜도는 -0.2306이다. 왼쪽 꼬리가 긴 분포일 것으로 예상할 수 있다.

(c) 표본첨도는 -0.6772이다. 정규분포에 비해 꼬리가 얇은 분포일 것으로 예상할 수 있다.

6.

(a) $\bar{x} = 9,\ \bar{y} = 15$

(b) $s_x^2 = 30,\ s_y^2 = 20$

(c) -10

(d) 표본상관계수는 $-\frac{1}{\sqrt{6}}$이다. 문제 풀이 시간이 짧을수록 정답을 많이 맞히는 경향이 있지만 직선관계가 강하지 않다.

9.

상자그림의 상자의 벽은 1사분위수, 3사분위수, 상자 가운데의 선은 2사분위수이고 가로선의 왼쪽 수염 끝은 1사분위수에서 사분위수범위의 1.5배 범위 내의 관측값 중 가장 작은 값, 오른쪽 수염 끝은 3사분위수에서 사분위수범위의 1.5배 범위 내의 관측값 중 가장 큰 값이다. 오른쪽 수염 끝 밖에는 세 개의 이상점이 있다. 주어진 그림은 1사분위수와 2사분위수가 가깝고 왼쪽 수염보다 오른쪽 수염의 길이가 길다. 따라서 값이 작은 쪽의 관측값 밀도가 상대적으로 높은 것을 알 수 있다.

10.

산점도에서 확인할 수 있는 바로는 x축의 변수 값이 증가함에 따라 y축의 변수 값도 증가한다. 두 변수 사이의 관계는 단순한 직선이 아닌 곡선 관계로 볼 수도 있다. 또, x축 변수 값이 증가함에 따라 y축 변수의 분산도 커지는 경향이 있다.

3장 연습문제 공개용 정답

2.

(a) 0

(b) $\frac{1}{4}$

3.

(a) $\mathrm{P}(E)=\frac{1}{2},\ \mathrm{P}(F)=\frac{1}{2},\ \mathrm{P}(G)=\frac{1}{2}$

(b) $\mathrm{P}(E|F)=\frac{1}{2},\ \mathrm{P}(E|G)=\frac{1}{2},\ \mathrm{P}(F|G)=\frac{1}{2}$

(c) 세 사건은 쌍으로 독립이다.

(d) 세 사건은 독립이 아니다.

5.

(a) 0.8

(b) 0.16

(c) 0.2

(d) 0.84

7.

(a) $c=1$

(b) $F(x)=\begin{cases}0 & x<0\\ \frac{x^2}{2} & 0\le x\le 1\\ 1-\frac{1}{2}(x-2)^2 & 1<x\le 2\\ 1 & 2<x\end{cases}$

(c) $\mathrm{E}(X)=1$

(d) $\mathrm{E}(X^2)=\frac{7}{6}$

(e) $\mathrm{Var}(X)=\frac{1}{6}$

9.

(a) $f(x)=\begin{cases}0.5-0.1x & x=1,2,3,4\\ 0 & \text{이외의 경우}\end{cases}$

(b)

x	$x<1$	$1\le x<2$	$2\le x<3$	$3\le x<4$	$4\le x$
$F(x)$	0	0.4	0.7	0.9	1

(c) $\mathrm{E}(X)=2$

(d) $\mathrm{E}(X^2)=5$

(e) $\mathrm{Var}(X) = 1$

(f) $\mathrm{Var}(3X-1) = 9$

11.

(a) $p_0 = 0.4,\ p_1 = 0.3,\ p_2 = 0.3,\ q_0 = 0.3,\ q_2 = 0.3,\ q_1 = 0.4$

(b) $\mathrm{E}(X) = 0.9,\ \mathrm{E}(Y) = 1.0$

(c) $\mathrm{E}(XY) = 0.9$

13.

(a) $\mathrm{E}(X) = 80,\ \mathrm{Var}(X) = 16$

(b) $\mathrm{P}(X \leq 70) = 0.006$

15.

(a) $c = 0.72$

(b) $c = 0.82$

16.

(a) $\mathrm{P}(V \geq 20) = 0.416$

(b) $v = 60.345$

17.

$$\mathrm{P}\left(\overline{X} \geq \sqrt{\frac{Q}{10}}\right) = 0.0075$$

4장 연습문제 공개용 정답

4.

(a) 10

(b) $\{1,\ 2\},\ \{1,\ 3\},\ \{1,\ 4\},\ \{1,\ 5\},\ \{2,\ 3\},\ \{2,\ 4\},\ \{2,\ 5\},\ \{3,\ 4\},\ \{3,\ 5\},\ \{4,\ 5\}$

(c)

표본	{1,2}	{1,3}	{1,4}	{1,5}	{2,3}	{2,4}	{2,5}	{3,4}	{3,5}	{4,5}
추출확률	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
표본평균	1.5	2	2.5	3	2.5	3	3.5	3.5	4	4.5

(d)

표본평균	1.5	2	2.5	3	3.5	4	4.5
확률	0.1	0.1	0.2	0.2	0.2	0.1	0.1

(e) $\mathrm{E}(\overline{X}) = 3$

(f) $\mathrm{Var}(\overline{X}) = \frac{3}{4}$

6.

(a) $\mathrm{E}(X) = 0,\ \mathrm{Var}(X) = \frac{1}{2}$

(b)

표본	{-1,-1}	{-1,0}	{-1,1}	{0,0}	{0,1}	{1,1}
추출확률	1/16	1/4	1/8	1/4	1/4	1/16

(c)

표본	{-1,-1}	{-1,0}	{-1,1}	{0,0}	{0,1}	{1,1}
추출확률	1/16	1/4	1/8	1/4	1/4	1/16
표본평균	-1	-1/2	0	0	1/2	1

(d) $\mathrm{E}(\overline{X}) = 0,\ \mathrm{Var}(\overline{X}) = \frac{1}{4}$

7.

(a) $\overline{X} \dot{\sim} N\left(2, \frac{16}{25}\right)$

(b) $\mathrm{P}(\overline{X} \geq 0) \approx 0.994$

9.

97

11.

(a) $\mathrm{E}(\overline{X}) = 1,\ \mathrm{Var}(\overline{X}) = \frac{16}{100}$

(b) $\overline{X} \dot{\sim} N(1,\ 0.4^2)$

(c) $\mathrm{P}(\overline{X} \geq 2) \approx 0.006$

(d) 16

(e) 39

5장 연습문제 공개용 정답

3.

20.5

4.

0.5

5.

(a) 비편향추정량이다.
(b) 비편향추정량이 아니다.
(c) 비편향추정량이다.
(d) 비편향추정량이 아니다.

6.

(a) $\mathrm{Var}(X_2) = \sigma^2$

(b) $\mathrm{Var}(X_1 + X_2) = 2\sigma^2$

(c) $\mathrm{Var}\left(\frac{X_1 + X_3}{4} + \frac{X_2}{2}\right) = \frac{3}{8}\sigma^2$

(d) $\mathrm{Var}(X_1 - X_2) = 2\sigma^2$

7.

(a) $\mathrm{MSE}(X_2) = \sigma^2$

(b) $\mathrm{MSE}(X_1 + X_2) = \mu^2 + 2\sigma^2$

(c) $\mathrm{MSE}\left(\frac{X_1 + X_3}{4} + \frac{X_2}{2}\right) = \frac{3}{8}\sigma^2$

(d) $\mathrm{MSE}(X_1 - X_2) = \mu^2 + 2\sigma^2$

8.

(a) 36

(b) $(33.648, 38.352)$

(c) $(33.11, 38.89)$

(d) $(33.256, 38.744)$

12.

(a) $(0.52,\ 0.68)$

(b) 1112

13.

$(-0.026, 0.626)$

14.

(a) $(0.110, 1.090)$
(b) $(0.012, 1.188)$

15.

$(-0.029, 0.229)$

6장 연습문제 공개용 정답

5.

(a) 귀무가설을 기각하지 못한다. 즉, 스마트워치가 정확하게 거리를 측정한다는 회사의 주장을 기각할 증거가 충분하지 않다.
(b) 거리를 정확하게 측정한다는 귀무가설을 기각할 뚜렷한 증거가 없다.

6.

(a) 귀무가설을 기각하지 못한다. 즉, 수질이 우수하다는 주장을 기각할 만한 뚜렷한 증거가 없다.
(b) 귀무가설을 기각한다. 즉, 수질이 우수하다는 주장을 기각할 만한 증거가 있다.

8.

$$\frac{5}{z_\alpha}$$

9.

유의수준 1%에서 관상동맥환자들의 맥압이 45라는 귀무가설을 기각한다.

10.

유의수준 1%에서 관상동맥환자들의 맥압이 45라는 귀무가설을 기각하고 맥압이 45보다 높다는 대립가설을 채택한다.

11.

유의수준 1%에서 사용된 단어에 따라 두 집단의 속도에 대한 평가가 달라질 수 있다는 대립가설을 채택한다.

7장 연습문제 공개용 정답

1.

(a) (참)
(b) (거짓)
(c) (거짓)
(d) (거짓)

3.

(a) $s_{xx} = 60$
(b) $s_{xy} = 90$
(c) $\widehat{\beta} = \frac{3}{2}$
(d) $\widehat{\alpha} = 0$

5.

(a) 0.6
(b) 19.5
(c) 0.002

6.

(a) $\widehat{\alpha} = -17.579,\ \widehat{\beta} = 3.932$
(b) 총제곱합 32539.0 = (회귀제곱합 21185.5) + (오차제곱합 11353.5)
(c)

요인	제곱합	자유도	평균제곱(제곱합/자유도)	F-값	p-값(유의확률)
회귀	21185.5	1	21185.5	89.6	0.00000
잔차	11353.5	48	236.5		
합	32539.0	49			

(d) 대립가설 $H_1: \beta \neq 0$을 채택한다.

8.

(a) 총제곱합 162, 회귀제곱합 101, 오차제곱합 61
(b)

요인	제곱합	자유도	평균제곱(제곱합/자유도)	F-값	p-값(유의확률)
회귀	101	1	101.0	43.9	0.0000001
잔차	61	26	2.3		
합	162	27			

10.

(a) $r^2 = 0.7176$

(b) $\hat{\beta} = 0.0412$

11.

(a) (참)

(b) (참)

(c) (참)

(d) (거짓)

8장 연습문제 공개용 정답

1.

(a) 범주 1로 분류한다.
(b) 범주 0으로 분류한다.

2.

②

4.

■ 꽃잎 너비 4.75cm 기준 : 두 노드의 개체 수를 가중치로 가중평균한 지니 지수는 0.119
■ 꽃잎 너비 4.85cm 기준 : 두 노드의 개체 수를 가중치로 가중평균한 지니 지수는 0.110
■ 꽃잎 너비 4.95cm 기준 : 두 노드의 개체 수를 가중치로 가중평균한 지니 지수는 0.086
■ 꽃잎 너비 5.05cm 기준 : 두 노드의 개체 수를 가중치로 가중평균한 지니 지수는 0.099
■ 꽃잎 너비 5.15cm 기준 : 두 노드의 개체 수를 가중치로 가중평균한 지니 지수는 0.105
적절한 분할 기준은 꽃잎 길이 4.95cm이다.

5.

(a) 정확도 0.95, 민감도 0.625, 특이도 0.978, 정밀도 0.714, F1 스코어 0.667

(b) 표 인공지능 분류모형의 MRI 이미지 분류 결과

	목표변수		
예측	악성종양	정상	계
악성종양	60	24	84
정상	20	896	916
계	80	920	1000

(c) 정확도 0.956, 민감도 0.75, 특이도 0.974, 정밀도 0.714, F1 스코어 0.732
특이도는 조금 낮아졌지만 정확도와 민감도, F1 스코어는 개선되었으므로 분류모형이 개선되었다고 볼 수 있다.

6.

(a) (거짓)
(b) (거짓)
(c) (거짓)

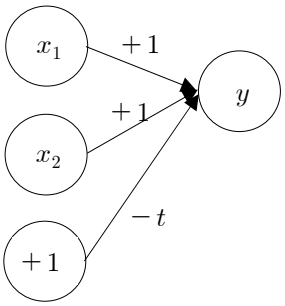
8.

(a) S-T, C-H, E-A, M
(b) (생략)

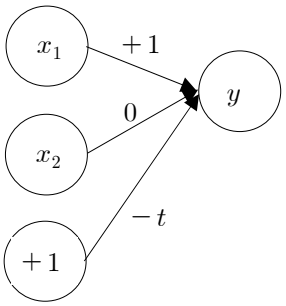
9장 연습문제 공개용 정답

1.

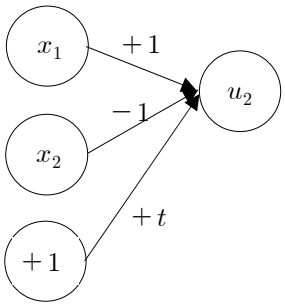
(a), **(d)**, **(e)**는 다음과 같은 퍼셉트론을 이용하여 경계를 설정할 수 있다.(아래에 주어진 풀이는 경계선에 대한 하나의 예이므로 기울기와 절편을 다르게 설정할 수도 있다.)



(a) $x_1 + x_2 \geq t$
(단, $1 < t < 2$)



(d) $x_1 \geq t$
(단, $0 < t < 0.5$)



(e) $x_1 - x_2 \geq -t$
(단, $0 < t < 1$)

(b)와 **(c)**는 퍼셉트론을 이용하여 분류할 수 없다.

3.

(a) 0

(b) 0.9855

(c) 교차 엔트로피 함수에 대해 다음과 같은 식이 성립한다.

$$\sum_{i=1}^{n+1} y_i \log p_i + \sum_{i=1}^{n+1} (1-y_i) \log (1-p_i)$$
$$= \left\{ \sum_{i=1}^{n} y_i \log p_i + \sum_{i=1}^{n} (1-y_i) \log (1-p_i) \right\} + \{ y_i \log p_i + (1-y_i) \log (1-p_i) \}$$

이 식에서 y_i는 0 또는 1이고 $0 \leq p_i \leq 1$이므로 $y_i \log p_i + (1-y_i) \log (1-p_i) \leq 0$이다.
따라서 주어진 식이 성립한다.

5.

(a) $L'(\beta | x_1, x_2, \cdots, x_n) = -2 \sum_{i=1}^{n} x_i y_i + 2\beta \sum_{i=1}^{n} x_i^2$

(b) $\beta_k = \beta_{k-1} - \alpha L'(\beta_{k-1} | x_1, x_2, \cdots, x_n)$

(c) $\beta_1 = 2.012$

7.

(a)

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$$

(b) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 3 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 1 & 1 \end{bmatrix}$