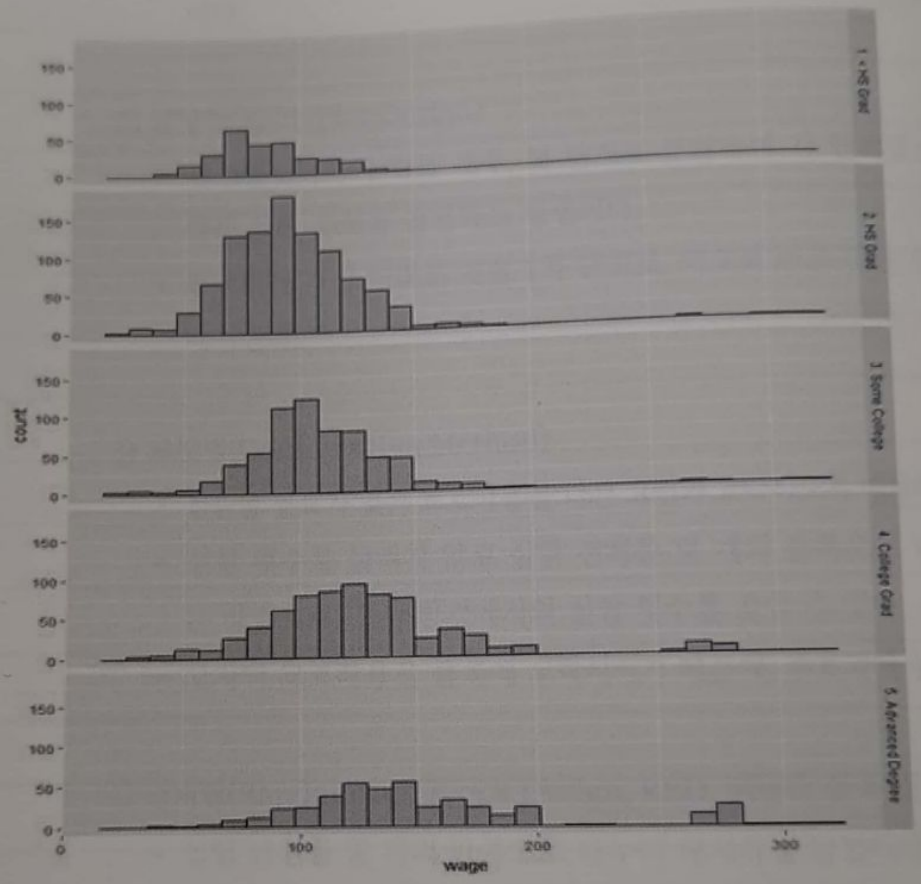
통계 분석

✔16회 기출

01 아래는 근로자의 임금(wage)과 교육수준(1. <HS Grad 2, HS Grad. 3,some College, 4. College Grad, 5. Advanced Degree)의 관계를 나타낸 그래프이다. 다음 중 아래에 대한 설명으로 부적절한 것은?



① 각 학력 수준에 따라 임금의 분포를 나타낸다.

② 학력 수준이 높아질수록 임금은 높아지는 경향이 있다.

③ 각 막대의 높이는 임금 수준을 나타낸다.

④ 5. Advanced Degree 그룹의 임금 분포는 낙도형이다.

✔12회 기출

02 다음 중 모집단에서 표본을 추출하는 방법이 아닌 것은 무엇인가?

① 단순랜덤추출법 ② 계통추출법

③ 층화추출법 ④ 김스추출법

CamScanner로 스캔하기

✓15회 기출

03 확률이란 "특정사건이 일어날 가능성의 척도"라고 정의할 수 있다. 통계적 실험을 실시할 때 나타날 수 있는 모든 결과들의 집합을 표본공간이라고 하고, 사건이란 표본공간의 부분집합을 말한다. 다음 중 확률 및 확률분포에 대한 설명으로 가장 부적절한 것은?

① 모든 사건의 확률값은 0과 1사이에 있다.
② 서로 배반인 사건들의 합집합의 확률은 각 사건들의 확률의 합이다.
③ 두 사건 A, B가 독립이라면 사건 B의 확률은 A가 일어난다는 가정하에서의 B의 조건부 확률과 동일하다.
④ 확률변수 X가 구간 또는 구간들의 모임인 숫자 값을 갖는 확률분포함수를 이산형확률밀도함수라 한다.

✓13회 기출

04 자료의 정보를 이용해 집단에 관한 추측, 결론을 이끌어내는 과정인 통계적 추론에 대한 설명으로 가장 부적절한 것은?

① 전수조사가 불가능하면 모집단에서 표본을 추출하고 표본을 근거로 확률론을 활용하여 모집단의 모수들에 대해 추론하는 것을 추정이라 한다.
② 점 추정은 표본의 정보로부터 모집단의 모수를 하나의 값으로 추정하는 것이다.
③ 통계적 추론은 제한된 표본을 바탕으로 모집단에 대한 일반적인 결론을 유도하려는 시도이므로 본질적으로 불확실성을 수반한다.
④ 구간추정은 모수의 참값이 포함되어 있다고 추정되는 구간을 결정하는 것이며, 실제 모집단의 모수는 신뢰구간에 포함되어야 한다.

✓25회 기출

05 모집단내에서 모집단의 특성을 잘 타나낼 수 있는 일부를 추출하여 이들로부터 자료를 수집하고 수집된 자료를 토대로 모집단의 특성을 추정하게 된다. 이 때 조사하는 모집단의 일부분을 표본(sample)이라 한다. 다음 중 표본조사에 대한 설명으로 가장 부적절한 것은?

① 표본오차(sampling error)는 모집단을 대표할 수 있는 표본 단위들이 조사대상으로 추출되지 못함으로서 발생하는 오차를 말한다.
② 표본편의(sampling bias)는 모수를 작게 또는 크게 할 때 추정하는 것과 같이 표본추출방법에서 기인하는 오차를 의미한다.
③ 표본편의는 확률화(randomization)에 의해 최소화하거나 없앨 수 있다. 확률화란 모집단으로부터 편의되지 않은 표본을 추출하는 절차를 의미하며 확률화 절차에 의해 추출된 표본을 확률표본(random sample)이라 한다.
④ 비표본오차(non-sampling error)는 표본오차를 제외한 모든 오차로 조사 과정에서 발생하는 모든 부주의나 실수, 알 수 없는 원인 등 모든 오차를 의미하며 조사대상이 증가한다고 해서 오차가 커지지는 않는다.

CamScanner로 스캔하기

✓19회 기출

06 표본공간은 어떤 실험이나 시도의 결과로 나올 수 있는 모든 가능한 결과의 집합이다. 사건(표본공간의 부분집합을 말한다. 다음 중 확률 및 확률분포에 관한 설명으로 부적절한 것은

① (사건 A가 일어나는 경우의 수)/(일어날 수 있는 모든 경우의 수)를 P(A)라 할 때 이를 수학적 확률이라 한다.

② 한 사건 A가 일어날 확률을 P(A)라 할 때 n번의 반복시행에서 사건 A가 일어난 횟수 하면, 상대도수는 r/n는 n이 커짐에 따라 확률 P(A)에 가까워짐을 알 수 있다. P(A) A의 통계적 확률이라 한다.

③ 두 사건 A, B가 독립일 때, 사건 B의 확률은 A가 일어났다는 가정 하에서의 B의 조 률과는 다르다.

④ 표본공간에서 임의의 사건 A가 일어날 확률 P(A)는 항상 0과 1사이에 있다.

✓14회 기출

07 다음 중 표본조사의 유의점에 대한 설명으로 가장 부적절한 것은?

① 표본편의는 표본추출 과정에서 특정 대상이 다른 대상에 비해 우선적으로 추출될 오차를 의미한다.

② 표본편의(sampling bias)는 모형 추론 방법으로 최소화하거나 없앨 수 있다.

③ 표본값으로 모집단의 모수를 추정할 때 표본오차의 비표본오차가 발생할 수 있다

④ 응답오차, 유도질문 등은 표본조사에서 유의할 점이다.

✓19회 기출

08 표본조사나 실험을 하는 과정에서 추출된 원소나 관측 자료를 얻는 것을 측정이라고 의 측정수준에 따라 통계에 이용해야 할 통계량이나 검정법이 다르다. 자료는 분류 자료로 나눌 수 있는데 다음 중 자료의 측정 수준에 대한 설명으로 부적절한 것은?

① 명목척도(nominal scale)는 단순한 번호로 차례의 의미는 없다.

② 순서척도(ordinal scale)는 순서가 의미를 가지는 번호이다.

③ 구간척도(interval scale)는 순서뿐만 아니라 그 간격도 의미가 있으며 0이 절대 가진다.

④ 비율척도(ratio scale)는 0을 기준으로 하는 절대적 척도를 간격뿐만 아니라 비 가 있다.

CamScanner로 스캔하기

✓18회 기출

09 귀무가설이 사실인데도 불구하고 사실이 아니라고 판정할 때 (귀무가설을 기각하는 오류) 이를 제 1종 오류라고 한다. 이때 우리가 내린 판정이 잘못되었을 실제 확률은 무엇으로 나타낼 수 있는가?

① α (알파)
② p-vlaue
③ 검정통계량
④ 1-α

✓15회 기출

10 확률변수 X가 확률질량함수 f(x)를 갖는 이산형 확률변수인 경우 그 기댓값으로 옳은 식은?

① $E(X)=\sum xf(x)$
② $E(X)=\int xf(x)dx$
③ $E(X)=\sum x^2f(x)$
④ $E(X)=\int x^2f(x)dx$

✓14회 기출

11 아래 조건부 확률에서 사건 A가 일어났다는 가정하의 사건 B의 확률을 조건부 확률이라고 하고 아래의 식으로 표현한다. 다음 중 의 계산식을 표현하기 위해 (가)에 들어갈 식으로 적절한 것은?

$$P(B|A)=\frac{(가)}{P(A)}$$

① $P(A\cap B)$ ② $P(A)$ ③ $P(B)$ ④ $P(A\cup B)$

✓14회 기출

12 다음 중 모분산의 추론에 대한 설명으로 가장 부적절한 것은?

① 모집단의 변동성 또는 퍼짐의 정도에 관심이 있는 경우, 모분산이 추론의 대상이 된다.
② 정규모집단으로부터 n개를 단순임의 추출한 표본의 분산은 자유도가 n-1인의 t 분포를 따른다.
③ 모집단이 정규분포를 따르지 않더라도 중심극한정리를 통해 정규모집단으로부터의 모분산에 대한 검정을 유사하게 시행할 수 있다.
④ 이 표본에 의한 분산비 검정은 두 표본의 분산이 동일한지를 비교하는 검정으로 검정통계량은 F분포를 따른다.

CamScanner로 스캔하기

✓25회 기출

13 통계적 추론이란 표본으로부터 모집단에 관한 정보를 얻고 도출하는 과정으로, 추정과 가설검정을 통하여 이루어진다. 표본을 이용하여 모집단의 특성치에 대한 추측값을 제공하고 오차한계를 제시하는 과정을 추정이라고 한다. 다음 중 추정에 대한 설명으로 부적절한 것은?

① 추정의 목적은 표본통계량에 기초하여 모수의 근사값을 결정하는 것이다. 표본 평균을 활용해서 모평균을 추정하는 것 등을 예로 들 수 있다.

② 추정량 $\hat{\mu}$를 사용하여 μ의 추정값과 그 오차한계를 제시할 때, 오차한계의 기본이 되는 것은 추정량 $\hat{\mu}$의 표준편차인 $\sigma/\sqrt{n}$이므로 이를 $\hat{\mu}$의 표준오차(standard error)라고 한다.

③ 신뢰수준 95%의 의미는 추정값이 신뢰구간 내에 존재할 확률이 95%라는 것이다.

④ 구간추정은 모수의 참값이 포함되어 있으리라고 추정되는 구간을 결정하는 것이며 실제 모집단의 모수는 신뢰구간에 포함되지 않을 수도 있다.

✓13회 기출

14 다음 중 아래의 표가 나타내는 확률질량함수를 가진 확률변수 x의 기댓값 E(x)로 가장 적절한 것은?

x	1	2	3	4
f(x)	0.2	0.3	0.2	0.075

① 1 ② 1.7 ③ 2.5 ④ 10

✓23회 기출

15 다음 중 이산형 확률분포에 해당하지 않는 것은?

① 기하분포 ② 이항분포
③ 지수분포 ④ 초기하분포

✓25회 기출

16 표본조사나 실험을 실시하는 과정에서 추출된 원소들이나 실험 단위로부터 주어진 목적에 적합하도록 관측해 자료를 얻는 것을 측정(measurement)이라 한다. 다음 중 자료의 종류에 대한 설명으로 부적절한 것은?

① 명목척도 - 측정 대상이 어느 집단에 속하는지 분류할 때 사용하는 척도로 성별구분 등이 해당한다.

② 순서척도 - 측도 대상의 특성이 가지는 서열관계를 관측하는 척도로 특정 서비스의 선호도 등이 해당된다.

③ 비율척도 - 절대적 기준인 원점이 존재하지 않으며 모든 사칙연산이 가능하고 제일 많은 정보를 가지고 있는 척도로 나이, 무게 등이 해당된다.

④ 구간척도 - 측정 대상이 갖는 속성의 양을 측정하는 것으로 온도 등이 해당된다.

CamScanner로 스캔하기

✓17회 기출

17 히스토그램은 표로 되어 있는 도수분포표를 그래프로 나타낸 것이다. 다음 중 히스토그램에 대한 설명으로 부적절한 것은?

① 히스토그램에서는 가로축이 계급, 세로축이 도수를 나타낸다. 계급은 보통 변수의 구간이며, 서로 겹치지 않는다.

② 히스토그램은 표본의 크기가 작아도 각 막대의 높이가 데이터 분포의 형상을 잘 표현해낸다.

③ 그래프의 모양이 치우쳐있거나 봉우리가 여러개 있는 그래프는 비정규 데이터일 수 있다.

④ 봉우리가 여러개 있는 데이터는 일반적으로 2개 이상의 공정이나 조건에서 데이터가 수집되는 경우 발생한다.

✓17회 기출

18 Wage 데이터셋에 대한 아래 요약통계량에 대한 설명으로 가장 부적절한 것은 무엇인가?

```
> summary(Wage[,c("wage","education")])
      wage                       education
 Min.   : 20.09   1. < HS Grad      :268
 1st Qu.: 85.38   2. HS Grad        :971
 Median :104.92   3. Some College   :650
 Mean   :111.70   4. College Grad   :685
 3rd Qu.:128.68   5. Advanced Degree:426
 Max.   :318.34
```

① wage의 최소값은 20.09 이다.

② 교육수준의 5개의 그룹으로 구분된다.

③ wage는 범주형 변수이다.

④ education은 순서형 변수이다.

CamScanner로 스캔하기

✔23회 기출

19 아래는 chickwts 데이터프레임을 분석한 것이다. 다음 중 결과에 대한 해석이 잘못된 것은

```
> t.test(chickwts$weight)

        One Sample t-test

data:  chickwts$weight
t = 28.202, df = 70, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 242.8301 279.7896
sample estimates:
mean of x
 261.3099
```

① 전체 관측치 수는 70개 이다.
② 99% 신뢰구간을 구하기 위해서는 "conf.level=0.99"라는 옵션을 사용할 수 있다.
③ 닭 무게의 점 추정량은 261.3이며, 95% 신뢰구간은 242.8에서 279.8이다.
④ 닭 무게에 대한 p-value는 p-value〈2.2e-16이므로 귀무가설이 기각된다.

✔10회 기출

20 다음 제1종 오류에 대한 설명 중 올바른 것은?

① H_0가 사실일 때, H_0가 사실이라고 판정
② H_0가 사실이 아닐 때, H_0가 사실이라고 판정
③ H_0가 사실일 때, H_0가 사실이 아니라고 판정
④ H_0가 사실이 아닐 때, H_0가 사실이 아니라고 판정

✔18회 기출

21 통계적 추론에서 모집단의 모수를 검증하기 위해 사용하는 모수적 방법과 비교하여 비모수적 방법의 특징으로 가장 부적절한 것은?

① 비모수적 검정은 모집단의 분포에 대해 아무런 제약을 가하지 않는다.
② 관측된 자료가 특정 분포를 따른다고 가정할 수 없는 경우에 이용된다.
③ 분포의 모수에 대한 가설을 설정하지 않고 분포의 형태에 대해 가설을 설정한다.
④ 비모수 검정에서는 관측값의 절대적 크기에 의존하여 평균, 분산 등을 이용해 검정을 실시한

CamScanner로 스캔하기

✓15회 기출

22 다음 중 표본을 도표화함으로써 모집단 분포의 개형을 파악하는 방법에 대한 설명으로 가장 부적절한 것은?

① 히스토그램은 도수분포표를 이용하여 표본자료의 분포를 나타낸 그래프이다. 수평축 위에 계급구간을 표시하고 그 위로 각 계급의 상대도수에 비례하는 넓이의 직사각형을 그린 것이다.

② 줄기잎그림은 각 데이터의 점들을 구간단위로 요약하는 방법으로써 계산량이 많다.

③ 산점도는 두 특성의 값이 연속적인 수인 경우, 표본자료를 그래프로 나타내는 방법으로써 각 이차원 자료에 대하여 좌표가 (특성 1의 값, 특성 2의 값)인 점을 좌표평면 위에 찍은 것이다.

④ 파레토그림(pareto diagram)은 명목형 자료에서 "중요한 소수"를 찾는데 유용한 방법이다.

✓17회 기출

23 Wage 데이터에서 wage에 대한 t-test를 실시하였다 다음 설명 중 부적절한 것은?

```
> t.test(Wage$wage,mu=100)

        One Sample t-test

data:  Wage$wage
t = 15.362, df = 2999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
 110.2098 113.1974
sample estimates:
mean of x
 111.7036
```

① 한 집단의 평균에 대한 t-test(one sample t-test) 결과이다.

② 양측검정 결과를 보여주고 있다.

③ t-test 의 자유도는 2999이다.

④ 평균에 대한 95% 신뢰구간은 귀무가설에서 설정한 평균의 참값을 포함한다.

CamScanner로 스캔하기

✓25회 기출

24 이상값 탐색을 위해 상자그림(boxplot)을 사용하려 한다. 아래와 같은 데이터 요약 결과가 있을 때, 다음 중 이상값을 판단하는 하한선, 상한선으로 옳은 것은?

```
> summary( x )
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0       4       7   9.615      12      39
```

① (-12, 36) ② (4, 12) ③ (-2, 30) ④ (-8, 24)

✓14회 기출

25 Carseats 데이터프레임은 400개 상점에서 판매 중인 유아용 키시트에 대한 자료이다. 다음 중 아래의 결과물에 대한 설명으로 가장 부적절한 것은?

```
> summary(Carseats)
     Sales          CompPrice       Income        Advertising      Population
 Min.   : 0.000   Min.   : 77   Min.   : 21.00   Min.   : 0.000   Min.   : 10.0
 1st Qu.: 5.390   1st Qu.:115   1st Qu.: 42.75   1st Qu.: 0.000   1st Qu.:139.0
 Median : 7.490   Median :125   Median : 69.00   Median : 5.000   Median :272.0
 Mean   : 7.496   Mean   :125   Mean   : 68.66   Mean   : 6.635   Mean   :264.8
 3rd Qu.: 9.320   3rd Qu.:135   3rd Qu.: 91.00   3rd Qu.:12.000   3rd Qu.:398.5
 Max.   :16.270   Max.   :175   Max.   :120.00   Max.   :29.000   Max.   :509.0
     Price       ShelveLoc        Age          Education     Urban       US
 Min.   : 24.0   Bad    : 96   Min.   :25.00   Min.   :10.0   No :118   No :142
 1st Qu.:100.0   Good   : 85   1st Qu.:39.75   1st Qu.:12.0   Yes:282   Yes:258
 Median :117.0   Medium:219   Median :54.50   Median :14.0
 Mean   :115.8                Mean   :53.32   Mean   :13.9
 3rd Qu.:131.0                3rd Qu.:66.00   3rd Qu.:16.0
 Max.   :191.0                Max.   :80.00   Max.   :18.0
```

① ShelveLoc은 명목척도에 해당된다.

② ShelveLoc은 Good인 카시트의 비율은 0.21이다.

③ US 변수는 구간척도에 해당된다.

④ US가 No인 카시트가 Yes인 카시트보다 적다.

✓18회 기출

26 Chickwts는 71마리의 병아리들에게 서로 다른 모이(feed)를 6주간 먹인 후 무게(weight)를 측정한 자료이다. 아래는 첨가물 그룹 간 평균 무게에 차이가 있는지 검정하기 위해 분산분석을 한 결과이다. 설명이 가장 부적절한 것은?

CamScanner로 스캔하기

```
> summary(aov(weight~feed, chickwts))
            Df Sum Sq Mean Sq F value   Pr(>F)
feed         5 231129   46226   15.37 5.94e-10 ***
Residuals   65 195556    3009
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

① 귀무가설은 "첨가물 그룹 간의 평균이 모두 동일하다"이다.

② 첨가물의 개수는 5개다.

③ 유의수준 0.05하에서 첨가물 그룹 간의 무게 평균이 동일하지 않다는 통계적으로 유의한 증거가 있다.

④ 위의 가설검정은 F-통계량을 기반으로 한다.

✓19회 기출

27 아래 데이터는 두 종류의 수면 유도제(group)를 무작위로 선정된 20명의 환자를 대상으로 수명시간 증감(extra)을 측정한 자료이다. 아래 결과에 대한 설명으로 잘못된 것은?

```
> head(sleep)
  extra group ID
1   0.7     1  1
2  -1.6     1  2
3  -0.2     1  3
4  -1.2     1  4
5  -0.1     1  5
6   3.4     1  6
> t.test(extra~group, sleep)

        Welch Two Sample t-test

data:  extra by group
t = -1.8608, df = 17.776, p-value = 0.07939
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.3654832  0.2054832
sample estimates:
mean in group 1 mean in group 2
           0.75            2.33
```

① 유의수준 1%하에서 수면유도제 2가 수면유도제 1보다 통계적으로 유의하게 평균 수면시간을 증가시킨다고 결론지을 수 있다. 즉, 수면유도제 2가 수면유도제 1보다 더 효과적이다.

② 수면유도제 1에 의해 평균적으로 0.75시간의 수면시간이 증가하였다.

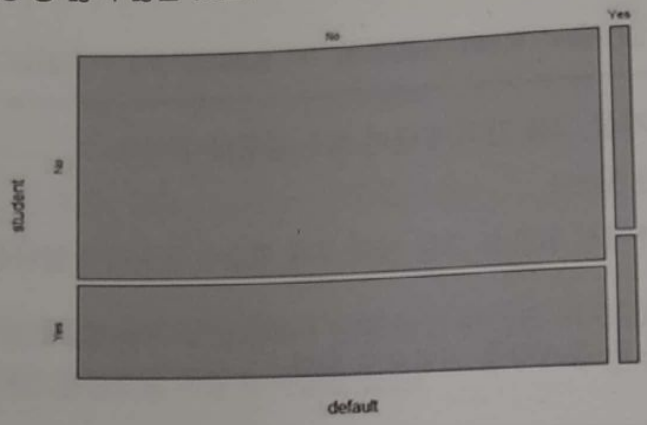
③ 수면유도제 2에 의해 평균적으로 2.33시간의 수면시간이 증가하였다.

④ 두 수면유도제에 의해 증가된 평균 수면시간의 차이는 -3.37시간에서 0.21시간 사이에 있다고 95% 확신할 수 있다.

CamScanner로 스캔하기

✓25회 기출

28 Default 데이터셋은 10000명의 신용카드 고객에 대한 카드대금 연체여부(default=Yes/No)와 학생여부(student=Yes/No)를 포함한다. 아래는 default와 student 간의 관계를 나타내는 그림이다. 보기의 설명 중 옳지 않은 것은?



① 학생인 고객이 학생이 아닌 고객보다 많다.
② 연체 고객이 연체하지 않은 고객에 비해 적다.
③ 연체하지 않은 고객 중 학생의 비율이 연체한 고객 중 학생의 비율보다 적다.
④ 학생 여부와 연체 여부는 서로 독립이 아닐 것으로 추측된다.

✓17회 기출

29 다음 중 스피어만 상관계수에 대한 설명으로 부적절한 것은?

① 비선형적인 상관관계는 나타내지 못한다.
② 서열척도로 측정된 변수간 관계를 측정한다.
③ -1과 1사이의 값을 가진다.
④ 0은 상관관계가 없음을 의미한다.

✓10회 기출

30 다음 중 회귀분석의 가정으로 부적절한 것은?

① 독립성 ② 선형성 ③ 정규성 ④ 이분산성

✓15회 기출

31 다음 중 상관계수에 대한 설명으로 가장 부적절한 것은?

① 피어슨 상관계수는 두 변수 간의 선형관계의 크기를 측정한다.
② 스피어만 상관계수는 두 변수 간의 비선형적인 관계도 측정 가능하다.
③ 피어슨 상관계수와 스피어만 상관계수는 -1과 1사이의 값을 가진다.
④ 피어슨 상관계수는 두 변수를 순위로 변환시킨 후 두 순위 사이의 스피어만 상관계수로 정의된다.

CamScanner로 스캔하기

✓18회 기출

32 상관분석에 대한 설명으로 가장 부적절한 것은?

① 등간 척도 및 비율척도로 측정된 변수들 간의 상관계수를 측정하는데 피어슨 상관계수를 이용한다.

② 서열 척도로 측정된 변수들 간의 상관계수를 측정하는데 스피어만 상관계수를 이용한다.

③ 상관분석은 변수들 간의 연관성을 파악하기 위해 사용하는 분석 기법 중 하나로 변수 간의 선형 관계 정도를 분석하는 통계기법이다.

④ 상관분석은 종속변수에 미치는 영향력의 크기를 파악하여 독립변수의 특정한 값에 대응하는 종속 변수값을 예측하는 선형모형을 산출하는 방법이다.

✓19회 기출

33 다음 중 추정된 다중회귀모형이 통계적으로 유의미한지 확인하는 방법으로 적절한 것은 ?

① F-통계량을 확인한다.

② 결정계수를 확인한다.

③ t-통계량을 확인한다.

④ 잔차를 그래프로 그리고 회귀진단을 한다.

✓10회 기출

34 데이터 프레임 attitude에대해 아래와 같이 R 명령을 적용하고 결과를 얻었다. 다음 설명 중 가장 부적절한 것은?

```
> cor(attitude)
             rating complaints privileges learning   raises   critical  advance
rating     1.0000000  0.8254176  0.4261169 0.6236782 0.5901390 0.1564392 0.1550863
complaints 0.8254176  1.0000000  0.5582882 0.5967358 0.6691975 0.1877143 0.2245796
privileges 0.4261169  0.5582882  1.0000000 0.4933310 0.4454779 0.1472331 0.3432934
learning   0.6236782  0.5967358  0.4933310 1.0000000 0.6403144 0.1159652 0.5316198
raises     0.5901390  0.6691975  0.4454779 0.6403144 1.0000000 0.3768830 0.5741862
critical   0.1564392  0.1877143  0.1472331 0.1159652 0.3768830 1.0000000 0.2833432
advance    0.1550863  0.2245796  0.3432934 0.5316198 0.5741862 0.2833432 1.0000000
```

① 모든 변수들 사이에 양(+)의 상관관계가 존재한다.

② rating과 complaints 사이에 가장 강한 상관관계가 존재한다.

③ critical과 learning 사이의 상관관계가 가장 약하다.

④ 모든 변수의 분산이 1이다.

CamScanner로 스캔하기

✓22회 기출

35 아래는 남학생과 여학생이 좋아하는 과일에 대한 빈도교차표이다. 전체에서 1명을 뽑았을 때, 그 학생이 남학생일 때 사과를 좋아할 확률은 얼마인가 ?

	사과	딸기
남학생	30	40
여학생	10	20

① 3/10 ② 4/10 ③ 3/7 ④ 6/10

✓18회 기출

36 아래는 200개의 특정 제품의 sales(단위: 1천개)와 TV, Radio, Newspaper 광고예산 (단위 :1천달러) 간의 pearson 상관계수 행렬이다. 설명이 가장 부적절한 것은?

	TV	Radio	Newspaper	Sales
TV	1.000	0.054	0.057	0.793
Radio	0.054	1.000	0.333	0.543
Newspaper	0.057	0.333	1.000	0.222
Sales	0.793	0.543	0.222	1.000

① 3가지 매체의 광고예산은 Sales와 양의 상관관계를 가지고 있다.
② Sales와 가장 상관관계가 높은 변수는 TV이다.
③ Radio 광고예산이 증가할 때 Newspaper 광고 예산이 증가하는 경향이 있다.
④ TV 광고 예산을 늘릴 경우 Sales가 증가하는 인과관계를 가진다.

CamScanner로 스캔하기

19회 기출

37 Carseats 데이터프레임은 400개 상점에서 판매 중인 유아용 카시트에 대한 자료이다. 이 데이터의 일부 변수들의 상관분석 결과로 가장 부적절한 것은?

```
> rcorr(as.matrix(Carseats[,c(1:6,8)]),type="pearson")
            Sales CompPrice Income Advertising Population Price   Age
Sales        1.00      0.06   0.15        0.27       0.05 -0.44 -0.23
CompPrice    0.06      1.00  -0.08       -0.02      -0.09  0.58 -0.10
Income       0.15     -0.08   1.00        0.06      -0.01 -0.06  0.00
Advertising  0.27     -0.02   0.06        1.00       0.27  0.04  0.00
Population   0.05     -0.09  -0.01        0.27       1.00 -0.01 -0.04
Price       -0.44      0.58  -0.06        0.04      -0.01  1.00 -0.10
Age         -0.23     -0.10   0.00        0.00      -0.04 -0.10  1.00

n= 400

P
            Sales  CompPrice Income Advertising Population Price  Age
Sales               0.2009    0.0023 0.0000      0.3140     0.0000 0.0000
CompPrice   0.2009            0.1073 0.6294      0.0584     0.0000 0.0451
Income      0.0023 0.1073           0.2391      0.8752     0.2579 0.9258
Advertising 0.0000 0.6294    0.2391             0.0000     0.3743 0.9276
Population  0.3140 0.0584    0.8752 0.0000                 0.8087 0.3948
Price       0.0000 0.0000    0.2579 0.3743      0.8087            0.0411
Age         0.0000 0.0451    0.9258 0.9276      0.3948     0.0411
```

① Sales와 CompPrice 간의 상관계수는 유의하지 않다.

② Sales와 가장 강한 상관관계를 보이는 변수는 Price이다.

③ Price가 올라갈수록 Sales는 낮아지는 경향이 있다.

④ Sales와 Price는 양의 선형관계를 가진다.

CamScanner로 스캔하기

✓10회 기출

38 아래는 단순회귀분석의 결과이다 다음 설명 중 부적절한 것은?

```
Call:
lm(formula = Height ~ BodyWeight)

Residuals:
     Min       1Q       Median       3Q      Max
-3.56937  -0.96341  -0.09212      1.04255  5.12382

Coefficients:
                Estimate    Std. Error    t value    Pr(>|t|)
(Intercept)      0.5            1          -0.5        0.610
Bodyweight       3.2            0.2        16        <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.452 on 142 degrees of freedom
Multiple R-squared:  0.6466,     Adjusted R-squared:  0.6441
F-statistic:  259.8 on 1 and 142 DF,  p-value: <2.2e-16
```

① 종속변수는 Height이다.

② 독립변수는 Bodyweight이다.

③ 모형의 설명력은 약 64.66%이다.

④ 모형의 적합도는 통계적으로 유의하지 않다.

✓18회 기출

39 다중 회귀분석에서 가장 적합한 회귀모형을 찾기 위한 과정의 설명으로 가장 부적절한 것은

① 독립변수의 수가 많아지면 모델의 설명력이 증가하지만 모형이 복잡해지고, 독립변수들 간에 서로 영향을 미치는 다중공선성의 문제가 발생하므로 상대적인 조정이 필요하다.

② 회귀식에 대한 검정은 독립변수의 기울기(회귀계수)가 0이 아니라는 가정을 귀무가설, 기울기가 0인 것을 대립가설로 놓는다.

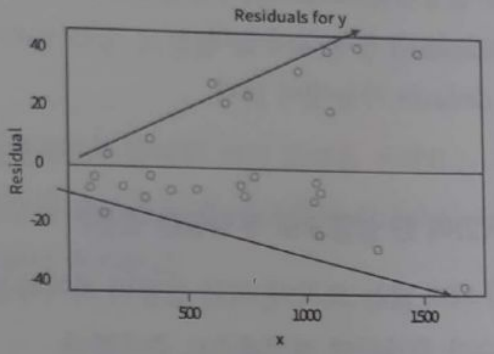
③ 잔차의 독립성, 등분산성 그리고 정규성을 만족하는지 확인해야 한다.

④ 회귀분석의 가설검정에서 p값이 0.05보다 작은 값이 나와야 통계적으로 유의한 결과로 받아들일 수 있다.

CamScanner로 스캔하기

✓12회 기출

40 아래는 결과를 생성한 잔차도이다. 다음 중 어떤 회귀분석의 가정이 위배되었다고 판단할 수 있을지 고르시오.



① 선형성 ② 독립성 ③ 등분산성 ④ 비상관성

✓17회 기출

41 Default 데이터셋은 10,000명의 신용카드 고객에 대한 연체여부(default:1-default,0-not default), 카드대금 납입 후 남은 평균 카드잔고(balance), 연봉(income)을 포함하고 있다. 아래는 연체 가능성을 95% 신뢰수준으로 모형화한 결과이다. 다음 설명이 부적절한 것은 무엇인가?

```
> model<-glm(default~balance+income, data=Default, family="binomial")
> summary(model)

Call:
glm(formula = default ~ balance + income, family = "binomial",
    data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4725  -0.1444  -0.0574  -0.0211   3.7245

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.154e+01  4.348e-01 -26.545  < 2e-16 ***
balance      5.647e-03  2.274e-04  24.836  < 2e-16 ***
income       2.081e-05  4.985e-06   4.174 2.99e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1579.0  on 9997  degrees of freedom
AIC: 1585

Number of Fisher Scoring iterations: 8
```


CamScanner로 스캔하기