# 11장

11.2 분산이 같은 세 정규모집단으로부터 확률표본을 다음과 같이 추출하였다.

| | 모집단1 | 모집단2 | 모집단3 |
|---|---|---|---|
| | 3.1<br>4.3<br>2.2 | 1.1<br>0.2<br>2.0 | 5.4<br>3.6<br>4.0<br>3.0 |
| $Y_{i+}$ | 9.6 | 3.3 | 16 |
| $n_i$ | 3 | 3 | 4 |
| $\overline{y_i}$ | 3.2 | 1.1 | 4 |

(1) CT와 TSS를 구하여라.

$$Y_{++} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij} = 28.9$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 = 105.31$$

$$CT = Y_{++}^2/m = 28.9^2/10 = 83.521$$

$$TSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - CT$$
$$= 105.31 - 83.521 = 21.789$$

(2) SSTR과 MSTR을 구하여라.

$$SSTR = \sum_{i=1}^{k} n_i(\overline{Y_i} - \overline{Y})^2 = \sum_{i=1}^{k} n_i \overline{Y_i}^2 - CT = \sum_{i=1}^{k} Y_{i+}^2 / n_i - CT$$

$$= \frac{9.6^2}{3} + \frac{3.3^2}{3} + \frac{16^2}{4} - 83.521 = 98.35 - 83.521 = 14.829$$

$$MSTR = \frac{SSTR}{k-1} = \frac{14.829}{2} = 7.4145$$

(3) SSE와 MSE를 구하여라.

$$SSE = TSS - SSTR = \sum_{i=1}^{k} \sum_{j=1}^{n} Y_{ij}^2 - \sum_{i=1}^{k} n_i \overline{Y_i}^2 = 21.789 - 14.829 = 6.96$$

$$MSE = \frac{SSE}{m-k} = \frac{6.96}{7} = 0.9943$$

(4) 분산분석표를 작성하여 세 모집단의 평균이 같은지 5%유의수준에서 검정하여라.

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F |
|---|---|---|---|---|
| 처리(모형) | 2 | 14.829 | 7.4145 | 7.457 |
| 오차 | 7 | 6.96 | 0.9943 | |
| 전체 | 9 | 21.789 | | |

가설: $H_0$ :세 모집단 평균이 모두 같다.

검정통계치: F=7.457

기각역 $F \geq F_{0.05,2,7} = 4.74$

결론 : F=7.457 > 4.74 이므로 5% 유의수준에서 귀무가설을 기각할 수 있다. 최소한 한 모집단 평균이 다르다고 할 수 있다.

(5) 모집단 3의 평균에 대한 95% 신뢰구간을 구하여라.

$$\overline{X} \pm t_{0.025,3}\frac{s}{\sqrt{n}} = 4 \pm 3.182\frac{1.0198}{\sqrt{4}} \quad (2.377, 5.623)$$

(6) Fisher의 LSD 방법을 이용하여 세 모집단 평균에 대해 5% 유의수준으로 다중비교하여라.

$$\left|\overline{y}_i - \overline{y}_j\right| > t_{\alpha/2, m-k}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$ 이면 $H_0$기각

$t_{0.025,7} = 2.365, \quad MSE = 0.9943$

| 종류 | 평균 | 종류 | 평균 | 차 | $t_{\alpha/2,m-k}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$ | LSD |
|---|---|---|---|---|---|---|
| 1 | 3.2 | 2 | 1.1 | 2.1 | 1.924 | ○ |
| | | 3 | 4 | -0.8 | 1.802 | × |
| 2 | 1.1 | 3 | 4 | -2.9 | 1.802 | ○ |

모집단 1과 모집단 2, 그리고 모집단 2와 모집단 3은 차이가 있다고 할 수 있고, 모집단 1과 모집단 3은 차이가 없다. 즉 모집단 2가 모집단 1과 모집단 3과 차이가 있다고 볼 수 있다.

11.4 다음 분산분석표에서 부분정보가 주어졌을 때 나머지 부분을 채우고 처리 효과가 있는지를 5%유의수준에서 검정하여라.

(1)

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F |
|---|---|---|---|---|
| 처리(모형) | 2 | 14 | 7 | 3.00 |
| 오차 | 12 | 28 | 2.333 | |
| 전체 | 14 | 42 | | |

기각역: F ≥ $F_{0.05,2,12} = 3.89$

결론 : 귀무가설을 기각할 수 없다. 그러므로 처리 효과가 있다고 할 수 없다.

(2)

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F |
|---|---|---|---|---|
| 처리(모형) | 5 | 35 | 7 | 2 |
| 오차 | 10 | 35 | 3.5 | |
| 전체 | 15 | 70 | | |

기각역: F ≥ $F_{0.05,5,10} = 3.33$

결론 : 귀무가설을 기각할 수 없다. 그러므로 처리 효과가 있다고 할 수 없다.

(3)

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F |
|---|---|---|---|---|
| 처리(모형) | 3 | 30 | 10 | 5 |
| 오차 | 7 | 14 | 2 | |
| 전체 | 10 | 44 | | |

기각역: F ≥ $F_{0.05,3,7} = 4.35$

결론 : 귀무가설을 기각할 수 있다. 그러므로 처리 효과가 있다고 할 수 있다.

(4) (5) 세 가지 다이어트 방법에 따라 체중 감소량(kg)에 차이가 있는지를 알아보기 위해 12명의 사람을 3그룹으로 4명씩 무자위로 배치한 후 시험을 진행하였다. 12명의 체중 감소량 합이 60이고 Ⓖ=175, Ⓕ=470이라고 한다.

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F |
|---|---|---|---|---|
| 처리(모형) | 2 | 350 | 175 | 13.125 |
| 오차 | 9 | 120 | 13.333 | |
| 전체 | 11 | 470 | | |

가설 $H_0$: 세가지 다이어트 방법의 체중 감소량에 차이가 없다.

기각역: F ≥ $F_{0.05,2,9} = 4.26$

결론 : 귀무가설을 기각할 수 있다. 그러므로 세 가지 다이어트 방법의 체중감소량에 차이가 있다고 볼 수 있다.

(6) 세 가지 다이어트 방법에 따라 체중 감소량에 차이가 있는지 알아보기 위해 12명의 사람을 3그룹으로 4명씩 무작위로 배치한 후 실험을 진행하였다. 12명의 체중 감소량 합이 60, 감소량의 제곱합이 770이고, Ⓓ=350라고 한다.

$$\sum_i \sum_j y_{ij} = 60, \quad \sum_i \sum_j y_{ij}^2 = 770$$

$TSS = 770 - \frac{60^2}{12} = 470$ => (4),(5)번 분산분석표와 같다.

11.5 다음 표는 세 기관의 연수자들의 학업성과를 비교하기 위해 무작위로 3명씩 선발하여 시험을 본 결과를 분석한 것이다.

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F | p-값 |
|---|---|---|---|---|---|
| 처리(모형) | 2 | 150.0 | 75.0 | 5.0 | 0.053 |
| 오차 | 6 | 90.0 | 15.0 | | |
| 전체 | 8 | 240.0 | | | |

(1) 세 기관의 연수자들의 학업성과에 차이가 있다는 주장을 입증하고 싶다. 적절한 가설을 기술하여라.

$H_0$ : 세 기관의 연수자들의 학업성과에 차이가 없다.

$H_1$ : 세 기관의 연수자들의 학업성과에 차이가 있다.

(2) 유의수준 1%, 5%, 10%에서 세 국방부 직할부대 및 기관의 연수자들의 학업성과에 차이가 있는지를 각각 검정하여라.
=> 유의수준 1%,5%에서는 p-값이 유의수준보다 크므로 귀무가설을 기각할 수 없다. 즉 세 기관이 연수자들의 학업성과에 차이가 있다고 할 수 없다.
유의수준 10%에서는 p-값이 유의수준보다 작으므로 귀무가설을 기각할 수 있다. 즉 세 기관이 연수자들의 학업성과에 차이가 있다고 할 수 있다.

11.10 빵을 만드는 세 방법에 따라 빵의 밀도에 차이가 있는가를 알아보기 위해 실험을 진행하여 다음과 같은 자료를 얻었다.

| 방법 | 관측값 |
|---|---|
| 1 | 0.95 0.86 0.71 0.72 |
| 2 | 0.71 0.85 0.62 0.72 0.64 |
| 3 | 0.69 0.68 0.51 0.73 0.44 |

(1) 5% 유의수준에서 세 방법 간에 평균밀도 차이가 있는지를 검정하여라.

| 변인 | 자유도 | 제곱합(SS) | 평균제곱(MS) | F |
|---|---|---|---|---|
| 처리(모형) | 2 | 0.089 | 0.045 | 3.567 |
| 오차 | 11 | 0.137 | 0.012 | |
| 전체 | 13 | | | |

기각역: F ≥ $F_{0.05,2,11} = 3.98$
결론 : 귀무가설을 기각할 수 없다. 그러므로 세 방법 간에 평균밀도 차이가 있다고 할 수 없다.

(2) 만약 방법 1은 새로운 방법이고 방법2와 3은 기존방법이라고 하면 새로운 방법과 기존방법 간에 평균밀도 차이가 있는지를 알아보려면 어떻게 하면 되는 알아보아라.

=> 새로운 방법에 대한 자료가 4개, 기존의 방법 자료가 10개로 하고 독립표본 두 모집단의 평균 비교를 하면 된다.

(3) Fisher LSD와 본페로니 방법을 이용하여 방법 간 평균밀도에 차이가 있는지 다중비교하여라.

LSD : $\left|\bar{y}_i - \bar{y}_j\right| > t_{\alpha/2, m-k}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$ 이면 $H_0$기각

$$t_{0.025,11} = 2.201, \quad MSE = 0.012$$

본페로니: $\left|\bar{y}_i - \bar{y}_j\right| > t_{\alpha/2p, m-k}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$ 이면 $H_0$기각

$t_{0.00833,1}1 = 2.82$ (0.025/3=0.00833)

| 종류 | 평균 | 종류 | 평균 | 차 | LSD | 본페로니 |
|---|---|---|---|---|---|---|
| 1 | 0.81 | 2 | 0.708 | 0.102 | 0.162 | 0.207 |
| | | 3 | 0.61 | 0.2 | 0.162 | 0.207 |
| 2 | 0.708 | 3 | 0.61 | 0.098 | 0.152 | 0.195 |

LSD : 방법1과 방법3의 빵의 밀도에 차이가 있다고 할 수 있다.
본페로니는 세 방법에 따라 빵의 밀도에 차이가 없다.

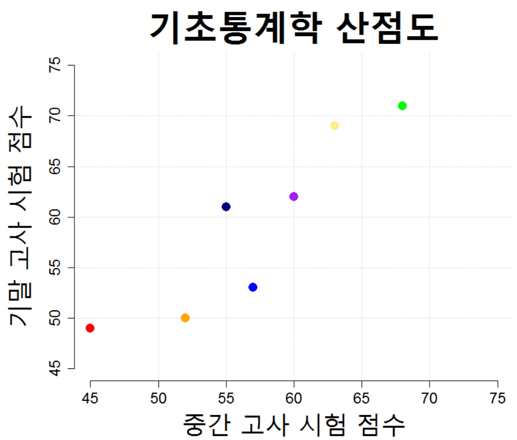
12장

12.2 다음 자료는 기초통계학을 수강한 학생들 중 임의로 선택한 7명의 중간시험점수와 기말시험점수이다.

| 학생 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 중간시험(x) | 45 | 52 | 63 | 68 | 57 | 55 | 60 |
| 기말시험(y) | 49 | 50 | 69 | 71 | 53 | 61 | 62 |
| 적합값 | 46.3 | 53.8 | ①65.5 | 70.9 | 59.1 | ③ 56.97 | 62.3 |
| 잔차 | 2.7 | ②−3.8 | | 0.1 | −6.1 | ④ 4.03 | −0.3 |

참고 : $\sum x_i y_i = 24076,\ \sum x_i^2 = 23196,\ \sum y_i^2 = 25077$

(1) 중간시험과 기말시험의 점수들 간 산점도를 그려보아라.




(2) 중간시험점수와 기말시험점수의 상관계수를 구하여라.

$$r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}} = \frac{24076 - 7(57.143)(59.286)}{\sqrt{23196 - 7(57.143^2)}\sqrt{25077 - 7(59.286^2)}}$$

$$= \frac{361.541}{18.405 \times 21.753} = 0.903$$

(3) 중간시험점수가 설명변수, 기말시험점수가 반응변수인 단순선형회귀모형의 회귀계수를 최소제곱법으로 추정하여라.

$\widehat{\beta_1} = \frac{s_{xy}}{s_{xx}} = \frac{361.541}{338.744} = 1.067$

$\widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x} = -1.712$

추정된 회귀직선: $\hat{y} = -1.712 + 1.067x$

(4) (3)에서 구한 결과와 표의 값을 이용하여 ①②③④를 계산하여라.

=> 위의 표에 정답

(5) 중간시험점수가 기말시험점수에 영향을 주는지를 5%유의수준에 확인하여라.

$H_0 : \beta_1 = 0, \quad H_1 : \beta_1 \neq 0$

-검정통계치 $t = \frac{1.067}{\sqrt{17.463/338.744}} = 4.699$ > $t_{0.025,5}$ =2.571이므로

유의수준 5%에서 $H_0$ 기각 할 수 있다.

여기서 $MSE = \dfrac{SSE}{n-2} = \left(s_{yy} - \dfrac{s_{xy}^2}{s_{xx}}\right)/n-2 = \left(21.753^2 - \dfrac{361.541^2}{18.405^2}\right)/5 = 17.463$

$\Rightarrow$ 5% 유의수준에서 중간시험점수가 기말시험점수에 영향을 준다고 할 수 있음

(6) 중간시험점수가 55점 받은 학생의 평균기말시험점수에 대한 95% 신뢰구간을 구하여라.

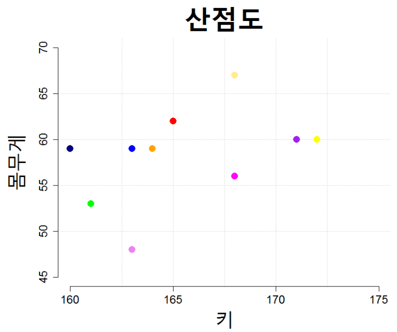
$\hat{y} = -1.712 + 1.067(55) = 56.973$

$t_{0.025,5}$ =2.571, $\bar{x} = 57.143$, $s_{xx} = 18.405^2$

$$\hat{y}_i \pm t_{\alpha/2, n-2}\sqrt{MSE}\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{s_{xx}}} = 56.973 \pm 4.249$$

12.3 다음 자료는 임의로 선발된 20명의 신장과 체중 자료이다.

(1) 번호1에서 10까지 사람들에 대한 신장과 체중의 산점도를 그리고 신장과 체중 간의 관계를 간략하게 설명하여라.




키와 몸무게 간에는 양의 상관관계를 보인다.

(2) 피어슨의 표본상관관계를 계산하여라.

$$r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}} = \frac{206138 - 20(167.55)(61.35)}{\sqrt{561941 - 20(167.55^2)}\sqrt{76543 - 20(61.35^2)}} = 0.710$$

$s_{xy}$ =554.15, $s_{xx}$ =480.95, $s_{yy}$ = 1266.55

(3) 신장을 설명변수, 체중을 반응변수라고 할 때, 최소제곱법을 이용하여 단순 선형회귀식을 추정하여라.

$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{554.15}{480.95} = 1.152$

$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = -131.701$

추정된 회귀직선: $\hat{y} = -131.701 + 1.152x$

(4) (3)의 결과에서 절편의 값에는 관계없이 기울기가 1인지 아닌지를 5%유의수준에서 검정하여라.

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) (3)에서 추정한 회귀식을 이용하여 신장이 185cm인 학생의 체중에 대한 95% 예측구간을 구하여라.

$$MSE = \frac{SSE}{n-2} = \frac{628.059}{18} = 34.8922 \qquad \sqrt{MSE} = 5.9070$$

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

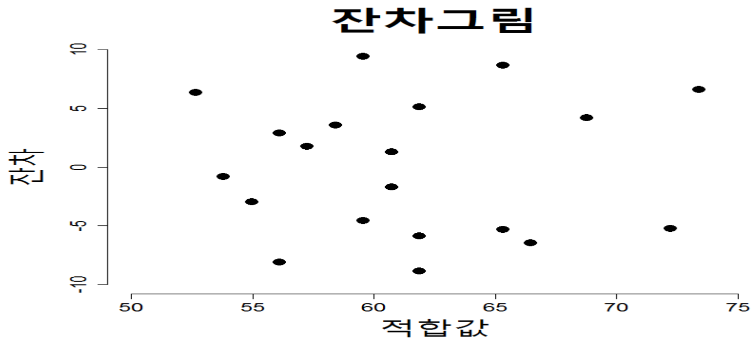
$$\bar{x} = 167.55, \quad t_{0.025,18} = 2.101$$

⇨ $x = 185$에서 $y$(몸무게) 예측값의 표준오차:

$$\sqrt{MSE}\sqrt{1 + \frac{1}{20} + \frac{(185 - 167.55)^2}{S_{xx}}} = 7.6635$$

⇒ 95% 신뢰구간: $81.4561 \pm 2.101 \times 7.6635 = 81.4561 \pm 16.101$

(6) 잔차그림을 통해 등분산 가정을 만족하는지 확인하여라.




등분산 만족

12.5 다음 표는 제품인지도의 제곱과 판매량이 선형관계를 가지는 것으로 나타나 설명변수가 제품인지도의 제곱, 반응변수가 판매량인 선형회귀모형에 대한 결과이다. 회귀모형을 추정하는 데 사용된 자료의 수는 20개이다.

| 변수 | 추정값 | 표준오차 |
| --- | --- | --- |
| 절편 | 20 | 1.5 |
| X | 1.2 | 0.3 |

(1) $S_{xx} = 100/9$일 때 분산추정량인 $MSE$를 구하여라.

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

(2) 위의 결과를 이용하여 제품인지도의 제곱이 판매량에 영향을 주는지 5%유의수준에서 검정하여라.

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) 제품인지도 제곱의 평균이 10이라고 할 때 회귀식의 절편에 대한

95% 신뢰구간을 구하여라.

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

- $\widehat{\beta}_0 \pm t_{0.025,18} \times \widehat{s.e}(\widehat{\beta}_0) = 20 \pm 2.101 \times 3.0083 = 20 \pm 6.32$

(4) 위의 결과를 이용하여 평균 판매량이 50이 되는 제품인지도의 값은 얼마나 될지 유도하여라.

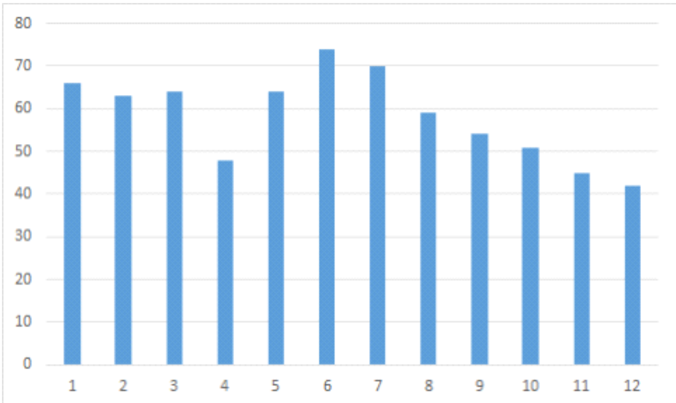
$\hat{y} = 20 + 1.2x^2$ => $50 = 20 + 1.2x^2$ => $x = \sqrt{25} = 5$

13장

13.2 다음 표는 어느 대학에 재학 중인 700명의 여학생을 대상으로 언제 출생했는지를 조사한 자료이다.

| | | | | | | |
|---|---|---|---|---|---|---|
| 월 | 1 | 2 | 3 | 4 | 5 | 6 |
| 명 | 66 | 63 | 64 | 48 | 64 | 74 |
| 월 | 7 | 8 | 9 | 10 | 11 | 12 |
| 명 | 70 | 59 | 54 | 51 | 45 | 42 |

(1) 막대그래프를 이용하여 월별 출생도수를 비교하여라.




(2) 일수와 관계없이 월별 출생비율이 같은지 5%유의수준에서 검정하고 같지 않다면 어떤 월에서 차이가 있는지를 설명하여라.

가설 : $H_0 : \theta_1 = \theta_2 = \cdots = \theta_{12} = \dfrac{1}{12}$

검정통계치 : $\chi^2 = \frac{(66-58.333)^2}{58.333} + \cdots + \frac{(42-58.333)^2}{58.333} = 19.726$

기각역 : $\chi^2 \geq \chi^2_{0.05,11} = 19.68$

결론: 귀무가설을 기각할 수 있다. 일수와 관계없이 월별 출생비율이 같다고 할 수 없다.

피어슨 잔차를 구한 결과는 다음 표와 같다.

| 월 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 명 | 66 | 63 | 64 | 48 | 64 | 74 |
| 잔차 | 1.004 | 0.611 | 0.742 | −1.353 | 0.742 | 2.051 |
| 월 | 7 | 8 | 9 | 10 | 11 | 12 |
| 명 | 70 | 59 | 54 | 51 | 45 | 42 |
| 잔차 | 1.528 | 0.087 | −0.567 | −0.960 | −1.746 | −2.139 |

대략적으로 6월, 12월에서 차이가 있었다.

(3) 3~5월을 봄, 6~8월을 여름, 9~11월을 가을, 12~2월을 겨울이라고

할 때,

① 계절에 따른 출생비율이 차이가 있는지를 5% 유의수준에서 검정하여라.

| 계절 | 봄 | 여름 | 가을 | 겨울 |
|---|---|---|---|---|
| 관측도수 | 176 | 203 | 150 | 171 |
| 기대도수 | 175 | 175 | 175 | 175 |

가설 : $H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4 = \frac{1}{4}$

검정통계치 : $\chi^2 = \frac{(176-175)^2}{175} + \cdots + \frac{(171-175)^2}{175} = 8.149$

기각역 : $\chi^2 \geq \chi^2_{0.05,3} = 7.81$

결론: 귀무가설을 기각할 수 있다. 계절에 따른 출생비율이 차이가 있다.

② 봄에 태어날 확률에 대한 95% 신뢰구간을 구하여라.

$$p=\frac{176}{700}=0.251,$$

$$p\pm z_{0.025}\sqrt{\frac{p(1-p)}{n}}=0.251\pm 1.96(0.016)=0.251\pm 0.031$$

③ 봄 또는 여름에 태어날 확률이 가을 또는 겨울에 태어날 확률보다 높은지를 5% 유의수준에서 검정하여라.

$$H_0:\theta_1\le 0.5\ ,\quad H_1:\theta_1>0.5$$

$$p_1=\frac{176+203}{700}=0.541$$

검정통계치 $z=\dfrac{0.541-0.5}{\sqrt{0.5\times 0.5/700}}=2.17$

기각역 : $z>z_{0.05}=1.645$

결론 : 귀무가설을 기각할 수 있다. 봄 또는 여름에 태어날 확률이 가을 또는 겨울에 태어날 확률보다 높다고 할 수 있다.

13.3 다음 자료는 무작위로 선정된 남성 아이돌 그룹 18팀, 여성 그룹 17팀 멤버의 혈액형 자료를 정리한 것이다.

| 성별 | 혈액형 | | | | 합계 |
|---|---|---|---|---|---|
| | A | B | O | AB | |
| 남 | 52 | 28 | 26 | 11 | 117 |
| 여 | 26 | 23 | 23 | 16 | 88 |
| 합계 | 78 | 51 | 49 | 27 | 205 |

(1) 우리나라 혈액형 분포는 아래와 같다고 한다. 남자그룹 멤버의 혈액형 분포가 우리나라 혈액형 분포와 같은지를 유의수준 5%에서 검정하고 차이가 있다면 어떤 부분에서 차이가 있는지를 설명하여라.

| 혈액형 | A | B | O | AB |
|---|---|---|---|---|
| 비율 | 34% | 27% | 28% | 11% |

| 성별 | 혈액형 | | | | 합계 |
|---|---|---|---|---|---|
| | A | B | O | AB | |
| 남 | 52 | 28 | 26 | 11 | 117 |
| 기대도수 | 39.78 | 31.59 | 32.76 | 12.87 | 117 |

가설 : $H_0 : \theta_A = 0.34,\ \theta_B = 0.27,\ \theta_o = 0.28,\ \theta_{AB} = 0.11$

검정통계치 : $\chi^2 = 5.828$

기각역 : $\chi^2 \geq \chi^2_{0.05,3} = 7.81$

결론: 귀무가설을 기각할 수 없다. 남자그룹 멤버의 혈액형 분포가 우리나라 혈액형 분포와 같다고 할 수 있다.

(2) 남녀 멤버 간에 혈액형 비율에 차이가 있는지를 알아보고자 한다.
① 어떤 통계적 모형으로 설명해야 하는지 기술하고 분석목적에 맞게 가설을 설정하여라.
- 동질성 검정
- $H_0 : \theta_{11} = \theta_{21}, \theta_{12} = \theta_{22}, \theta_{13} = \theta_{23}, \theta_{14} = \theta_{24}$
② 남녀 멤버 간에 혈액형 분포 차이여부에 대해 5%유의수준으로 검정하고 차이가 있다면 어떤 차이가 있는지를 수정잔차를 이용하여 설명하여라.
<기대도수>

| 성별 | 혈액형 | | | | 합계 |
|---|---|---|---|---|---|
| | A | B | O | AB | |
| 남 | 44.517 | 29.107 | 27.966 | 15.410 | 117 |
| 여 | 33.483 | 21.893 | 21.034 | 11.590 | 88 |
| 합계 | 78 | 51 | 49 | 27 | 205 |

검정통계치 : $\chi^2 = 6.29$

기각역 : $\chi^2 \geq \chi^2_{0.05,3} = 7.81$

결론 : 귀무가설을 기각할 수 없다. 남녀 멤버 간에 혈액형 분포가 같다고 할 수 있다.

13.6 다음 자료는 림프종을 치료하기 위해 시토산+프레드니손(CP)와 BCNU + 프레드니손(BP)을 처치한 임상시험 결과이다. 138명에게 CP를 처치하고 135명에게 BP를 처리한 후 환자의 종량의 크기변화를 기준으로 4그룹으로 나누었다고 하자. 이 임상시험이 목적은 두 처지 간 처리효과에 차이가 있는지를 검정하는 것이다.

| 종량크기 | 현저히 작아짐 | 부분적으로 작아짐 | 변화없음 | 더 진행됨 | 합 |
|---|---|---|---|---|---|
| CP | 26 | 51 | 21 | 40 | 138 |
| BP | 31 | 59 | 11 | 34 | 135 |
| 합계 | 57 | 110 | 32 | 74 | 273 |

(1) 어떤 통계모형으로 설명해야 하는지 기술하고 그 모형을 근거로 위의 분할표에 비율을 추가하여라.

- 동질성 검정 - $H_0 : \theta_{11} = \theta_{21}, \theta_{12} = \theta_{22}, \theta_{13} = \theta_{23,} \theta_{14} = \theta_{24}$

| 종량크기 | 현저히 작아짐 | 부분적으로 작아짐 | 변화없음 | 더 진행됨 | 합 |
|---|---|---|---|---|---|
| CP | 0.188 | 0.370 | 0.152 | 0.290 | 1 |
| BP | 0.230 | 0.437 | 0.081 | 0.252 | 1 |

(2) 두 처치 간 처리효과에 차이가 있는지를 검정하고 차이가 있는 경우

어떤 부분에서 차이가 있는지를 수정잔차를 이용하여 분석하여라.

<기대도수>

| 종량크기 | 현저히 작아짐 | 부분적으로 작아짐 | 변화없음 | 더 진행됨 | 합 |
|---|---|---|---|---|---|
| CP | 28.813 | 55.604 | 16.176 | 37.407 | 138 |
| BP | 28.187 | 54.396 | 15.824 | 36.593 | 135 |
| 합계 | 57 | 110 | 32 | 74 | 273 |

검정통계치 : $\chi^2 =$ 4.599

기각역 : $\chi^2 \geq \chi^2_{0.05,3} = 7.81$

결론: 귀무가설을 기각할 수 없다. 두 처치 간 처리효과에 차이가 없다고 할 수 있다.

13.10 다음 자료는 애완용 새를 키우는 것이 폐암에 대한 위험요소가 되는지를 알아보기 위한 조사결과이다. 표에서 사례는 폐암에 걸린 사람이고 대조는 건강인을 의미하며 경험이 있다는 것은 성인 때 애완용 새를 키웠다는 것을 의미한다.

| 애완용 새 | 사례 | 대조 |
|---|---|---|
| 경험있음 | 98 | 101 |
| 경험없음 | 141 | 328 |
| 합계 | 239 | 429 |

(1) 분석목적에 적합한 통계적 모형을 기술하고 가설을 설정하여라.

- 독립성 검정

가설:

$H_0$ : 애완용 새를 키우는 것과 폐암에 대한 위험요소는 관계가 없다.

(2) 이 자료를 이용하여 애완용 새를 키우는 것이 폐암에 대한 위험요소

가 되는지를 5%유의수준에서 검정하여라.

<기대도수>

| 애완용 새 | 사례 | 대조 |
|---|---|---|
| 경험있음 | 71.1991 | 127.8009 |
| 경험없음 | 167.8009 | 301.1991 |
| 합계 | 239 | 429 |

검정통계치 : $\chi^2$ = 22.374 기각역 : $\chi^2 \geq \chi^2_{0.05,1} = 3.84$

결론 : 귀무가설을 기각할 수 있다. 애완용 새를 키우는 것은 폐암에 대한 위험요소와 관계가 있다.

수정잔차

| 애완용 새 | 사례 | 대조 |
|---|---|---|
| 경험있음 | 4.730134 | -4.73013 |
| 경험없음 | -4.73013 | 4.730134 |

비율

| 애완용 새 | 사례 | 대조 |
|---|---|---|
| 경험있음 | 0.492462 | 0.507538 |
| 경험없음 | 0.30064 | 0.69936 |

사례의 경우 경험있음이 경험없음 보다 상대적으로 비율이 높을 것이고 대조는 경험없음이 경험있음 보다  상대적으로 비율이 높을 것을 알 수 있다.