

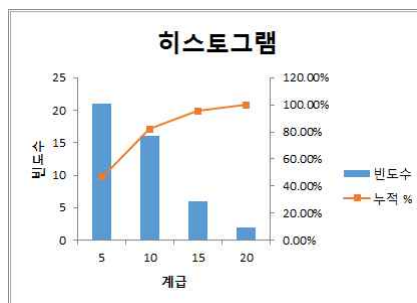
<EXCEL을 이용한 의학데이터의 통계분석 제3판> 해답

- 3장 -

1-①

계급	빈도수	누적 %
5	21	46.67%
10	16	82.22%
15	6	95.56%
20	2	100.00%

②



③ 46.67%

3.



5-① 평균 : 22.25, 중앙값 : 20, 최빈수 : 17

② 분산 : 44.93182, 표준편차 : 6.70312

③ 변동계수 : 30.1264

7-① 평균 : 4.074359, 표준편차 : 0.696502

② 중앙값 : 4, 사분위범위 : 1.3

③ 왜도 : 0.077225, 첨도 : 1.101834, 평균을 중심으로 오른쪽으로 긴 꼬리를 가지는 형태를 하고 있다.

- 4장 -

1-① $18/100=0.18$

② $8/40=0.2$

③ $8/100=0.08$

3-① $\binom{7}{5}(0.85)^5(0.15)^2 = 0.209651$

② $\binom{7}{5}(0.85)^5(0.15)^2 + \binom{7}{6}(0.85)^6(0.15)^1 + \binom{7}{7}(0.85)^7(0.15)^0 = 0.92635$

5-① $P(Z \geq \frac{200-140}{50}) = 0.11507$

② $P(\frac{100-140}{50} \leq Z \leq \frac{200-140}{50}) = 0.673075$

9. 표본의 비율의 분포는 근사적으로 $\hat{p} \sim N(0.35, 0.001138)$ 이다.

그러므로 $P(\hat{p} \geq 0.4) = P(Z \geq \frac{0.4-0.35}{\sqrt{0.001138}}) = P(Z \geq 1.48) = 0.069$

- 5장 -

1. $(147.4 - 1.96 \frac{2.45}{\sqrt{31}}, 147.4 + 1.96 \frac{2.45}{\sqrt{31}}) = (146.54, 148.26)$

3-① $(33.3 - 2.58 \frac{12.14}{\sqrt{242}}, 33.3 + 2.58 \frac{12.14}{\sqrt{242}}) = (31.49, 35.51)$

② 양측 검정 시, p-value가 <0.0001로 모평균이 30이라는 가설을 기각한다.

5-① 모평균 점추정치 : 113.28, 모분산 점추정치 : 200.09

② 모집단이 정규분포를 따른다고 가정하자. 그러면 95% 신뢰구간은

$(113.28 - t_{0.025}(17) \frac{14.15}{\sqrt{18}}, 113.28 + t_{0.025}(17) \frac{14.15}{\sqrt{18}}) = (106.24, 120.32)$ 이다.

③ 포함되어 있다. 95% 신뢰구간의 의미는 이와 같은 조사를 1,000번 시행해서 이러한 방법으로 1,000개의 신뢰구간을 구했을 때 이중 950개 정도의 구간은 실제로 참값인 모평균을 포함한다는 의미이다.

7. 평균 생존률 : $6/52=0.1154$

95% 신뢰구간 :

$(0.1154 - 1.96 \sqrt{\frac{0.1154*0.8846}{52}}, 0.1154 + 1.96 \sqrt{\frac{0.1154*0.8846}{54}}) = (0.0252, 0.2022)$

9-① $H_0 : p = 0.6$ vs. $H_1 : p > 0.6$

② 귀무가설을 기각한다.

③ 0.0008

- 6장 -

$$1-\textcircled{1} \left((2.1-1.6)-1.96\sqrt{1.1^2/79+1.2^2/74}, (2.1-1.6)+1.96\sqrt{1.1^2/79+1.2^2/74} \right) \\ = (0.1345, 0.8655)$$

② X 를 비당뇨 환자의 심전반사 수치, Y 를 당뇨 환자의 심전반사 수치로 하였을 때,

$$Z = \frac{(\bar{x} - \bar{y}) - \delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{2.1 - 1.6}{\sqrt{1.1^2/79 + 1.2^2/74}} = 2.68 \text{으로 } Z_{0.01} = 2.33 \text{보다 크다. 그러므}$$

로 유의수준 1% 하에서 비당뇨 환자에 비해 당뇨환자의 심전반사 수치가 유의하게 낮다.

3. 먼저, 두 그룹의 모분산에 대한 검정을 실시한다. $F = 0.8527$ 로 $F_{0.975}(7, 7) = 0.2002 < F < F_{0.025}(7, 7) = 4.9949$ 이므로 유의수준 5%하에서, 분산이 동일하다는 귀무가설을 기각하지 못하였다. 그러므로 등분산을 가정한 t-검정을 실시한다. $t = -0.7160$ 로 $t_{0.025}(14) = 2.1448$ 보다 작으므로 유의수준 5% 하에서 동맥압이 각 단계별로 동일하다는 귀무가설을 기각하지 못한다.

5. 먼저, 두 그룹의 모분산에 대한 검정을 실시한다. $F = 1.5182$ 로 $F_{0.975}(19, 8) = 0.3383 < F < F_{0.025}(19, 8) = 4.0158$ 이므로 유의수준 5% 하에서 분산이 동일하다는 귀무가설을 기각하지 못하였다. 그러므로 등분산을 가정한 t-검정을 실시한다. $t = 1.2827$ 로 $t_{0.025}(27) = 2.0518$ 보다 작으므로 유의수준 5% 하에서 평균 두께가 동일하다는 귀무가설을 기각하지 못하였다.

7. 먼저, 두 그룹의 모분산에 대한 검정을 실시한다. $F = 1.5625$ 로 $F_{0.975}(11, 10) = 0.2836 < F < F_{0.025}(11, 10) = 3.6649$ 이므로 유의수준 5% 하에서 분산이 동일하다는 귀무가설을 기각하지 못하였다. 그러므로 등분산을 가정한 t-검정을 실시한다.

$$t = \frac{120 - 70}{9.1026 \sqrt{\frac{1}{12} + \frac{1}{11}}} = 13.1592 \text{로 } t_{0.025}(21) = 2.0796 \text{보다 크므로 유의수준 5\% 하에서 두}$$

집단의 평균 체중감소량은 유의하게 다르다.

$$9. Z = \frac{0.7143 - 0.3548}{\sqrt{\frac{0.4211(1-0.4211)}{35} + \frac{0.4211(1-0.4211)}{155}}} = 3.8902 \text{로 } Z_{0.025} = 1.96 \text{보다 크므로 유}$$

의수준 5% 하에서 두 집단의 발병비율은 유의하게 다르다.

- 7장 -

1-① $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. $H_1 : \mu_i$ 들이 모두 같지는 않다.

② 4개의 모집단은 독립이며 정규분포를 따른다. 또한 분산이 동일하다.

③ 유의수준 5% 하에서 각 처리의 평균은 유의하게 다르다.

변동의 요인	제곱합	자유도	제곱 평균	F 비	P-값	F 기각치
처리	15865083	3	5288361	27.36453	4.09E-06	3.343889
잔차	2705585	14	193256.1			
계	18570669	17				

3-① 이원분산분석을 실시한다. $F = 21.3846$ 로 $F_{0.05}(2, 8) = 4.4590$ 보다 크므로 유의수준 5% 하에서 방법에 따라 사용방법을 숙지하는 시간이 유의하게 다르다.

② 이원분산분석을 실시한다. $F = 14.3846$ 로 $F_{0.05}(4, 8) = 3.8379$ 보다 작으므로 유의수준 5% 하에서 나이에 따라 사용방법을 숙지하는 시간이 동일하다는 귀무가설을 기각하지 못하였다.

5. 반복있는 이원 배치 분산분석을 실시한다. 약물에 대한 효과를 살펴보면 $F = 7$ 로 $F_{0.05}(2, 18) = 3.5546$ 보다 크므로 유의수준 5% 하에서 약물에 따라 유의하게 다르다. 상황에 따른 효과를 살펴보면 $F = 16.7778$ 로 $F_{0.05}(2, 18) = 3.5546$ 보다 크므로 유의수준 5% 하에서 상황에 따라 유의하게 다르다. 약물과 상황의 교호작용을 살펴보면 $F = 3.7778$ 로 $F_{0.05}(4, 18) = 2.9277$ 보다 크므로 유의수준 5% 하에서 상황과 약물의 교호작용에 따라 유의하게 다르다.

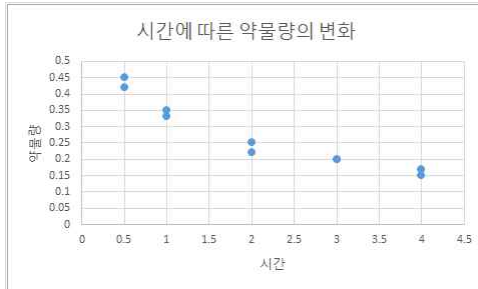
- 8장 -

1-① 0.8573

② $t = -0.8573 \sqrt{\frac{12-2}{1-(-0.8573)^2}} = -5.2660$ 로 $t_{0.025}(10) = 2.2281$ 보다 절대값이 크므로 유의수준 5% 하에서 칼슘과 콜레스테롤 간에는 상관계수가 0이라는 귀무가설을 기각한다(대립가설: $\rho \neq 0$).

3. 0.5813

5. $Y = 0.4301 - 0.0743X$

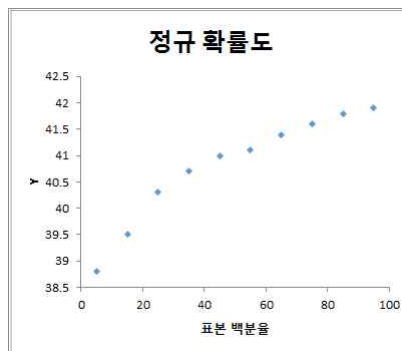
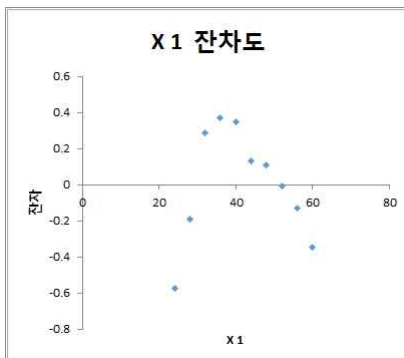


7-① $t = -0.9519 \sqrt{\frac{10-2}{1-(-0.9519)^2}} = 8.7869$ 로 $t_{0.025}(8) = 2.3060$ 보다 크므로 유의수준 5% 하에서 시간과 체온 간에는 유의한 직선관계가 존재한다.

② $Y = 37.4564 + 0.0798X$

절편의 $t = 94.6104$ 이고 X 회귀계수의 $t = 8.7821$ 로, $t_{0.025}(8) = 2.3060$ 보다 크므로 유의수준 5% 하에서 회귀계수는 0과 유의하게 다르다.

③ 잔차도를 보았을 때, 오차항의 독립성 위반에 대한 의심을 할 수 있다. 더빈-왓슨 통계량을 계산하면 0.5818로 0에 상당히 가까운 것을 확인할 수 있다. 또한 정규 확률도상에서도 곡선의 경향이 나타난다.



9. 적합된 회귀식은 다음과 같다.

$$PEmax = 153.0385 - 2.1145 * \text{나이} - 0.3948 * \text{키} + 2.8349 * \text{체중}$$

$$+ 1.7416 * BMP + 1.2651 * FEV + 0.1779 * RV - 0.2483 * FRC + 0.2084 * TLC$$

이 때, 단계적선택법(stepwise selection)으로 설명변수를 선택하여 적합된 최종 회귀식은 아래와 같다.

$$PEmax = 63.9467 + 1.7489 * \text{체중} - 1.3772 * BMP + 1.5477 * FEV + 0.1257 * RV$$

- 9장 -

- 1-① 상대위험률 : 2.1635, 오즈비 : 2.4405
② 상대위험률 95% 신뢰구간 : (0.7734, 6.0517)
오즈비 95% 신뢰구간 : (0.7235, 8.2322)
③ 상대위험률과 오즈비 모두 신뢰구간에 1을 포함하기 때문에, 위험요인과 산후우울 사이에 관계가 없다는 귀무가설을 기각할 수 없다.
3. 오즈비 : 1.0481, 오즈비 95% 신뢰구간 : (0.9054, 1.2133)
오즈비의 신뢰구간이 1을 포함하기 때문에 구강피임약의 복용과 유방암 사이에 관계가 없다는 귀무가설을 기각할 수 없다.
5. 카이제곱 검정을 실시한다. $\chi^2 = 20.6061$ 로 $\chi^2_{0.05}(2) = 5.9917$ 보다 크므로 유의수준 5% 하에서 HPV 감염 여부와 HIV 감염 여부사이에는 관계가 유의하게 존재한다.
7. 카이제곱 검정을 실시한다. $\chi^2 = 8.5748$ 로 $\chi^2_{0.05}(4) = 9.4877$ 보다 작으므로 유의수준 5% 하에서 동의 수준의 분포와 성별사이에는 관계가 없다는 귀무가설을 기각할 수 없다..
- 9-① 민감도 : 0.6279, 특이도 : 0.8230
② 양성예측도 : 0.2827, 음성예측도 : 0.9522

- 10장 -

1. 순위합 검정을 실시한다. 수정된 $Z = -2.5841$ 로 $Z < Z_{0.95} = -1.65$ 이므로 유의수준 5% 하에서 흡연자 집단의 카드뮴 수치가 비흡연자 집단의 수치보다 유의하게 높다.
3. 부호순위 검정을 실시한다. 수정된 $Z = -1.1679$ 로 $Z_{0.975} = -1.96 < Z$ 이므로 유의수준 5% 하에서 처리전과 처리도중의 좌심실의 크기가 동일하다는 귀무가설을 기각하지 못하였다.
5. 크루스칼-왈리스 검정을 실시한다. 수정된 $\chi^2 = 5.7254$ 로 $\chi^2 < \chi_{0.05}^2(2) = 5.9915$ 이므로 유의수준 5% 하에서 각 처리 방법에 따른 두통의 경감정도가 동일하다는 귀무가설을 기각하지 못하였다.
7. 0.7544