$\chi^2_{0.01}(19) = 36.1908$이므로, 검정통계량은 기각역에 포함되어 귀무가설을 기각한다.
따라서 기업 생산성의 분산이 과거보다 커졌다고 볼 수 있다.

22 귀무가설을 기각한 경우, 대립가설이 참인데도 귀무가설을 채택한 제2종의 과오를 범했을 수는 없다.

23 귀무가설을 채택한 경우, 대립가설이 참일 때도 해당 표본관찰의 결과가 나올 확률이 있었다면 제2종의 과오를 범했을 가능성이 있다.

24 $f(x_1, x_2;\theta) = \theta^2 x_1^{\theta-1} x_2^{\theta-1},\ 0 < x_1 < 1,\ 0 < x_2 < 1$

검출력함수 $K(\theta) = P[(X_1, X_2) \in R] = 1 - \int_0^1 \int_0^{\frac{3}{4}x_1} \theta^2 x_1^{\theta-1} x_2^{\theta-1} dx_2 dx_1$

$$= 1 - \frac{1}{2}\left(\frac{3}{4}\right)^{\theta}$$

본문에서 정의한 검출력 $1-\beta = 1 - P\left[(X_1, X_2) \in A | H_0 \text{ is false}\right]$

$$= P\left[(X_1, X_2) \in R | H_a\right]$$

$\therefore 1-\beta = K(2) = 1 - \frac{1}{2}\left(\frac{3}{4}\right)^2 = \frac{23}{32}$

25 $\overline{X} \sim N\left(\mu, \frac{5{,}000^2}{n}\right)$

$$K(\mu) = P[(X_1, X_2, \cdots, X_n) \in R] = P\left[\frac{\sqrt{n}(\overline{X}-\mu)}{5{,}000} > \frac{\sqrt{n}(c-\mu)}{5{,}000}\right]$$

$$= 1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{5{,}000}\right)$$

$1 - K(30{,}000) = P\left[Z < \frac{\sqrt{n}(c-30{,}000)}{5{,}000}\right] = 0.99$에서 $\sqrt{n}(c-30{,}000) = 5{,}000 \cdot (2.33)$

$1 - K(35{,}000) = P\left[\frac{\sqrt{n}(c-35{,}000)}{5{,}000}\right] = 0.02$에서 $\sqrt{n}(c-35{,}000) = 5{,}000 \cdot (-2.05)$

$\sqrt{n} = 4.38$에서 $n = 19$또는20(개) $c = 32{,}659.82$

제10장 두 모집단에 대한 검정

1 1) $D = \overline{X} - \overline{Y}$라 할 때,

$\sigma_D^2 \approx \frac{5^2}{100} + \frac{5^2}{120} = 0.4583,\ Z = \frac{20-24}{\sqrt{0.4583}} = -5.91,\ -Z_{0.025} = -1.96$

$\therefore H_0 : \mu_X - \mu_Y = 0$을 기각

2) $u = \frac{1}{\sigma^2}\{(10-1)5^2 + (12-1)5^2\} = \frac{1}{\sigma^2}(500),$

$\frac{u}{10+12-2} \cdot \sigma^2\left(\frac{1}{10} + \frac{1}{12}\right) = 4.583$

$$t=\frac{20-24}{\sqrt{4.583}}=-1.868,\ -t_{0.025,\,21}=-2.0796 \quad \therefore H_0:\mu_X-\mu_Y=0\text{을 채택}$$

2 $D=X_1-X_2$라 할 때,

$$\sigma_D^2\approx\frac{15^2}{100}+\frac{10^2}{120}=3.0833,\ Z=\frac{75-80}{\sqrt{3.0833}}=-2.85,\ -Z_{0.005}=-2.58$$

$\therefore H_0:\mu_1-\mu_2=0$을 기각

두 반 사이에 우열이 없다고 할 수 없다.

3 $S_Y^2/S_X^2=2.7778,\ F_{19,\,14,\,0.01}=3.51,\ F_{19,\,14,\,0.05}=2.44$

유의수준 0.01에서는 두 모집단의 분산이 동일하다고 할 수 있으나, 유의수준 0.05에서는 그렇게 말할 수 없다.

4 $$\hat{p}=\frac{250+170}{400+350}=0.56,\ 0.56(1-0.56)\left(\frac{1}{400}+\frac{1}{350}\right)=0.00132$$

$$Z=\frac{0.625-0.486}{\sqrt{0.00132}}=3.82,\ Z_{0.005}=2.58$$

$\therefore H_0:p_1=p_2$를 기각, 선호율이 동일하다고 할 수 없다.

5 두 수송방법에 의한 표본비율 중 큰 쪽을 $\hat{p}_1$, 다른 쪽을 $\hat{p}_2$라 하고,

$H_0:p_1-p_2=0.01, H_a:p_1-p_2<0.01$을 유의수준 0.05로 검정할 때,

$p_1=p_2=p,\ n_1=n_2=n$이면

$$Z=\frac{\frac{X_1}{n_1}-\frac{X_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{\frac{X_1-X_2}{n}}{\sqrt{p(1-p)\left(\frac{2}{n}\right)}}\text{이다. 또한}$$

$\sigma=\sqrt{p(1-p)\frac{2}{n}}$는 $p=\frac{1}{2}$일 때 극대화되어 $\sigma=\sqrt{\frac{1}{2n}}$이다.

따라서 $n\geq\frac{4(z_{0.025})^2\sigma^2}{E^2}=\frac{(4)(1.96)^2\left(\frac{1}{2n}\right)}{(0.01)^2}$이며 $n\geq 277.2$가 되므로 표본크기는 278개 이상이어야 한다.

6 두 모집단의 평균 μ_A, μ_B를 비교하기 위해, 귀무가설 $H_0:\mu_A=\mu_B$를 대립가설 $H_a:\mu_A\neq\mu_B$에 대하여 유의수준 $\alpha=0.05$로 검정한다고 하자.

	가족경영기업	비 가족경영기업
수익률	13.24	16.77
분산	35.33	44.76
표본 숫자	42	98

먼저 위 가설을 검정하기 위하여 동분산 가정을 사용할 수 있는지 여부를 알아본다. 이를 위해 아래의 가설을 검정한다. 가설검정을 위한 귀무가설 및 대립가설은 다음과 같다.

$$\begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_a : \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1 \end{cases}$$

위 가설을 검정하려면, F-분포를 이용해야 한다. 집단 1(가족경영기업)의 표본수는 $140 \times 0.3 = 42$이므로, 집단 1의 자유도는 41이며, 집단 2의 자유도는 $98 - 1 = 97$이 된다. 10%의 유의수준에서 기각역이 되는$F-$ 값은 다음과 같다.

$F(0.05, 41, 97) = 0.635$, $F(0.95, 41, 109) = 1,501$.

한편 F통계량의 값은1.266912이므로, 기각역에 속하지 않는다. 따라서 귀무가설을 기각하지 않는다. 즉, 동분산 가정 하에서 문제를 풀게 된다.

두 그룹의 표본이 $n_A = 42, n_B = 98$이고, 각 표본의 $\overline{X}_A = 13.24, \overline{X}_B = 16.77, S_A^2 = 35.33, S_B^2 = 44.76$이므로 이로부터 검정통계량의 값은 다음과 같다.

$$z = \frac{(X_A - X_B) - 0}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}}$$

이를 계산하면 값이 -3.9849이고, 유의수준이 $\alpha = 0.05$인 양측검정이므로, 기각역은 $z > 1.96, z < -1.96$임을 알 수 있다. z값이 기각역에 속하므로, 귀무가설이 기각된다. 이 결과는 가족경영기업의 수익률이 전문경영인이 경영하는 기업보다 떨어짐을 의미한다.

7 해당 문제의 귀무가설과 대립가설은 다음과 같다.

$$\begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_a : \dfrac{\sigma_1^2}{\sigma_2^2} > 1 \end{cases}$$

대립가설을 볼 때 단측검정을 한다. $F-$분포에서 유의수준 5%에서의 기각역은 다음과 같다.

$F(0.95, 29, 29) = 1,861$. 위의 표본분산의 비율을 구하면 $F-$값 $= 6.349/4.681 = 1.356$이므로, 이는 임계값보다 작은 값이므로, 귀무가설을 기각하지 않는다.

이는 30일 간의 관측치로부터는 A사의 주식이 B사의 주식보다 더 위험한 (수익률의 변동이 크다고) 주식이라고 말할 수 없음을 의미한다.

8 귀무가설 및 대립가설은 다음과 같다.

$$H_0 : p_1 - p_2 = 0$$
$$H_a : p_1 - p_2 > 0$$

검정통계량은 다음과 같다.

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})(\dfrac{1}{n_1} + \dfrac{1}{n_2})}}$$

각 표본비율이 0.1991 및 0.1493이며, 통합 모비율 추정치는 다음과 같다.

$$\hat{p} = \frac{180+155}{904+1,038} = \frac{335}{1,942} = 0.1725$$

이 경우 검정통계량을 계산하면,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{(0.1991-0.1493)}{\sqrt{0.1725*(1-0.1725)(\frac{1}{904}+\frac{1}{1,038})}} = 2.90$$이 되므로,

5% 및 1% 유의수준에서의 기각역인 1.65와 2.58을 넘어서는 수치가 나옴을 알 수 있다. 즉 귀무가설을 기각한다.

9 뉴저지에 속한 주유소의 집단을 A, 펜실베이니아 주에 속한 주유소의 집단을 B라고 하자. 두 모집단의 평균 μ_A, μ_B를 비교하기 위해, 귀무가설 $H_0 : \mu_A = \mu_B$를 대립가설 $H_a : \mu_A < \mu_B$에 대하여 유의수준 $\alpha = 0.01$로 검정한다고 하자.
먼저 두 집단의 동분산성을 검정하면, 다음과 같다.
모집단 A와 B로부터 표본크기 $n_A = 50$, $n_B = 50$인 표본을 추출하였다.
귀무가설은 $H_0 : \sigma_A^2 = \sigma_B^2$이고, 대립가설은 $H_a : \sigma_A^2 \neq \sigma_B^2$이다. 유의수준은 $\alpha = 0.05$로 검정한다고 하자. 검정통계량은

$$F(n_B-1, n_A-1) = \frac{S_B^2}{S_A^2}$$이고, 이 식에 대입하여 F값을 구하면,

$F = \frac{4.5526}{6.7656} = 0.672904$이다. F분포표로부터, 기각역이 $F(0.025, 50, 120) = 1.762$이고, $F(0.975, 50, 120) = 0.567$이다. F값이 기각역의 범위 사이에 있으므로, 귀무가설을 기각하지 않는다. 즉, 동분산 가정 하에 문제를 풀 수 있다.
두 그룹의 표본이 $n_A = 50, n_B = 50$이고, 각 표본의 $\overline{X}_A = 3.2, \overline{X}_B = 6.3, S_A^2 = 6.7656, S_B^2 = 4.5526$이므로 이로부터 검정통계량의 값은 다음과 같다.

$$z = \frac{(\overline{X}_A - \overline{X}_B) - 0}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

이를 계산하면 값이 -6.516이고, 유의수준이 $\alpha = 0.01$인 단측이므로, 기각역은 $z < -2.33$이다. z값이 기각역의 값 -2.33보다 작으므로, 귀무가설이 기각된다. 이는 뉴저지 주에 비해 펜실베이니아 주에 소재한 주유소에 셀프주유기가 평균적으로 많음을 시사한다. 즉 펜실베이니아 주의 주유소가 최저임금의 상승으로 인해 자동화를 선택했을지도 모른다는 가설을 지지한다고 볼 수 있다.

제11장 상관분석과 단순회귀분석

1 $\sum X_i^2 - n\overline{X}^2 = 44, \sum Y_i^2 - n\overline{Y}^2 = 34, \sum X_i Y_i - n\overline{X}\,\overline{Y} = 24$
$\gamma_{XY} = 24/\sqrt{44 \cdot 34} \approx 0.621$

2 1) $\gamma_{XY} = 1,600/\sqrt{1,600 \times 1,800} \approx 0.943$

2) $b=\dfrac{1{,}600}{1{,}600}=1,\ \hat{Y}-60=1\cdot(X-50),\ \hat{Y}=10+X$

3 1) $\hat{Y}-25=-5(X-6),\ \hat{Y}-55=-5X\quad R^2=\dfrac{(-5)(-400)}{860}=\dfrac{100}{43}$

2) $\hat{X}-6=-\dfrac{20}{43}(Y-25),\ \hat{X}=\dfrac{758}{43}-\dfrac{20}{43}Y,\ R^2=\left(-\dfrac{20}{43}\right)(-400)/80=\dfrac{100}{43}$

3) $bB=(-5)\left(-\dfrac{20}{43}\right)=\dfrac{100}{43}$

4) $bB=\dfrac{\Sigma(X_i-\overline{X})(Y_i-\overline{Y})}{\Sigma(X_i-\overline{X})^2}\cdot\dfrac{\Sigma(Y_i-\overline{Y})(X_i-\overline{X})}{\Sigma(Y_i-\overline{Y})^2}=\dfrac{[\Sigma(X_i-\overline{X})(Y_i-\overline{Y})]^2}{\Sigma(X_i-\overline{X})^2\ \Sigma(Y_i-\overline{Y})^2}=R^2$

4 1) $\hat{p}=(24-6\cdot 4)+6W=6W$

2) $SST=\ \Sigma(Y_i-\overline{Y})^2=\ \Sigma Y_i^2-n\overline{Y}^2=1{,}048$
$SSR=b\,\Sigma(X_i-\overline{X})(Y_i-\overline{Y})=b(\Sigma X_iY_i-n\,\overline{X}\,\overline{Y})=1{,}008$
$SSE=1{,}048-1{,}008=40$

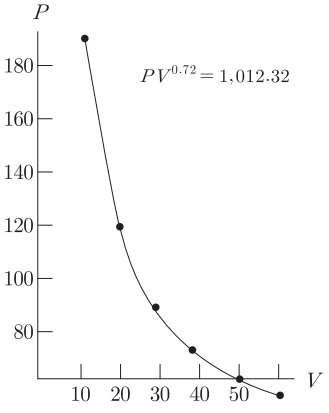
3) $R^2=\dfrac{1{,}008}{1{,}048}=1-\dfrac{40}{1{,}048}\approx 0.962$

5 1) 양변에 자연로그를 취하면 $\ln(PV^b)=\ln a,\ \ln P+b\ln V=\ln a$ 즉, $p=\alpha+\beta v$의 형태로 볼 수 있다.(이때 $p=\ln P,\ v=\ln V,\ \alpha=\ln a,\ \beta=-b$)

p	3.99	4.12	4.28	4.49	4.78	5.27
v	4.11	3.90	3.63	3.35	2.97	2.31

$\hat{p}-4.49=-0.72(v-3.38),\ \ \hat{p}=6.92-0.72v\quad\therefore a=e^{6.92}=1{,}012.32,\ b=0.72$

2)



3) $\hat{p}=6.92-0.72(\ln 100)=3.60\quad\hat{P}=e^{3.60}=36.60$

6 1) $(1,5.5),\ (2,7.5),\ (3,12.5),\ (4,15),\ (5,18.5)$

2) $\sum_{i=1}^{n}(X_i^2-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2=10$

$\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})=\sum_{i=1}^{n}X_iY_i-n\overline{X}\,\overline{Y}=33.5$

$b=3.35,\ a=\overline{Y}-3.35\overline{X}=1.75$

3) $\sum_{i=1}^{n}e_i^2=6,\ S_e^2=\frac{6}{3}=2,\ S_b^2=2/10=0.2$

$H_0:\beta=0,\ H_1:\beta\neq 0$의 검정에서 $t_{0.005,\,3}=5.8409,\ t=\frac{b}{S_b}=7.4908$이므로 유의수준 1%에서 귀무가설을 기각

$\therefore$ 체계적인 관계가 있다고 할 수 있다.

7 $Y=100,\ C=80,\ \Sigma(Y_i-\overline{Y})^2=200,\ \Sigma(C_i-\overline{C})^2=234,\ \Sigma(Y_i-\overline{Y})(C_i-\overline{C})=120$

1) $b=120/200=\frac{3}{5}\quad a=80-\frac{3}{5}\cdot 100=20$

2) $SSE=\Sigma e_i^2=234-\left(\frac{3}{5}\right)^2\cdot 200=162,\ S_e^2=162/2=81\ \ S_e=9$

3) $S_b^2=81/200,\ S_b=\frac{9}{10\sqrt{2}}=0.636\quad S_a^2=81\cdot\frac{40{,}200}{4\cdot 200}=4{,}070.25,\ S_a=63.799$

4) $R^2=\left(\frac{3}{5}\right)^2\left(\frac{200}{234}\right)=0.3077$

5) $t=\frac{b-0.5}{S_b}=0.157,\ t_{0.025,\,2}=4.3027$

$\therefore$ 유의수준 5%에서 $H_0:\beta=0.5$를 채택

8 문제 7의 전제에 의해 $S_i=-\alpha+(1-\beta)Y_i-\mu_i$ 역시 회귀모델 5개 가정을 모두 만족시킨다.

$\overline{S}=20,\ \Sigma(S_i-\overline{S})^2=194,\ \Sigma(Y_i-\overline{Y})(S_i-\overline{S})=80$

1) $(1-\hat{\beta})=80/200=\frac{2}{5},\ \ (-\hat{\alpha})=20-\frac{2}{5}\cdot 100=-20$

2) $\Sigma(e_i')^2=194-\left(\frac{2}{5}\right)^2\cdot 200=162,\ S_{e'}^2=162/2=81,\ S_{e'}=9$

3) $(R^2)'=\left(\frac{2}{5}\right)^2\left(\frac{200}{194}\right)=0.1649$

4) $\overline{S}=\overline{Y}-\overline{C},\ \Sigma(Y_i-\overline{Y})(S_i-\overline{S})=\Sigma(Y_i-\overline{Y})(Y_i-C_i-\overline{Y}+\overline{C})$

$=\Sigma(Y_i-\overline{Y})^2-\Sigma(Y_i-\overline{Y})(C_i-\overline{C})$

$-\hat{\alpha}=\overline{S}-(1-\hat{\beta})\overline{Y}=(\overline{Y}-\overline{C})-(1-\hat{\beta})\overline{Y}=-\overline{C}+\hat{\beta}\overline{Y}=-\hat{\alpha}$

$\Sigma e_i^2=\Sigma(S_i-\overline{S})^2-(1-\hat{\beta})^2\Sigma(Y_i-\overline{Y})^2$

$=\Sigma(Y_i-\overline{Y})^2-2\Sigma(Y_i-\overline{Y})(C_i-\overline{C})+\Sigma(C_i-\overline{C})^2-(1-2\hat{\beta}+\hat{\beta}^2)\Sigma(Y_i-\overline{Y})^2$

$=\Sigma(C_i-\overline{C})^2-\hat{\beta}^2\Sigma(Y_i-\overline{Y})=\Sigma e_i^2$

에서 $S_e^{'} = S_e$

하지만, $\Sigma(S_i - \overline{S})^2 \neq \Sigma(C_i - \overline{C})^2$ $\left(\text{즉, } \hat{\beta} \neq \frac{1}{2}\right)$일 때,

$\frac{1-R^2}{1-R^{2'}} = \frac{\Sigma(S_i - \overline{S})^2}{\Sigma(C_i - \overline{C})^2} \neq 1$에서 $R^2 \neq R^{2'}$

9 $Y_i = 20 + \frac{3}{5}X_i + \mu_i$에서 점예측치 $\widehat{C_0} = 20 + \frac{3}{5} \cdot 120 = 92$

$\mu_\epsilon + E(\varepsilon) = 0,\ \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon) = \sigma_u^2\left[1 + \frac{1}{4} + \frac{1}{200}(120-100)^2\right] = \frac{13}{4}\sigma_u^2$

$\therefore S_\varepsilon^2 = \frac{13}{4}S_e^2 = 263.25 \quad S_\varepsilon = 16.225$

$t_{0.025,\,2} = 4.3027,\ \ P(|\frac{\varepsilon}{S_\varepsilon}| < 4.3027) = P(|C_0 - 92| < 69.81) = 0.95$

$\therefore C_0$에 대한 95% 예측구간은 (22.19, 161.81)

10 1) $\widehat{Y} = a + b\overline{X} = (Y - b\overline{X}) + bX,\ \widehat{Y} - Y = b(X - \overline{X}),\ \hat{y} = bx$

$\overline{b} = \frac{\Sigma X_i Y_i - X\,\Sigma Y_i}{\Sigma X_i^2 - \overline{X}\Sigma X_i} = \frac{\Sigma X_i Y_i - n\overline{X}\,\overline{Y}}{\Sigma X_i^2 - n\overline{X}^2} = \frac{\Sigma(X_i - X)(Y_i - Y)}{\Sigma(X_i - X)^2}$

$= \frac{\Sigma(X_i - X)Y_i}{\Sigma(X_i - \overline{X})^2} = \frac{\Sigma X_i(Y_i - Y)}{\Sigma(X_i - \overline{X})^2}$

$(\because \Sigma(X_i - \overline{X})(\overline{Y_i} - \overline{Y}) = \Sigma X_i Y_i - \overline{X}\Sigma Y_i - \overline{Y}\Sigma X_i + n\overline{X}\,\overline{Y},\ \Sigma X_i = n\overline{X},\ \Sigma Y_i = n\overline{Y})$

따라서, $y = \frac{\Sigma x_i y_i}{\Sigma x_i^2}x = \frac{\Sigma x_i Y_i}{\Sigma x_i^2}x = \frac{\Sigma X_i y_i}{\Sigma x_i^2}x$

2) $\widehat{Y} - Y = b(X - 0),\ b = \frac{\Sigma X_i Y_i - 0 \cdot \Sigma Y_i}{\Sigma X_i^2 - 0 \cdot \Sigma X_i} \qquad \therefore \widehat{Y} = Y + \frac{\Sigma X_i Y_i}{\Sigma X_i^2}X$

11 1) $\frac{b}{S_b} = 0.05,\ S_b = 18$

H_0이 참일 때, $-t = \frac{b-\beta}{S_b} = \frac{0.9 - 0.83}{18} = 0.039$ 한편, $t_{0.025,\,20} = 2.0860$

$\therefore\ \beta = 0.83$은 채택

2) $P\left(\left|\frac{0.9-\beta}{18}\right| < 2.086\right) = 0.9$에서 β에 대한 95% 신뢰구간은 $(-36.648, 38.448)$

12 $Y_i^{'} = a^{'} + b^{'}X_i + e_i{'}$에 $X_i^{'} = X_i + A, Y_i^{'} = Y_i + B$를 대입하면

$Y_i + B = a^{'} + b^{'}(X_i + A) + e_i{'} - Y_i = (a^{'} - b^{'}A + B) + b^{'}X_i + e_i{'}$

정규방정식 $n(a^{'} - b^{'}A + B) + b^{'}\Sigma X_i = \Sigma Y_i$

$(a^{'} - b^{'}A + B)\Sigma X_i + b^{'}\Sigma X_i^2 = \Sigma X_i Y_i$를 풀면

$b^{'} = \frac{\Sigma X_i Y_i - X\Sigma X_i}{\Sigma X_i^2 - X\Sigma X_i} = b$

13 $Y_i = \alpha + \beta X_i + \epsilon_i$의 회귀계수는 $b = \dfrac{\sum_{i=1}^{n} Y_i(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$

여기에 실제 회귀식 $Y_i = \beta X_i + \epsilon_i$를 대입하고 기댓값을 구하면 다음과 같다.

$$E[b] = E\left[\frac{\sum_{i=1}^{n}(\beta X_i + \epsilon_i)(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right] = \frac{\sum_{i=1}^{n}\beta X_i(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} + \frac{\sum_{i=1}^{n}(X_i - \overline{X})E(\epsilon_i)}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \beta$$

따라서 b는 불편추정량이다.

14 정규방정식에 의해서 $\sum_{i=1}^{n} e_i = 0$이고, $\sum_{i=1}^{n} e_i X_i = 0$이다.

따라서 $\sum_{i=1}^{n} e_i \hat{Y}_i = \sum_{i=1}^{n} e_i(a + bX_i) = a\sum_{i=1}^{n} e_i + b\sum_{i=1}^{n} e_i X_i = 0$

15 1) $\Sigma(X_i - \overline{X})^2 = 10.5, \Sigma(X_i - \overline{X})(Y_i - \overline{Y}) = 21$에서 $b = 2$

$\Sigma e_i^2 = 84 - 2 \cdot 21 = 42 \quad S_e^2 = 42/200 = 0.21 \quad S_b^2 = 0.21/10.5 = 0.02$에서 $S_b = \dfrac{\sqrt{2}}{10}$

$\dfrac{2-\beta}{\frac{\sqrt{2}}{10}} \sim t(198),\ t_{0.025,\,198} \approx Z_{0.025} = 1.96$에서 $\dfrac{\sqrt{2}}{10} t_{0.025,\,198} = 0.277$

$\therefore \beta$에 대한 95% 신뢰구간은 $(1.723, 2.277)$

2) $\dfrac{(n-2)S_e^2}{\sigma_u^2} = \dfrac{\Sigma e_i^2}{\sigma_u^2} \sim \chi^2(200-2)$

(증명) $\Sigma u_i^2 = \Sigma(Y_i - a - bX_i + a + bX_i - \alpha - bX_i)^2$

$= n\overline{u}^2 + (b-\beta)^2\Sigma(X_i - \overline{X})^2 + \Sigma e_i^2$

$\Sigma\left(\dfrac{u_i}{\sigma_u}\right)^2 \sim \chi^2(n)$이며, $\left(\dfrac{\overline{u}-0}{\sigma_u/\sqrt{n}}\right)^2, \left(\dfrac{b-\beta}{\sigma_b}\right)^2 \sim \chi^2(1)$이므로

$\therefore \dfrac{1}{\sigma_u^2}\Sigma e_i^2 \sim \chi^2(n-2)$

$P(u < \dfrac{\Sigma e_i^2}{\sigma_u^2} | H_0) = P_r(U < 210) = 0.734(p\text{-value}) > 0.05$

$\therefore$ 유의수준 5%에서 $H_0 : \sigma_u^2 = 0.2$를 채택

16 $\hat{Y} = a + bX = (Y - b\overline{X}) + bX,\ \hat{Y} - \overline{Y} = b(X - \overline{X})$

17 1) $Y_i = \alpha + \beta X_i + u_i,\ \ X_i = \gamma + \delta Y_i + v_i$의 두 회귀모델에서

$\hat{\beta} = \dfrac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{\Sigma(X_i - \overline{X})^2},\ \ \hat{\delta} = \dfrac{\Sigma(Y_i - \overline{Y})(X_i - \overline{X})}{\Sigma(Y_i - \overline{Y})^2}$

$$\hat{\beta}\,\hat{\delta}=\frac{[\Sigma(X_i-\overline{X})(Y_i-\overline{Y})]^2}{\Sigma(X_i-\overline{X})^2\,\Sigma(Y_i-\overline{Y})^2}=(r_{XY})^2=R^2.\;\; R^2=1\text{인 경우}\;\hat{\delta}=\frac{1}{\hat{\beta}}$$

18 $a=\overline{Y}-b\overline{X}$에서 $\overline{X}\neq 0$이면 추정되는 절편은 기울기의 추정치와 반대방향으로 움직인다.

이때, $Cov(a,b)=-\overline{X}\sigma_b^2=-\dfrac{\overline{X}}{\Sigma(X_i-\overline{X})^2}\sigma_u^2<0\;(10.2.15)$

하지만, $\overline{X}=0$이면 $a=\overline{Y}$로 회귀선은 항상 $(0,\overline{Y})$를지나며 절편의 추정은 $\widehat{Y}_i-\overline{Y}=b(X_i-\overline{X})$인 기울기의 추정치 b와 무관하다.

이때, $Cov(a,b)=-0\cdot\sigma_b^2=0$

19 $E(X_iu_i)\neq 0$이면, $Cov(X_i,u_i)=E(X_iu_i)-E(X_i)E(u_i)=E(X_iu_i)\neq 0$

즉, 설명변수 X와 오차항 u사이에 상관관계가 존재한다.

한편, 최소이승법의 정규방정식은 X의 Y에 대한 영향이 오차항에 의한 영향과 선형적으로 분리되어 있다는 기본모형(PRF)을 전제로 오차항에 대응하는 전차항의 제곱을 최소화함으로써 X의 영향을 최대한 반영할 수 있는 회귀계수들을 구한다. 따라서 X와 u사이에 상관관계가 존재하여 각각의 Y에 대한 영향이 분리될 수 없는 경우에는 최소이승추정방법이 의미가 없다. (이 경우 회귀모수에 대한 OLS 추정치는 불편성 및 일치성을 갖지 못한다.)

즉 식(10.2.3)에 의하면 $b=\beta+\Sigma w_iu_i$이며 $E(b)=\beta\rightarrow E(\Sigma w_iu_i)$인데 $E(X_iu_i)\neq 0$이면 $E(w_iu_i)\neq 0$이고 $E(b)\neq\beta$가 되어 b가 불편추정량이 아니다. 따라서 $E(X_iu_i)=0$일 때 최소이승추정량 b가 불편추정량이 됨을 알 수 있다.

20 $\widehat{Y}_i=a+bX_i=\overline{Y}-b\overline{X}+bX_i$이므로

$$R^2=\frac{\sum_{i=1}^{n}(\widehat{Y}_i-\overline{Y})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}=\frac{\sum_{i=1}^{n}(\overline{Y}-b\overline{X}+bX_i-\overline{Y})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}=\frac{b^2\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}$$

여기서 $b=\dfrac{\sum_{i=1}^{n}(Y_i-\overline{Y})(X_i-\overline{X})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$를 대입하면,

$$R^2=\frac{b^2\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}=\frac{\left[\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})\right]^2}{\left[\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right]\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]}=\gamma_{XY}^2$$

1 1) $\widehat{M_t} = -276.10 + 1.84\,Y_t - 0.25\gamma_t$

2) $R^2 = 0.951$

$\Sigma e_t^2 = 28.998$

3) $S_e^2 = \dfrac{\Sigma e_i^2}{n-3} = 3.222$

$S_a = 214.79$

$S_{b_1} = 0.27$

$S_{b_2} = 9.04$

4) $T = \dfrac{b_1}{S_{b_1}} = \dfrac{1.84}{0.27} = 6.81$

유의수준 0.05에서 자유도는 $n-3=9$이므로 $t_{0.025} = 2.26$이다.

따라서 $t = 6.70$은 임계치 2.26보다 크므로 $\beta_1 = 0$의 가설을 기각한다.

5) $T = \dfrac{b_2}{S_{b_2}} = \dfrac{-0.25}{9.04} = -0.027$

$-t_{0.025} = -2.26 \le T = -0.027 \le 2.26$이므로

$\beta_2 = 0$을 기각하지 못한다.

2 1) $\Sigma E_i = 95.6 \quad \Sigma Y_i = 48,994 \quad \Sigma E_i Y_i = 526,872.4$

$\Sigma E_i^2 = 1,024.16 \quad \Sigma Y_i^2 = 273,829,152 \quad n = 9$

$\overline{E} = 10.622 \quad \overline{Y} = 5,443.778$

$$b_1 = \frac{526,872.4 - 9(10.622)(5,443.778)}{1,024.16 - 9(10.622)^2}$$

$$= \frac{6,458.11076}{8.718044} = 740.775$$

$a_1 = 5,443.778 - 740.775(10.622) = 2,424.734$

$\therefore Y = -2,424.734 + 740.775E$

$$\Sigma e_i^2 = \left[\Sigma Y_i^2 - n\overline{Y}^2\right] - b_1\left[\Sigma E_i Y_i - n\,\overline{E}\,\overline{Y}\right]$$

$$= [293,829,152 - 9 \times (5,443.778)^2]$$

$$= 740.775[526,872.4 - 9(10.622)(5,443.778)]$$

$$= 7,116,681.79 - 4,784,006.998$$

$$= 2,332,674.791$$

$s_e^2 = \dfrac{1}{7}\Sigma e_i^2 = 333,239.2559$

$s_e = 577.269$

$s_{b_1}^2 = s_e^2 / \Sigma(E_i - \overline{E})^2 = s_e^2 / 8.718044 = 38,224.08512$

$s_{b_1} = 195.51$

$$s_{a_1}^2 = s_e^2\left[\frac{1}{n}+\frac{\overline{E}^2}{\Sigma(E_i-\overline{E})^2}\right]$$

$$= 333,239.2559\left[\frac{1}{9}+\frac{(10.622)^2}{8.718044}\right]$$

$$= 4,349,731.001$$

$$R^2 = 1-\frac{\Sigma e_i^2}{\Sigma(Y_i-\overline{Y})^2} = 1-\frac{2,332,674.791}{7,116,681.79} = 0.6722$$

2) $\Sigma F_i = 75.9 \quad \Sigma F_i Y_i = 378,268.6$

$\Sigma F_i^2 = 934.27 \quad \overline{F} = 8.43$

$$b_2 = \frac{\Sigma F_i Y_i - n\overline{F}\,\overline{Y}}{\Sigma F_i^2 - n\overline{F}^2} = \frac{378,268.6-9(8.43)(5,443.778)}{934.27-9(8.43)^2}$$

$$= \frac{-34,750.83686}{294.6859} = -117.93$$

$$a_2 = 5,443.778-(-117.93)(8.43) = 6,473.93$$

$$\therefore \overline{Y} = 6,473.93-117.93\overline{F}$$

$$\Sigma e_i^2 = [\Sigma Y_i^2 - n\overline{Y}^2] - b_2[F_i Y_i - n\overline{F}\,\overline{Y}]$$

$$= 7,116,681.79+117.93[378,268.6-9(8.43)(5,443.778)]$$

$$= 3,018,515.599$$

$$s_e^2 = \frac{1}{7}\Sigma e_i^2 = 431,216.5141$$

$$s_e^2 = 656.67$$

$$s_b^2 = s_e^2/\Sigma(F_i-F)^2 = s_e^2/294.6859 = 1,463.31$$

$$s_b = 38.25$$

$$s_a^2 = s_e^2\left[\frac{1}{n}+\frac{\overline{F}^2}{\Sigma(F_i-\overline{F})^2}\right]$$

$$= s_e^2\left[\frac{1}{a}+\frac{(8.43)^2}{294.6589}\right]$$

$$= 151,912.3787$$

$$s_a = 389.76$$

$$R^2 = 1-\frac{\Sigma e_i^2}{\Sigma(Y_i-\overline{Y})^2} = 1-\frac{3,018,515.599}{7,116,681.79} = 0.5759$$

3 $Y_i = 479.775+523.506E-70.766F_i$

$R^2 = 0.8214,\ s_e = 460.2261,\ \Sigma e_i^2 = 270,848.24$

$s_a = 2,121.803,\ s_b = 184.667,\ s_c = 31.713$

두 변수는 모두 평균소득(Y)과 중요한 관계를 가지고 있음을 알 수 있다. 문제 2의 추정치는 편의(bias)가 있다.

4 $Y_i = 5.00+2.00X_i+3.00W_i$

$R^2 = 0.865 \quad s_e = 2.236$

$s_{b_1} = 0.71 \quad s_{b_2} = 0.91$

5 $Y_i = 5.50 + 2.00X_i + 2.83\,W_i$
$R^2 = 0.941 \quad s_e = 2.217$
$s_{b_1} = 1.11 \quad s_{b_1} = 1.43$

6 문제 4와 문제 5의 자료를 비교해 보면 X의 자료는 동일하나 Y와 W의 자료는 표본평균은 동일하지만 각각의 관찰치는 약간 다르다. 따라서 β_1의 추정치는 문제 6, 문제 7에서 동일하게 $b_1 = 2.0$으로 나타났으며 a, b_2는 약간의 차이를 보인다.
X의 자료가 동일하여도 Y와 W의 표본평균이 달랐다면 b_1의 값도 서로 다르게 추정되었을 것이다.

7 1) $H_0 : \beta_1 = 0$을 $\alpha = 0.05$에서 검정하는 문제이다. $s_{b_1} = 0.08$로 주어져 있으므로 $t = \dfrac{b_1}{s_{b_1}} = \dfrac{0.34}{0.08} = 4.25$

한편 자유도는 $(n-k-1) = 20$이므로 $\alpha = 0.05$에서 임계치는 $t_{a/2} = t_{0.025} = 2.09$이다. $t > t_{0.025}$이므로 귀무가설이 기각되어 가처분소득이 닭고기 소비에 영향력이 있는 것으로 나타났다.

2) 95%의 신뢰구간은 $(b_2 - t_{0.025}\ S_{b_2},\ b_2 + t_{0.025}\ S_{b_2})$
$b_2 = -0.5,\ s_{b_2} = 0.11,\ t_{0.025} = 2.09$이므로 신뢰구간: $(-0.73,\ -0.27)$

3) $H_0 : \beta_3 = \beta_4 = 0$은 F-비율을 이용하여 검정할 수 있다.

$$F = \frac{(SSE_R - SSE_u)/m}{SSE_R/(n-k-1)} = \frac{(R_u - R_R)/m}{(1-R_R)/(n-k-1)}$$

$$= \frac{(0.98-0.97)/2}{(1-0.97)/20} = 3.33$$

임계치는 $F_{0.05} = 3.49$이므로 $F < F_{0.05}$가 되어 귀무가설을 기각할 수 없다.

4) $H_0 : \beta_2 = -1$에 대한 검정은 t-비율을 이용하면 된다.

$$t = \frac{b_2 - (-1)}{s_{b_2}} = \frac{-0.5+1}{0.11} = 4.55$$

$\alpha = 0.01$에서 임계치는 $t_{0.005} = 2.85$이므로 귀무가설이 기각된다.

8 1) $T_1 = \dfrac{4.3}{1.72} = 2.5$로, $t_{60,\,0.025} = 2$보다 크므로, 귀무가설을 기각한다.

2) $T_2 = \dfrac{-0.04}{0.015} = -2.22$로, $t_{60,\,0.025} = 2$보다 크므로, 귀무가설을 기각한다.

3) 1)과 2)를 통해서 두 회귀계수(나이, 나이^2)가 모두 유의함을 알 수 있다. 따라서 교육연수가 고정되어 있다면, 나이가 들어감에 따라 연봉이 증가하다가 정점을 찍고, 이후에 나이가 들어가면서 연봉이 감소함을 알 수 있다.

이때 $-\dfrac{4.3}{2\times -0.04} = 53.75$이므로, 연봉을 가장 많이 받는 나이는 대략 53세라는 것을 알 수 있다.

9 SSE_R은 제약모형의 잔차의 제곱합이며 SSU_u는 무제약모형의 잔차의 제곱합이다. 예를 들어 $\beta_1 = 0$이라는 제약이 사실과 다르다면 제약모형의 잔차는 매우 클 것이고 따라서 SSE_R도 SSU_u보다 클 것이다. 만약 이 제약이 사실이라면 $\beta_1 = 0$이라는 제약을 부가하지 않는 무제약모형에서는 β_1을 0으로 추정할 것이고 이 경우, 제약모형과 무제약모형의 차이가 없어진다. 따라서 $SSE_R = SSE_u$가 된다.

10 제약 하에서의 모델은 $Y_i = \alpha + u_i$이므로, $a = \overline{Y}$이다.

따라서 $SSE_R = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = SST$이다.

$$F = \frac{(SSE_R - SSE_U)/k}{SSE_U/(n-k-1)} = \frac{(SST - SSE_U)/k}{SSE_U/(n-k-1)} = \frac{(SST - SSE_U)/SST}{SSE_U/SST}\ \frac{n-k-1}{k}$$

$$= \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

제13장 분산분석

1 24개 표본관찰치의 변동은 첫째는 4개의 모집단이 서로 다른 특성을 갖고 있기 때문에 있을 수 있다. 예를 들어 4개 모집단 평균이 서로 다르면 표본관찰치도 서로 다를 수가 있다. 둘째 변동의 원천은 확률변수이기 때문에 발생하는 표본오차이다.
즉, 총변동(SST)은 표본간 변동(SSB)과 표본내 변동(SSW)의 합이며 표본간 변동은 모집단 평균이 다르기 때문에 발생하는 변동을 나타내며 표본내 변동은 표본오차 때문에 변동을 나타낸다.

2 1) 표본내의 5개 관찰값이 모두 같으므로 각 관찰값은 표본평균과 같다. $x_{ij} = \overline{x}_i$

따라서 $SSW = \sum_{i=1}^{3}\sum_{j=1}^{5}(x_{ij} - \overline{x}_i)^2 = 0$

2) $SST = SSB + SSW$인데 $SSW = 0$이므로 $SST = SSB$이다. 즉 총변동은 표본평균간 변동에 의해서만 발생한다.

3 1) $n = 16,\ \overline{x}_1 = 62,\ \overline{x}_2 = 53,\ \overline{x}_3 = 52$

$SSW = 1{,}028 + 1{,}044 + 1{,}536 = 3{,}608$

$SSB = 7(62 - 56.0625)^2 + 6(53 - 56.0625)^2 + 6(52 - 56.0625)^2 = 340.94$

$SST = SSB + SSW = 3{,}948.94$

2) $H_0 : \mu_1 = \mu_2 = \mu_3$

분산분석표

변동의 원천	변동	자유도	분산	F값
표본간	340.94	2	170.47	0.61
표본내	3,608.0	13	277.54	
총	3,948.94	15		

$\alpha = 0.05$에서 $f_{2,\,13} = 3.81$이므로 귀무가설을 기각하지 못한다.

4 1) $n = 15,\ \overline{x}_1 = 17.4,\ \overline{x}_2 = 16.5,\ \overline{x}_3 = 17.3$

분산분석표

변동의 원천	변동	자유도	분산	F값
표본간	2.43	2	1.22	0.40
표본내	36.10	12	3.01	
총	38.53	14		

2) $H_0 : \mu_1 = \mu_2 = \mu_3$
$\alpha = 0.05$에서 $f_{2,12} = 3.89$이므로 귀무가설을 기각하지 못한다.

5 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$$n = \sum_{i=1}^{k} = 21,\ SSB = SST - SSW = 150$$

$$MSW = \frac{SSW}{n-k} = \frac{50}{21-4} = 2.94$$

$$MSB = \frac{SSB}{k-1} = \frac{150}{4-1} = 50$$

$$f = \frac{MSB}{MSW} = \frac{50}{2.94} = 17.01$$

$\alpha = 0.05$에서 $f_{3,17} = 3.20$이므로 귀무가설을 기각한다.

6 SSB의 자유도는 $k-1 = 4$이며 $MSB = 10$이므로 $SSB = (k-1)MSB = 40$이다.
따라서 $SSW = SST - SSB = 60$이며 자유도는 $n-k = 23$이다.

$$MSW = \frac{SSW}{n-k} = \frac{60}{23} = 2.61$$

$$f = \frac{MSB}{MSW} = \frac{10}{2.61} = 3.83$$

분산분석표

변동의 원천	변동	자유도	분산	F값
표본간	(40)	(4)	10	(3.83)
표본내	(60)	(23)	(2.61)	
총	100	(27)		

7 $H_0 : \mu_1 = \mu_2 = \cdots = \mu_5$

$$X = \frac{1}{5}\sum_{i=1}^{5} X_i = 74.2$$

$$SSW = \sum_{i=1}^{5} S_i^2 = 37,\ MSW = \frac{37}{60} = 0.62$$

$$SSB = 13\sum_{i=1}^{5}(X_i - X)^2 = 322.4,\ MSB = \frac{322.4}{4} = 80.6$$

$$f = \frac{MSB}{MSW} = \frac{80.6}{0.62} = 130$$

$\alpha = 0.01$에서 $f_{4,60} = 3.70$이므로 귀무가설을 기각한다.

8 $n = 30,\ k = 5$

$SSW = (n-k)MSW = 150$

$SSB = SST - SSW = 250,\ MSB = \dfrac{250}{5-1} = 62.5$

$f = \dfrac{MSB}{MSW} = \dfrac{62.5}{6} = 10.42$

$\alpha = 0.05$에서 $f_{4,25} = 2.76$이므로 모집단 평균이 같다는 귀무가설을 기각한다.

9 1) $n = 9,\ k = 3$

$X_1 = 3,\ X_2 = 2,\ X_3 = 4,\ X = 3$

$SSB = 3[(3-3)^2 + (2-3)^2 + (4-3)^2] = 6$

$SSW = 12,\ SST = 18$

$MSB = \dfrac{SSB}{k-1} = \dfrac{6}{2} = 3$

$MSW = \dfrac{SSW}{n-k} = \dfrac{12}{6} = 2$

$f = \dfrac{MSB}{MSW} = 1.5$

2) $SST = \sum_{i=1}^{3}\sum_{j=1}^{3}(5X_{ij} - 5X)^2 = 25\sum_{i=1}^{3}\sum_{j=1}^{3}(X_{ij} - X)^2$

$SSB = \sum_{j=1}^{3}3(5X_i - 5X)^2 = 25\sum_{i=1}^{3}3(X_i - X)^2$

F-값은 MSW와 MSB의 비율인데 두 값 모두 (a)의 25배인데 서로 상쇄되므로 F-값은 (a)와 차이가 없다.

3) $X_1 = 7,\ X_2 = 6,\ X_3 = 8,\ X = 7$

$SSB = 6,\ MSB = \dfrac{6}{2} = 3$

$SSW = 12,\ MSW = \dfrac{12}{6} = 2$이므로 $f = 1.5$이다.

4) F값은 (a), (b), (c)에서 모두 같다.

10 총변동은 모집단에 따른 변동, 블럭에 따른 변동, 표본오차에 따른 변동으로 구성되어 있는데 블럭변수를 고려하지 않으면, 블럭에 따른 변동을 무시하는 결과이므로 정확한 분석이 될 수 없다.

11 블럭변수가 총변동에 영향을 미치는 요인인데 고려를 하지 않으면 블럭변수에 의한 변동이 표본오차에 의한 변동에 혼합되어 나타난다. 따라서 표본오차에 의한 변동, 즉 표본내 변동이 실제보다 크게 나타난다.

12

변동의 원천	변동	자유도	분산	F값
표본간	(52)	2	(26)	1
블럭간	(88)	3	(29.33)	(1.13)
오차	(260)	10	26	
총	400	(15)		

1) $SSE = 26 \times 10 = 260$

2) $F = \dfrac{MSB_S}{MSE} \Rightarrow MSB_S = F \cdot MSE = 1 \cdot 26 = 26$

3) $SSB_S =$ 자유도 $\cdot MSB_S = 2 \cdot 26 = 52$

4) $SSB_B = SST - SSE - SSB_S = 88$

5) 총자유도$= (k-1)+(h-1)+(k-1)(h-1)$

$= 2+3+10 = 15$

6) $MSB_B = \dfrac{SSB_B}{\text{자유도}} = \dfrac{88}{3} = 29.33$

7) $F = \dfrac{MSB_B}{MSE} = \dfrac{29.33}{26} = 1.13$

13

1) $f_1 = \dfrac{MSB_S}{MSE} = \dfrac{127.33}{3.58} = 35.57$

$f_2 = \dfrac{MSB_B}{MSE} = \dfrac{29.25}{3.58} = 8.17$

2) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$\alpha = 0.05$에서 $f_{3,6} = 4.76$이며 $f_1 = 35.57$로 임계치보다 크므로 모집단 평균이 같다는 가설은 기각된다.

3) $\alpha = 0.05$에서 $f_{2,6} = 5.14$이며 $f_2 = 8.17$로 임계치보다 크므로 블럭간 모집단 평균의 차가 없다는 가설은 기각된다.

14

변동의 원천	변동	자유도	분산	F값
표본간	(80)	(3)	26.7	8.9
블럭간	75	(5)	15	5
오차	(45)	(15)	3	
총	200	(23)		

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$는 $f_1 = 8.9$이고 $f_{0.01} = 5.42$이므로 기각된다.

15

표본별 평균: $X_{1.} = \dfrac{22}{4} = 5.5,\ X_{2.} = \dfrac{24}{4} = 6,\ X_{3.} = \dfrac{18}{4} = 4.5$

블럭별 평균: $X_{.1} = \dfrac{23}{3} = 7.67,\ X_{.2} = \dfrac{20}{3} = 6.67,\ X_{.3} = \dfrac{6}{3} = 2,\ X_{.4} = \dfrac{15}{3} = 5$

전체표본 평균: $X = \dfrac{64}{12} = 5.33$

$$SST = \sum_{i=1}^{3}\sum_{j=1}^{4}(X_{ij}-X)^2 = (9-5.33)^2+(5-5.33)^2+\cdots+(1-5.33)^2 = 94.67$$

$$SSB_S = h\sum_{i=1}^{k}(X_{i.}-X)^2$$
$$= (4)[(5.5-5.33)^2+(6-5.33)^2+(4.5-5.33)^2] = 5.44$$

$$SSB_B = k\sum_{j=1}^{h}(X_{.j}-X)^2$$
$$= 3[(7.67-5.33)^2+(6.67-5.33)^2+(2-5.33)^2+(5-5.33)^2] = 55.41$$

$$SSE = SST - SSB_S - SSB_B = 33.82$$

$$MSB_S = \frac{SSB_S}{k-1} = \frac{5.44}{2} = 2.72$$

$$MSB_B = \frac{SSB_B}{h-1} = \frac{55.41}{3} = 18.47$$

$$MSE = \frac{SSE}{(k-1)(h-1)} = \frac{33.82}{6} = 5.64$$

$$f_1 = \frac{MSB_S}{MSE} = 0.48$$

$$f_2 = \frac{MSB_B}{MSE} = 3.27$$

변동의 원천	변동	자유도	분산	F값
표본간	5.44	2	2.72	0.48
블럭간	55.41	3	18.47	3.21
오차	33.82	6	5.64	
총	94.67	11		

16 $H_0 : \mu_1 = \mu_2 = \cdots = \mu_5$을 검정하기 위해서는 SSB_S를 우선적으로 알아야 한다.

$$SSE = (k-1)(h-1)MSE = (5-1)(4-1)\cdot 5 = 60$$

$$SSB_S = SST - SSB_B - SSE$$
$$= 500 - 240 - 60 = 200$$

$$MSB_S = \frac{SSB_S}{k-1} = \frac{200}{5-1} = 50$$

$$f_1 = \frac{MSB_S}{MSE} = \frac{50}{5} = 10$$

이때 $\alpha = 0.05$에서 자유도가 4,12인 임계값은 $f_{4,12} = 3.26$이다. 따라서 귀무가설은 기각된다.

17 $k = 3,\ h = 4$

$$MSB_S = \frac{SSB_S}{k-1} = \frac{95.2}{3-1} = 47.6$$

$$MSE = \frac{SSE}{(k-1)(h-1)} = \frac{79.3}{(3-1)(4-1)} = 13.22$$

$$f = \frac{MSB_S}{MSE} = \frac{47.6}{13.22} = 3.60$$

한편 $\alpha = 0.05$에서 자유도가 2,6인 임계값은 5.14이므로 세 연속극에 대한 시청자의 평균만족도는 동일하다. 귀무가설 기각 못함.

18 1) 표본별 평균: $X_{1.} = \frac{49}{5} = 9.8,\ X_{2.} = \frac{49}{5} = 9.8,\ X_{3.} = \frac{44}{5} = 8.8,\ X_{4.} = \frac{61}{5} = 12.2$

블럭별 평균: $X_{.1} = \frac{40}{4} = 10,\ X_{.2} = \frac{38}{4} = 9.5,\ X_{.3} = \frac{41}{4} = 10.25,$

$$X_{.4} = \frac{43}{4} = 10.75,\ X_{.5} = \frac{41}{4} = 10.25$$

전체표본 평균: $X = \frac{203}{20} = 10.15$

$SSB_S = 31.35,\ SSB_B = 3.3,\ SST = 66.55,\ SSE = 31.9$

변동의 원천	변동	자유도	분산	F값
표본간	31.35	3	10.45	3.93
블럭간	3.3	4	0.83	0.31
오차	31.9	12	2.66	
총	66.55	19		

2) 각 정유회사의 순이익증가율에 대한 평균예상치가 같다는 귀무가설을 검정하는 데 $\alpha = 0.05$에서 자유도가 4,12인 임계값은 $f_{4,12} = 3.26$이다. 한편 F값은 0.31이므로 귀무가설을 기각할 수 없다.

19 1) $SSB_S = 218.19,\ SSB_B = 197.63,\ SSE = 71.4,\ SST = 487.23$

변동의 원천	변동	자유도	분산	F값
표본간	218.19	3	72.73	6.11
블럭간	197.63	2	98.82	8.3
오차	71.4	6	11.9	
총	487.23	11		

2) 비료의 차에 따라 평균경작결과가 같다는 귀무가설을 검정해보자. $\alpha = 0.05$에서 자유도가 3,6이므로 임계값은 $f_{3,6} = 4.76$이므로 F값 6.11보다 작기 때문에 귀무가설은 기각된다.

제14장 적합도 검정과 분할표

1 χ^2적합도 검정통계량은 관할도수와 기대도수의 차의 제곱을 기대도수로 나누어준 것이므로 차가 클수록 χ^2검정통계치는 커진다.

2 1) $df = (r-1)(c-1) = (6-1)(3-1) = 10,\ \chi^2_{0.06} = 18.31$

2) $df=(5-1)(4-1)=12,\ \chi^2_{0.01}=26.22$

3) $df=(2-1)(6-1)=5,\ \chi^2_{0.05}=11.07$

3 p_1 = 4년만에 졸업하는 남학생비율

p_2 = 4년만에 졸업하는 여학생비율이라고 할 때, 귀무가설은 $H_0: p_1=p_2$

귀무가설이 사실이라면 $\hat{p}=\dfrac{345}{820}=0.42$이며, 기대빈도는

$$E_{11}=\frac{(445)(345)}{820}=187.23,\ E_{21}=\frac{(375)(345)}{820}=157.77$$

$$E_{12}=\frac{(445)(475)}{820}=257.77,\ E_{22}=\frac{(375)(475)}{820}=217.23$$

따라서 아래와 같은 표를 만들 수 있다.

칸	O_{ij}	E_{ij}	$\dfrac{O_{ij}^2}{E_{ij}}$
$(1,1)$	160	187.23	136.73
$(1,2)$	285	257.77	315.11
$(2,1)$	185	157.77	216.93
$(2,2)$	190	217.23	166.18
합	820		834.95

$$\chi^2=\Sigma\Sigma\frac{O_{ij}^2}{E_{ij}}-n=834.95-820=14.95$$

한편 자유도는 $(r-1)(e-1)=1$이므로 $\alpha=0.05$에서 임계치는 3.84이다. 따라서 남녀 간의 4년내 졸업비율이 같다는 귀무가설은 기각된다.

4 $$E_{11}=\frac{(168)(211)}{220}=161.13,\ E_{12}=\frac{(168)(9)}{220}=6.87$$

$$E_{21}=\frac{(41)(211)}{220}=39.32,\ E_{22}=\frac{(41)(9)}{220}=1.68$$

$$E_{31}=\frac{(11)(211)}{220}=10.55,\ E_{32}=\frac{(11)(9)}{220}=0.45$$

칸	O_{ij}	E_{ij}	$\dfrac{O_{ij}^2}{E_{ij}}$
$(1,1)$	165	161.13	168.96
$(1,2)$	3	6.87	1.31
$(2,1)$	39	39.32	38.68
$(2,2)$	2	1.68	2.38
$(3,1)$	7	10.55	4.64
$(3,2)$	4	0.45	35.56
합	220		251.53

$$\chi^2 = \Sigma\Sigma\frac{O_{ij}^2}{E_{ij}} - n = 251.53 - 220 = 31.53$$

자유도는 $(r-1)(c-1) = 2$이며 $\alpha = 0.05$이므로 임계치는 5.99이다. χ^2통계치가 임계치보다 크므로 귀무가설은 기각된다.

5

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	1	2.69	0.37
$(1,2)$	3	3.28	2.74
$(1,3)$	4	2.03	7.88
$(2,1)$	100	98.31	101.72
$(2,2)$	120	119.72	120.28
$(2,3)$	72	73.97	70.08
합	300		303.07

검정통계치는 3.07이며, $\alpha = 0.01$에서 자유도는 2이므로 임계치는 9.21이 되어 귀무가설이 기각되지 않는다.

6 X의 기대도수는 $E_X = P(X) \cdot n = (0.3)(200) = 60$

Y의 기대도수는 $E_Y = P(Y) \cdot n = (0.25)(200) = 50$

7

				합
	40	(50)	(10)	100
	(40)	50	(10)	(100)
합	(80)	(100)	20	200

8

					합
	(3)	(3.75)	(6.75)	(1.5)	15
	(8)	(10)	(18)	(4)	40
	(8)	(10)	(18)	(4)	40
	(1)	(1.25)	(2.25)	(0.5)	5
합	20	25	45	10	100

9 $P(A) = \frac{61}{152} = 0.40,\ P(B) = 0.33,\ P(C) = 0.27$

P(국민차) $= 20$, P(소형) $= 0.26$, P(중형) $= 0.33$, P(대형) $= 0.21$

식(15.3.1)에 의해 기대도수를 구하여 정리하면 아래와 같다.

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	15	12.04	18.69
$(1,2)$	14	16.05	12.22

$(1,3)$	18	20.07	16.14
$(1,4)$	14	12.84	15.26
$(2,1)$	9	9.87	8.21
$(2,2)$	17	13.16	21.96
$(2,3)$	15	16.45	13.68
$(2,4)$	9	10.53	7.69
$(3,1)$	6	8.09	4.45
$(3,2)$	9	10.79	7.51
$(3,3)$	17	13.49	21.42
$(3,4)$	9	8.63	9.39
합	152		156.62

검정통계치는 4.62이며, $\alpha = 0.05$에서 자유도는 6이므로 임계치는 12.59이므로 귀무가설을 기각할 수 없다.

10 업종과 수익상태는 서로 독립적이라는 가설을 검정하여 보자. 귀무가설 하에서의 기대도수를 정리하면 아래와 같다.

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	31	24.6	39.07
$(1,2)$	330	314.61	346.14
$(1,3)$	459	480.79	438.20
$(2,1)$	10	10.8	9.26
$(2,2)$	200	138.12	289.60
$(2,3)$	150	211.08	106.59
$(3,1)$	43	24.9	74.26
$(3,2)$	321	318.44	323.58
$(3,3)$	466	486.66	446.22
$(4,1)$	6	29.7	1.21
$(4,2)$	300	379.83	236.95
$(4,3)$	684	580.47	806.0
합	3,000		3,117.08

χ^2통계치는 117.08이며, $\alpha = 0.05$에서임계값은 12.59이므로 귀무가설은 기각된다.

11 과학성적과 과학에 관심정도가 독립적이라는 귀무가설을 검정하여 보자. 귀무가설 하에서의 기대도수를 식(15.3.1)에 의해 계산하고 정리하면 아래와 같다.

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	100	120	83.33
$(1,2)$	150	195	115.38
$(1,3)$	200	135	296.3
$(2,1)$	50	106.67	23.44

$(2,2)$	200	173.33	230.77
$(2,3)$	150	120	187.5
$(3,1)$	250	173.33	360.58
$(3,2)$	300	281.67	319.52
$(3,3)$	100	195	51.28
합	1,500		1,668.1

χ^2검정통계치는168.1이며,$\alpha = 0.05$에서 임계치인9.49보다 크므로 귀무가설은 기각된다.

12

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	12	15.05	9.57
$(1,2)$	7	8.81	5.56
$(1,3)$	4	7.34	2.18
$(1,4)$	21	17.25	25.57
$(1,5)$	14	9.54	20.55
$(2,1)$	11	9.08	13.33
$(2,2)$	4	5.32	3.01
$(2,3)$	6	4.43	8.13
$(2,4)$	9	10.41	7.78
$(2,5)$	5	5.76	4.34
$(3,1)$	18	16.87	19.21
$(3,2)$	13	9.87	17.12
$(3,3)$	10	8.23	12.15
$(3,4)$	17	19.34	14.94
$(3,5)$	7	10.70	4.58
합	158		168.02

χ^2검정통계치는10.02인 한편,$\alpha = 0.05$에서 임계치는15.51이므로 흡연량과 혈압은 독립적이라고 할 수 있다.

13 1)

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	27	17.61	41.40
$(1,2)$	22	18.50	26.16
$(1,3)$	47	59.89	36.88
$(2,1)$	173	182.39	164.09
$(2,2)$	188	191.50	184.56
$(2,3)$	633	620.11	646.27
합	1,090		1,099.36

$$\chi^2 = \Sigma\Sigma\frac{O_{ij}^2}{E_{ij}} - n = 1{,}099.36 - 1{,}090 = 9.36$$

2) $\alpha = 0.05$이며 자유도는 $(r-1)(c-1) = (2-1)(3-1) = 2$이므로 임계치는 5.99이다.

3) 통계치가 임계치보다 크므로 두 변수가 독립적이라는 귀무가설은 기각된다.

14 $H_0 : p_1 = p_2 = p_3$

칸	O_{ij}	E_{ij}	$\frac{O_{ij}^2}{E_{ij}}$
$(1,1)$	46	74.12	28.55
$(1,2)$	137	196.16	95.68
$(1,3)$	237	149.72	375.16
$(2,1)$	104	75.88	142.54
$(2,2)$	260	200.84	336.59
$(2,3)$	66	153.28	28.42
합	850		1,006.94

$\chi^2 = 156.94$이므로 H_0 기각

15 관찰도수, 기대도수 각각의 합은 총관찰수인 표본의 크기 n과 같다.

$$\sum_i\sum_j O_{ij} = \sum_i\sum_j E_{ij} = n$$

$$\sum_{i=1}^{r}\sum_{j=1}^{c}(O_{ij} - E_{ij}) = \sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij} - \sum_{i=1}^{r}\sum_{j=1}^{c} E_{ij} = 0$$

16
$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2 - 2O_{ij}E_{ij} + E_{ij}^2}{E_{ij}}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - 2\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij} + \sum_{i=1}^{r}\sum_{j=1}^{c} E_{ij}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - 2n + n$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - n$$

제15장 비모수통계학

1 P는 A맥주가 B맥주보다 선호되는 비율이라 할 때 가설을 설정하면

$H_0 : P = 0.5$ $H_a : P > 0.5$

$n=20,\ M=16$이므로 식 (16.2.2)에 의해

$$z=\frac{2M-n}{\sqrt{n}}=\frac{32-20}{\sqrt{20}}=\frac{6}{\sqrt{5}}=2.68$$

$\alpha=0.05$에서 임계치가 $z_{0.05}=1.65$이므로 검정통계치가 임계치보다 커서 귀무가설을 기각할 수 있다. 즉, A맥주가 B맥주보다 선호된다고 볼 수 있다.

2 $H_0: p=0.5\ \ H_a: p>0.5$

p는 A보석이 더 좋게 평가되는 비율.

$n=10,\ M=8$이므로

$$z=\frac{2M-n}{\sqrt{n}}=\frac{16-10}{\sqrt{10}}=\frac{6}{\sqrt{10}}=1.90$$

$\alpha=0.05$에서 $z_{0.05}=1.65$이므로 귀무가설이 기각되어 A보석이 더 좋다고 평가할 수 있다.

3 $H_0: \mu_X-\mu_Y=0\ \ H_a: \mu_X-\mu_Y>0$

X는 통계학 성적, Y는 철학 성적이라 할 때 $(X-Y)$의 +부호는 10개 −부호는 5개이다. 따라서 $M=10$이다. 한편 9번째 학생은 통계학 성적과 철학 성적이 같아 부호화할 수 없으므로 $n=15$이다.

$$z=\frac{2M-n}{\sqrt{n}}=\frac{20-15}{\sqrt{15}}=1.29$$

검정통계치가 $\alpha=0.05$에서 임계치인 1.65보다 작으므로 귀무가설을 기각할 수 없다. 즉 경제학과 학생의 통계학 성적이 철학 성적보다 더 높다고 볼 수 없다.

4 1) $H_0: \mu-54=0\ \ H_a: \mu-54\neq 0$

양의 부호는 9개, 음의 부호는 8개이다. $M=9,\ n=17$.

$$z=\frac{2M-n}{\sqrt{n}}=\frac{18-17}{\sqrt{17}}=0.24$$

$\alpha=0.05$에서 임계치인 $z_{0.025}=1.96$이다. 부호검정에 의하면 귀무가설을 기각할 수 없다.

2)

연령(X)	$X-54$	순위	음의 순위	양의 순위
26	−28	17	17	
49	−5	5	5	
55	1	1.5		1.5
55	1	1.5		1.5
60	6	6		6
61	7	7		7
68	14	13		13
57	3	3		3
33	−21	15	15	
32	−22	16	16	
68	14	13		13
62	8	8.5		8.5

62	8	8.5		8.5
50	−4	4	4	
41	−13	10.5	10.5	
41	−13	10.5	10.5	
40	−14	13	10.5	
합			88.5	62

$T^+ = 62,\ T^- = 88.5,\ T = \max(T^+, T^-) = 88.5$

$\mu_T = \frac{n(n+1)}{4} = \frac{(17)(17+1)}{4} = 76.5,\ \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = 21.12$

$z = \frac{T-\mu_T}{\sigma_T} = \frac{88.5-76.5}{21.12} = 0.57.$

$-z_{0.025} < z < z_{0.025}$이므로 귀무가설을 기각할 수 없다.

5

통계학 성적:	75	88	96	88	94	55	65	53	85	75	82	72
순위:	5.5	9.5	12	9.5	11	2	3	1	8	5.5	7	4

6 부호화된 순위검정

t − 검정은 모집단분포의 정규성 가정이 필요하며 이 가정이 사실과 다르다면 정확치 못한 검정이 된다. 반면 부호화된 순위검정은 모집단분포에 대한 가정이 필요없다.

7

표본:	26.2	18.6	19.3	26.2	8.9	12.3	26.2	32.8	8.9	18.6
순위:	8	4.5	6	8	1.5	3	8	10	1.5	4.5

순위합 = 55

8 $H_0 : \mu_A - \mu_B = 0 \quad H_a : \mu_A - \mu_B \neq 0$

기능공	A	B	$(A-B)$	순위	음의 순위	양의 순위
1	15.2	14.5	0.7	10		10
2	13.6	13.7	−0.1	2	2	
3	14.6	14.8	−0.2	4	4	
4	13.3	12.9	0.4	6		6
5	14.2	13.8	0.4	6		6
6	15.7	15.6	0.1	2		2
7	14.9	15.3	−0.4	6	6	
8	15.2	14.3	0.9	12		12
9	15.6	15.5	0.1	2		2
10	15.0	15.0	0	−	−	−
11	16.2	15.6	0.6	9		9
12	15.7	15.2	0.5	8		8
13	15.6	14.8	0.8	11		11
합					12	66

$T^+ = 66,\ T^- = 12,\ T = \max(T^+,\ T^-) = 66$

$\mu_T = \dfrac{n(n+1)}{4} = \dfrac{(12)(12+1)}{4} = 39, \sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{24}} = 12.74$

따라서 검정통계치는

$z = \dfrac{T-\mu_T}{\sigma_T} = \dfrac{66-39}{12.74} = 2.12$

$\alpha = 0.01$에서 임계치 $z_{0.005} = 2.575$이므로 $-z_{0.005} < z < z_{0.005}$이어서 귀무가설을 기각할 수 없다.

9 $H_0 : \mu_A - \mu_B = 0 \qquad H_a : \mu_A - \mu_B \neq 0$

$T^+ = 55, T^- = 50$이므로 $T = 55$이며 $n = 14$이다.

$\mu_T = \dfrac{n(n+1)}{4} = 52.5,\ \sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{4}} = 39$

$z = \dfrac{T-\mu_T}{\sigma_T} = \dfrac{55-52.5}{39} = 0.064$

따라서 귀무가설을 기각할 수 없다.

10 1) 순위합검정을 이용하여 보자.

국산차의 순위합을 W라 하면 $W = 67$이다.

$U = W - \dfrac{n_1(n_1+1)}{2} = 67 - \dfrac{(10)(10+1)}{2} = 12$

한편 U의 표본평균과 표본표준편차는 식 (16.2.7)과 식(16.2.8)을 이용하면 구할 수 있다.

$\mu_U = \dfrac{n_1 n_2}{2} = 50$

$\sigma_U = \sqrt{\dfrac{n_1 n_2 (n_1+n_2+1)}{2}} = 32.4$

U의 표본분포가 정규분포에 근접하므로 z통계량을 이용하면 된다.

$z = \dfrac{u-\mu_u}{\sigma_u} = \dfrac{12-50}{32.4} = -1.17$

$\alpha = 0.05$에서 임계값이 1.65이므로 평균이 같다는 귀무가설을 기각할 수 없다.

2) 비모수법, 모수법을 이용하려면 모집단의 분포를 정규분포로 가정해야 한다. 또한 모집단의 분산을 알수 없을 때 $t-$검정을 이용할 것이다. $t-$검정은 표본의 크기가 작을 때는 정규분포 가정에 매우 민감한 약점을 갖고 있다. 반면 1)에서 사용한 비모수법은 모집단 분포에 관한 어떤 가정도 필요로 하지 않는다는 장점을 가진다.

11 1) $H_0 : \mu_A - \mu_B = 0 \quad H_a : \mu_A - \mu_B \neq 0$

$\overline{X}_A = 15,\ \overline{X}_B = 17,\ s_A = s_B = 2$이므로 식 (9.3.16)을 이용하면

$t = \dfrac{\overline{X}_A - \overline{X}_B}{\sqrt{\dfrac{(n-1)s_A^2 + (m-1)s_B^2}{n+m-2}\dfrac{m+n}{mn}}} = \dfrac{15-17}{\sqrt{\dfrac{(15)(4)+(15)(4)}{16+16-2}\dfrac{16+16}{(16)(16)}}} = -2.828$

이때 자유도는 $n+m-2 = 16+16-2 = 30$이므로 $t_{0.025} = 2.04$이다.

즉 $t < -t_{0.025}$이므로 귀무가설이 기각된다.

2)

표본관찰치	순위	표본	표본관찰치	순위	표본
10.5	(1)	A	15.1	17	B
11.0	(2.5)	A	15.2	18	B
11.0	(2.5)	A	15.5	19.5	B
11.9	4	B	15.5	19.5	B
12.0	(5.5)	A	15.8	(21)	A
12.0	5.5	B	16.1	22	B
12.1	(7)	A	16.6	(23)	A
12.3	(8)	A	16.7	(24)	A
12.6	9	B	16.8	25	B
13.1	(10)	A	16.9	26	B
13.3	(11)	A	17.0	27	B
13.8	(12)	A	18.3	(28.5)	A
14.1	13	B	18.3	28.5	B
14.2	(14)	A	18.8	30	B
14.4	(15)	A	19.5	31	B
14.6	16	B	20.5	(32)	A

A회사의 순위합인 $W=217$이다.

식 (16.2.6)에 의해 검정통계량 U의 값을 구하면

$$U=W-\frac{n_1(n_1+1)}{2}=217-\frac{(16)(16+1)}{2}=81$$

한편 $\mu_U=\frac{n_1n_2}{2}=128,\ \sigma_U=\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}=26.53$

$$z=\frac{\mu-\mu_U}{\sigma_U}=\frac{81-128}{26.53}=-1.77$$

$\alpha=0.05$이므로 $z_{0.025}=1.96$이어서 $-z_{0.025}<z<z_{0.025}$이므로 귀무가설을 기각할 수 없다.

3) 1)에서는 귀무가설이 기각되었으나 2)에서는 기각되지 않는다. 1)은 모집단이 정규분포를 갖는다는 가정하에서 t − 검정을 하였는데 만약 정규분포를 갖지 않는다면 잘못된 결과가 도출될 가능성이 크다. 반면 2)에서는 모집단 분포에 대한 가정을 하지 않았으므로 1)에 비해 적합한 결론이 도출되었다고 볼 수 있다.

12 $H_0=\mu_1=\mu_2=\mu_3=\mu_4$

KW검정을 이용하여 학생들의 성적을 순위별로 정리하면 아래와 같다.

성적	순위	학교	성적	순위	학교
73	1	C	81	11.5	B
76	2	C	81	11.5	C
77	3	C	81	11.5	D
78	4	C	82	15	A
79	5	D	82	15	A
80	7.5	A	82	15	B

80	7.5	D	83	17	B
80	7.5	D	84	18.5	A
80	7.5	D	84	18.5	B
81	11.5	A	85	20	B

따라서 각 학교별 순위합은

$R_A = (7.5 + 11.5 + 15 + 15 + 18.5) = 67.5$

$R_B = (11.5 + 15 + 17 + 18.5 + 20) = 82$

$R_C = (1 + 2 + 3 + 4 + 11.5) = 21.5$

$R_D = (5 + 7.5 + 7.5 + 7.5 + 11.5) = 39$

$n = 20$이므로 검정통계량 H를 식 (16.3.2)에 의해 구하면

$$H = \frac{2}{n(n+1)} \sum_{i=1}^{4} \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{20(20+1)} \frac{1}{5} (67.5^2 + 82^2 + 21.5^2 + 39^2) - 3(20+1)$$

$$= 12.8$$

자유도는 $(k-1) = 3$이고 유의수준 $\alpha = 0.01$이므로 임계값은 $\chi^2_{0.01} = 11.35$이다. 따라서

$H = 12.8 > \chi^2_{0.01} = 11.35$가 되어 귀무가설 H_0를 기각한다. 즉 서울시내 4개 고교의 성적은 평준화되었다고 말할 수 없다.

13 $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$

KW검정을 이용하면

$R_1 = 206.5,\ R_2 = 201,\ R_3 = 212.5,\ R_4 = 200$

$n_1 = n_2 = n_3 = n_4 = 10$이며 $n = n_1 + n_2 + n_3 + n_4 = 40$이다.

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{4} \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{(40)(40+1)} (4,264.23 + 4,040.1 + 4,515.63 + 4,000) - 3(40+1)$$

$$= 123.07 - 123 = 0.07$$

$\alpha = 0.05$에서 자유도는 $(k-1) = 3$이며 임계치 $\chi^2_{0.05} = 7.81$이므로 귀무가설을 기각할 수 없다.

14 1) $H_0 = \mu_1 = \mu_2 = \mu_3$

$n = 31,\ k = 3$

$SST = 254.06$

$SSB = (10)(10 - 11.55)^2 + (11)(12 - 11.55)^2 + (10)(14 - 11.55)^2$

$= 24.03 + 2.23 + 60.03 = 86.29$

$SSW = SST - SSB = 167.77$

$MSW = \frac{SSW}{n-k} = \frac{167.77}{31-3} = 5.99$

$MSB = \frac{SSB}{k-1} = \frac{86.29}{3-1} = 43.15$

$F = \frac{MSB}{MSW} = \frac{43.15}{5.99} = 7.20$

$\alpha = 0.05$에서 $f_{2,\,28} = 3.34$이므로 귀무가설은 기각된다.

2) 각 표본의 순위합을 구하면

$R_1 = 80.5,\ R_2 = 158.5,\ R_3 = 257$

$n = 10 + 11 + 10 = 31$이므로 검정통계량 H의 값을 구하면

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{3} \frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{(31)(31+1)}(648.03 + 2,283.84 + 6,604.9) - 3(31+1)$$

$$= 115.4 - 96 = 19.4$$

$\alpha = 0.05$, 자유도 $= 2$이므로 임계치 $\chi^2_{0.05} = 5.99$이므로 귀무가설이 기각된다.

3) 1), 2) 모두 귀무가설을 기각한다.

단, 1)에서 이용한 분산분석은 모집단 분포에 대한 가정을 필요로 하나 KW검정은 분포에 대한 가정이 필요없는 장점을 지닌다. 반면 분산분석은 표본관찰값을 그대로 이용하지만 KW검정에서는 표본관찰값을 순위로 전환하면서 일부 정보의 손실을 보는 약점이 있다.

15 1) $R = 17,\ n_1 = 14,\ n_2 = 14,\ \mu_R = \dfrac{2n_1n_2}{n_1+n_2} + 1 = 15$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{6.74}$$

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{17-15}{\sqrt{6.74}} = 0.77$$

2) $R = 18,\ n_1 = 14,\ n_2 = 14,\ n_2 = 14,\ \mu_R = 15.48$

$$\sigma_R = \sqrt{6.97}$$

$$z = \frac{18 - 15.48}{\sqrt{6.97}} = 0.95$$

3) $R = 16,\ n_1 = 10,\ n_2 = 12,\ \sigma_R = 11.91$

$$\sigma_R = \sqrt{5.15}$$

$$z = \frac{16 - 11.91}{\sqrt{5.15}} = 1.80$$

4) $R = 13,\ n_1 = 12,\ n_2 = 12,\ \mu_R = 13$

$$\sigma_R = \sqrt{5.74}$$

$$z = \frac{13 - 13}{\sqrt{5.74}} = 0$$

16 $R = 16,\ n_1 = 13,\ n_2 = 37$

$\mu_R = 20.24,\ \sigma_R = \sqrt{7.16}$

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{16 - 20.24}{\sqrt{7.16}} = -1.58$$

$\alpha = 0.01$에서 $z_{0.005} = 2.575$이므로 표본은 무작위로 추출되었다는 귀무가설은 기각될 수 없다.

17 $R = 11,\ n_1 = 12,\ n_2 = 13$

$\mu_R = 13.48,\ \sigma_R = \sqrt{5.97}$

$z = \dfrac{11 - 13.48}{\sqrt{5.97}} = -1.01$

$-z_{0.025} < z < z_{0.025}$이므로 올해의 기온이 무작위형태를 띤다고 할 수 있다.

18 1) 순위차 D_i를 구하면 아래와 같다.

학생	A	B	C	D	E	F	G	H	I	J
D_i	-2	-1	-2	1	-2	3	1	1	1	0
D_i^2	4	1	4	1	4	9	1	1	1	0
$\sum D_i^2 = 26$										

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} = 1 - \frac{(6)(26)}{10(100-1)} = 084$$

2) 검정통계치를 식 (16.5.2)를 이용하여구할 수 있다.

$z = \rho_s \sqrt{n-1} = 0.84\sqrt{10-1} = 2.52$

$\alpha = 0.05$에서 임계값은 $z_{0.025} = 1.96$이며 검정통계치가 더 크므로 귀무가설이 기각된다.

19 1) 중간고사성적과 기말고사성적의 순위를 각각 구하고 순위차 D_i를 구하면

$\sum_{i=1}^{14} D_i^2 = 68$이다.

Spearman 상관계수는

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} = 1 - \frac{(6)(68)}{14(196-1)} = 0.85$$

2) $z = \rho_s \sqrt{n-1} = 0.85\ \sqrt{14-1} = 3.06$

따라서 상관관계가 없다는 귀무가설이 기각된다.

20 1) $\sum_{i=1}^{10} D_i^2 = 215.25$

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} = 1 - \frac{(6)(215.25)}{10(100-1)} = -0.30$$

2) $H_0 : \rho_s = 0\ \ H_a : \rho_s \neq 0$

$z = \rho_s \sqrt{n-1} = -0.30\sqrt{10-1} = -0.90$이며 $\alpha = 0.01$에서 임계값 $z_{0.005} = 2.57$이므로 귀무가설을 기각할 수 없다.