1. $E(X_1)=E(X_2)=E(X_3)=\mu$, $Var(X_1)=7$, $Var(X_2)=13$, $Var(X_3)=20$일 때, 다음 추정량을 이용하여 모평균 μ를 점추정하고자 한다.

$$\hat{\mu}_1=\frac{1}{3}(X_1+X_2+X_3),\quad \hat{\mu}_2=\frac{X_1}{4}+\frac{X_2}{2}+\frac{X_3}{5},\quad \hat{\mu}_3=\frac{X_1}{3}+\frac{X_2}{4}+\frac{X_3}{5}+2$$

(1) 불편추정량과 편의추정량을 구하여라.

(2) $\hat{\mu}=aX_1+bX_2+cX_3$이 불편추정량일 때, 최소분산추정량이 되는 a, b, c값을 구하여라.

(sol)

(1) $E[\widehat{\mu_1}]=E[\frac{1}{3}(X_1+X_2+X_3)]=\frac{1}{3}(E[X_1]+E[X_2]+E[X_3])=\mu \Rightarrow \widehat{\mu_1}$ 불편추정량이다.

$E[\widehat{\mu_2}]=E[\frac{X_1}{4}+\frac{X_2}{2}+\frac{X_3}{5}]=\frac{1}{4}E[X_1]+\frac{1}{2}E[X_2]+\frac{1}{5}E[X_3]\mu=\frac{19}{20}\mu \Rightarrow \widehat{\mu_2}$은 불편추정량이다.

$E[\widehat{\mu_3}]=E[\frac{X_1}{3}+\frac{X_2}{4}+\frac{X_3}{5}+2]=\frac{1}{3}E[X_1]+\frac{1}{4}E[X_2]+\frac{1}{5}E[X_3]+E[2]=\frac{47}{60}\mu+2 \Rightarrow \widehat{\mu_3}$은 불편추정량이다.

(2) $\hat{\mu}=aX_1+bX_2+cX_3$은 불편추정량이므로,

$$E[\hat{\mu}]=E[aX_1+bX_2+cX_3]=aE[X_1]+bE[X_2]+cE[X_3]=(a+b+c)\mu=\mu \Rightarrow a+b+c=1$$

이다. $\hat{\mu}$의 분산을 구하면

$$Var[\hat{\mu}]=Var[aX_1+bX_2+cX_3]=a^2\,Var[X_1]+b^2\,Var[X_2]+c^2\,Var[X_3]=7a^2+13b^2+20c^2$$

이므로, 제한조건 $g(a,b,c)=a+b+c-1=0$에서 $f(a,b,c)=7a^2+13b^2+20c^2$의 최솟값을 구하면

$$F(a,b,c,\lambda)=(7a^2+13b^2+20c^2)+\lambda(a+b+c-1)$$
$$F_a=14a+\lambda=0$$
$$F_b=26b+\lambda=0$$
$$F_c=40c+\lambda=0$$
$$F_\lambda=a+b+c-1=0$$

$\Rightarrow a=-\frac{1}{14}\lambda,\ b=-\frac{1}{26}\lambda,\ c=-\frac{1}{40}\lambda$이므로, $-\frac{1}{14}\lambda-\frac{1}{26}\lambda-\frac{1}{40}\lambda=1$ 이다.

$\Rightarrow \lambda=-\frac{3640}{491}$이므로, $a=\frac{260}{491}$, $b=\frac{140}{491}$, $c=\frac{91}{491}$

따라서 최소분산불편추정량은 $\hat{\mu}=\frac{260}{491}X_1+\frac{140}{491}X_2+\frac{91}{491}X_3$이다.

2. $X_1, X_2, \cdots, X_{10}$은 $(0,b)$에서 균등분포를 이루는 모집단으로부터 추출된 확률표본이고, 관찰값이 다음 표와 같다고 한다.

10, 7, 11, 12, 8, 8, 9, 10, 9, 13

(1) 모평균 μ, 모분산 σ^2의 불편추정값과 모표준편차 σ의 추정값을 구하여라.

(2) 모수 b의 불편추정값을 구하여라.

(sol)

(1) 표본평균 $\overline{x}$은 모평균 μ의 불편추정량이다.

$$\hat{\mu}=\overline{x}=\frac{1}{10}\sum_{i=1}^{10}x_i=\frac{1}{10}(10+7+11+12+8+8+9+10+9+13)=9.7$$

표본분산 s^2은 모분산 σ^2의 불편추정량이다.

$$\hat{\sigma}^2=s^2=\frac{1}{9}\sum_{i=1}^{10}(x_i-9.7)=\frac{32.1}{9}=3.5667$$

표본표준편차 s는 모표준편차 σ의 편의추정량이다.

$$\hat{\sigma}=s=\sqrt{s^2}=\sqrt{\frac{32.1}{9}}=1.889$$

(2) 연속균등분포를 이루는 모집단의 모평균은 $\mu=\frac{b+0}{2}$이고, 표본평균에 의한 모평균의 불편추정값은 $\hat{\mu}=9.7$이다.

b에 대한 불편추정값을 $\hat{b}$라고 하면, $\hat{\mu}=\frac{\hat{b}-0}{2}=9.7 \Rightarrow \hat{b}=19.4$ 이다.

3. 상호대화식의 컴퓨터 시스템은 대단위 장치에서 사용이 가능하다. 시간당 수신된 신호수 X는 모수 λ인 푸아송분포에 따른다고 한다. 이때 수신된 신호 수에 대하여 다음 표본을 얻었다.

31, 28, 28, 25, 17, 26, 22, 10, 4, 27

(1) 모수 λ의 불편추정값을 구하여라.

(2) 30분당 수신된 평균 신호수의 불편추정값을 구하여라.

(sol)

(1) $X \sim Poisson(\lambda)$일 때, 모평균은 $\mu = \lambda$이고 표본평균 $\overline{x}$은 모평균 μ의 불편추정량이다.

$$\hat{\mu} = \overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}(31+28+28+17+26+22+10+4+27) = 21.8$$

(2) 1시간당 수신된 평균 신호수가 21.8로 추정되므로, 30분당 평균 10.9회 신호가 수신될 것으로 추정된다.

4. $X_1, X_2, \cdots, X_{10}$은 $n=10$이고 미지의 p를 모수로 갖는 이항분포로부터 추출된 확률표본이라 한다.

1, 8, 2, 5, 7, 6, 2, 9, 4, 7

$\hat{p} = \dfrac{\overline{X}}{10}$이 모수 p에 대한 불편추정량임을 보이고, 관찰된 표본을 이용하여 모수 p에 대한 불편추정값을 구하여라.

(sol)

모집단 $X \sim B(10, p)$이므로

모든 $i\,(1 \le i \le 10)$에 대하여, $X_i \sim B(10, p)$이고 $E[X_i] = 10p$이다.

$\overline{X} = \dfrac{1}{10}\displaystyle\sum_{i=1}^{10} X_i$이므로, $\hat{p} = \dfrac{\overline{X}}{10} = \dfrac{1}{100}\displaystyle\sum_{i=1}^{10} X_i$이다. 따라서

$$E[\hat{p}] = E\left[\frac{1}{100}\sum_{i=1}^{10} X_i\right] = \frac{1}{100}\left\{E[X_1] + \cdots + E[X_{10}]\right\} = \frac{1}{100}\left\{10p + \cdots + 10p\right\} = p$$

이다.

관찰된 표본에 대한 모수 p의 불편추정값은 $\hat{p} = \dfrac{\overline{X}}{10} = \dfrac{1}{10}\left\{\dfrac{1}{10}\displaystyle\sum_{i=1}^{10} x_i\right\} = \dfrac{1}{100}(1+8+2+5+7+6+2+9+4+7) = 0.51$이다.

5. 강의실 옆의 커피 자판기에서 컵 한잔에 나오는 커피의 양을 조사하기 위하여 101잔을 조사한 결과 평균 0.3ℓ, 표본표준편차 0.06ℓ이였다. 다음 각 조건 아래서 이 자판기에서 나오는 커피 한잔의 평균 양에 대한 95% 신뢰구간을 구하여라.

(1) 커피의 양은 정규분포에 따르며, 모표준편차 $\sigma = 0.05\ell$로 알려져 있는 경우

(2) 커피의 양은 정규분포에 따르며, 모표준편차를 모르는 경우

(sol)

(1) 모분산 $\sigma^2 = 0.05^2$을 알고 있으므로, $z-$추정을 한다.

$z-$분포를 이용하여 95% 오차한계와 신뢰구간을 구하면 다음과 같다.

$$e_{95\%} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = z_{0.025}\frac{0.05}{\sqrt{101}} = (1.96)(0.004975) = 0.00975$$

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = (0.3 - 0.00975, 0.3 + 0.00975) = (0.29025, 0.30975)$$

(2) 모분산을 모르고 있으므로, $t-$추정을 한다.

$t-$분포를 이용하여 이때 95% 오차한계와 신뢰구간은 다음과 같다.

$$e_{95\%} = t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} = t_{100,0.025}\frac{0.06}{\sqrt{101}} = (1.984)(0.00597) = 0.011845$$

$$\left(\overline{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right) = (0.3 - 0.11845, 0.3 + 0.11845) = (0.288155, 0.311845)$$

※ 대표본이므로, $z-$추정으로 근사 신뢰구간을 구하면 다음과 같다.

$$e_{95\%} = z_{\alpha/2}\frac{s}{\sqrt{n}} = z_{0.025}\frac{0.06}{\sqrt{101}} = (1.96)(0.00597) = 0.0117$$

$$\left(\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right) = (0.3 - 0.0117, 0.3 + 0.0117) = (0.2883, 0.3117)$$

6. 다음은 남자와 여자의 생존 연령을 조사한 자료이다.

남자 : 52, 60, 55, 46, 33, 75, 58, 57, 88, 56

여자 : 62, 58, 65, 56, 53, 45, 64, 77

(1) 두 그룹의 모분산이 동일하다는 조건 아래서 남자와 여자의 평균 생존 연령의 차에 대한 90% 신뢰구간을 구하여라.

(2) 두 그룹의 모분산의 비 ${\sigma_1}^2/{\sigma_2}^2$에 대한 90% 신뢰구간을 구하여라.

(sol)

(1) 남자와 여자의 평균생존연령과 표본분산을 구하면,

$$\overline{x}=\frac{1}{10}\sum_{i=1}^{10}x_i=58, \qquad {s_1}^2=\frac{1}{9}\sum_{i=1}^{10}(x_i-58)^2=223.5556,$$

$$\overline{y}=\frac{1}{8}\sum_{i=1}^{8}y_i=60, \qquad {s_2}^2=\frac{1}{7}\sum_{i=1}^{8}(y_i-60)^2=89.7143$$

이고, 합동표본분산은 ${s_p}^2=\dfrac{1}{n_1+n_2-2}\{(n_1-1){s_1}^2+(n_2-1){s_2}^2\}=\dfrac{1}{16}\{9(223.5556)+7(89.7143)\}=165$이다.

이때 90% 오차한계와 신뢰구간은 다음과 같다.

$$e_{90\%}=t_{n_1+n_2-2,\alpha/2}\,s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}=t_{16,0.05}\sqrt{165}\sqrt{\frac{1}{10}+\frac{1}{8}}=(1.745)(12.8452)(0.004975)=10.6323$$

$$(\overline{x}-\overline{y}-e_{90\%},\overline{x}-\overline{y}+e_{90\%})=(-2-10.6323,-2+10.6323)=(-12.6323,8.6323)$$

(2) $\dfrac{{S_1}^2/{\sigma_1}^2}{{S_2}^2/{\sigma_2}^2}\sim F(n_1-1,n_2-1)$이므로, 90% 신뢰구간은 다음과 같다.

$$\left(\frac{{s_1}^2/{s_2}^2}{f_{\alpha/2}(n_1-1,n_2-1)},\frac{{s_1}^2/{s_2}^2}{f_{1-\alpha/2}(n_1-1,n_2-1)}\right)=\left(\frac{223.5556/89.7143}{f_{0.05}(9,7)},\frac{223.5556/89.7143}{f_{0.95}(9,7)}\right)$$

$$=\left(\frac{223.5556/89.7143}{3.6767},\frac{223.5556/89.7143}{0.3037}\right)=(0.6777,8.2051)$$

7. 어떤 공장에서 생산되는 전자부품의 수명은 평균 14개월, 표준편차 1.4개월인 정규분포를 따른다고 한다. 이 공장에서 생산되는 부품 7개를 선택하여 관측된 수명이

19.5, 23.2, 31.2, 34.5, 41.1, 31.2, 24.7

일 때, 모분산 σ^2에 대한 신뢰도 95% 신뢰구간을 구하여라.

(sol)

표본평균 : $\overline{x}=\dfrac{1}{7}\sum_{i=1}^{7}x_i=29.3429$, 표본분산 : $s^2=\dfrac{1}{6}\sum_{i=1}^{7}(x_i-\overline{x})^2=54.6495$

$V=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$이므로, 95% 신뢰구간은 다음과 같다.

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}},\frac{(n-1)s^2}{\chi^2_{n-1,1-(\alpha/2)}}\right)=\left(\frac{6(54.6495)}{\chi^2_{6,0.025}},\frac{6(54.6495)}{\chi^2_{6,0.975}}\right)$$

$$=\left(\frac{6(54.6495)}{14.4494},\frac{6(54.6495)}{1.2373}\right)=(22.6928,265.0007)$$

※ 주어진 표본의 신뢰구간은 모수의 참값을 포함하지 않는다.

8. 한 포대에 $1\,\mathrm{kg}$인 설탕 16포대를 조사하여 평균 $1.053\,\mathrm{kg}$, 표본표준편차 $0.058\,\mathrm{kg}$을 얻었다.

(1) 모평균 μ의 99% 신뢰구간을 구하여라.

(2) 모표준편차 σ의 99% 신뢰구간을 구하여라.

(3) 모평균 μ의 99% 신뢰구간을 0.05이하로 하기 위한 표본의 크기를 구하여라.

(sol)

(1) 모분산을 모르고 있으므로, $t-$추정을 한다. 이때 99% 오차한계와 신뢰구간은 다음과 같다.

$$e_{99\%}=t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}=t_{15,0.005}\frac{0.058}{\sqrt{16}}=(2.9467)(0.0145)=0.04273$$

$$(\overline{x}-t_{n-1,\alpha/2}\frac{s}{\sqrt{n}},\overline{x}+t_{n-1,\alpha/2}\frac{s}{\sqrt{n}})=(1.053-0.04273,1.053+0.04273)=(1.010273,1.095727$$

(2) 표본의 개수가 10이상이므로, 모표준편차 σ의 99% 신뢰구간은 다음과 같다.

$$(s\sqrt{\frac{n-1}{\chi^2_{n-1,\alpha/2}}},s\sqrt{\frac{n-1}{\chi^2_{n-1,1-(\alpha/2)}}})=(0.058\sqrt{\frac{15}{\chi^2_{15,0.005}}},0.058\sqrt{\frac{15}{\chi^2_{15,0.995}}})$$

$$=(0.058\sqrt{\frac{15}{4.6009}},0.058\sqrt{\frac{15}{32.8013}})=(0.0392,0.1047)$$

(3) 모평균 μ의 99% 신뢰구간의 길이는 $L=2t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$이므로,

$L\le 0.05$이기 위해서는 $n\ge\left(\frac{2\cdot t_{15,0.005}}{0.05}\right)^2 s^2=\left(\frac{2(2.947)}{0.05}\right)^2(0.058)^2=46.745$이다.

따라서 16개를 사전에 조사하였으므로 추가로 31개를 더 조사하면 된다.

9. 다음 표는 어느 대학에서 두 전공 A, B를 선택한 여학생 수 (x_1,x_2)를 조사한 자료이다.

전공 A : $n_1=500$, $x_1=155$

전공 B : $n_2=350$, $x_2=92$

이 자료를 기초로 두 전공을 선택하는 여성비율의 차 p_1-p_2에 대한 99% 신뢰구간을 구하여라.

(sol)

전공 A와 B의 여성비율을 각각 p_1, p_2이라 하면, 두 전공의 여성비율에 대한 추정값은

$$\hat{p}_1=\frac{155}{500}=0.31,\ \hat{p}_2=\frac{92}{350}=0.2628$$

이다. 이때 99% 오차한계와 신뢰구간은 다음과 같다.

$$e_{99\%}=z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}=2.58\sqrt{\frac{(0.31)(0.69)}{500}+\frac{(0.2628)(0.7372)}{350}}=0.0808$$

$$(\hat{p}_1-\hat{p}_2-e_{99\%},\hat{p}_1-\hat{p}_2+e_{99\%})=(0.0472-0.0808,0.0472+0.0808)=(-0.0336,0.1280)$$

10. 어떤 도시에서 임의로 500가구를 관찰한 결과 98가구가 심야전기를 사용하고 있다고 한다.

(1) 이 자료를 기초로 이 도시의 심야전기를 사용하는 비율에 대한 99% 신뢰구간을 구하시오.

(2) 표본의 비와 모비율의 차가 0.05안에 있을 신뢰도가 97%가 되는 표본의 크기를 정하여라.

(sol)

(1) 모비율 p의 추정값은 $\hat{p}=\frac{98}{500}=0.196$이고, 이때 99% 오차한계와 신뢰구간은 다음과 같다.

$$e_{99\%}=z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}=2.58\sqrt{\frac{(0.196)(0.804)}{500}}=0.0458$$

$$(\hat{p}-e_{99\%},\hat{p}+e_{99\%})=(0.196-0.0458,0.196+0.0458)=(0.1502,0.2418)$$

(2) $|p-p|\le 0.05$이므로, 모비율 p의 97% 신뢰구간의 길이는 $L=2z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\le 0.1$ 이어야 한다.

$L\le 0.1$이기 위해서는 $n\ge 4\left(\frac{z_{0.015}}{0.1}\right)^2\hat{p}\,\hat{q}=4\left(\frac{2.17}{0.1}\right)^2(0.196)(0.804)=296.8189$이다.

따라서 $n=297$가구를 조사해야 한다.