数理統計：レポート２略解

担当： 熊谷 隆

１。(平均) $= E[\sum_{i=1}^n X_i^2] = nE[X_1^2] = n(\mathrm{Var}\,(X_1) + 0) = n$。
（分散）$= \mathrm{Var}\,(\sum_{i=1}^n X_i^2) = n\mathrm{Var}\,(X_1^2) = n(E[X_1^4] - E[X_1^2]^2) = n(3-1^2) = 2n$。ただし、$E[X_1^4] = 3$ は以下の計算による。

$$\begin{aligned} E[X_1^4] &= \frac{2}{\sqrt{2\pi}}\int_0^\infty y^4 e^{-y^2/2}dy = \frac{2}{\sqrt{2\pi}}\Big([-y^3e^{-y^2/2}]_0^\infty + 3\int_0^\infty y^2 e^{-y^2/2}dy\Big) \\ &= \frac{6}{\sqrt{2\pi}}\Big([-ye^{-y^2/2}]_0^\infty) + \int_0^\infty e^{-y^2/2}dy\Big) = \frac{6}{\sqrt{2\pi}} \times \frac{\sqrt{2\pi}}{2} = 3 \end{aligned}$$

（χ_n^2 の平均と分散を求めるという問題だから、定理 2.4 を用いて計算することもできる。）

２。 標本平均を $\bar{X}_{10}$、不偏分散を V_{10} とおくと、簡単な計算により $\bar{X}_{10} = 27.18, V_{10} = \sum_{i=1}^9 (X_i - \bar{X}_{10})^2/9 = \frac{2.896}{9} = 0.32177\cdots$ となる。また、χ_9^2 に従う μ について、$\mu([x(0.975), x(0.025)]) = 0.95$ を満たすのは（χ^2 分布表から）$x(0.975) = 2.70, x(0.025) = 19.02$ だから、求める信頼区間は

$$\Big[\frac{\sum_{i=1}^9 (X_i - \bar{X}_{10})^2}{x(0.025)}, \frac{\sum_{i=1}^9 (X_i - \bar{X}_{10})^2}{x(0.975)}\Big] = [0.1523, 1.0726]$$

（小数第５位を四捨五入した）である。

３。帰無仮説を「H: 母分散が等しい」とおく。2000 年代の不偏分散は $V'_{10} = 3.565/9 = 0.39611\cdots$ である。また、F_9^9 に従う μ について、$\mu([a,\infty)) = 0.05$ を満たすのは（F 分布表から）$a = 3.18$ だから、求める信頼区間は、小数第４位を四捨五入すると

$$\Big[\frac{V'_{10}}{aV_{10}}, \frac{aV'_{10}}{V_{10}}\Big] = [0.3871, 3.9147]$$

となる。（なお、V_{10} と V'_{10} を入れ替えて計算した場合、上の値の逆数が出る。）この区間は 1 を含むから、仮説 H を棄却できない。つまり、「平均気温の年による変動の大きさに差が出ているとは言えない」。

４。帰無仮説を「H: 母平均が等しい」とおく。問３の結果から、以下母分散は等しいと仮定する。2000 年代の標本平均は $\bar{X}'_{10} = 28.25$、不偏分散は $V'_{10} = 3.565/9 = 0.39611\cdots$ である。また、T_{18} に従う μ について、$\mu([a,\infty)) = 0.025$ を満たすのは（t 分布表から）$a = 2.101$ だから、求める信頼区間は、小数第５位を四捨五入すると

$$\Big[\bar{X}_{10} - \bar{X}'_{10} - a\sqrt{\frac{20(9V_{10} + 9V'_{10})}{10\times 10\times 18}}, \bar{X}_{10} - \bar{X}'_{10} + a\sqrt{\frac{20(9V_{10} + 9V'_{10})}{10\times 10\times 18}}\Big] = [-1.633, -0.507]$$

となる。（なお、$(\bar{X}_{10}, V_{10})$ と $(\bar{X}'_{10}, V'_{10})$ を入れ替えて計算した場合、上の値の逆符号が出る。）この区間は 0 を含まないから、仮説 H を棄却できる。つまり、「平均気温が上昇したと言える」。

５。標準正規分布に従う μ について、$x(\alpha/2)$ を $\mu([-x(\alpha/2), x(\alpha/2)]) = 1-\alpha$ を満たすものとすると、講義 3.2 節の例２より

$$P([\bar{X}_n - x(\alpha/2)\sqrt{\frac{p(1-p)}{n}}, \bar{X}_n + x(\alpha/2)\sqrt{\frac{p(1-p)}{n}}]) = 1-\alpha$$

である。したがって信頼率 $100(1-\alpha)$ %の信頼区間の幅が１%以下になるように推定するには $2x(\alpha/2)\sqrt{\frac{p(1-p)}{n}} \le 0.01$ となればよい。ここで $p(1-p) \le 1/4$ を用いると、$2x(\alpha/2)\sqrt{\frac{p(1-p)}{n}} \le x(\alpha/2)/\sqrt{n}$ となるので、$x(\alpha/2)/\sqrt{n} \le 0.01$ つまり $(100\cdot x(\alpha/2))^2 \le n$ となるように取れば十分である。$x(0.025) = 1.9600, x(0.005) = 2.5758$ だから、信頼率 95 %の場合は $(196)^2 = 38416$ 人以上、信頼率 99 %の場合は $(257.58)^2 = 66347.5$ なので 66348 人以上の標本を抽出するとよい。