問題 1 9月8日のプロ野球公式戦は全5試合が行われ、10チームの得点と安打数は以下の通りでした：

得点	5	1	9	6	1	10	1	3	0	3
安打数	11	4	10	12	6	15	5	4	6	3

得点データ（これを X とします）の平均値・分散、安打数データ（これを Y とします）の平均値・分散を求め、 X と Y の相関係数を求めて下さい。なお、必要であれば

$$3.33^2 \sim 11.09, \quad 3.88^2 \sim 15.05$$

を参考にして下さい。

配点：20点 | シラバス達成度目標：ア、イ

解答例 まず平均値と分散は

$$E[X] = \frac{5 + 1 + 9 + \cdots + 3}{10} = \frac{39}{10}$$

$$\begin{aligned} \text{Var}[X] &= \frac{5^2 + 1^2 + 9^2 + \cdots + 3^2}{10} - \left(\frac{39}{10}\right)^2 \\ &= \frac{25 + 1 + 81 + 36 + 1 + 100 + 1 + 9 + 9}{10} - \frac{39^2}{100} \\ &= \frac{2630 - 1521}{100} \\ &= 11.09 \end{aligned}$$

$$E[Y] = \frac{11 + 4 + 10 + \cdots + 3}{10} = \frac{76}{10}$$

$$\begin{aligned} \text{Var}[Y] &= \frac{11^2 + 4^2 + 10^2 + \cdots + 3^2}{10} - \left(\frac{76}{10}\right)^2 \\ &= \frac{121 + 16 + 100 + 144 + 36 + 225 + 25 + 16 + 36 + 9}{10} - \frac{76^2}{100} \\ &= \frac{7280 - 5776}{100} \\ &= 15.04 \end{aligned}$$

ですから、次に共分散は

$$\begin{aligned} \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] \\ &= \frac{55 + 4 + 90 + 72 + 6 + 150 + 5 + 12 + 9}{10} - \frac{39}{10} \cdot \frac{76}{10} \\ &= \frac{4030 - 2964}{100} \\ &= 10.66 \end{aligned}$$

になります。従って相関係数は

$$\text{Cor}[X, Y] = \frac{10.66}{\sqrt{11.09}\sqrt{15.04}} \sim \frac{10.66}{3.33 \cdot 3.88} = \frac{10.66}{12.92} \sim 0.825$$

となります。

問題 2 確率変数 X が平均 50、分散 9 の正規分布に従うとき、標準正規分布表を参照して確率 $P[55 \leq X \leq 57]$ を求めて下さい。

配点：15点 | シラバス達成度目標：エ

解答例

$$\begin{aligned} P[55 \leq X \leq 57] &= P[5 \leq X - 50 \leq 7] \\ &= P\left[\frac{5}{3} \leq \frac{X - 50}{3} \leq \frac{7}{3}\right] \\ &= P\left[\frac{5}{3} \leq N(0, 1) \leq \frac{7}{3}\right] \\ &= P[1.667 \leq N(0, 1) \leq 2.333] \\ &= P[0 \leq N(0, 1) \leq 2.333] - P[0 \leq N(0, 1) \leq 1.667] \\ &\sim 0.4901 - 0.452 \\ &\sim 0.0381 \end{aligned}$$

問題 3 分布密度関数が次の $f(x)$:

$$f(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

で与えられている確率変数 X に対して、確率 $P[-2 \leq X \leq 0.5]$ 、平均値 $E[X]$ を求めて下さい。

配点：20点 | シラバス達成度目標：ウ

解答例

$$P[-2 \leq X \leq 0.5] = \int_{-2}^{0.5} f(x) dx$$

$$= \int_0^{0.5} 3x^2 dx$$

$$= [x^3]_0^{0.5}$$

$$= 0.125$$

$$E[X] = \int_{-\infty}^{\infty} xf(x) dx$$

$$= \int_0^1 3x^3 dx$$

$$= \left[\frac{3}{4}x^4\right]_0^1$$

$$= \frac{3}{4}$$

$$= 0.75$$

問題 4 2次元の確率変数 (X, Y) の分布密度関数が $h(x, y)$ であるとはどう云う事か説明して下さい。ただし、2重積分を使った正確な数式を示す必要はなく、『体積』と云う言葉を使って簡単に説明して下さい。

配点：5点 | シラバス達成度目標：オ

解答例 2次元の確率変数 (X, Y) の分布密度が $h(x, y)$ であるとは、任意の領域 D に対して、

$$P[(X, Y) \in D] = \left(\begin{array}{l} \text{領域 } D \text{ を床とし、この領域の境界線上に垂直な壁を立て、} \\ \text{曲面 } z = h(x, y) \text{ を屋根とした立体の体積} \end{array} \right)$$

が成り立つことです。

問題 5 母集団 A は平均値と分散をもち、 $E[A] = m, \text{Var}[A] = v > 0$ であると仮定します。このとき、この母集団から取った大きさ n の標本平均 $\bar{A}$ に関して、

$$E[\bar{A}] = m, \quad \text{Var}[\bar{A}] = \frac{v}{n}$$

となる事を証明して下さい。

配点：10点 | シラバス達成度目標：カ、キ

解答例 n 個の独立な確率変数 $A_1, A_2, \dots, A_n$ は全て A と同分布であるとし、このとき

$$\bar{A} = \frac{A_1 + A_2 + \dots + A_n}{n}$$

を大きさ n の標本平均と呼ぶわけですが、この期待値は

$$\begin{aligned} E[\bar{A}] &= E\left[\frac{A_1 + A_2 + \dots + A_n}{n}\right] \\ &= \frac{1}{n}E[A_1 + A_2 + \dots + A_n] \\ &= \frac{1}{n}(E[A_1] + E[A_2] + \dots + E[A_n]) \\ &= \frac{mn}{n} \\ &= m \end{aligned}$$

となり、また分散は独立性により

$$\begin{aligned} \text{Var}[\bar{A}] &= \text{Var}\left[\frac{A_1 + A_2 + \dots + A_n}{n}\right] \\ &= \frac{1}{n^2}\text{Var}[A_1 + A_2 + \dots + A_n] \\ &= \frac{1}{n^2}(\text{Var}[A_1] + \text{Var}[A_2] + \dots + \text{Var}[A_n]) \\ &= \frac{nv}{n^2} \\ &= \frac{v}{n} \end{aligned}$$

となります。

問題 6 ある母集団から2万個のサンプルを取って調査したところ、サンプルの平均は157.9、サンプルの不偏分散は 5.35^2 でした。母平均の95%信頼区間を求めて下さい。

配点：15点 | シラバス達成度目標：ク

解答例 まず状況を整理すると、

	分布： 不明		サイズ： 20000
【母集団】	平均： 不明	【サンプル】	平均： 157.9
	分散： 不明		不偏分散： 5.35^2

です。

母平均は不明ですが、計算の都合上仮に m としておきます。また、大きさ20000の大きなサンプルを取っているので母分散はサンプル不偏分散で代用出来ます。

従って中心極限定理によれば、この母集団から取った大きさ20000の標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{5.35^2}{20000}\right)$ で近似されます。

まず

$$P[|\bar{X} - m| \leq d] = 0.95$$

となるような正数 d を求めますが、正規分布で近似して更に標準化すれば

$$\begin{aligned} 0.95 &= P[|\bar{X} - m| \leq d] \\ &\sim P\left[\left|N\left(m, \frac{5.35^2}{20000}\right) - m\right| \leq d\right] \\ &= P\left[\left|\frac{N\left(m, \frac{5.35^2}{20000}\right) - m}{\sqrt{\frac{5.35^2}{20000}}}\right| \leq \frac{d}{\sqrt{\frac{5.35^2}{20000}}}\right] \\ &= P\left[|N(0, 1)| \leq \frac{d}{\frac{5.35}{100\sqrt{2}}}\right] \\ 0.475 &= P\left[0 \leq N(0, 1) \leq \frac{d}{\frac{5.35}{100\sqrt{2}}}\right] \end{aligned}$$

ですから、正規分布表により $\frac{d}{\frac{5.35}{100\sqrt{2}}} \sim 1.96$ です。これは $d \sim 0.0744$ を意味しますから結局

$$P[|\bar{X} - m| \leq 0.0744] = 0.95$$

が得られた事になり、今回のサンプル調査の平均値が $\bar{X}$ の1つの実現値である事に注意すれば、信頼度95%で

$$|157.9 - m| \leq 0.0744, \quad \text{すなわち} \quad 157.83 \leq m \leq 157.97$$

が成り立つと言えます。従って求める信頼区間は $[157.8, 158.0]$ です。

問題 7 ある工場で作られる電球の寿命は標準偏差 100 時間の正規分布に従っているようです。

この工場で製造された多数の電球の中から 25 個を抽出して寿命時間を測定したところ、その平均値は 1835 時間でした。寿命時間の平均値は 1800 時間より長いと言えるでしょうか。有意水準 5% で検定して下さい。

配点：15点 | シラバス達成度目標：ケ

解答例 仮説 H_0 : 『寿命の平均値は 1800 時間である』が正しいと仮定します。するとこの母集団から取った大きさ 25 の標本平均 $\bar{X}$ は正規分布 $N\left(1800, \frac{100^2}{25}\right)$ に従います。

平均値が 1800 より長いかどうかの問題になっているので右片側検定として棄却域を求めます。

$$\begin{aligned} 0.05 &= P[\bar{X} - 1800 \geq d] \\ &= P\left[N\left(0, \frac{100^2}{25}\right) \geq d\right] \\ 0.45 &= P[0 \leq N(0, 400) \leq d] \\ &= P\left[0 \leq N(0, 1) \leq \frac{d}{20}\right] \end{aligned}$$

ですから、正規分布表から $\frac{d}{20} \sim 1.65$ 、つまり、 $d \sim 33$ が分かります。これは

$$0.05 \sim P[\bar{X} - 1800 \geq 33]$$

を意味し、今回の具体値 1835 はこの棄却域に入っています。従って仮説 H_0 は棄却され、寿命時間の平均値は 1800 時間より長いと判断されます。