問題 **1** 全国一斉にある教科のテストが行われました。受験生から１００名を抽出し、その得点の平均と標準偏差を求めたところそれぞれ 58.3 点、12.4 点でした。全受験生の平均得点の９５％信頼区間を求めて下さい。

配点：２０点	シラバス達成度目標：エ、ク

解答例

母分布：	unknown
母平均：	m : unknown
母分散：	unknown
サンプルサイズ：	100（large）
サンプル平均：	58.3
サンプル分散：	12.4^2

大きさ 100 の大きなサンプルを取っているので母分散は標本分散で代用出来ます。また母平均は仮に m としておきます。すると中心極限定理によれば標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{12.4^2}{100}\right)$ で近似されます。

まず

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

となる様な $d > 0$ を求めます。少し変形すれば

$$0.95 = P\left[\left|\frac{\bar{X} - m}{\sqrt{\frac{12.4^2}{100}}}\right| \le \frac{d}{\sqrt{\frac{12.4^2}{100}}}\right] \sim P\left[|N(0,1)| \le \frac{d}{\sqrt{\frac{12.4^2}{100}}}\right] = 2P\left[0 \le N(0,1) \le \frac{d}{\sqrt{\frac{12.4^2}{100}}}\right]$$

$$0.475 = P\left[0 \le N(0,1) \le \frac{d}{\sqrt{\frac{12.4^2}{100}}}\right]$$

となりますが、

標準正規分布表に依れば、

$$P[0 \le N(0,1) \le 1.96] \sim 0.475$$

ですから、

$$\frac{d}{\sqrt{\frac{12.4^2}{100}}} = 1.96, \quad d = \frac{1.96 \times 12.4}{10} \sim 2.43$$

であれば良い事が分かります。

従って

$$P[|\bar{X} - m| \le 2.43] \sim 0.95$$

が分かりました。

標本平均の具体値が分かっていますのでこれを m に関する条件に読み替えれば

$$|58.3 - m| \le 2.43 \quad \text{すなわち} \quad 58.3 - 2.43 \le m \le 58.3 + 2.43$$

であることが 95 パーセントの信頼度で正しいわけです。

従って元データの平均値 m の、信頼度９５％の信頼区間は $[55.87, 60.73]$ になります。

問題 **2** 次のデータの平均値と分散を求めて下さい。

$$\{3, 6, 4, 3, 6, 6, 5, 9, 5, 2, 5, 6\}$$

配点：１５点	シラバス達成度目標：ア

解答例

$$（平均値）= \frac{3+6+4+3+6+6+5+9+5+2+5+6}{12} = \frac{60}{12} = 5$$

$$\begin{aligned}（分散）&= \frac{(-2)^2+1^2+(-1)^2+(-2)^2+1^2+1^2+0^2+4^2+0^2+(-3)^2+0^2+1^2}{12} \\ &= \frac{4+1+1+4+1+1+0+16+0+9+0+1}{12} \\ &= \frac{38}{12} \\ &= \frac{19}{6} \\ &\sim 3.167\end{aligned}$$

問題 **3** ３次元確率変数 $\mathbf{X} = (X_1, X_2, X_3)$ の各成分 X_1, X_2, X_3 は独立であって、どれも同じ分布に従い、平均値・分散は存在しているとします。

　このとき直交行列 U によって変換：

$$U = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{pmatrix}, \quad \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = U \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} \frac{X_1+X_2+X_3}{\sqrt{3}} \\ \frac{X_1-X_2}{\sqrt{2}} \\ \frac{X_1+X_2-2X_3}{\sqrt{6}} \end{pmatrix}$$

して出来る Y_1, Y_2 の共分散 $Cov[Y_1, Y_2]$ を求めて下さい。

配点：１０点	シラバス達成度目標：イ、オ

解答例　まず

$$E[Y_2] = \frac{1}{\sqrt{2}} E[X_1 - X_2] = \frac{1}{\sqrt{2}} \left(E[X_1] - E[X_2]\right) = 0$$

であることに注意しておきます。

　すると、独立性などにより

$$\begin{aligned}Cov[Y_1, Y_2] &= E[Y_1 Y_2] - E[Y_1]E[Y_2] \\ &= \frac{1}{\sqrt{6}} E[(X_1+X_2+X_3)(X_1-X_2)] \\ &= \frac{1}{2\sqrt{3}} \left(E[X_1^2] - E[X_2^2] + E[X_3X_1] - E[X_3X_2]\right) \\ &= \frac{1}{2\sqrt{3}} \left(E[X_1^2] - E[X_2^2] + E[X_3]E[X_1] - E[X_3]E[X_2]\right) \\ &= 0\end{aligned}$$

が得られます。

問題 **4** 確率変数 R が正規分布 $N(10, 5^2)$ に従うとき、正規分布表を利用して $P[5.5 \le R \le 8.5]$ を（近似）計算して下さい。

配点：１５点	シラバス達成度目標：エ

解答例　正規化すると

$$
\begin{aligned}
P[5.5 \le R \le 8.5] &= P[5.5-10 \le R-10 \le 8.5-10] \\
&= P\left[-\frac{4.5}{5} \le \frac{R-10}{5} \le -\frac{1.5}{5}\right] \\
&= P[-0.9 \le N(0,1) \le -0.3] \\
&= P[0.3 \le N(0,1) \le 0.9] \\
&= P[0 \le N(0,1) \le 0.9] - P[0 \le N(0,1) \le 0.3]
\end{aligned}
$$

ですから、正規分布表から

$$
\begin{aligned}
&\sim 0.3159 - 0.1179 \\
&= 0.198
\end{aligned}
$$

が得られます。

問題 **5** 確率変数 W の分布密度関数が次の $f(x)$ で与えられるとき：

$$
f(x) = \begin{cases} e^{-x} & 0 \le x < \infty \\ 0 & -\infty < x < 0 \end{cases}
$$

期待値 $E[W]$ を求めて下さい。

配点：１０点	シラバス達成度目標：ウ

解答例

$$
\begin{aligned}
E[W] &= \int_0^\infty x\mathrm{e}^{-x} dx \\
&= \left[x\left(-\mathrm{e}^{-x}\right)\right]_0^\infty - \int_0^\infty \left(-\mathrm{e}^{-x}\right) dx \\
&= -\lim_{x\to\infty} \frac{x}{\mathrm{e}^x} + \left[-\mathrm{e}^{-x}\right]_0^\infty \\
&= 1
\end{aligned}
$$

問題 **6** ある工場で生産している製品の重さは、通常は平均値が 80g、標準偏差が 5g の正規分布をしています。

正しく生産されているか調べるために、ある日の製品の中から１００個の標本を抽出して測定したところ、重さの平均値が 80.9g でした。この日の製品は平常と比べて重いと言えるでしょうか。

この日も標準偏差は変わらず 5g であるとした上で有意水準５％で検定して下さい。

配点：２０点	シラバス達成度目標：ケ

解答例　仮説 H_0 :『この日も重さの平均値は 80g であった』が正しいと仮定します。

するとこの日の製品の中から取り出す大きさ１００の標本平均 $\bar{X}$ は正規分布 $N\left(80, \frac{5^2}{100}\right)$ に従います。

今回問題になっているのは平常より重いかどうかですから片側検定として有意水準５％の棄却域を求めます。つまり、

$$0.05 = P[d \leq \bar{X}]$$

となる $d > 0$ を求めます。標準化すれば

$$\begin{aligned} &= P\left[\frac{d-80}{\frac{5}{10}} \leq N(0,1)\right] \\ &= 0.5 - P\left[0 \leq N(0,1) \leq \frac{d-80}{\frac{1}{2}}\right] \\ 0.45 &= P\left[0 \leq N(0,1) \leq \frac{d-80}{\frac{1}{2}}\right] \end{aligned}$$

ですから正規分布表を参照して $\frac{d-80}{\frac{1}{2}} \sim 1.645$、つまり、$d \sim 80.82$ が分かります。

これは

$$0.05 \sim P[80.82 \leq \bar{X}]$$

を意味し、棄却域は $80.82 \leq \bar{X}$ であり、今回の具体値 80.9 はこの棄却域に入っているため仮説は棄却されます。

従ってこの日は平常よりも重かったと言えます。

問題 7 n 個の確率変数 $X_1, \ldots, X_n$ は独立であり、全て同じ分布に従うものとします。共通の平均値を m、共通の分散を v とします。このときこれらの平均を表す確率変数 $\bar{X}$：

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

の平均値と分散を m, n, v で表して下さい。

配点：１０点	シラバス達成度目標：カ、キ

解答例

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{X_1 + \cdots + X_n}{n}\right] \\ &= \frac{E[X_1] + \cdots + E[X_n]}{n} \\ &= \frac{m + \cdots + m}{n} \\ &= m \\ Var[\bar{X}] &= Var\left[\frac{X_1 + \cdots + X_n}{n}\right] \\ &= \frac{Var[X_1 + \cdots + X_n]}{n^2} \\ &= \frac{Var[X_1] + \cdots + Var[X_n]}{n^2} \\ &= \frac{v}{n} \end{aligned}$$