

問題 1 次の表はある地域で産出される赤ワインについて、年度ごとの育成期の平均気温 D 、平均価格（1961 年産の価格との比の対数） P をまとめたものです：

	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
D	16.5	17.3	16.3	15.7	17.3	15.4	16.5	16.2	16.2	16.6
P	-2.0	0	-1.1	-1.8	-1.2	-2.2	-0.7	-1.7	-2.2	-2.1

(1) D, P の平均値 $E[D], E[P]$ を計算して下さい。ここで計算ミスをするると以下の全てを間違える事になりますから嚴重に注意して下さい。

(2) 上で求めた平均値を使って得られる派生データ $D - E[D]$ と $P - E[P]$ の具体的な値を下の表にまとめ、これを利用して D の分散 $Var[D]$ と D と P の共分散 $Cov[D, P]$ を求めて下さい。

	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
$D - E[D]$										
$P - E[P]$										

(3) P の D への回帰直線が

$$y - E[P] = \frac{Cov[D, P]}{Var[D]}(x - E[D])$$

となることを使って平均気温が 16.0 だった場合の価格の概算値を求めて下さい（ただし、価格の形式は P の値と同じ形式で結構です）。

配点：(1) 20 点、(2) (3) 20 点 シラバス達成度目標：ア、イ

【解答例】 (1)

$$E[D] = \frac{16.5 + 17.3 + 16.3 + 15.7 + 17.3 + 15.4 + 16.5 + 16.2 + 16.2 + 16.6}{10} = 16.4$$

$$E[P] = \frac{-2.0 + 0 + -1.1 + -1.8 + -1.2 + -2.2 + -0.7 + -1.7 + -2.2 + -2.1}{10} = -1.5$$

(2)

	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
$D - E[D]$	0.1	0.9	-0.1	-0.7	0.9	-1.0	0.1	-0.2	-0.2	0.2
$P - E[P]$	-0.5	1.5	0.4	-0.3	0.3	-0.7	0.8	-0.2	-0.7	-0.6
積	-0.05	1.35	-0.04	0.21	0.27	0.7	0.08	0.04	0.14	-0.12

$Var[D]$

$$\begin{aligned} & \frac{(16.5 - 15.4)^2 + (17.3 - 15.4)^2 + (16.3 - 15.4)^2 + (15.7 - 15.4)^2 + (17.3 - 15.4)^2 + (15.4 - 15.4)^2 + (16.5 - 15.4)^2 + (16.2 - 15.4)^2 + (16.2 - 15.4)^2 + (16.6 - 15.4)^2}{10} \\ &= \frac{(0.1)^2 + (0.9)^2 + (-0.1)^2 + (-0.7)^2 + (0.9)^2 + (-1.0)^2 + (0.1)^2 + (-0.2)^2 + (-0.2)^2 + (0.2)^2}{10} \\ &= \frac{0.01 + 0.81 + 0.01 + 0.49 + 0.81 + 1 + 0.01 + 0.04 + 0.04 + 0.04}{10} \\ &= \frac{3.26}{10} \\ &= 0.326 \end{aligned}$$

$Cov[D, P]$

$$\begin{aligned} &= E[(D - E[D])(P - E[P])] \\ &= \frac{0.1 \cdot (-0.5) + 0.9 \cdot 1.5 + (-0.1) \cdot 0.4 + (-0.7) \cdot (-0.3) + 0.9 \cdot 0.3 + (-1.0) \cdot (-0.7) + 0.1 \cdot 0.8 + (-0.2) \cdot (-0.2) + (-0.2) \cdot (-0.7) + 0.2 \cdot (-0.6)}{10} \\ &= \frac{-0.05 + 1.35 - 0.04 + 0.21 + 0.27 + 0.7 + 0.08 + 0.04 + 0.14 - 0.12}{10} \\ &= \frac{2.58}{10} \\ &= 0.258 \end{aligned}$$

(3) 回帰直線は

$$y + 1.5 = \frac{0.258}{0.326}(x - 16.4)$$

ですから、 $x = 16.0$ を代入すると

$$\begin{aligned} y + 1.5 &= \frac{0.258}{0.326}(16.0 - 16.4) \\ y &= -\frac{0.258 \cdot 0.4}{0.326} - 1.5 \\ &\sim -1.8166 \end{aligned}$$

となりますから、平均気温が 16.0 だった場合の平均価格は概算で -1.82 となります。

問題 2 密度関数が

$$h(x) = \begin{cases} x & 0 \leq x \leq 1 \\ -x + 2 & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

で与えられる確率変数 X の期待値を計算して下さい。

配点：10 点 シラバス達成度目標：ウ

【解答例】

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xh(x)dx \\ &= \int_0^1 x^2 dx + \int_1^2 x(-x+2)dx \\ &= \left[\frac{1}{3}x^3 \right]_0^1 + \left[-\frac{1}{3}x^3 + x^2 \right]_1^2 \\ &= \frac{1}{3} - \frac{8}{3} + 4 + \frac{1}{3} - 1 \\ &= 1 \end{aligned}$$

問題 3 確率変数 A は正規分布 $N(0, 1)$ に従い、 B は正規分布 $N(1, 3)$ に従っており、これらは互いに独立であるとします。このとき標準正規分布表を使って確率 $P[-1 \leq A + B \leq 2]$ を求めて下さい。ただし、独立な正規分布の和がまた正規分布に従うことは既知とします。

配点：10 点 シラバス達成度目標：エ、オ

【解答例】 $A + B$ は正規分布 $N(1, 4)$ に従うので、

$$\begin{aligned} P[-1 \leq A + B \leq 2] &= P[-1 \leq N(1, 4) \leq 2] \\ &= P[-2 \leq N(0, 4) \leq 1] \\ &= P\left[-1 \leq N(0, 1) \leq \frac{1}{2}\right] \\ &= P[0 \leq N(0, 1) \leq 1] + P\left[0 \leq N(0, 1) \leq \frac{1}{2}\right] \\ &\sim 0.3413 + 0.1915 \\ &= 0.5328 \end{aligned}$$

となります。

問題 4 次の文章が正しいかどうか、○/×のみで答えて下さい（登場する平均値、分散、密度関数等は全て存在するものとします）。

（1）確率変数の和の分散 $Var[A+B]$ はそれぞれの分散の和 $Var[A]+Var[B]$ に一致します。

（2）標本平均の平均値は母集団の平均値より少し小さい。

（3）2次元の確率変数 (X, Y) の密度関数が分かっているならば、成分確率変数 X, Y の密度関数も分かります。

配点：各5点 | シラバス達成度目標：イ、オ、カ、キ

【解答例】 （1）×、（2）×、（3）○。

問題 5 ある工場で生産しているケース入洗剤の内容量は、従来の測定によって母分散が $(3.5\text{g})^2$ の正規分布に従う事が知られています。

ある単位生産時間の製品の中から 25 個を無作為に抽出して測定したところ、平均値が 202.8g でした。内容量の母平均の信頼度 95 %の信頼区間を求めて下さい。

配点：10点 | シラバス達成度目標：カ、ク

【解答例】 この母集団からとった大きさ 25 の標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{3.5^2}{25}\right)$ に従います。

まず

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

となる正数 d を求めます。

$$\begin{aligned} 0.95 &= P\left[\left|N\left(m, \frac{3.5^2}{25}\right) - m\right| \leq d\right] \\ &= P\left[|N(0, 1)| \leq \frac{d}{\frac{3.5}{5}}\right] \\ &= 2P\left[0 \leq N(0, 1) \leq \frac{d}{0.7}\right] \\ 0.475 &= P\left[0 \leq N(0, 1) \leq \frac{d}{0.7}\right] \end{aligned}$$

ですから、標準正規分布表によれば

$$\frac{d}{0.7} \sim 1.96, \quad \text{従って} \quad d \sim 1.372$$

が得られます。従って

$$P[|\bar{X} - m| \leq 1.372] \sim 0.95$$

ですから、今回のサンプル平均値が 202.8 だったことから、信頼度 95 %で

$$|202.8 - m| \leq 1.372, \quad \text{従って} \quad 201.428 \leq m \leq 204.172$$

が成立しており、これが求める信頼区間になります。

問題 6 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ます。計画では寿命の平均値は 1800 時間になるように製造されている筈ですがこの平均値に関して疑義が生じています。

そこで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かどうか有意水準 5 % で仮説検定して下さい。

配点：15 点	シラバス達成度目標：カ、ケ
---------	---------------

【解答例】 帰無仮説 H_0 ：『電球の寿命時間の平均値は 1800 時間である』が正しいと仮定します。

すると調査に使った 25 個のサンプルは、正規分布 $N(1800, 100^2)$ に従う母集団からの大きさ 25 のサンプルと考えられ、この母集団からとった大きさ 25 の標本平均を $\bar{X}$ と書く事にすれば $\bar{X}$ は正規分布 $N\left(1800, \frac{100^2}{25}\right)$ に従うことが分かります。

この場合は平均値が 1800 時間より長いかどうかを問題にしているので、

対立仮説 H_1 ：『電球の寿命時間の平均値は 1800 時間より長い』

と考え、右片側検定としてまず棄却域を求めます。

$$0.05 = P[\bar{X} - 1800 \geq d]$$

となる $d > 0$ を求めると、

$$\begin{aligned} &= P\left[N\left(0, \frac{100^2}{25}\right) \geq d\right] \\ &= P\left[N(0, 1) \geq \frac{d}{20}\right] \\ &= 0.5 - P\left[0 \leq N(0, 1) \leq \frac{d}{20}\right] \\ 0.45 &= P\left[0 \leq N(0, 1) \leq \frac{d}{20}\right] \end{aligned}$$

ですから正規分布表を参照して $\frac{d}{20} \sim 1.645$ 、つまり、 $d \sim 32.9$ が分かります。これは

$$0.05 \sim P[\bar{X} - 1800 \geq 32.9]$$

を意味し、求める右片側棄却域は $[1832.9, \infty)$ です。

今回の具体値 1835 はこの棄却域に入っていますから、帰無仮説は棄却され、寿命は 1800 時間より長いと言えます。