✍ ✍ ✍ 統計学 II の課題問題（10 月 28 日）と解答 [ref;x.y] は×章ｙ番 ✍ ✍ ✍

1. [ref;**5.2**] X が $N(3,1)$ に従うとき、次の値を求めよ．

(a) $P(X<3)$ (b) $P(X<4.93)$

(c) $P(X>3.06)$ (d) $P(1<X<4.2)$

(e) $P(1.5<X<2.5)$ (f) $P(|X-2|<1)$

(g) $P(X>c)=0.1$ を満たす c の値 (h) $P(X<c)=0.2$ を満たす c の値

【解】確率変数 $X\sim N(\mu,\sigma^2)$ を確率変数 Z に $Z=\dfrac{X-\mu}{\sigma}$ と標準化すると、$Z\sim N(0,1)$ にしたがうから、正規分布表をもちいることで確率事象の算出、あるいは確率の数値からパーセント点の値を求めることができる。$X=\sigma Z+\mu$ を代入して Z の不等式に変形する。正規分布表は 値 s と確率 $\Phi(s)$ を表す：$P(Z<s)=P(Z\leqq s)=\displaystyle\int_{(-\infty,s)}\phi(t)\,dt=\int_{-\infty}^{s}\phi(t)\,dt=\Phi(s)$. 連続型の確率変数は等号、不等号の有無でも同じ値であること．また差の確率や補事象の性質 $P(a<X<b)=P(\{a<X\}\cap\{X<b\})=P(\{X<b\}\cap\overline{\{X\leq a\}})=P(X<b)-P(X\leq a)=\Phi(b)-\Phi(a)$ に注意.

(a) 条件 $\mu=3,\sigma=1$ から、確率事象の領域を定める関係は $\{X<3\}=\left\{\dfrac{X-3}{1}<\dfrac{3-3}{1}\right\}=P(Z<3)$ となる。$P(X<3)=P(Z<0)=\Phi(0)=0.5000$

(b) $P(X<4.93)=P\left(Z<\dfrac{4.93-3}{1}\right)=P(Z<1.93)=\Phi(1.93)=0.972$

(c) $P(X>3.06)=P\left(Z>\dfrac{3.06-3}{1}\right)=P(Z>0.06)=1-P(Z\leqq 0.06)=1-P(Z<0.06)=1-\Phi(0.06)=1-0.2239=0.4761$

(d) $P(1<X<4.2)=P\left(\dfrac{1-3}{1}<Z<\dfrac{4.2-3}{1}\right)=P(-2<Z<1.2)=P(Z<1.2)-P(Z>-2)=\Phi(1.2)-[1-P(Z<2)]=\Phi(1.2)-[1-\Phi(2)]=0.8621$

(e) $P(1.5<X<2.5)=P(1.5-3<Z<2.5-3)=P(-1.5<Z<-0.5)=P(0.5<Z<1.5)=\Phi(1.5)-\Phi(0.5)=0.2417$

(f) $P(|X-2|<1)=P(-1<X-2<1)=P(-1<(1\cdot Z+3)-2<1)=P(-1<Z+1<1)=P(-2<Z<0)=P(0<Z<2)=\Phi(2)-\Phi(0)=0.4772$

(g) 以下は確率値からパーセント点を求める．このときには、正規分布表に掲げた数値にぴったりと当てはまらない場合が多いが、求める数値の桁数を増やすためには、2 カ所の数値を結んで作られる三角形の相似を利用した近似法がよく用いられる。本文にもその説明式があるので参考にされたい。
$P(X>c)=0.1\Rightarrow P(X<c)=0.9\Rightarrow P\left(Z<\dfrac{c-3}{1}\right)=0.9\Rightarrow\Phi(c-3)=0.9\Rightarrow$ 確率値 $\Phi(z)$ からパーセント値 z の逆引きをすると $\alpha=1.28$ のとき $\Phi(\alpha)=0.8997$, $\beta=1.29$ のとき $\Phi(\beta)=0.9015$ $\Rightarrow 1.28<c-3<1.29\Rightarrow 4.28<c<4.29\Rightarrow$ 三角形の近似で 約 $c\fallingdotseq 2.282$ を得る.

(h) 前問と同様に考える．$P(X<c)=0.2\Rightarrow\Phi(c-3)=0.2\Rightarrow$ 用いる正規分布表は 0.5 以上の値にたいするものだから、$1-\Phi(c-3)=1-0.2=0.8$ であるから、 $\Phi(3-c)=0.8$ を解く $\Rightarrow$ 逆引きから $\alpha=0.84$ のとき $\Phi(\alpha)=0.7995$, $\beta=0.85$ のとき $\Phi(\beta)=0.8023\Rightarrow 0.84<3-c<0.85\Rightarrow$ $c\fallingdotseq 2.158$

2. [ref;**5.7**] **以下の数値はある生徒の 10 日間の通学時間 (分) を示したものである.**

36 32 26 22 44 38 34 32 42 34

通学時間の平均と標準偏差を求めよ.

この生徒の通学時間はこれらの平均と標準偏差をもつ正規分布に従うとして、(a) 生徒のある日の通学時間が 38 分以上となる確率を求めよ. (b) 通学時間がある時間を越えることは高々 10 回に 1 回にしたいとすれば、その時間は何分か.

【解】個数 10 の数値の算術平均をとると $\overline{x} = \dfrac{1}{10}(36+32+26+\cdots+42+34) = 34$ また標準偏差の計算は 2 乗値の算術平均と平均値の 2 乗とから計算できる。$s^2 = \dfrac{1}{10}(36^2+32^2+26^2+\cdots+42^2+34^2) - 34^2 = 40$, $\therefore s \fallingdotseq 6.3$ となる。(a) 毎日の通学時間を確率変数 X で表すと、仮定から $X \sim N(34, 40)$ となる。標準化をすると、

$$P(X > 38) = P\left(Z > \frac{38-40}{\sqrt{40}}\right) = 1 - \Phi(0.63) = 0.2643$$

(b) 求める時間を c とすると、$P(X > c) = 1/10$ を求めればよい。正規分布表の範囲 $0.5000 \sim 0.9999$ にするために変形して、$1 - \dfrac{1}{10} = P(X < c) = P(Z < \dfrac{c-34}{\sqrt{40}}) = \Phi\left(\dfrac{c-34}{\sqrt{40}}\right) = 0.9$ を解く。付表 3 をもちいると、若干正確な値があるからこれをもちいる。$\Phi(1.282) = 0.9$ から、$\dfrac{c-34}{\sqrt{40}} = 1.282 \therefore c \fallingdotseq 42.1$

3. [ref;**5.3**] **ある電球の寿命 $\overline{X}$ は，過去の経験から平均 1500 時間，標準偏差 25 時間の正規分布に従うことが知られている．寿命 X が**

(a) 1500 時間以上

(b) 1489 時間未満

(c) 1475 時間から 1550 時間の間

にある電球の割合を求めよ.

【解】平均 $\mu = 1500$ と標準偏差 $\sigma = 25$ が与えられているから、これから標準形への線形変換をすると、$Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$ となり、正規分布表をもちいればよい。

(a) “数値より大きい”という右裾の確率であるから、$P(Z > \alpha) = 1 - \Phi(\alpha), (\alpha > 0)$ の形．$P(X > 1530) = P\left(\dfrac{X-1500}{25} > \dfrac{1530-1500}{25}\right) = P(Z > 1.2) = 1 - \Phi(1.2) = 0.1151$

(b) この問いは“数値より小さい”という確率であるが、“数値が負”であるから、対称性をもちいた反転の部分を計算すればよい．$P(X < 1480) = P\left(\dfrac{X-1500}{25} < \dfrac{1480-1500}{25}\right) = P(Z < -0.8) = 1 - \Phi(0.8) = 0.2119$

(c) 数値の範囲が負から正までにわたるから、つぎのようになる.

$$\begin{aligned} P(1475 < X < 1550) &= P\left(\frac{1475-1500}{25} < \frac{X-1500}{25} < \frac{1550-1500}{25}\right) \\ &= P(-1 < Z < 2) = \Phi(2) - (1-\Phi(1)) = \Phi(2) + \Phi(1) - 1 = 0.8185 \end{aligned}$$

4. [ref;**6.3**] X と Y は独立な確率変数で

$$X \sim N(5,9), \quad Y \sim N(7,16)$$

のとき，**(a)** $X-Y$ **(b)** $3X+Y$ **(c)** $\overline{X}$ **(d)** $\overline{X}-\overline{Y}$ の分布を求めよ．ただし $\overline{X}$ は $N(5,8)$ からとった大きさ **25** 個の無作為標本の平均で、$\overline{Y}$ は $N(7,16)$ からとった大きさ **16** 個の無作為標本の平均である．

【解】(a) $E(X-Y)=E(X)-E(Y)=5-7=-2$, (平均は ± の符号はそのまま) $V(X-Y)=V(X)+V(Y)=9+16=25$ （分散は差であっても和となることに注意）$\therefore X-Y \sim N(-2,25)$ (b) $E(3X+Y)=3E(X)+E(Y)=3\times 5+7=22$ (c) $E(\overline{X})=E(X)=5, V(\overline{X})=\dfrac{\sigma^2}{n}=\dfrac{9}{25}=0.36$, $\therefore \overline{X}\sim N(5,0.36)$ (d) $E\left(\overline{X}-\overline{Y}\right)=E\left(\overline{X}\right)-E(\overline{Y})=E(X)-E(Y)=-2$, また差であっても分散は和になることに注意して $V(\overline{X}-\overline{Y})=V(\overline{X})+V(\overline{Y})=\dfrac{9}{25}+\dfrac{16}{16}=1.36$ となる．したがって平均と分散が定まれば、対応する分布は正規分布であるから、 $\overline{X}-\overline{Y}\sim N(-2,1.36)$ である．

5. [ref;**6.5**] 高校３年生の身長は次のような正規分布に従うことが知られている．

男子： **169.3 cm, 標準偏差 4.6 cm** 女子： **156.6 cm, 標準偏差 4.2 cm**

これら２つの正規母集団からとった大きさ **64** と **36** の標本平均をそれぞれ $\overline{X},\overline{Y}$ とするとき、次の値を求めよ．

(a) $\overline{X},\overline{Y}$ の平均と標準偏差

(b) $\overline{X}-\overline{Y}$ の平均と標準偏差

(c) $\overline{X}$ が $\overline{Y}$ を少なくとも **12** ㎝超える確率

【解】男子、女子の身長をそれぞれ、X,Y と表すと、$X\sim N(169.3,4.6^2), Y\sim N(156.6,4.2^2)$ であり、その母集団からの標本平均、$\overline{X},\overline{Y}$ とおく．(a) 平均は $E(\overline{X})=E(X)=169.3$ (cm), 分散は $V(\overline{X})=\dfrac{V(X)}{64}=\dfrac{4.6^2}{64}$ (cm^2)．したがって、標準偏差は分散の平方根であるから,$\sqrt{V(\overline{X})}=\dfrac{4.6}{8}=0.575$ (cm). 女子についても同様に、平均は $E(\overline{Y})=E(Y)=156.6$ (cm), 分散は $V(\overline{Y})=\dfrac{4.2^2}{36}$ (cm^2). したがって、標準偏差は,$\sqrt{V(\overline{X})}=\dfrac{4.2}{6}=0.7$ (cm).

(b) 差の平均は $E(\overline{X}-\overline{Y})=E(X)-E(Y)=169.3-156.6=12.7$ (cm), 分散は $V(\overline{X}-\overline{Y})=\dfrac{V(X)}{64}+\dfrac{V(Y)}{36}=\dfrac{4.6^2}{64}+\dfrac{4.2^2}{36}$ (cm^2)．したがって、標準偏差は $\sqrt{V(\overline{X}-V(\overline{Y}))}=\sqrt{\dfrac{V(X)}{64}+\dfrac{V(Y)}{36}}=\sqrt{\dfrac{4.6^2}{64}+\dfrac{4.2^2}{36}}=0.906$ (cm).

(c) 前問の結果から, 差の分布が $\overline{X}-\overline{Y}\sim N(12.7,0.906^2)$ であるから、求める確率はこれを標準化した確率変数 $Z=\dfrac{(\overline{X}-\overline{Y})-12.7}{0.906}$ で計算する。$P(\overline{X}-\overline{Y}>12)=P(Z>\dfrac{12-12.7}{0.906})=P(Z>-0.773)=1-\Phi(0.773)=1-0.7803=0.2197.$

※本文テキストの訂正をしています。

6. [ref;**6.9**] 次の問いに答えよ．

(a) 平均 μ ，分散 σ^2 の正規母集団から大きさ 4 の無作為標本の平均を $\overline{X}$ とするとき、確率

$$P(|\overline{X}-\mu|<\sigma)$$

の値を求めよ．

(b) 平均 80, 標準偏差 10 の母集団から大きさ **64** の無作為標本をとるとき、標本平均 $\overline{X}$ が **82** を越える確率を求めよ．

【解】(a) 標本数の大きさが 4 であるから、分散の値が $\dfrac{\sigma^2}{4}$ となる．よって標本平均 $\overline{X}$ の分布は $N\left(\mu, \dfrac{\sigma^2}{4}\right)$ となる。これを標準正規分布になるよう線形変換するには $Z=\dfrac{\overline{X}-\mu}{\sigma/2}$ とおけばよい。したがって

$$\begin{aligned} P\left(|\overline{X}-\mu|<\sigma\right) &= P\left(\mu-\sigma<\overline{X}<\mu+\sigma\right) \\ &= P\left(\frac{(\mu-\sigma)-\mu}{\sigma/2}<Z<\frac{(\mu+\sigma)-\mu}{\sigma/2}\right) \\ &= P\left(-2<Z<2\right)=\Phi(2)-(1-\Phi(2))=2\,\Phi(2)-1=0.9545 \end{aligned}$$

✐ ✐ ✐ まとめと補足 ✎ ✎ ✎

I. 確率変数の平均と分散, 共分散

標本データ (確率変数)： $X_1, X_2, \cdots, X_n, \cdots$ ；その分布を母集団分布 (Population distribution)

(i) 無作為標本 ⇔ 独立同一分布 (確率変数が独立で、同じ分布にしたがう)

(ii) 標本サイズ (size) とは、個数のこと．n, m, などが一般にもちいられる．インデックスは i, j, k など．標本平均、あるいはデータの平均値（average）とは算術平均で、変数値の和を個数で割った比率あるいは割合 $\overline{X}=\dfrac{1}{n}\sum_i X_i=\dfrac{1}{n}(X_1+X_2+\cdots+X_n$.

(iii) 期待値 (expectation) は変数値にその確率を掛けた総和値、あるいは積分値．$E(X)=\mu$ は平均であり、平均を引いて 2 乗した値 (2 乗偏差) の期待値が分散 $V(X)=E(X-\mu)^2=E(X^2)-\mu^2$ たとえば、無作為抽出されたデータの場合、標本平均の期待値は $E(\overline{X})=\mu$ なぜなら、すべての $i=1,2,\cdots,n$ について、$E(X_i)=E(X)=\mu$ で総和あるいは積分の線形性から、もとの集団での確率変数 X の分布と同じであるから、等しい値となる．ただし、この標本平均の線形性の性質は、「独立でなくても成立する」性質であることに注意する．分散はついては共分散の項が加わり、それぞれの分散の和とはならない．

(iv) 分散 (variance) について変数値を a 倍し、数値 b を加えると、 $V(aX+b)=a^2V(X)$ となる．b には依存しないが、定数 a の 2 乗が係数にかかる。たとえば、$V(b-X)=V(-X+b)=V(X)$ となる．分散は相互のズレ違いを表す尺度であり、平均値が影響しない。つまりデータ相互のすれ具合を表すから、データのそれぞれを平行移動しても同じ値となる．$n=3$, X_1, X_2, X_3 の分散と $10-X_1, 10-X_2, 10-X_3$ の分散は同じである．

一般に数式で表現すると c は定数、$X, Y, Z, \cdots$ は確率変数、独立を仮定しない。(a) $E(c)=c$, $E(cX)=cE(X)$, $E(X+Y+\cdots+Z)=E(X)+E(Y)+\cdots+E(Z)$ (b) 平均値 (期待値)$\mu_X=E(X)$, $\mu_Y=E(Y)$ などとするとき、X の分散の定義は $V(X)=E(X-\mu_X)^2=E(X^2)-(\mu_X)^2$, X, Y の共分散 (covariance) $Cov(X,Y)$ とは、$Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]=E[X\,Y]-\mu_X\,\mu_Y$ 2 個の和 $X+Y$ に対し

ては、$V(aX+bY)=a^2V(X)+2abCov(X,Y)+b^2V(Y)$, 3 個の和であれば、$V(aX+bY+cZ)=a^2V(X)+b^2V(Y)+c^2V(Z)+2\{abCov(X,Y)+acCov(X,Z)+bcCov(Y,Z)\}$, 3 項の和を 2 乗した関係式から得られる。もしこれら確率変数が独立であれば、積の期待値がそれぞれの期待値の積となるから、$E(XY)=E(X)E(Y)$ などとなり、$V(aX+bY+cZ)=a^2V(X)+b^2V(Y)+c^2V(Z)$ となって、共分散の項が消える．ただし、期待値の積と積の期待値が等しくなる場合は、独立であれば成り立つが、逆に共分散がゼロとなっても独立でない場合があることに注意する．

II. 表計算ソフトによる正規分布の確率計算

(i) 標準正規分布 $Z\sim N(0,1)$ の確率計算：

“norm.s.dist”関数 (s は standard の意味）　入力形式と引数 = norm.s.dist (セル番号（値）,true/false), true は 領域 $(-\infty,z)$ での確率（分布関数値）$P(Z<z)$ を求める．false は　z での密度関数 $\phi(z)=\dfrac{1}{\sqrt{2\pi}}e^{-z^2/2}$.

$$z\to\alpha;\quad P(Z<z)=\Phi(z)=\int_{-\infty}^{z}\phi(t)dt=\alpha$$

例; 値 (z) から確率 (α):

`=norm.s.dist(0, true)` $\Leftrightarrow \Phi(0)=0.5$

`=norm.s.dist(0, false)` $\Leftrightarrow \phi(0)=0.3989$

`=norm.s.dist(1.2, true)` $\Leftrightarrow \Phi(1.2)=0.8849$

`=norm.s.dist(1.2, false)` $\Leftrightarrow \phi(1.2)=0.1942$

`=norm.s.dist( -2.3, false)` $\Leftrightarrow P(Z<-2.3)=1-P(Z>2.3)=1-\Phi(2.3)=0.028327$

(ii) 一般の正規分布 $X\sim N(\mu,\sigma^2)$ ：“norm.dist” 関数 (平均 μ と標準偏差 σ を入力)

入力形式と引数 = norm.dist (セル番号（値）,μ, σ, true/false)

例; $X\sim N(50,100)$ の場合；値 (x) から確率 (α): $P(X<x)=\displaystyle\int_{-\infty}^{x}f_X(t)dt=\alpha,\ Z=\frac{X-\mu}{\sigma}$

`=norm.dist(50, 50, 10, true)` $\Leftrightarrow P(X<50)=\Phi(0)=0.5$

`=norm.dist(50, 50, 10, false)` $\Leftrightarrow f_X(50)=0.0398=\dfrac{\phi(0)}{10}$

`=norm.dist(55, 50, 10, true)` $\Leftrightarrow P(X<55)=\Phi\left(\dfrac{55-50}{10}\right)=0.6915$

`=norm.dist(55, 50, 10, false)` $\Leftrightarrow f_X(55)=0.0352$

(iii) 逆引き、逆関数 (inverse) の数値計算：$P(X<z)=\alpha$ となる z を求める．$\Phi^{-1}(\alpha)=z$

例; 確率 (α) から値 (z) を求めること:

`=norm.inv(50, 50, 10)` $\Leftrightarrow P(X<c)=0.5$ となる c の値； $c=50$

`=norm.inv(55, 50, 10)` $\Leftrightarrow P(X<c)=0.6915$ となる c の値；

$c=50.00107$ ※近似計算の誤差が含まれ、前問の例 $P(X<55)=0.6915$ とは多少のずれが起こる．