

# シミュレーション (クラスS/E)

小池 行弘

2006 年 7 月 30 日

## 目次

<b>0</b>	<b>複利法 (1 講)</b>	<b>2</b>
0.1	Excel の利用 . . . . .	2
0.2	72 の法則 . . . . .	3
<b>1</b>	<b>金利のシミュレーション (2-4 講)</b>	<b>5</b>
1.1	平均寿命 . . . . .	5
1.2	年金のシミュレーション . . . . .	6
1.3	利息制限 . . . . .	7
<b>2</b>	<b>線形計画法 (5-7 講)</b>	<b>9</b>
2.1	連立 1 次方程式 . . . . .	9
2.2	図解法 . . . . .	11
2.3	Simplex 法 . . . . .	12
2.4	潜在価格 (shadow price) . . . . .	14
2.5	被約費用 (reduced cost) . . . . .	14
<b>3</b>	<b>回帰分析 (8-10 講)</b>	<b>17</b>
3.1	平均値・分散・標準偏差 . . . . .	17
3.2	共分散・相関係数 . . . . .	18
3.3	回帰分析 . . . . .	19
3.4	決定係数 . . . . .	22
3.5	重回帰分析 . . . . .	22
3.6	Excel の行列関数 . . . . .	23
<b>4</b>	<b>確率分布 (11-13 講)</b>	<b>25</b>
4.1	度数分布表と確率分布 . . . . .	25
4.2	二項分布 . . . . .	26
4.3	標準正規分布 . . . . .	29
4.4	二項分布の正規近似 . . . . .	29
4.5	乱数 . . . . .	30
4.6	在庫管理 . . . . .	31
<b>5</b>	<b>定期試験とレポートの概要</b>	<b>33</b>

## 0 複利法 (1 講)

### 0.1 Excel の利用

まずはじめに基本的な Excel の操作をまとめておこう。

- (数式の入力) セルに数式を書き込んで自動的に計算させることができる。数式は必ず = で始める。

Excel の数式	通常の数式	Excel の画面
=5+2	5 + 2	7
=5-2	5 - 2	3
=5*2	5 × 2	10
=5/2	5 ÷ 2	2.5
=5^2	5 <sup>2</sup>	25

指数になるのは、表示形式が“標準”の場合である。

- (指数表現) 桁数が非常に大きい数、あるいは 0 に非常に近い数は、Excel 画面上では指数表示に改められる。

Excel の表示	指数表示	通常の手書き
1.23E+05	$1.23 \times 10^5$	123000
1.23E-05	$1.23 \times 10^{-5}$	0.0000123
-1.23E+04	$-1.23 \times 10^4$	-12300
-1.23E-04	$-1.23 \times 10^{-4}$	-0.000123

表示形式を“数値”等に変更するとよい。

- (セルの書式設定) 思ったおりの表示にならないときは、セルの書式設定を行う。たとえば、1.23E+05 と表示されたセルを 123000 に変更したいときは、該当セルを右クリックして“セルの書式設定 (F)...”を選択する。
- (相対参照と絶対参照) セル参照を含む数式が書かれたセルを、別の場所にコピーしたとき、通常はコピー元とコピー先のセル参照は異なっている (相対参照)。コピーしても参照位置がずれないように書くこともできる (絶対参照)。セルの行番号または列番号の前に \$ 記号を挿入すると、絶対参照となる。

相対参照	絶対参照
=A1	=\$A\$1
絶対参照 (行のみ)	絶対参照 (列のみ)
=A\$1	=\$A1

セル位置入力中に、F4 キーを押すと絶対参照に切り替わる。

ソルバーは、金利シミュレーション (2-4 講)、線形計画法 (5-7 講) で利用する。

- (ソルバー) 目的セル、変化させるセル等を指定して、最適解を求めるツールである。最適解は、Excel が自動的に行う数値計算によって求められるもので、いつも同じ値を返すとは限らない。ソルバーを実行するには、メニューバーから“ツール” → “ソルバー (V)...”を選択する。メニューにないときは、“ツール” → “アドイン (I)...”を選択し、“ソルバーアドイン”がチェックされているか確認する。

## 0.2 72の法則

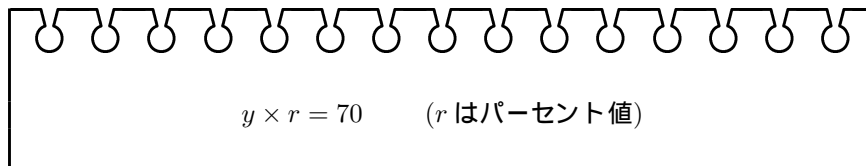
利率が  $r$  のとき、元利合計 (元金+利息) がどのように増加していくかを考える。利率が 1% の場合は  $r = 0.01$  とおく。  $x$  円の元金が 2 倍 ( $2x$  円) に増えるまでの年数を  $y$  とおき、この  $y$  を求めたい。

ここでは 1 年単位の複利法を用いる。つまり 1 年分の利息が次年度の元金に組み入れられるものとする。次表は  $x = 10000$ ,  $r = 0.1$  (10%) の場合の変化のようすである。

年次	元金	利息	合計
0	10,000	1,000	11,000
1	11,000	1,100	12,100
2	12,100	1,210	13,310
3	13,310	1,331	14,641
4	14,641	1,464	16,105
5	16,105	1,611	17,716
6	17,716	1,772	19,487
7	19,487	1,949	21,436
8	21,436	2,144	23,579

7 年後に 19,487 円とおよそ 2 倍になるから、 $x = 10000$ ,  $r = 0.1$  のとき、およそ  $y = 7$  である。

与えられた利率  $r$  に対して、元金が 2 倍に増加するまでの年数  $y$  を知る方法として、次にあげる 2 つの公式がよく知られている。1 つめは “70 の法則” である。



$$y \times r = 70 \quad (r \text{ はパーセント値})$$

この公式の説明 元金が  $x$  のとき、利息は  $x \times r$  なので、1 年後の元利合計は  $x(1+r)$  である。

$$\begin{aligned} \text{元 + 利} &= x + xr \\ &= x(1+r) \end{aligned}$$

である。つまり  $1+r$  を掛けると、1 年後の元利合計が求まる。

同様に 2 年後の元利は  $x(1+r)^2$  であり、さらにくりかえすと、 $y$  年後の元利は  $x(1+r)^y$  となることがわかる。  $y$  年後に 2 倍 ( $= 2x$ ) になっていることから、

$$x(1+r)^y = 2x$$

両辺を  $x$  で割って、 $x$  を消去すると次式が成り立っている。

$$(1+r)^y = 2$$

この式から  $y$  を求めればよい。対数表示に改め、

$$y \log(1+r) = \log 2$$

近似式を求めるためには微積分の知識が必要であるが、この講義では微積分は必要でない。

$\log(1+r)$  の 2 次までの近似式を用いて、

$$\log(1+r) = r \times \frac{2-r}{2}$$

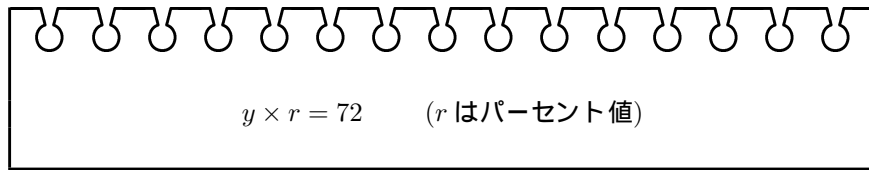
$\log 2 = 0.693$  を代入すると、次のような関係が導かれる。

$$yr = \frac{2 \times 0.693}{2-r} \quad (*)$$

利率が 0.5 ~ 3.5% 程度するとき、(\*) 式にパーセント値の  $r$  を適用すると、次の“70 の法則”が得られる。

$$y \times r = 70$$

2 つめの公式は“72 の法則”である。



この公式の説明 利率が 6 ~ 9% 程度するとき、(\*) 式にパーセント値の  $r$  を適用すると、“72 の法則”が得られる。

$$y \times r = 72$$

2 つの公式を比べると“72 の法則”のほうがずっと有名である。しかし、法則の導き方と現在の金利と比べてみれば、“70 の法則”のほうがよりよい近似が得られるはずである。

#### 問題 0.1

- 年利が 7% のとき、元利合計が 2 倍になるまで何年かかるか。
- 年利が 0.1% のとき、元利合計が 2 倍になるまで何年かかるか。

#### 問題 0.2

- 3% の年利で 100 万円を預金する。複利期間を 1 年間としたとき、1 年後の元利はいくらになるか。
- 複利期間を 1 ヶ月に短縮したとき、1 年後の元利はいくらになるか。
- 複利期間を 1 日に短縮したとき、1 年後の元利はいくらになるか。

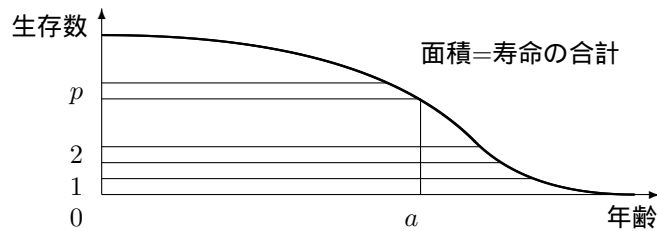
# 1 金利のシミュレーション (2-4 講)

## 1.1 平均寿命

下の表はある年の男性の死亡率を年齢ごとにまとめたものである。生存数および死亡数は人口十万人あたりの数を記してある。たとえば 75 歳時の生存者は 66727 人で、そのうちの 0.03784、即ち 2525 人がこの 1 年間に死亡したことを表している。

年齢	死亡率	生存数	死亡数
0	.00345	100000	345
1	.00051	99655	51
2	.00038	99604	38
⋮	⋮	⋮	⋮
75	.03784	66727	2525
76	.04162	64202	2672
77	.04606	61530	2834
⋮	⋮	⋮	⋮
110	.54918	4	2
111	.57147	2	1
112	.59378	1	0

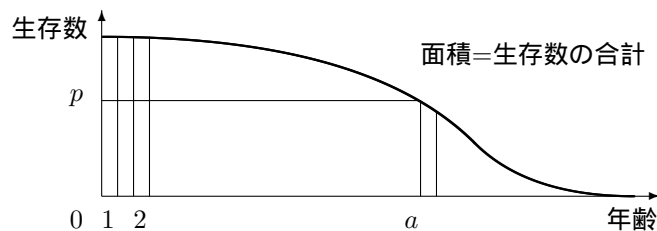
平均寿命とは、各人の誕生してから死亡するまでの年数 (寿命) をすべて合計し、その人数で割ったものである。



グラフは年齢に対応して生存数が減少していくようすを示したものである。p 番目の人の寿命は、a 歳であることが読み取れる。すべての人の寿命を合計すると、曲線と軸とで囲まれた図形の面積になっている。平均寿命は、寿命の合計を対象人数で割ると求まるので、次が成り立つ。

$$(\text{平均寿命}) = \frac{(\text{面積})}{100000} = \frac{(\text{寿命の合計})}{100000}$$

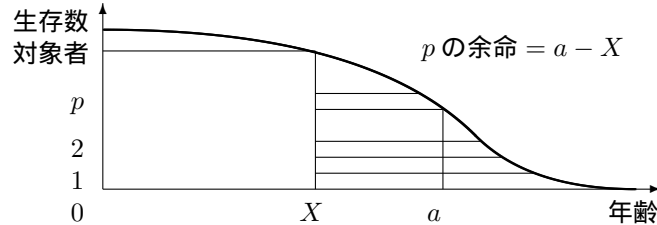
今度は同じグラフを縦方向に切ってみよう。年齢 a の時点で、p の人数だけ生存していることが読み取れる。各年齢について、その時点での生存数を合計すると、曲線と軸で囲まれた部分の面積になる。



$$(\text{平均寿命}) = \frac{(\text{面積})}{100000} = \frac{(\text{生存数の合計})}{100000}$$

$$\text{平均寿命} = \frac{\text{寿命の合計}}{\text{出生数}} = \frac{\text{生存数の合計}}{\text{出生数}}$$

平均余命とは、ある年齢において生存している人について、それぞれの人が死亡するまでの残りの年数(余命)を合計し、その人数で割ったものである。



75 歳における平均余命を求めるには、75 歳以降の生存数をすべて合計し、75 歳人口 (66727) で割ればよい。

$$\begin{aligned} \text{(75 歳の平均余命)} &= \frac{\text{(75 歳の余命合計)}}{66727} \\ &= \frac{\text{(75 歳以降の生存数合計)}}{66727} \end{aligned}$$

$$X \text{ 歳の平均余命} = \frac{X \text{ 歳の余命合計}}{X \text{ 歳人口}} = \frac{X \text{ 歳以降の生存数合計}}{X \text{ 歳人口}}$$

“平均寿命” は “0 歳における平均余命” に等しい。

## 1.2 年金のシミュレーション

年金制度の見直し計画について、その概要を確認しておこう。

- 国民年金保険料は 2004 年度まで月額 13,300 円であった。その後毎年月額 280 円ずつ値上げし、2017 年度以降は月額 16,900 円で固定する。(2006 年度は 13,860 円)
- 老齢基礎年金は 65 歳から受給が始まり、給付額は年額 794,500 円である。
- 基礎年金の財源に占める国庫負担の割合は 3 分の 1 であった。これを段階的に 2 分の 1 まで引き上げる。
- 厚生年金の保険料負担は 2004 年度まで収入の 13.58% であったが、2017 年度以降は 18.30% で固定する。(保険料は事業者と本人の折半)
- 厚生年金は 65 歳から受給し、収入の 50% を上回るものとする。
- 消費税の税収は 1% あたり 2 兆円ある。

### 1.3 利息制限

利息の上限金利は次の通り。(利息制限法)

- 元本が 100,000 円未満のとき, 年利 20%
- 元本が 100,000 円以上 1,000,000 円未満のとき, 年利 18%
- 元本が 1,000,000 円以上のとき, 年利 15%

この利率により計算した金額を超える利息は, その超過分について無効となる.

100 万円 (制限利率 15%) を 2006 年 1 月 1 日に借り, 利息 年 20% をもって同年 12 月 31 日に返す契約をしたとすると, 超過した  $20\% - 15\% = 5\%$  (5 万円) 分の利息は無効となる. 貸主は, 元本 100 万円と利息 15 万円分しか請求できない.

いわゆる利息の天引きについて. 次のような事例では利率が再計算される.

200 万円 (制限利率 15%) を 2006 年 1 月 1 日に借り, 利息 年 15% をもって同年 12 月 31 日に返す契約をした. ただし利息相当額 (30 万円) は天引きされる. 借主は 1 月 1 日に 170 万円を受け取り, 12 月 31 日に 200 万円を返済することになっていた.

この場合, 170 万円を借り, 200 万円を返す契約とみなされる. 上限金利は 15% なので, 貸主は 25 万 5 千円までの利息しか請求できない. 即ち, 4 万 5 千円分の利息は無効となる.

#### 問題 1.1 (利息制限)

21 万円を借入し, 1 ヶ月後に返済する. ただし利息相当額の 1 万円は借入時に天引きされていた. この場合の利率 (年利) は何%であるか. 利息制限法に従うとき, 無効とされる利息はいくらであるか.

## 課題 1.1 (平均寿命)

各年齢ごとの死亡率が男女別に与えられている。平均寿命 (0 歳の平均余命) を求めよ。

## 課題 1.2 (積立年金 1)

20 歳時の月収は 20 万円, 月収が毎年 1 万円ずつ上昇すると仮定して, 20 歳から 59 歳まで収入の 13.58% (負担率) を積み立てる。年利は 0.1% とする。60 歳から 81 歳まで積立金から一定額を受け取るとき, 給付額および給付率を求めよ。給付率は 40 歳時点の収入 (月収 40 万円) を基準とする。

“ソルバー:パラメータ設定” で, 以下の値を設定して “実行” ボタンを押す。

目的セル	給付後残額
目標値	“値” を選択して “0” を記入
変化させるセル	給付額

## 課題 1.3 (積立年金 2)

上記と同じ条件で, 65 歳から 81 歳まで受け取るとする。給付率が 50% となるようにするには, 負担率は何% にすればよいか。

目的セル	給付後残額
目標値	“値” = “0”
変化させるセル	負担率

## 課題 1.4

3000 万円を借り入れ, 35 年ローンで返済したい。借入金利は 3%, 複利期間は 1 年間, 年賦での返済とすると, 年間の返済額はいくらになるか。また 35 年間の返済総額はいくらになるか。

目的セル	35 年後の借入残高
目標値	“値” = “0”
変化させるセル	年賦額

## 課題 1.5

借入金利は 3% のまま, 複利期間を 1 ヶ月に短縮し, 月賦とした場合, 1 ヶ月間の返済額および返済総額はいくらになるか。

## 課題 1.6

3000 万円を借り入れ, 複利期間を 1 ヶ月として月賦で返済したい。計画によると 35 年後に返済が完了し, 返済総額はちょうど 6000 万円になるという。金利は何パーセントであるか。

月賦額に  $60000000 \div 35 \div 12$  を記入して, ソルバーを実行する。

目的セル	35 年後の借入残高
目標値	“値” = “0”
変化させるセル	金利

## 2 線形計画法 (5-7 講)

### 2.1 連立 1 次方程式

次の連立 1 次方程式を解く.

$$2x - 4y + 6z = 12 \quad (1)$$

$$3x + y + 2z = 11 \quad (2)$$

$$x - 3y - 3z = -14 \quad (3)$$

まず, (1) 式を 2 で割る.

$$x - 2y + 3z = 6 \quad (4)$$

(2) 式から (4) 式の 3 倍を引くと, (2) 式から  $x$  が消去でき, (3) 式から (4) 式を引くと, (3) 式から  $x$  が消去できる.

$$7y - 7z = -7 \quad (5)$$

$$-y - 6z = -20 \quad (6)$$

(5) 式を 7 で割る.

$$y - z = -1 \quad (7)$$

(6) 式に (7) 式を加えると, (6) 式から  $y$  が消去できるので,

$$-7z = -21 \quad (9)$$

$$z = 3 \quad (10)$$

これで  $z$  が求められた.  $z$  の値を (7) 式に代入すると  $y$  の値が, さらに (4) 式に代入すると  $x$  の値が決定できる.

$$y = 2 \quad (12)$$

$$x = 1 \quad (11)$$

以上の操作を図式的に行ってみよう. この計算方法は掃き出し法とよばれる.

$$\begin{array}{cccc|l}
 2 & -4 & 6 & 12 & (1) & 2x - 4y + 6z = 12 \\
 3 & 1 & 2 & 11 & (2) & 3x + y + 2z = 11 \\
 1 & -3 & -3 & -14 & (3) & x - 3y - 3z = -14 \\
 \hline
 1 & -2 & 3 & 6 & (4) \leftarrow (1) \div 2 & x - 2y + 3z = 6 \\
 3 & 1 & 2 & 11 & & \\
 1 & -3 & -3 & -14 & & \\
 \hline
 1 & -2 & 3 & 6 & & \\
 0 & 7 & -7 & -7 & (5) \leftarrow (2) - (4) \times 3 & 7y - 7z = -7 \\
 0 & -1 & -6 & -20 & (6) \leftarrow (3) - (4) & -y - 6z = -20 \\
 \hline
 1 & -2 & 3 & 6 & & \\
 0 & 1 & -1 & -1 & (7) \leftarrow (5) \div 7 & y - z = -1 \\
 0 & -1 & -6 & -20 & & \\
 \hline
 1 & 0 & 1 & 4 & (8) \leftarrow (4) + (7) \times 2 & x + z = 4 \\
 0 & 1 & -1 & -1 & & \\
 0 & 0 & -7 & -21 & (9) \leftarrow (6) + (7) & -7z = -21 \\
 \hline
 1 & 0 & 1 & 4 & & \\
 0 & 1 & -1 & -1 & & \\
 0 & 0 & 1 & 3 & (10) \leftarrow (9) \div (-7) & z = 3 \\
 \hline
 1 & 0 & 0 & 1 & (11) \leftarrow (8) - (10) & x = 1 \\
 0 & 1 & 0 & 2 & (12) \leftarrow (7) + (10) & y = 2 \\
 0 & 0 & 1 & 3 & & \\
 \hline
 \end{array}$$

重複を省き, 次のように計算してもよい.

$$\begin{array}{cccc|l}
 2 & -4 & 6 & 12 & (1) & 2x - 4y + 6z = 12 \\
 3 & 1 & 2 & 11 & (2) & 3x + y + 2z = 11 \\
 1 & -3 & -3 & -14 & (3) & x - 3y - 3z = -14 \\
 \hline
 1 & -2 & 3 & 6 & (4) \leftarrow (1) \div 2 & \\
 0 & 7 & -7 & -7 & (5) \leftarrow (2) - (4) \times 3 & \\
 0 & -1 & -6 & -20 & (6) \leftarrow (3) - (4) & \\
 \hline
 1 & 0 & 1 & 4 & (8) \leftarrow (4) + (7) \times 2 & \\
 0 & 1 & -1 & -1 & (7) \leftarrow (5) \div 7 & \\
 0 & 0 & -7 & -21 & (9) \leftarrow (6) + (7) & \\
 \hline
 1 & 0 & 0 & 1 & (11) \leftarrow (8) - (10) & x = 1 \\
 0 & 1 & 0 & 2 & (12) \leftarrow (7) + (10) & y = 2 \\
 0 & 0 & 1 & 3 & (10) \leftarrow (9) \div (-7) & z = 3 \\
 \hline
 \end{array}$$

### 問題 2.1

次の 3 元連立 1 次方程式を掃き出し法で解いてみよ.

$$\begin{cases} 3x - 3y + 6z = 15 \\ -x + 4y + z = 10 \\ 2x + y + 6z = 21 \end{cases}$$

## 2.2 図解法

次の問題を考える. このような問題は線形計画問題とよばれる. ここでは図を用いて解いてみよう.

## 問題 2.2

原料  $A$  と原料  $B$  を加工し, 製品  $X$  と製品  $Y$  を製造する工場がある. 1 単位の製品  $X$  を製造するためには, 2 キログラムの原料  $A$  と 2 キログラムの原料  $B$  を消費し, 1 単位の製品  $Y$  を製造するためには, 6 キログラムの原料  $A$  と 1 キログラムの原料  $B$  を消費するものとする. 製品  $X$  と製品  $Y$  の販売価格は 1 単位あたりそれぞれ 9 万円と 16 万円である. また原料  $A$  は 60 キログラム, 原料  $B$  は 20 キログラムまでしか調達できない. 販売価格の合計を最大にするには, 製品  $X$  と製品  $Y$  を何単位ずつ製造すればよいか.

問題の解答 製品  $X$  を  $x$  単位, 製品  $Y$  を  $y$  単位製造するとき, 原料  $A$  の消費量は  $2x + 6y$  となる. しかし原料  $A$  は 60 キログラムまでしか調達できないから, 次式の制約条件を満たす必要がある.

$$2x + 6y \leq 60$$

原料  $B$  についての制約条件は,

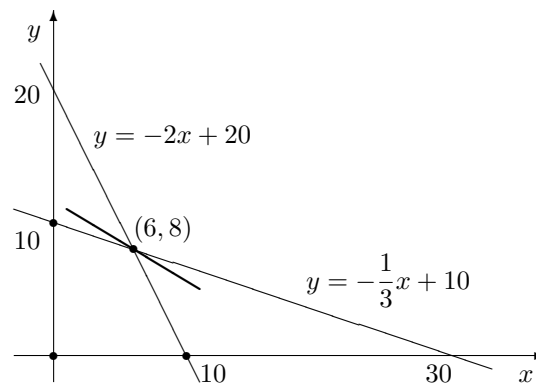
$$2x + y \leq 20$$

製造単位数  $x, y$  は負値をとれないから, 次式 of 非負条件を満たす.

$$x \geq 0, \quad y \geq 0$$

以上の制約条件, 非負条件の下で, 販売価格の最大値を求めることが, この線形計画問題の目的である. 販売価格  $p$  は次のように表されている.

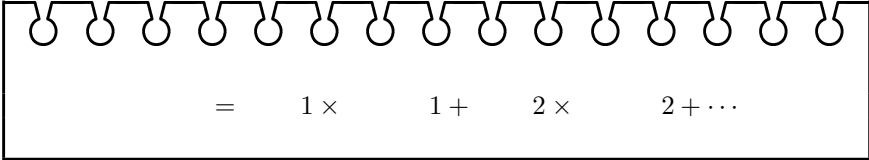
$$p = 9x + 16y$$



上記の制約条件を座標平面に書き込むと, 図のようになる. これらの条件を満たすのは図の四角形領域で, これを可能領域という. 可能領域の内部または境界上で,  $p$  が最大となる点を探せばよい. この問題では頂点  $(6, 8)$  において  $p$  が最大となる.

$$\max p = 9 \cdot 6 + 16 \cdot 8 = 182$$

したがって販売価格の最大値は 182 万円で,  $X$  を 6 単位,  $Y$  を 8 単位製造すればよい.



$$\text{最大値} = \text{単価 } 1 \times \text{製造数 } 1 + \text{単価 } 2 \times \text{製造数 } 2 + \dots$$

次に原料  $A, B$  がどれだけ消費されたか調べてみよう.

$$2x + 6y = 2 \cdot 6 + 6 \cdot 8 = 60$$

$$2x + y = 2 \cdot 6 + 8 = 20$$

となっていて, 原料をすべて使い切ったことがわかる.

### 2.3 Simplex 法

前節では図を描いて求めたが, 以下では, 変数が多い (図が描けない) 場合でも解くことができる Simplex 法とよばれる方法を紹介しよう. 製品  $X$  を  $x$  単位, 製品  $Y$  を  $y$  単位製造するとき, 制約条件は次のものである.

$$2x + 6y \leq 60$$

$$2x + y \leq 20$$

ここで, 新たに負でない変数 (スラック変数という)  $a \geq 0, b \geq 0$  を導入し, 不等式を等式におきかえてみよう. スラック変数  $a, b$  は, 使い切れずに残ってしまった原料  $A, B$  の余剰量を表している.

$$2x + 6y + a = 60 \quad (A)$$

$$2x + y + b = 20 \quad (B)$$

販売価格  $p = 9x + 16y$  を次の形に変形する.

$$-9x - 16y + p = 0 \quad (P)$$

Simplex 法とは, このようにして作った 3 つの式 (A), (B), (P) に掃き出し法を行って線形計画問題を求める方法のことである. 連立 1 次方程式を解く場合と同じようにして変形していく. (変形方法の詳細は後述する)

$x$	$y$	$a$	$b$	$p$	
2	6	1	0	0	60 (A)
2	1	0	1	0	20 (B)
-9	-16	0	0	1	0 (P)
1/3	1	1/6	0	0	10
5/3	0	-1/6	1	0	10
-11/3	0	8/3	0	1	160
0	1	1/5	-1/5	0	8 (Y) $\leftarrow y + \frac{1}{5}a - \frac{1}{5}b = 8$
1	0	-1/10	3/5	0	6 (X) $\leftarrow x - \frac{1}{10}a + \frac{3}{5}b = 6$
0	0	23/10	11/5	1	182 (**) $\leftarrow \frac{23}{10}a + \frac{11}{5}b + p = 182$

Simplex 法で変形すると, 最下行が次式のようなになる.

$$\frac{23}{10}a + \frac{11}{5}b + p = 182 \quad (**)$$

販売価格  $p$  の最大値を求めたい。スラック変数は負でない ( $a \geq 0, b \geq 0$ ) ので、 $a, b$  をできるだけ小さな値にすれば、その分だけ  $p$  を大きな値にすることができる。したがって  $a = b = 0$  のとき、販売価格  $p$  が最大となる。

$$\max p = 182$$

次の (X), (Y) 式において,

$$y + \frac{1}{5}a - \frac{1}{5}b = 8 \quad (Y)$$

$$x - \frac{1}{10}a + \frac{3}{5}b = 6 \quad (X)$$

$a = b = 0$  を用いると、製品  $X, Y$  の製造量が求められる。

$$x = 6, \quad y = 8$$

原料の使い残しは  $a = b = 0$  であるから、すべての原料を使い切ったことも確認できる。

変形方法 線形計画問題においては、単純に掃き出し法を行うだけでは最適解にたどりつけない。正であるべき数(制約量等)を正に保ったまま変形していくのがコツである。下に示した手順 2) は、制約量等が位置する右端の列を、常に正値に保つために必要となるものである。

$x$	$y$	$a$	$b$	$p$	
2	<b>6</b>	1	0	0	60 (10)
2	<b>1</b>	0	1	0	20 (20)
-9	<b>-16</b>	0	0	1	0
<b>1/3</b>	1	1/6	0	0	10 (30)
<b>5/3</b>	0	-1/6	1	0	10 (6)
<b>-11/3</b>	0	8/3	0	1	160
0	1	1/5	-1/5	0	8
1	0	-1/10	3/5	0	6
0	0	23/10	11/5	1	182

次の手順で変形する。最下行から負値がなくなれば変形を取りやめる。

- 1) 最下行にある負数 (-9, -16) のうち、どちらか適当な列を選択する。この変形では、絶対値が最大であるほう(第 2 列)を選択している。
- 2) 制約量を、選択した列の数で割り、その値が最小となる行(第 1 行)を選択する。
- 3) 選択した行と列(第 1 行 第 2 列)を軸に変形する。
- 4) 最下行に負数が残っていれば、手順 1) に戻って繰り返す。

うまく変形できない場合は、手順 1) に従わなくても構わない。別の列を軸に変形を試してみる。

どうしても負値が消えない場合、 $p$  の最大値が存在しないことが考えられる。

## 2.4 潜在価格 (shadow price)

変数  $x, y$  は各製品の製造量, スラック変数  $a, b$  は各原料の使い残しを表している.  $p$  は販売価格の合計である.

$$2x + 6y + a = 60 \quad (A)$$

$$2x + y + b = 20 \quad (B)$$

$$-9x - 16y + p = 0 \quad (P)$$

$$\frac{23}{10}a + \frac{11}{5}b + p = 182 \quad (**)$$


製品をまったく作らない ( $x = 0, y = 0$ ) とき, (A) 式, (B) 式から,  $a$  は原料 A の制約量に,  $b$  は原料 B の制約量に等しい. (P) 式から  $p = 0$  である.

$$a = 60, \quad b = 20, \quad p = 0$$

この結果と (\*\*) 式を用いると, 次式が得られる.

$$\frac{23}{10} \times 60 + \frac{11}{5} \times 20 = 182$$

つまり, 販売価格の最大値は, 係数  $\frac{23}{10}, \frac{11}{5}$  と, 各原料の制約量から求められるのである. この係数  $\frac{23}{10}, \frac{11}{5}$  のことを 潜在価格 (shadow price) という.



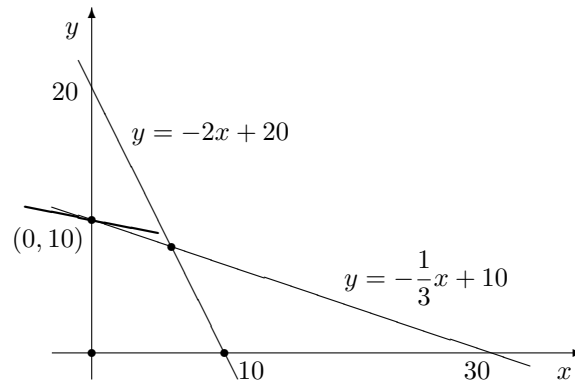
最大値 = 潜在価格 1 × 制約量 1 + 潜在価格 2 × 制約量 2 + …

## 2.5 被約費用 (reduced cost)

### 問題 2.3

問題 2.2 において, 製品  $X, Y$  の販売価格を 6 万円, 20 万円に変更した場合, 販売価格の合計を最大にする  $X, Y$  の製造量を求めよ.

図解による解答  $p = 6x + 20y$  であるから, 点  $(0, 10)$  において  $p$  が最大値をとる.



最適解は  $(x, y) = (0, 10)$ .  $p$  の最大値は  $\max p = 6 \cdot 0 + 20 \cdot 10 = 200$ .

$$2 \cdot 0 + 6 \cdot 10 = 60, \quad 2 \cdot 0 + 10 = 10$$

なので, 原料 B が 10 キログラム残っている.

Simplex 法による解答 次のように変形する. 変形の手順は 1 回だけで終わっているのに注意すること.

$x$	$y$	$a$	$b$	$p$		
2	<b>6</b>	1	0	0	60	(10)
2	<b>1</b>	0	1	0	20	(20)
-6	<b>-20</b>	0	0	1	0	
1/3	1	1/6	0	0	10	
5/3	0	-1/6	1	0	10	
2/3	0	10/3	0	1	200	

最下行の形から, 次のことがわかる.

$$\frac{2}{3}x + \frac{10}{3}a + p = 200 \quad (**)$$

$p$  をできるだけ大きくするためには,  $x$  と  $a$  をできるだけ小さくすればよい. すなわち,  $x = 0, a = 0$  のとき,  $p$  が最大となる.

非負条件から  $x \geq 0, a \geq 0$  である.

$$x = 0, \quad a = 0, \quad \max p = 200$$

さらに変形後の形から,

$$\frac{1}{3}x + y + \frac{1}{6}a = 10, \quad \frac{5}{3}x - \frac{1}{6}a + b = 10$$

この式に  $x = 0, a = 0$  を代入すると,  $y, b$  の値も決定可能である.

$$y = 10, \quad b = 10$$

つまり, 最適解は  $x = 0, y = 10$  (単位) で, 販売価格の最大値は  $\max p = 200$  (万円), 原料  $A$  の余剰はなし, 原料  $B$  の余剰は 10 (キログラム) であることが確認できる.

潜在価格と被約費用 (\*\*\*) 式にあらわれる係数のうち,  $\frac{10}{3}$  が潜在価格 (shadow price) である. この問題においても, “潜在価格と制約量の関係” が成り立つ.

$$\frac{10}{3} \times 60 + 0 \times 20 = 200$$

一方の係数  $\frac{2}{3}$  は被約費用 (reduced cost) とよばれている. 製品  $X$  の製造量を 1 単位増やすと, 合計価格が  $\frac{2}{3}$  万円減ってしまう. 逆にいうと, 製品  $X$  の製造を 1 単位取りやめるごとに,  $\frac{2}{3}$  (万円) だけ損失を減らせることを表している.

## 課題 2.1

製品  $X$  と製品  $Y$  は単位量あたり、原材料  $A$  をそれぞれ 2 キログラム、1 キログラムを要し、機械加工  $B$  をそれぞれ 1 時間、2 時間要する。原材料  $A$  と機械加工  $B$  の月間制約量はそれぞれ 120 キログラム、150 時間である。製品  $X$ 、製品  $Y$  による利益がそれぞれ単位量あたり 2 万円、3 万円であるとき、月間の利益を最大にする製品組合せを求めよ。

Simplex 法を用いる

## 課題 2.2

製品  $X$ 、製品  $Y$ 、製品  $Z$  は単位量あたり、原材料  $A$  をそれぞれ 1 キログラムずつ使い、原材料  $B$  をそれぞれ 2 キログラム、3 キログラム、0 キログラム使う。機械加工  $C$  はそれぞれ 1 時間、2 時間、1 時間を要する。また原材料  $A$ 、原材料  $B$ 、機械加工  $C$  の月間制約量はそれぞれ 10 キログラム、24 キログラム、16 時間である。製品  $X$ 、製品  $Y$ 、製品  $Z$  の単位量あたりの利益が 3 万円、5 万円、2 万円のとき、月間利益を最大にする製品組合せを求めよ。

ソルバーを用いる

目的セル	価格の合計
目標値	最大値
変化させるセル	$X, Y, Z$ の製造量
制約条件	“追加” ボタンを押し、“合計 $\leq$ 制約量” を記入

ソルバー実行後に、“レポート” の “感度” を選択  
感度レポートのラグランジュ乗数を “L 乗数” 欄にコピーする

## 課題 2.3 (非線形モデル)

ある工場の製造部門  $X$  と  $Y$  は異なる生産様式で同一売価の製品を、それぞれ  $x$  単位、 $y$  単位生産している。ただし  $Y$  製造部門では生産過程で単位あたりの売価が 4 千円である副産物  $Z$  が  $z$  単位、以下の割合で発生する。

$$z = 3y$$

したがって  $Y$  製造部門の製造原価からこれが控除されるものとする。また各部門の製造原価  $C_x, C_y$  は次のようである。(単位:千円)

$$C_x = 2x^{1.5} + 9x, \quad C_y = 0.3y^2 - 4z$$

$X, Y$  両部門とも、1 単位の製品を製造するのに同一の原料 1 トンを要するが、今月の原料調達量は 101 トンが限界であるという。

$$x + y \leq 101$$

- 当該製品を 101 単位製造するとき、総製造原価  $C = C_x + C_y$  を最小にするような製品組合せを求めよ。  
(解:  $x = 36.0, y = 65.0, C = 1243.5$ )
- 当該製品の単価が  $p = 26.1$  のとき、限界利益  $R$  を最大にする製品組合せを求めよ。  
(解:  $x = 32.5, y = 63.5, R = 1394.9$ )

### 3 回帰分析 (8-10 講)

#### 3.1 平均値・分散・標準偏差

多数の数値からなる集団をあつかうとき、集団全体のもつ傾向を 1 つの数値で代表させると便利なが多い。集団全体の代表させる数値を代表値という。代表値には次のようなものがある。(括弧内は Excel 関数の名称)

- 平均値 (AVERAGE)

$N$  個のデータ  $x_1, x_2, x_3, \dots, x_N$  があるとき、これらの合計を  $N$  で割った値を平均値といい、 $\bar{x}$  で表す。

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

- 分散 (VARP など)

データ  $x_i$  の平均値  $\bar{x}$  からの隔たり  $x_i - \bar{x}$  を偏差という。分散は、偏差の 2 乗の平均値のことで、集団全体の散らばりの大きさを表している。分散を  $V(x)$  と書く。

分散は負の値にならない。

体重のリストがあるとき、分散の単位は (kg<sup>2</sup>)

$$V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

- 標準偏差 (STDEVP など)

標準偏差は分散  $V(x)$  の平方根である。 $\sigma(x)$  で表す。

標準偏差の単位は (kg)

$$\sigma(x) = \sqrt{V(x)}$$

- 最大値 (MAX)

- 最小値 (MIN)

- 範囲 (MAX()-MIN()) で求める)

- 最頻値 (MODE)

数値が分布する範囲をいくつかの階級に分ける。もっとも多くのデータが属する階級の階級値を最頻値という。

- 中央値 (MEDIAN)

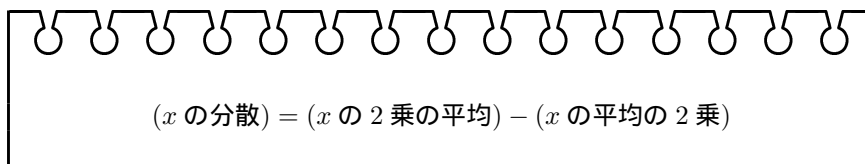
$N$  個のデータを大小の順にならべたときの中央の順位、 $(N + 1)/2$  位にあたる数値を中央値という。

$N$  が偶数のとき、 $N/2$  位と  $(N + 2)/2$  位の値を平均する。

分散の公式 分散値に関して、次の公式がよく知られている。

$$V(x) = \overline{x^2} - (\bar{x})^2$$

次のようにいいかえてもよい。



この公式を簡単に証明しよう。

$$\bar{x} = \frac{\sum x_i}{N}$$

なので,

$$\begin{aligned} V(x) &= \frac{\sum (x_i - \bar{x})^2}{N} \\ &= \frac{\sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2)}{N} \\ &= \frac{\sum x_i^2}{N} - 2\bar{x} \frac{\sum x_i}{N} + \bar{x}^2 \\ &= \frac{\sum x_i^2}{N} - 2\bar{x}\bar{x} + \bar{x}^2 \\ &= \overline{x^2} - \bar{x}^2 \end{aligned}$$

標準化とは  $x_i (i = 1, 2, \dots, N)$  の平均値, 標準偏差を  $\bar{x}, \sigma(x)$  とする.  $x$  に対して  $z$  を次のように定めよう.  $z$  を,  $x$  の標準化 (あるいは規準化) という.

$$z_i = \frac{x_i - \bar{x}}{\sigma(x)} \quad (x \text{ の標準化})$$

標準化は, 確率分布 (11-14 講) においてふたたび利用する.

標準化すると, 平均値は必ず 0 となり, 標準偏差は 1 となる.

$$\bar{z} = 0, \quad \sigma(z) = 1$$

いわゆる偏差値とは, このように標準化した値の 10 倍に 50 を加えたものをいう. 点数  $x_i$  に対する偏差値  $t_i$  は次式で求める.

$$\begin{aligned} t_i &= 50 + 10 z_i \\ t_i &= 50 + 10 \times \frac{x_i - \bar{x}}{\sigma(x)} \end{aligned}$$

偏差値の平均値は常に 50, 標準偏差は常に 10 である.

$$\bar{t} = 50, \quad \sigma(t) = 10$$

### 3.2 共分散・相関係数

同じ個数 ( $N$  個) からなる 2 種類のデータ  $x, y$  があるとき,  $x, y$  の共分散 (covariance) を次のように定める.

$$\begin{aligned} \text{Cov}(x, y) &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_N - \bar{x})(y_N - \bar{y})}{N} \end{aligned}$$

$x$  と  $x$  自身の共分散は,  $x$  の分散に等しい.  
 $\text{Cov}(x, x) = V(x)$

共分散  $\text{Cov}(x, y)$  においても, 分散と同様の関係式が成り立つ.

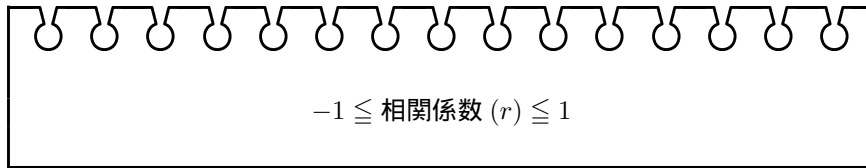
$$\text{Cov}(x, y) = \overline{xy} - \bar{x} \cdot \bar{y}$$

( $x, y$  の共分散) = ( $x \times y$  の平均) - ( $x$  の平均)  $\times$  ( $y$  の平均)

$x, y$  の共分散の値をそれぞれの標準偏差で割ったものを相関係数 (あるいは Pearson の積率相関係数) という。相関係数を  $r$  で表す。

$$\begin{aligned} r &= \frac{\text{Cov}(x, y)}{\sigma(x) \times \sigma(y)} \\ &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \end{aligned}$$

$x_i$  が大きい値をもつとき、それに対応する  $y_i$  も大きいとき、 $x$  と  $y$  は正の相関があるという。反対に  $y_i$  の値が小さくなる時は負の相関があるという。Cauchy-Schwarz の不等式  $|\sum pq| \leq \sqrt{\sum p^2} \sqrt{\sum q^2}$  より、相関係数  $r$  の絶対値は 1 以下である。



### 3.3 回帰分析

$N$  人分の身長と体重のデータがあるとする。身長を  $x_i$ 、体重を  $y_i$  とおくと、 $x_i$  と  $y_i$  には強い相関がある。つまり身長が高い人ほど体重が重いと考えられるので、ある人の身長がわかれば、体重もある程度予測できそうである。

身長が  $x_i$  である人の実際の体重が  $y_i$  である。いま、その人の体重の予測値を  $\hat{y}_i$  とおき、次のような形で表してみよう。

$$\hat{y}_i = a + bx_i$$

この式を回帰式という。この式は直線をあらわすので、回帰直線ともいう。回帰式で表現された  $\hat{y}$  を回帰値という。回帰分析の目的は、回帰値がもっともよい予測値を与えるような定数  $a, b$  を見つけることである。切片  $a$  を回帰定数、傾き  $b$  を回帰係数という。

回帰定数と回帰係数をあわせて単に回帰係数ということもある。

2 変数の平方完成について 実際には回帰定数、回帰係数を導く前に、2 つの変数からなる多項式の最小値を求める方法を確認しておきたい。下のようにして  $E$  の最小値を求めた。変形方法を確認してほしい。

$$\begin{aligned} E &= 2a^2 - 8ab + 11b^2 + 4a - 14b + 9 \\ &= 2(a^2 - 4ab + 2a) + 11b^2 - 14b + 9 \\ &= 2(a - (2b - 1))^2 - 2(2b - 1)^2 + 11b^2 - 14b + 9 \\ &= 2(a - 2b + 1)^2 + 3b^2 - 6b + 7 \\ &= 2(a - 2b + 1)^2 + 3(b - 1)^2 + 4 \end{aligned}$$

実数の平方は常に 0 以上であるから、 $E$  の最小値は 4 である。

$$\min E = 4$$

$E$  が最小となるのは

$$a - 2b + 1 = 0, \quad b - 1 = 0$$

のとき、すなわち  $a = b = 1$  のときである。

回帰係数の求め方 (1) 回帰値  $\hat{y}_i$  からの  $y_i$  のへだたりを  $e_i$  とおく.  $e_i$  を残差あるいは誤差という.

$$e_i = y_i - \hat{y}_i$$

最小 2 乗法では, 残差の平方和が最小となるように各係数を選ぶ.


回帰直線  $y = a + bx$  は残差の総和が最小となるように選ばなくてはならない. それには最小 2 乗法を用いる. 残差の平方和を  $E$  とおき, これが最小となるような  $a, b$  を求めたい.

$$\begin{aligned} E &= \sum_{i=1}^N (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - a - bx_i)^2 \\ &= Na^2 + 2ab \sum x_i + b^2 \sum x_i^2 - 2a \sum y_i - 2b \sum x_i y_i + \sum y_i^2 \\ &= N \left( a + \frac{b \sum x_i}{N} - \frac{\sum y_i}{N} \right)^2 \\ &\quad + N \left( \frac{\sum x_i^2}{N} - \frac{(\sum x_i)^2}{N^2} \right) \left( b - \frac{\frac{\sum x_i y_i}{N} - \frac{\sum x_i \sum y_i}{N^2}}{\frac{\sum x_i^2}{N} - \frac{(\sum x_i)^2}{N^2}} \right)^2 \\ &\quad + \text{Const.} \\ &= N (a + b\bar{x} - \bar{y})^2 + N V(x) \left( b - \frac{\text{Cov}(x, y)}{V(x)} \right)^2 + \text{Const.} \end{aligned}$$

ここで Const. は定数である. 実数の平方は常に非負の数となるから,

$$a + b\bar{x} - \bar{y} = 0, \quad b - \frac{\text{Cov}(x, y)}{V(x)} = 0$$

のとき, 残差の平方和は最小となる.  $E$  の最小値は Const. に等しい. 以上のことから, 回帰直線を求めるには次のようにすればよい.



$$\hat{y} = a + bx, \quad b = \frac{\text{Cov}(x, y)}{V(x)}, \quad a = \bar{y} - b\bar{x}$$

回帰係数の求め方 (2) 残差平方和  $E$  が最小となる条件は、偏微分して求めることもできる。

この講義では偏微分の知識は必要でない。

$$E = \sum (y_i - a - bx_i)^2$$

$E$  を  $a$  で偏微分する。

$$\frac{\partial E}{\partial a} = -2 \sum (y_i - a - bx_i)$$

$E$  が最小になるとき、偏微分の値は 0 になるから、上式を  $-2N$  で割ると、

$$\frac{\sum y_i}{N} - a - b \frac{\sum x_i}{N} = 0$$

したがって次式を得る。

$$a = \bar{y} - b\bar{x} \quad (*)$$

次に  $E$  を  $b$  で偏微分する。

$$\frac{\partial E}{\partial b} = -2 \sum x_i (y_i - a - bx_i)$$

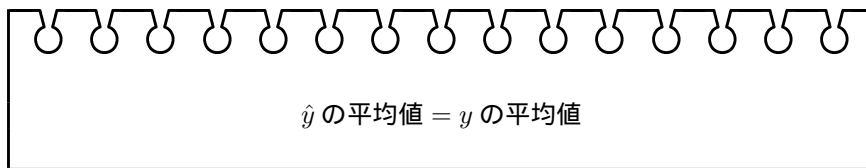
$E$  が最小のとき偏微分の値が 0 である。両辺を  $-2N$  で割ると、

$$\begin{aligned} 0 &= \frac{\sum x_i y_i}{N} - a \frac{\sum x_i}{N} - b \frac{\sum x_i^2}{N} \\ &= \frac{\sum x_i y_i}{N} - (\bar{y} - b\bar{x})\bar{x} - b \frac{\sum x_i^2}{N} \\ &= \left( \frac{\sum x_i y_i}{N} - \bar{x}\bar{y} \right) - b \left( \frac{\sum x_i^2}{N} - \bar{x}^2 \right) \\ &= \text{Cov}(x, y) - bV(x) \end{aligned}$$

これで回帰定数  $a$  と回帰係数  $b$  が求められた。

$$a = \bar{y} - b\bar{x}, \quad b = \frac{\text{Cov}(x, y)}{V(x)}$$

回帰値の平均 回帰値  $\hat{y}$  の平均値について、次のことがいえる。



なぜなら、

$$\overline{\hat{y}} = \overline{a + bx} = a + b\bar{x} = \bar{y}$$

となるからである。

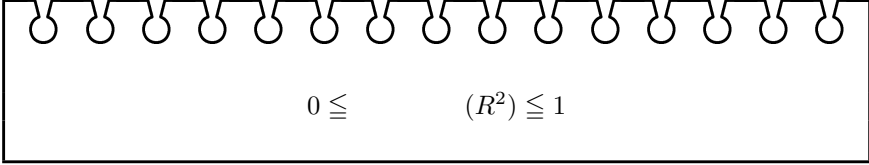
$\hat{y}$  は  $y$  の近似をあらわすので、同じ平均値をもつのは自然な結論といえる。

### 3.4 決定係数

回帰値  $\hat{y}_i$  の分散  $V(\hat{y})$  を, 実測値  $y_i$  の分散  $V(y)$  で割ったものを決定係数という. 決定係数を  $R^2$  という記号で表す.

$$R^2 = \frac{V(\hat{y})}{V(y)}$$

決定係数は“当てはめの精度”を表し, 回帰値が実測値のよい近似になっているほど  $R^2$  は 1 に近い値となる.



$0 \leq \text{決定係数 } (R^2) \leq 1$

単回帰において, 決定係数は相関係数の 2 乗に等しいことが知られている. なぜなら,

目的変数  $y$  が, 1 個の説明変数  $x$  であらわされているとき単回帰という.

説明変数が複数あるときは重回帰という.

$$\begin{aligned} (\text{決定係数 } R^2) &= \frac{1}{V(y)} \times \frac{\sum(\hat{y} - \bar{\hat{y}})^2}{N} \\ &= \frac{1}{V(y)} \times \frac{\sum(a + bx_i - a - b\bar{x})^2}{N} \\ &= \frac{1}{V(y)} \times b^2 \times \frac{\sum(x_i - \bar{x})^2}{N} \\ &= \frac{1}{V(y)} \times \frac{\text{Cov}(x, y)^2}{V(x)^2} \times V(x) \\ &= \frac{\text{Cov}(x, y)^2}{V(x)V(y)} \\ &= (\text{相関係数 } r)^2 \end{aligned}$$

となるからである. なお, この関係は重回帰の場合には利用できない.

### 3.5 重回帰分析

3 つの変数  $x_i, y_i, z_i$  について,  $z_i$  を  $x_i$  および  $y_i$  の式で近似してみよう. 1 個の目的変数を 2 個以上の説明変数によって回帰する場合, 特に重回帰という.  $z_i$  の回帰値を  $\hat{z}_i$  し, 次のような回帰式を作る.

$$\hat{z}_i = a + bx_i + cy_i$$

この回帰式は直線ではなく平面を表している. 単回帰と同様にして回帰定数  $a$  と偏回帰係数  $b, c$  を求めよう. 単回帰と同様に残差の平方和を  $E$  とおき, 最小 2 乗法を用いる.

単回帰の回帰係数に対して, 重回帰では偏回帰係数という名前になる.

$$\begin{aligned} E &= \sum(z_i - \hat{z}_i)^2 \\ &= \sum(z_i - a - bx_i - cy_i)^2 \end{aligned}$$

$a$  で偏微分すると,

$$\begin{aligned} \frac{\partial E}{\partial a} &= -2 \sum(z_i - a - bx_i - cy_i) \\ &= -2N(\bar{z} - a - b\bar{x} - c\bar{y}) \end{aligned}$$

$b$  で偏微分すると,

$$\begin{aligned}\frac{\partial E}{\partial b} &= -2 \sum x_i (z_i - a - bx_i - cy_i) \\ &= -2N (\bar{xz} - a\bar{x} - b\bar{x}^2 - c\bar{xy})\end{aligned}$$

$c$  で偏微分すると,

$$\begin{aligned}\frac{\partial E}{\partial c} &= -2 \sum y_i (z_i - a - bx_i - cy_i) \\ &= -2N (\bar{yz} - a\bar{y} - b\bar{xy} - c\bar{y}^2)\end{aligned}$$

$E$  が最小となるとき, それぞれの偏微分係数は 0 となるから,  $a, b, c$  は次の 3 式を満足する必要がある.

$$\begin{cases} \bar{z} - a - b\bar{x} - c\bar{y} = 0 \\ \bar{xz} - a\bar{x} - b\bar{x}^2 - c\bar{xy} = 0 \\ \bar{yz} - a\bar{y} - b\bar{xy} - c\bar{y}^2 = 0 \end{cases}$$

これを  $\bar{z}, \bar{xz}, \bar{yz}$  の式にして, 行列に置き換える.

$$\begin{bmatrix} \bar{z} \\ \bar{xz} \\ \bar{yz} \end{bmatrix} = \begin{bmatrix} 1 & \bar{x} & \bar{y} \\ \bar{x} & \bar{x}^2 & \bar{xy} \\ \bar{y} & \bar{yx} & \bar{y}^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

中央の行列の逆行列を左からかけると, 回帰定数  $a$  と偏回帰係数  $b, c$  が次のように得られる.

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & \bar{x} & \bar{y} \\ \bar{x} & \bar{x}^2 & \bar{xy} \\ \bar{y} & \bar{yx} & \bar{y}^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{z} \\ \bar{xz} \\ \bar{yz} \end{bmatrix}$$

重回帰係数を求める公式 目的変数  $y$  と, それを回帰する  $m$  個の説明変数  $x_j$  ( $j = 1, 2, \dots, m$ ) があり, 重回帰式が次のように与えられている.

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m$$

このとき, 重回帰係数  $a_j$  ( $j = 0, 1, 2, \dots, m$ ) は, 次の公式により求められる.

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} 1 & \bar{x}_1 & \cdots & \bar{x}_m \\ \bar{x}_1 & \bar{x}_1^2 & \cdots & \bar{x}_1\bar{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_m & \bar{x}_m\bar{x}_1 & \cdots & \bar{x}_m^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{y} \\ \bar{x}_1\bar{y} \\ \vdots \\ \bar{x}_m\bar{y} \end{bmatrix}$$

### 3.6 Excel の行列関数

Excel を利用して行列の積を求めるには MMULT() 関数, 逆行列を求めるには MINVERSE() 関数を利用する. 次の手順で実行する.

- 関数値が代入されるセル範囲を選択状態にする
- 関数名と引数を入力する
- Ctrl キーと Shift キーを押しながら, Enter キーを押す

**課題 3.1**

試験成績の資料が 25 件ある。成績の平均値・分散・標準偏差等の代表値を求めよ。さらに成績を標準化 (規準化) した値, 成績の偏差値について, 各種代表値を求めよ。

標準化は  $z = (x - \bar{x})/\sigma(x)$ , 偏差値は  $t = 10z + 50$ 。

**課題 3.2 (単回帰)**

入試成績  $X$  および取得単位  $Y$  の資料が 25 件ある。取得単位  $Y$  を入試成績  $X$  により回帰する回帰式を求め, この回帰における決定係数を求めよ。

$b = \text{Cov}(x, y)/V(x)$ ,  $a = \bar{y} - b\bar{x}$  を利用する。

**課題 3.3 (重回帰)**

15 件の資料  $X, Y, Z$  がある。 $Z$  を  $X$  および  $Y$  により回帰する重回帰式を求め, この回帰における決定係数を求めよ。

Excel 配列関数  $\text{MINVERSE}()$ ,  $\text{MMULT}()$  を利用する。

## 4 確率分布 (11-13 講)

### 4.1 度数分布表と確率分布

次の表は 100 件の資料を 7 つの階級にわけ、それぞれの頻度を調べた度数分布表/相対度数分布表である。

階級	階級値 ( $x_k$ )	度数 ( $f_k$ )	相対度数 ( $f_k/N$ )
39.5 ~ 42.5	41	7	0.07
42.5 ~ 45.5	44	12	0.12
45.5 ~ 48.5	47	18	0.18
48.5 ~ 51.5	50	24	0.24
51.5 ~ 54.5	53	21	0.21
54.5 ~ 57.5	56	14	0.14
57.5 ~ 60.5	59	4	0.04
計	—	100 ( $N$ )	1

階級の幅のちょうど中央にあたる値を階級値という。

各階級の度数  $f_k$  を資料の大きさ  $N$  (ここでは  $N = 100$  件) で割ったものを相対度数という。度数分布表を  $n$  個 (ここでは  $n = 7$  個) の階級に分けたとすると、平均  $\bar{x}$ 、分散  $V(x)$ 、標準偏差  $\sigma(x)$  は次式のようになる。

$$\begin{aligned}\bar{x} &= \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{N} \\ V(x) &= \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \cdots + (x_n - \bar{x})^2 f_n}{N} \\ &= \frac{x_1^2 f_1 + x_2^2 f_2 + \cdots + x_n^2 f_n}{N} - \bar{x}^2 \\ \sigma(x) &= \sqrt{V(x)}\end{aligned}$$

41 が 7 件, 44 が 12 件ある ... とみなして計算する。

$$V(x) = \overline{x^2} - (\bar{x})^2$$

度数ではなく相対度数を使って計算することもできる。相対度数を  $r_k = f_k/N$  と書くと、平均と分散は次式のようになる。

$$\begin{aligned}\bar{x} &= x_1 r_1 + x_2 r_2 + \cdots + x_n r_n \\ V(x) &= (x_1 - \bar{x})^2 r_1 + (x_2 - \bar{x})^2 r_2 + \cdots + (x_n - \bar{x})^2 r_n \\ &= (x_1^2 r_1 + x_2^2 r_2 + \cdots + x_n^2 r_n) - \bar{x}^2\end{aligned}$$

この相対度数分布表から平均値と分散を求めると、

$$\begin{aligned}\bar{x} &= 41 \times 0.07 + 44 \times 0.12 + \cdots + 59 \times 0.04 \\ &= 49.94 \\ V(x) &= 41^2 \times 0.07 + 44^2 \times 0.12 + \cdots + 59^2 \times 0.04 - 49.94^2 \\ &= 21.7764\end{aligned}$$

**確率分布** ある試行において実現値  $x_k$  が確率  $p_k$  で起こるものとする。起こりうるすべての場合について実現値と確率が与えられているとき、この実現値と確率の組を確率分布という。

実現値	$x_1$	$x_2$	...	$x_n$	計
確率	$p_1$	$p_2$	...	$p_n$	1

先の相対度数分布表において、階級値 41 をとる数値の割合は全体の 0.07 である。いいかえると、値が 41 となる確率が 0.07 であると考えてもよい。与えられた確率分布の平均値等は相対度数分布の場合と同様にして求める。

$$\begin{aligned}\bar{x} &= x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \\ V(x) &= (x_1 - \bar{x})^2 p_1 + (x_2 - \bar{x})^2 p_2 + \cdots + (x_n - \bar{x})^2 p_n \\ &= (x_1^2 p_1 + x_2^2 p_2 + \cdots + x_n^2 p_n) - \bar{x}^2 \\ \sigma(x) &= \sqrt{V(x)}\end{aligned}$$

## 4.2 二項分布

$n$  個のものから  $k$  個を選び出す組合せの数を  ${}_n C_k$  と書き, これを二項係数という. 二項係数は次のようにして求める.

$$\begin{aligned}{}_n C_k &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 2 \cdot 1} \\ &= \frac{n!}{k!(n-k)!}\end{aligned}$$

ここで  $k!$  は  $k$  の階乗, すなわち  $k! = k(k-1)\cdots 2 \cdot 1$  を表している. また  $0! = 1$  と定める. 独立試行において, 事象  $A$  が起こる ( $A$  が成功する) 確率を  $p$  とし,  $A$  の余事象  $\bar{A}$  ( $A$  が失敗する) の確率を  $q$  とする.

記号  $P(A)$  は, 事象  $A$  が起こる確率を表す.

$$P(A) = p$$

$$P(\bar{A}) = q \quad (= 1 - p)$$

$n$  回の試行を繰り返して事象  $A$  が  $k$  回起こったとする.  $n$  回のうち, 成功が  $k$  回, 失敗が  $n - k$  回であるから, この事象の確率は次式で表すことができる.

$${}_n C_k p^k q^{n-k}$$

試行回数  $n$ , 成功確率  $p$  の二項分布を  $B(n, p)$  と表す.

この確率分布を二項分布という.

$x_k :$	0	1	...	$k$	...	$n$	計
$p_k :$	${}_n C_0 q^n$	${}_n C_1 p q^{n-1}$	...	${}_n C_k p^k q^{n-k}$	...	${}_n C_n p^n$	1

“二項定理”を思い出していただきたい.

$$(p + q)^n = {}_n C_0 q^n + {}_n C_1 p q^{n-1} + {}_n C_2 p^2 q^{n-2} + \cdots + {}_n C_n p^n$$

二項定理を用いると,  $p + q = 1$  であるから,

$${}_n C_0 q^n + {}_n C_1 p q^{n-1} + \cdots + {}_n C_k p^k q^{n-k} + \cdots + {}_n C_n p^n = 1$$

確率  $p_k$  ( $k = 0, 1, \dots, n$ ) の総和が 1 となっていることが確認できる.

二項分布の平均値と分散を求めよう。平均値  $\bar{x}$  は、

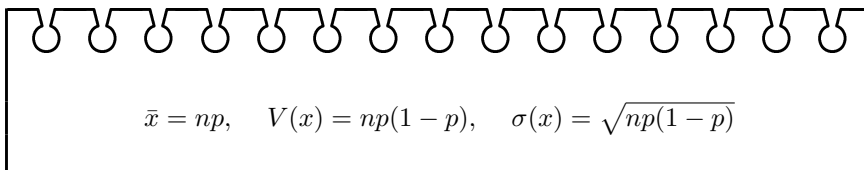
$$\begin{aligned}
 \bar{x} &= \sum_{k=0}^n x_k p_k \\
 &= \sum_{k=0}^n k {}_n C_k p^k q^{n-k} \\
 &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
 &= \sum_{k=0}^{n-1} (k+1) \frac{n!}{(k+1)!(n-k-1)!} p^{k+1} q^{n-k-1} \\
 &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1} \\
 &= np
 \end{aligned}$$

分散  $V(x)$  は、

$$\begin{aligned}
 V(x) &= \sum_{k=0}^n x_k^2 p_k - \bar{x}^2 \\
 &= \sum_{k=0}^n k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k} - (np)^2 \\
 &= np \sum_{k=0}^{n-1} (k+1) \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1} - n^2 p^2 \\
 &= np \sum_{k=1}^{n-1} k \frac{(n-1)!}{k!(n-k-1)!} p^k q^{n-k-1} + np - n^2 p^2 \\
 &= np(n-1)p \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-k-2)!} p^k q^{n-k-2} - np - n^2 p^2 \\
 &= n(n-1)p^2 - np - n^2 p^2 \\
 &= np(1-p)
 \end{aligned}$$

二項分布  $B(n, p)$  の平均値, 分散, 標準偏差は次式で求められる.

結果だけ確認しておけばよい.



$$\bar{x} = np, \quad V(x) = np(1-p), \quad \sigma(x) = \sqrt{np(1-p)}$$

COMBIN() 関数 二項係数  ${}_n C_s$  は、 $n$  個のものから  $s$  個のものを選び出す組合せの数に等しく、次式のように表せる。

$${}_n C_s = \frac{n!}{s!(n-s)!}$$

Excel では二項係数を求める関数 COMBIN が用意されている。

$${}_n C_s = \text{COMBIN}(n, s)$$

BINOMDIST() 関数 試行回数  $n$ 、成功確率  $p$  の二項分布において、成功回数  $s$  の場合の確率は次のように表される。

$${}_n C_s p^s (1-p)^{n-s}$$

Excel 関数 BINOMDIST ではこれを簡単に計算することができる。

$${}_n C_s p^s (1-p)^{n-s} = \text{BINOMDIST}(s, n, p, \text{FALSE})$$

さらに、成功回数が  $s$  回以下となる確率の累計も求めることができ、

$$\sum_{k=0}^s {}_n C_k p^k (1-p)^{n-k} = \text{BINOMDIST}(s, n, p, \text{TRUE})$$

とする。

#### 問題 4.1

ある大学では、1 人の学生が 4 年間で無事に卒業する確率は  $p = 0.9$  であるという。100 人の入学者のうち、95 人以上が 4 年間で卒業する確率を求めよ。

100 人のうち 95 人が 4 年間で卒業する確率は

$$\begin{aligned} {}_{100} C_{95} 0.9^{95} (1-0.9)^5 &= 75287520 \times 0.9^{95} \times 0.1^5 \\ &= 0.033865804 \end{aligned}$$

と書ける。95 人以上が 4 年間で卒業する確率は、同様の計算を 95 ~ 100 について行ったものを合計すればよい。求める確率を  $P$  とおくと、

$$\begin{aligned} P &= {}_{100} C_{95} 0.9^{95} 0.1^5 + {}_{100} C_{96} 0.9^{96} 0.1^4 + {}_{100} C_{97} 0.9^{97} 0.1^3 \\ &\quad + {}_{100} C_{98} 0.9^{98} 0.1^2 + {}_{100} C_{99} 0.9^{99} 0.1^1 + {}_{100} C_{100} 0.9^{100} \\ &= 75287520 \cdot 0.9^{95} 0.1^5 + 3921225 \cdot 0.9^{96} 0.1^4 + 161700 \cdot 0.9^{97} 0.1^3 \\ &\quad + 4950 \cdot 0.9^{98} 0.1^2 + 100 \cdot 0.9^{99} 0.1^1 + 1 \cdot 0.9^{100} \\ &= 0.057576886 \end{aligned}$$

### 4.3 標準正規分布

正規分布は後に述べるように非常に重要な確率分布である。平均値  $m$ 、標準偏差  $\sigma$  の正規分布は、確率密度関数  $f(x)$  をもち、

平均  $m$ 、標準偏差  $\sigma$  の正規分布を  $N(m, \sigma^2)$  と表す。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}$$

$X$  が  $a$  以上  $b$  以下となる確率は、この密度関数の定積分で表される。

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-(x-m)^2/2\sigma^2} dx$$

正規分布では、数表を参照して確率を求めるのが通例である。下の表には、平均 0、標準偏差 1 の標準正規分布について、0 以上  $x$  以下の値をとる確率が示されている。平均値や標準偏差の値が異なる場合は、標準化した値をもとに数表を参照する。

平均 0、標準偏差 1 の標準正規分布は  $N(0, 1)$  である。

$x$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

### 4.4 二項分布の正規近似

互いに独立で同一の分布を持つ確率変数の列  $X_1, X_2, \dots, X_n$  があり、それらの平均値を  $m$ 、標準偏差を  $\sigma$  とする。確率変数列  $X_i$  の和の平均値は  $nm$ 、標準偏差は  $\sqrt{n}\sigma$  となることが知られている。この和を標準化したものを  $X$  とおく。

$$X = \frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n}\sigma}$$

$X$  に対して、次の定理が成り立つ。

## 定理 4.1 (中心極限定理)

互いに独立で同一の分布を持つ確率変数の列  $X_1, X_2, \dots, X_n$  について, それぞれの平均値は  $m$ , 標準偏差は  $\sigma$  であるとする. このとき次式が成り立つ.

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + X_2 + \dots + X_n - nm}{\sqrt{n}\sigma} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$$

つまり標準化した確率変数の和の分布は, 標準正規分布  $N(0, 1)$  に収束する.

成功確率  $p$  の試行を  $n$  回の繰り返したときの二項分布  $B(n, p)$  は, 単独の試行における成功回数の和で表せる. 中心極限定理が適用できる場合に於けるので, 試行回数  $n$  の大きい二項分布は正規分布で近似できる.

## 問題 4.2

ある大学では, 1 人の学生が 4 年間で無事に卒業する確率は  $p = 0.9$  であるという. 100 人の入学者のうち, 95 人以上が 4 年間で卒業する確率を求めよ.

問題 4.1 と同一の問題である. 問題 4.1 ではすでに確率の正確な値を求めたが, 今回は“正規近似”を利用して求めてみよう.

$$n = 100, \quad p = 0.9$$

のとき, 二項分布の平均値は  $np$ , 分散は  $np(1-p)$  で求められる.

$$np = 100 \times 0.9 = 90$$

$$np(1-p) = 100 \times 0.9 \times (1-0.9) = 9$$

平均 90, 標準偏差 3 であるから,  $x = 95$  を標準化すると,

$$\frac{95 - 90}{3} = 1.67$$

標準正規分布の数表を参照すると, 卒業生数が 90 ~ 95 人となる確率は“およそ”0.4525 であることがわかる. 卒業生数が 90 人 (平均値) 以上である確率は 0.5 なので, 95 人以上である確率は,

$$0.5 - 0.4525 = 0.0475$$

として求められる.

## 4.5 乱数

コンピュータ上で擬似的に乱数を発生させることがよく行われる. 乱数関数では通常, 0 以上 1 未満の数を一様な確率で発生させるようになっているが, このような確率分布は一様分布とよばれている.  $a \leq x < b$  の範囲の値をとる一様分布があるとき, 平均値は,

$a \leq x < b$  の値を取る一様分布を  $U(a, b)$  で表す.

この講義では微積分の知識を仮定しない.

$$\begin{aligned} \bar{x} &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b \\ &= \frac{a+b}{2} \end{aligned}$$

分散は、

$$\begin{aligned} V(x) &= \int_a^b x^2 \frac{1}{b-a} dx - \left( \frac{a+b}{2} \right)^2 \\ &= \frac{1}{b-a} \left[ \frac{x^3}{3} \right]_a^b - \left( \frac{a+b}{2} \right)^2 \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

**RAND() 関数** Excel で一様疑似乱数を発生させるには RAND() 関数を利用するのが簡便である。RAND() は 0 以上 1 未満の範囲の適当な乱数を返す。表計算シートの内容を書き換えるか、F9 キーを押すたびに再計算され、値が変化する。

**分析ツール** メニューバーから、“ツール” → “分析ツール..” → “乱数発生”を選択する。

(もしメニューに“分析ツール”が表示されないときは、“ツール” → “アドイン..”で、“分析ツール”にチェックを入れておく)

0 以上 1 未満の乱数を 40 個発生させるとき、分析ツールの設定画面で次のように指定する。

変数の数	1
乱数の数	40
分布	“均一”
パラメータ	0 から 1 まで
出力オプション	“出力先”に乱数を書き込むセル位置を指定

分析ツールを用いると、様々な分布をもつ乱数を発生させることができる。

## 4.6 在庫管理

### 問題 4.3

ある商品の 7 日間あたりの平均需要量は 1500、標準偏差は 50 であった。商品は、在庫量が一定量 (発注点という) まで減少したときに発注される。また調達期間は 7 日間であるとする。

- 95% の確率で在庫切れを起こさない (危険率 5%) ようにするためには発注量はいくらにすればよいか。また発注点はいくらにすればよいか。
- 危険率を 1% としたときの発注量および発注点を求めよ。

次表の危険率と安全係数を利用して求めよ。この値は正規分布表から得られるものである。

危険率 (%)	安全係数
1.00	2.33
5.00	1.64

## 課題 4.1

ある大学の学生が 4 年間で無事卒業する確率は  $p = 0.9$  であるという. ある年度の入学生 400 人のうち, 351 人以上が 4 年間で卒業できる確率を求めよ.

## 課題 4.2

さいころを 72 回投げるとき, 2 または 3 の目が 20 回以上 28 回以下出る確率を求めよ.

## 課題 4.3

さいころを 450 回投げるとき, 2 または 3 の目が 142 回以上 163 回以下出る確率を求めよ.

正規近似を行い, 正規分布の数表から確率を求める

## 5 定期試験とレポートの概要

定期試験は通常教室において筆記試験を行う。コンピューターは使えないので注意。持ち込み許可物件は電卓のみ(計算機能以外を持つものは不可)とする。試験日は以下のとおり。

- Sクラス: 1月23日(火) 第4時限(15:00~16:00)
- Eクラス: 1月24日(水) 第4時限(15:00~16:00)

試験日程等は各学部事務室で確認のこと。