

10. 統計計算

- 1) A 列に出席番号 1～50 を付け、その成績を次表から B 列に入力する。[A1], [B1] に見出しとして”番号”, ”成績”と入力する。データ範囲は[B2..B51] となる。

	1	2	3	4	5	6	7	8	9	10
0	61	88	94	96	53	50	87	69	95	71
1	95	69	52	94	99	57	94	74	64	51
2	80	57	75	72	52	77	81	85	58	65
3	88	55	74	98	82	80	88	91	92	50
4	62	72	99	93	56	91	80	60	51	67

C 列に評価として、点数が 80 ～100, 70 ～79, 60～69, 0 ～59 に対してA, B, C, F (F は不可) の成績がつくようにすること。

B 列のデータの平均値、最大値・最小値、中央値、最頻値（モード）を求めること。

B 列のデータを 10 点刻みで各階層にいる人数をカウントし、ヒストグラムを作成すること。

B 列のデータを (A 列と C 列も含めて) 降順に並べ替えること。

B 列のデータの分散と標準偏差を求めよ。

[B2] 番地の値を判定するには =if(B2>=80,"A",if(B2>=70,"B",if(B2>=60,"C","F"))) である。

平均値は =average(範囲), 最大値は =max(範囲), 中央値は median, 最頻値は mode である。

A, B, C, F の頻度を集計する方法。

範囲	頻度
59	
69	
79	
100	

頻度を求めるには frequency を利用する。集計結果を表示したい範囲を、例えば見出しを含む左表 [D1..E5] に用意する。「D2..D5」には区間の最大数を用意しておく。ここで結果を入れる範囲 E2..E5 をマウスで範囲指定し、[E2] に =frequency(データ範囲,D2..D5) を入力し、Ctrl+Shift+Enter を行う。最後に、[D2..E5]を指定して、グラフウィザードを使って ヒストグラムを描く。

データ $x_1, x_2, \dots, x_n$ の平均を $mean_x$ とおくと、分散 var_x と標準偏差 s_x は

$$var_x = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - mean_x^2 \quad s_x = \sqrt{var_x}$$

で与えられる。Excelでは 標準偏差は =stdevp(範囲) とする。

Mathematica では <<Statistics`Master` のように統計パッケージを読み込み、データのリストを L とおき Mean[L], StandardDeviationMLE[L] とする。

答: 分散 258.56 標準偏差 16.09

- 2) n 個の数データの組 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ を入力し、次の式により共分散 s_{xy} と相関係数 C_{xy} を求める。

$mean_x = x_i$ の平均 $s_x = x_i$ の標準偏差

$mean_y = y_i$ の平均 $s_y = y_i$ の標準偏差

$$s_{xy} = (x_i - \text{mean}_x)(y_i - \text{mean}_y) \text{ の平均} \quad C_{xy} = \frac{s_{xy}}{s_x s_y}$$

データ:

	1	2	3	4	5	6	7	8	9	10
x	97	69	94	84	54	69	58	57	75	61
y	85	68	74	71	58	66	61	72	78	53

Excelでは共分散は =covar(範囲、範囲)、相関係数は =correl(範囲、範囲) とする。

Mathematicaでは統計パッケージを読み込み、xのデータリストをA,yのデータリストをBとおいて、CovarianceMLE[A,B], Correlation[A,B] とする。

答: 101.32, 0.76

- 3) n 個の数データの組 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ を入力して次の式により y の x についての回帰直線の方程式 $y = a + bx$ の a, b を求める。それは最小2乗法による。

$$b = C_{xy} \frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2}$$

$$a = \text{mean}_y - \text{mean}_x * b$$

次のデータの散布図を描き、それに回帰直線を添える。

データ: (2,53), (3,67), (4,79), (5,88), (6,97), (7,107), (8,101)

Excelでは b = slope(yの範囲, xの範囲), a = intercept(yの範囲, xの範囲) とする。

Mathematicaでは、Fit[{{2,53}, ..., {8,101}}, {1,x}, x] で x の一次式 $a + bx$ を得る。

答: 8.64, 41.36

- 4) 確率分布・確率密度関数と累積分布・累積密度関数を求め、そのグラフを描くこと。

2項分布

ポアソン分布

正規分布

2項分布は、ある事象Aの起こる確率が p のとき、n回の試行で事象Aがx回起こる確率で $f(x) = {}_n C_x p^x (1-p)^{n-x}$ ($x = 0, 1, 2, \dots, n$) で与えられる。 n, p を与えたとき、 $f(x)$ は =binomdist(x,n,p,False) で得る。また累積確率密度は =binomdist(x,n,p,True) で得る。

ポアソン分布はその平均 (=分散) を m とするとき $f(x) = \frac{m^x}{x!} e^{-m}$ で定義され、

=poisson(x, m, False/True) によって得る。

正規分布は $f(x) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{(x-m)^2}{2s^2}}$ で定義され、累積確率密度は =normsdist(x, μ , σ , False

/True) で得る。それぞれのグラフは、(累積) 分布の関数表からグラフウイザードを利用する。

Mathematica では、二項分布、ポアソン分布、正規分布はそれぞれ BinomialDistribution[n,p], PoissonDistribution[μ], NormalDistribution[μ , σ]である。確率密度と累積確率密度関数は、これらの分布 dist に対して PDF[dist,x], CDF[dist,x] であり、Plot を用いてグラフが描ける。例えば標準正規分布は次で得られる。

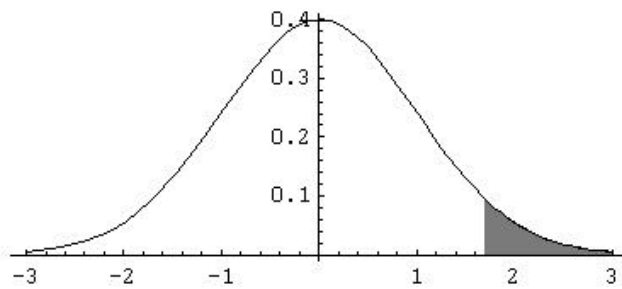
```
<<Statistics`Master`
```

```
Plot[PDF[NormalDistribution[0,1],x],{x,-3,3}]
```

5) 累積分布関数の逆関数の値を求める。

標準正規分布 $N(0, 1)$

$[x, \infty)$ の部分の面積が α のときの x を求める。



Excel では

=normsinv(1 - α)

Mathematica では

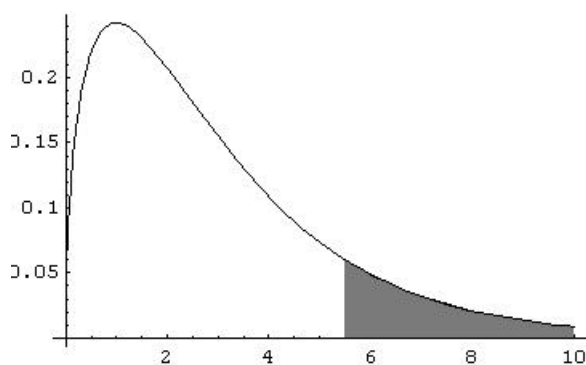
```
<<Statistics`Master`
```

```
Quantile[NormalDistribution [0,1], 1 -  $\alpha$  ]
```

$\alpha = 0.025$ のとき $x = 1.96$

自由度 n のカイ 2 乗分布 (下図は $n = 3$)

$[z, \infty)$ の部分の面積が α のときの z を求める。



Excel では

=chiinv(α , n)

Mathematica では

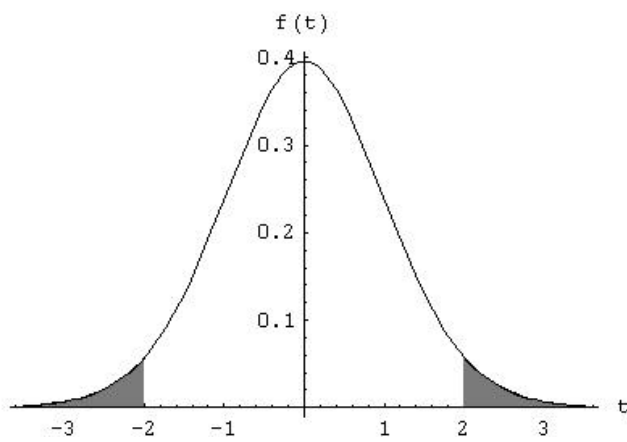
```
<<Statistics`Master`
```

```
Quantile[ChiSquareDistribution[n], 1 -  $\alpha$  ]
```

$n = 3, \alpha = 0.05$ のとき $z = 7.815$

自由度 n の t 分布 (下図は $n = 30$)

$(-\infty, z], [z, \infty)$ の部分の面積の和が α のときの t を求める。



Excel では

`=tinv( $\alpha$ , n)`

Mathematica では

`<<Statistics`Master``

`Quantile[StudentTDistribution`

`[n], 1 -  $\alpha$  / 2]`

$n = 30$, $\alpha = 0.05$ のとき $t = 2.042$