QC検定2級　受験前講習会

2.統計的方法の基礎(確率分析)

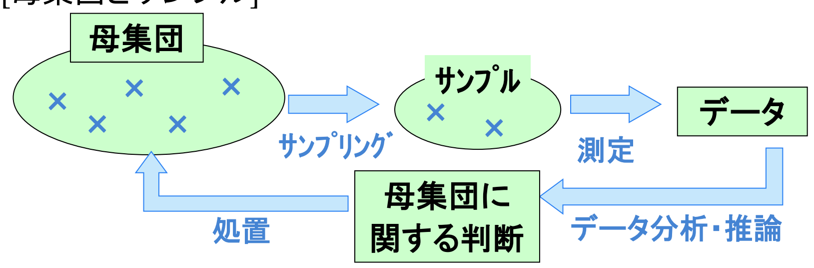
会津オリンパス株式会社
古川　英嗣

このテキストには著作権があり、許可無く資料のコピー・掲載はできません。
また、セミナーの録音もこれに該当し、録音できません。

統計的推測

統計的方法とは：沢山のデータ(母集団)を要約し、中に含まれている情報を把握しやすくするための手段

[母集団とサンプル]



神のみぞ知る量			得られるデータ
母数			**統計量**
・母平均	μ ←	x	・平均
・母分散	σ^2 ←	V	・分散
・母相関係数	ρ ←	r	・標本相関係数

統計的推測

例：100人の日本人について体重を測定した場合

100 個のデータを眺めただけでそこに含まれる情報を読み取るのは困難

↓

100 個のデータのほぼ真ん中を表す要約値または代表値として平均値を求める

↓

平均値が 60kg だった

↓

「100 人の日本人の体重はだいたい 60kg ぐらいである」という情報を読み取ることができる

↓

「日本人の体重はだいたい 60kg ぐらいである」と推測する

・平均値のような要約値 ⇒ 統計量
・個々のデータ ⇒ 確率変数

データの種類

テキストP.14～17

取得できるデータは大きく2種類

計量値

連続量として測られる品質特性の値。重量(kg)，寸法(mm)，作業時間(sec)，温度(deg)など。
⇒正規分布とみなす

計数値

離散的な数として測られる品質特性の値。キズの数(xヶ所)，不良品数(x個)，不良率(x%)など。
⇒二項分布，ポアソン分布となる

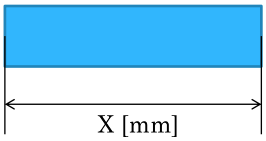
データの種類によって，推論する母集団の分布が異なる

計量値の例

テキストP.29～35

サンプルから母集団を推測

ある部品の寸法の長さ

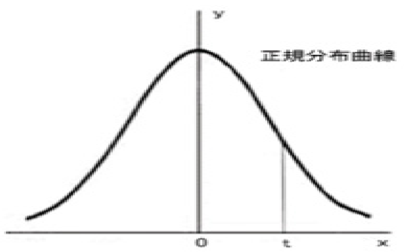


繰り返し加工した時の寸法

データ[mm]:7，8，5，7，6，7，9，6，8

平均値 x = 7.0
分散 V=1.5

} 統計量



5

基本統計量

テキストP.29～35

統計量とは

平均値，標準偏差など，得られた生のデータから加工(算出)したものを統計量という。

- **平均** $\bar{x}$

$$\bar{x} = \frac{\text{データの総和}}{\text{データ数}} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- **(偏差)平方和 S**

$$S = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

- **(不偏)分散 V**

$$V = \frac{S}{n-1}$$

- **標準偏差 s**

$$s = \sqrt{V}$$

6

その他の必要な計算式①

テキストP.29～35

・範囲 R

計量的な観測値の最大値と最小値の差

・メディアン $\tilde{x}$

観測値を大きさの順に並べたとき、ちょうど中央に当たる1つの値
（**偶数の場合**は中央の2つの値の平均）

・変動係数 CV

標準偏差を平均値で割ったもの。ばらつきを相対的に表す。

$$CV = \frac{s}{\bar{x}} \times 100(\%)$$

7

その他の必要な計算式②

テキストP.338～341

・工程能力指数 Cp,Cpk

平均値と標準偏差から，工程がどの程度安定して規格内の品物を生産することができるかを示す。

適用条件	工程能力指数
①平均値の偏りを考慮しない場合	$C_p = \frac{S_U - S_L}{6s}$
②平均値が規格の中心より下限規格側にずれている場合	$C_{pk} = \frac{\bar{x} - S_L}{3s}$
③平均値が規格の中心より上限規格側にずれている場合	$C_{pk} = \frac{S_U - \bar{x}}{3s}$

S_U；上限規格， S_L；下限規格， s；標準偏差，
$\bar{x}$；平均

8

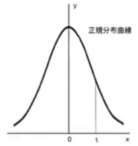
データの分布(連続変数) 計量値

テキストP.71～85

・正規分布

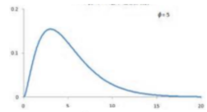
確率密度関数 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$

で与えられる分布。



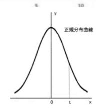
・x^2分布(カイ2乗分布)

分散の検定時に仮定される分布



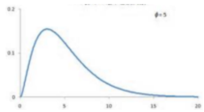
・t分布

標準偏差が未知の場合に仮定される分布



・F分布

母分散の比の検定時に仮定される分布



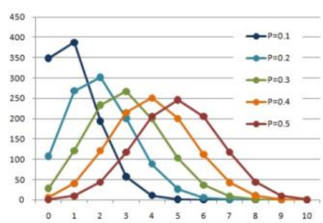
9

データの分布(離散変数) 計数値

テキストP.67～70

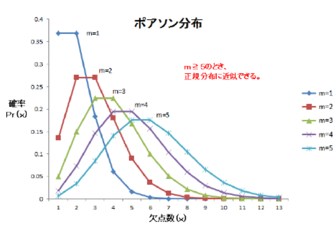
・二項分布

不良個数等の計数値に仮定される分布。n数が大きくなると，正規分布に近似される。



・ポアソン分布

欠点数等の計数値に仮定される分布。n数が大きくなると，正規分布に近似される。



・データの種類・条件から、母集団の分布を推測する。

・推測した分布により使う計算式が変わる

10

期待値と分散

テキストP.59～65

期待値：確率分布に対する平均値。意味上は、データの平均値と同じ。

データが計量値である確率密度関数で表せられる場合、

- **確率変数Xの期待値(平均)　E(X)**

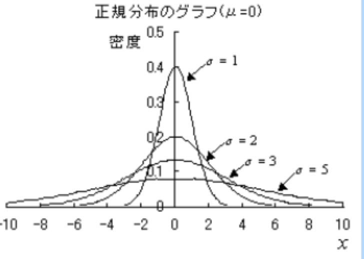
　分布の中心を表す量⇒母平均　μ

- **確率変数Xの分散　V(X)**

　ばらつき度合いを表す量

- 確率変数X，Yの共分散　Cov(X,Y)

　2つの確率変数の関係を表す量



11

平均(期待値)と分散の重要な性質

●E(X+Y)=E(X)+E(Y)
●E(X-Y)=E(X)-E(Y)
●E(aX+b)=aE(X)+b
●E(aX+bY)=aE(X)+bE(Y)

分散の加法性

●V(X+Y)=V(X)+V(Y)　ただし，XとYが独立のとき
●V(X-Y)=**V(X)+V(Y)**　ただし，XとYが独立のとき
●V(aX+b)=a^2V(X)
●V(aX+bY)=a^2V(X)+b^2V(Y)　ただし，XとYが独立のとき
※加法性が成り立つのは分散Vについてであり，標準偏差sと混同しないこと

12

共分散

XとYが独立でないとき(何らかの関係があるとき)

●$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$

●$V(X-Y)=$**$V(X)+V(Y)-2Cov(X,Y)$**

●$V(aX+bY)=a^2V(X)+b^2V(Y)+2ab\ Cov(X,Y)$

$Cov(X,Y)=E(XY)-E(X)E(Y)$

※XとYが独立の場合，$Cov(X,Y)=0$となる

13

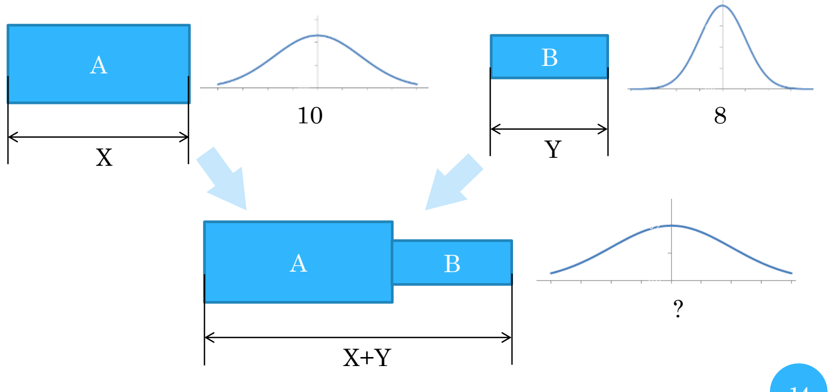
分散の加法性の具体例

部品Aと部品Bを組み合わせたX+Yの期待値と分散を求める

部品Aの寸法X:期待値$E(X)=10$,分散$V(X)=3$

部品Bの寸法Y:期待値$E(Y)=8$,分散$V(Y)=1$

XとYは独立とする。

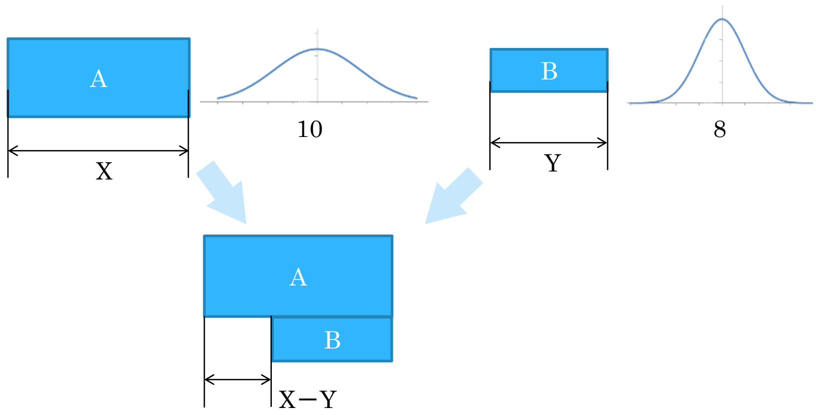


14

期待値　$E(X+Y)=E(X)+E(Y)=10+8=18$

分散　　$V(X+Y)=V(X)+V(Y)=3+1=4$

分散の加法性の具体例



期待値　$E(X-Y)=E(X)-E(Y)=10-8=2$

分散　　$V(X-Y)=V(X)+V(Y)=3+1=4$

演習問題

[問1,2,3]

正規分布(ガウス分布)

テキストP.71～78

- 計量値の分布として最もよく用いられる分布
- 正規分布の期待値

　(母平均)　$E(X)=\mu$，(母分散)　$V(X)=\sigma^2$

- 確率密度関数

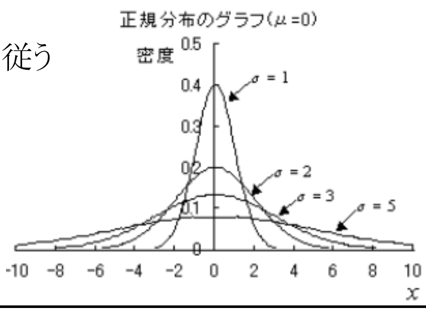
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

　μ, σ　⇒　正規分布の形状を決めるパラメータ(母数)

- 表記方法

　確率変数Xが正規分布に従う

　　⇒　$X\sim N(\mu,\sigma^2)$

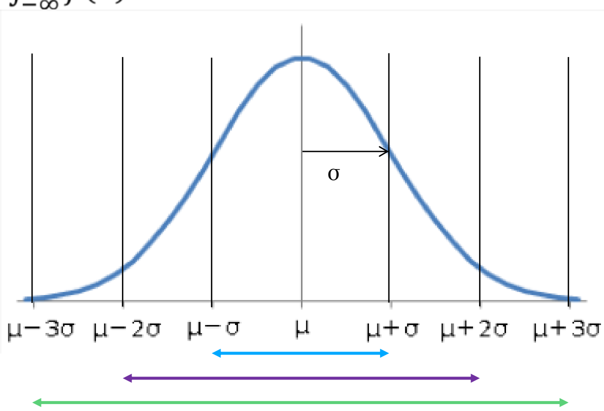


17

正規分布の特徴

- 平均値を中心にして左右対称
- 平均値の度数が最も多く，平均値から離れる程度数が減る
- 曲線下の面積の総和は1となる。

$$\int_{-\infty}^{\infty} f(x)dx=1$$



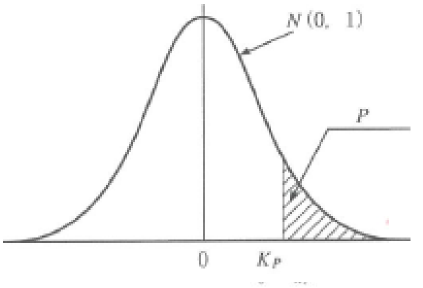
データの分布の範囲

- 約68.3%
- 約95.4%
- 約99.7%

18

標準正規分布

- 期待値 0 , 母分散1 の正規分布のこと
  X ~ N(0,1²)
- 確率の計算を簡単にするための基準
  ⇒付表から確率を求めることができる



（I） K_P から P を求める表

K_P	*=0	1	2	3	4	5	6	7	8	9
0.0*	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1*	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2*	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3*	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4*	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5*	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6*	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7*	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8*	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9*	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0*	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1*	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2*	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985

標準正規分布

- 計算時は標準正規分布に変換して計算する(標準化)
  確率pを算出する場合は，まず標準化
- 標準化の式

$$X \sim N(\mu, \sigma^2)$$

$$u = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu}{\sigma}$$

$$u \sim N(0\,, 1^2)$$

※後に似たような式が出てくるので注意。

$$u = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{\bar{\boldsymbol{X}} - \boldsymbol{\mu}}{\sqrt{\boldsymbol{\sigma}^2/\boldsymbol{n}}}$$

- 標準化後に、付表から確率を求める。

確率分布の計算

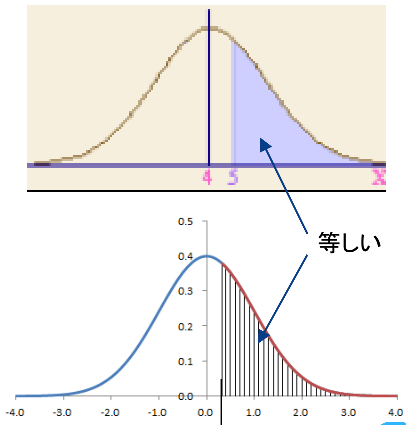
[例題1]
確率変数Xが正規分布 N(4,3²)に従うとき，Xが5以上となる確率P(=Pr(X>5))を求めよ。

N(4,3²) ⇒ μ=4 , σ=3
5以上 ⇒ X=5

標準化

$$u = \frac{X-\mu}{\sigma} = \frac{5-4}{3} = 0.333$$

P=Pr(X>5)=Pr(u(Kp)>0.333)



u=0.33
(Kp=0.33)

確率分布の計算

P=Pr(X>5)=Pr(u(Kp)>0.333)

付表

(K_p から P を求める表)

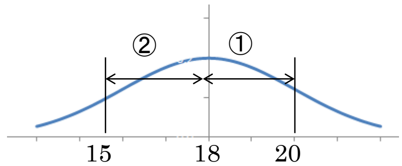
K_p	.00	.01	.02	.03	.04
.0	.50000	.49601	.49202	.48803	.48405
.1	.46017	.45620	.45224	.44828	.44433
.2	.42074	.41683	.41294	.40905	.40517
.3	.38209	.37828	.37448	.37070	.36693
.4	.34458	.34090	.33724	.33360	.32997
.5	.30854	.30503	.30153	.29806	.29460
.6	.27425	.27093	.26763	.26435	.26109
.7	.24196	.23885	.23576	.23270	.22965
.8	.21186	.20897	.20611	.20327	.20045
.9	.18406	.18141	.17879	.17619	.17361
1.0	.15866	.15625	.15386	.15151	.14917
1.1	.13567	.13350	.13136	.12924	.12714

解答　P=Pr(X>5)=0.3707 (約37%)

確率分布の計算

[例題2]
確率変数Xが正規分布 N(18,2²)に従うとき，Xが15以上20以下となる確率P(=Pr(15<X<20))を求めよ。

$N(18,2^2) \Rightarrow \mu=18 , \sigma=2$
確率 P=Pr(15<X<20)
=Pr(18<X<20) +Pr(15<X<18)

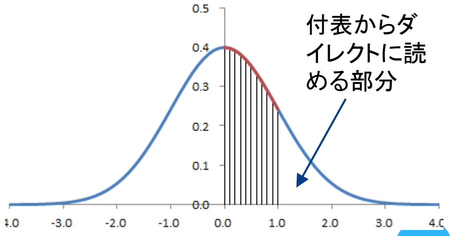


①Pr(18<X<20)

$$u=\frac{X-\mu}{\sigma}=\frac{20-18}{2}=1.0$$

P=Pr((0<)Kp<1.0)

付表より，Pr(Kp>1.0)=0.15866
右半分の総確率は0.5なので，
Pr(Kp<1.0)=0.5-Pr(Kp>1.0)
=0.5-0.1587
=0.3413



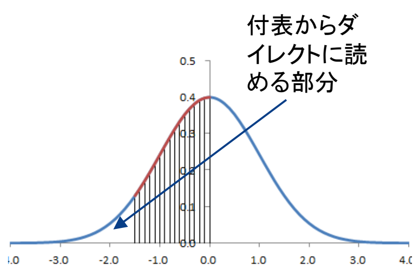
曲線下部の総面積は1,左右対称の形状なので，右半分の総面積は0.5

確率分布の計算

②Pr(15<X<18)

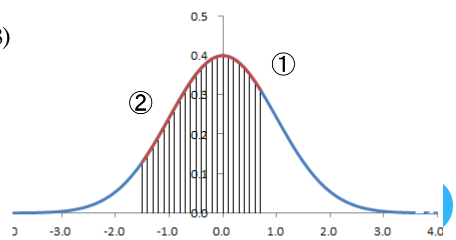
$$u=\frac{X-\mu}{\sigma}=\frac{15-18}{2}=-1.5$$

P=Pr((0>)Kp>-1.5)
=Pr(0<Kp<1.5) ←左右対称なので
=0.5-Pr(Kp>1.5)
=0.5-0.0668=0.4332



左右対称なので，Kpの符号を変えても値は同じ

①+② =Pr(18<X<20) +Pr(15<X<18)
=0.3413+0.4332
=0.7745



演習問題

[問4]

正規分布の特徴

・正規分布に基づくデータx の**平均値の分布**

データにばらつきがあるので，その平均値にもばらつきがある

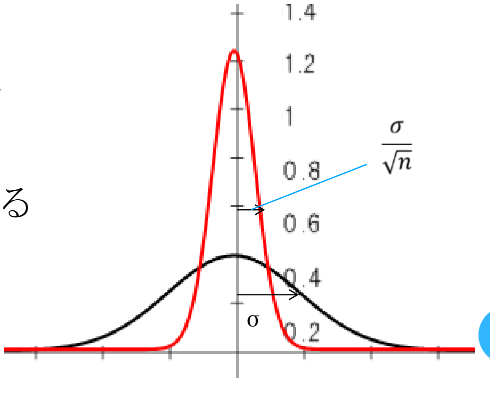
データが正規分布ならば，その平均値も正規分布になる

平均値のばらつきは，データのばらつきよりも小さい(1/√n)。

⇒大数の法則

$X \sim N(\mu, \sigma^2)$のとき、
平均値の分布は

$\boldsymbol{\bar{X}} \sim N\left(\mu, \boldsymbol{\sigma^2/n}\right)$となる



平均値の分布の場合の標準化式

・変換式

$X \sim N(\mu, \sigma^2)$

$$u = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu}{\sigma}$$

$u \sim N(0, 1^2)$

$\bar{X} \sim N(\mu, \boldsymbol{\sigma^2/n})$

$$u = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{\boldsymbol{\bar{X} - \mu}}{\boldsymbol{\sqrt{\sigma^2/n}}}$$

$u \sim N(0, 1^2)$

この後の検定推定時の計算では、平均値について
考える手法の為、右式が多く登場します。

27

演習問題

[問5,6,7]

28

まとめ

・データには，計量値(連続変数)，計数値(離散変数)があり，データの種類によって分布が異なる。

計量値：正規分布，カイ2乗分布，t分布，F分布
計数値：二項分布，ポアソン分布

・分布の形を決めるのが期待値と分散

分散には**加法性**が成り立つ

・正規分布は計量値の分布として最もよく用いられる分布

期待値　$E(X) = \mu,\ V(X) = \sigma^2$
平均値を中心にして左右対称
曲線下の面積の総和は1となる

29

まとめ

・平均値のばらつきは，データのばらつきよりも小さい(1/√n)。
⇒大数の法則

・期待値 0 , 母分散1 の正規分布を標準正規分布
・標準正規分布に変換することを標準化という

$$u = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu}{\sigma} \qquad u = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$$

30