【母集団と標本調査】

全数調査…調査の対象とする集団のすべてを調査

例）学校での身体測定

標本調査…調査の対象とする集団の中から一部を抜き出して調査

例）世論調査、TVの視聴率調査

【平均、分散、標準偏差】

確率変数 X の確率分布が下の表１のように与えられているとき、

X	x_1	x_2	・・・	x_n	計
P	p_1	p_2	・・・	p_n	1

表１

確率変数 X の

平均（期待値）$E(X)=x_1p_1+x_2p_2+\cdots+x_np_n=\sum_{i=1}^{n}x_ip_i$

分散　$V(X)=\sum_{i=1}^{n}(x_i-m)^2p_i=E(X^2)-\{E(X)\}^2$

標準偏差　$\sigma(X)=\sqrt{V(X)}$

a,b を定数とするとき、確率変数 X の１次式の

平均　$E(aX+b)=aE(X)+b$

分散　$V(aX+b)=a^2V(X)$

標準偏差　$\sigma(aX+b)=|a|\sigma(X)$

確率変数の和の平均　$E(X+Y)=E(X)+E(Y)$

独立な確率変数の

積の平均　$E(XY)=E(X)\cdot E(Y)$

和の分散　$V(X+Y)=V(X)+V(Y)$

【二項分布】

ある試行において、事象 A の起こる確率を p、$\overline{A}$ が起こる確率を $q=1-p$ とする。この試行を n 回くり返すとき、事象 A が起こる回数を X とすると、X は確率変数であり、$X=r$ となる確率は

$P(X=r)={}_nC_r\,p^r\,q^{1-r}\,(r=0,1,2,\cdot\ \cdot\ \cdot,n)$となり、

X の確率分布は

X	0	1	…	r	…	n	計
P	${}_nC_0q^n$	${}_nC_1p^1q^{n-1}$	…	${}_nC_rp^rq^{n-r}$	…	${}_nC_np^n$	1

表２

X の確率分布が上の表２のようになるとき、この分布を確率 p に対する次数 n の二項分布といい、$B(n,p)$ で表す。

確率変数 X が二項分布 $B(n,p)$ に従うとき

平均　$E(X)=np$

分散　$V(X)=npq$　　ただし、$q=1-p$

1　赤球４個と白球３個が入った袋から２個の球を同時に取り出すとき、これに含まれている白玉の個数を X とする。確率変数 X の平均 $E(X)$ と分散 $V(X)$ および標準偏差 $\sigma(X)$ を求めよ。

２　確率変数 X の平均点が 57 、標準偏差が ８ であるとき、次の問いに答えよ。

(1) 確率変数 $Y=\dfrac{X-57}{8}$ の平均と分散および標準偏差を求めよ。

(2) Aさんの点数が 65 点であるとき、Aさんの偏差値を求めよ。

３　箱 A には赤球２個、白球３個、箱 B には赤球３個、白球２個が入っている。箱 A,B からそれぞれ２個の球を取り出すとき、それぞれに含まれる白球の個数を X,Y とする。このとき、確率変数 X,Y に対して、 $E(3X+2Y)$ 、 $E(XY)$、 $V(3X+2Y)$ をそれぞれ求めよ。

４　次の問いに答えよ

(1)　１個のさいころを 300 回投げるとき、４の目が出る回数を X とする。 X の平均と標準偏差を求めよ。

(2)　確率変数 X の分布が二項分布 $B(40,p)$ であり、 X の分散が 10 であるとき、 p の値および X の平均を求めよ。

【連続型確率変数と確率密度関数】

一般に、実数のある区間全体において連続な値をとる確率変数 X に対して、次のような性質(ⅰ)〜(ⅲ)をもつ１つの関数 $y=f(x)$ が対応する。

(ⅰ)　$f(x)\geqq 0$

(ⅱ)　$P(a\leqq X\leqq b)=\int_a^b f(x)dx$

(ⅲ)　曲線 $y=f(x)$ と x 軸の間の面積は１

このとき、X を連続型確率変数、関数 $f(x)$ を X の確率度関数、関数 $y=f(x)$ のグラフを X の分布曲線という。

【連続型確率変数の平均と分散、標準偏差】

連続型確率変数 X のとり得る値の範囲が $a\leqq X\leqq b$ で、その確率密度関数が $f(x)$ であるとき、

X の平均 $E(X)=\int_a^b xf(x)dx$

X の分散 $V(X)=\int_a^b (x-m)^2 f(x)dx$　(ただし、$m=E(X)$)

X の標準偏差 $\sigma(X)=\sqrt{V(X)}$

【正規分布】

連続型確率変数 X の確率密度関数が $f(x)$ が

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

で与えられるとき、X は正規分布 $N(m,\sigma^2)$ に従うといい、$y=f(x)$ のグラフを正規分布曲線という。

確率変数 X が正規分布 $N(m,\sigma^2)$ に従うとき、

$E(X)=m$

$V(X)=\sigma^2$

$\sigma(X)=\sigma$

【標準正規分布】

確率変数 X が正規分布 $N(m,\sigma^2)$ に従うとき、

$$Z=\frac{X-m}{\sigma}$$

とすると、Z は平均０、標準偏差１の正規分布 $N(0,1)$ に従う。このZを、X を標準化した確率変数という。

【二項分布の正規分布による近似】

一般に、二項分布 $B(n,p)$ に従う確率変数 X は、n が十分に大きいとき、$q=1-p$ とおくと、近似的に $N(np,npq)$ に従うことが知られている。

したがって、$P(a\leqq X\leqq b)$ を求めるとき、確率変数 X の標準化を考えて、

$$Z=\frac{X-np}{\sqrt{npq}}$$

とおくと、確率変数 Z は標準正規分布 $N(0,1)$ に従うとみなして計算してよい。

１　確率変数 X の確率密度関数 $f(x)$ が、k を定数として、$f(x)=kx^2\ (0\leqq x\leqq 1)$ で与えられている。

(1)　定数 k を求めよ。

(2)　確率 $P\left(\frac{1}{3}\leqq X\leqq\frac{1}{2}\right)$ を求めよ。

(3)　X の平均 $E(X)$ を求めよ。

(4)　X の分散 $V(X)$ を求めよ。

（　　）組（　　）番　名前（　　　　　　　　　　）

２　確率変数 X が正規分布 $N(50,100)$ に従うとき、次の確率を求めよ。

(1)　$P(X \geqq 65)$ を求めよ。

(2)　$P(X \leqq 43)$ を求めよ。

(3)　$P(45 \leqq X \leqq 60)$ を求めよ。

３　T 高校の２年生の女子 200 人の身長の平均は 158 cm 、標準偏差 7 cm の正規分布と見なせるという。次の問いに答えよ。

(1)　身長が 155 cm 以上 165 cm 以下である生徒は約何％いるか。

(2)　身長が高い方から 40 番目の生徒の身長は約何 cm か。

４　さいころを 100 回投げる時、２の目が出る回数を X とする。このとき、次のようになる確率を正規分布による近似によって求めよ。

(1)　$X \geqq 12$

(2)　$9 \leqq X \leqq 15$

【標本平均】

母集団から無作為抽出をする大きさ n の標本の変量を $X_1, X_2, \cdot\cdot\cdot, X_n$ とする。これらの平均

$\overline{X} = \dfrac{X_1 + X_2 + \cdot\cdot\cdot + X_n}{n}$ を標本平均という。

【標本平均の平均と標準偏差】

母平均 m 、母標準偏差 σ の母集団から大きさ n の無作為標本を復元抽出するとき、

標本平均 $\overline{X}$ の平均　$E(\overline{X}) = m$

標準偏差　$\sigma(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$

【標本平均の分布と正規分布】

母平均 m 、母分散 σ^2 の母集団から無作為抽出された大きさ n の標本平均 $\overline{X}$ の分布は、 n が大きければ

正規分布 $N\left(m, \dfrac{\sigma^2}{n}\right)$ とみなすことができる。

【母平均の推定】

母分散 σ^2 が分かっている母集団から、大きさ n の標本を抽出したとき、この標本平均 $\overline{X}$ より母平均 m を推定すると、

m に対する信頼度 95 %の信頼区間は

$$\overline{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leqq m \leqq \overline{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

m に対する信頼度 99 %の信頼区間は

$$\overline{X} - 2.58 \times \frac{\sigma}{\sqrt{n}} \leqq m \leqq \overline{X} + 2.58 \times \frac{\sigma}{\sqrt{n}}$$

【母比率の推定】

母集団の中で、ある性質 A をもつ個体の割合を p とする。この p を、性質 A をもつ個体の母集団における母比率という。大きさ n の標本を抽出したとき、この標本における性質 A をもつ個体の割合 p_0 より母比率 p を推定すると、

母比率 p に対する信頼度 95 %の信頼区間は

$$p_0 - 1.96\sqrt{\frac{p_0(1-p_0)}{n}} \leqq m \leqq p_0 + 1.96\sqrt{\frac{p_0(1-p_0)}{n}}$$

母比率 p に対する信頼度 99 %の信頼区間は

$$p_0 - 2.58\sqrt{\frac{p_0(1-p_0)}{n}} \leqq m \leqq p_0 + 2.58\sqrt{\frac{p_0(1-p_0)}{n}}$$

【仮説検定】

ある事象に対して、

統計的に有意な差がないとする仮説（帰無仮説）
統計的に有意な差があるとする仮説（対立仮説）

を立て、データにもとづいて確率的に判断することを、仮説検定という。

棄却…帰無仮説が成り立たないと判断すること。
有意水準…めったに起こらないと判断する基準となる確率。
　　　　5 %、 1 %を利用することが多い。
棄却域…帰無仮説が棄却される基準となる値の範囲

【片側検定･両側検定】

片側検定…仮定した母平均の片側にだけ棄却域をとって考える検定。
両側検定…仮定した母平均よりも大きい側と小さい側の両側に棄却域をとって考える検定。

1　次の問いに答えよ。

(1)　T 高校の男子の身長の平均は 171 cm 、標準偏差は 7 cm である。 T 高校の男子 90 人を無作為に選ぶとき、この 90 人の身長の平均 $\overline{X}$ の平均と標準偏差を求めよ。

(2)　ある母集団から復元抽出された大きさ 4 の標本の変量を X_1, X_2, X_3, X_4 とするとき、標本平均 $\overline{X}$ の平均と標準偏差を求めよ。ただし、 X_2 の確率分布は、下の表の通りとする。

X_2	1	2	3	4
P	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	$\frac{1}{16}$

２　下の表のように、数字１、２、３、４、５を記入した 20 枚のカードを箱に入れ、これから大きさ 30 の標本を復元抽出する。

数字X	1	2	3	4	5	計
枚数	4	6	3	5	2	20

(1)　標本平均 $\overline{X}$ の平均と標準偏差を求めよ。

(2)　標本平均 $\overline{X}$ が 3.6 以上の値をとる確率を求めよ。

３　ある工場で、缶詰を製造している。製品の中から任意に 200 個を抽出し、その質量の平均を測定したところ、 105 g であった。母標準偏差を 1.44 g として、この工場の缶詰の質量の平均 m に対する信頼度 95 %の信頼区間を、小数第 3 位を四捨五入して求めよ。

４　過去の資料から、 高校 2 年生男子のハンドボール投げの記録の標準偏差は、6.6 m であることが知られている。高校 2 年生男子のハンドボール投げの記録の平均を推定するとき、信頼区間の幅を 2 m 以下、信頼度 95 %で推定するためには、何人以上調査する必要があるか求めよ。

５　ある川辺には、A,B 2種類の蛍が生息している。任意に300匹捕まえたところ、A 種が140匹いた。A 種の蛍は、この川辺に何%生息していると考えられるか。信頼度99%で推定せよ。ただし、$\sqrt{7}=2.65$ とし、小数第2位を四捨五入して求めよ。

６　T 国の金貨は、投げた時、表が出やすいという噂がある。この噂の真偽を判断するため、実験を行ったところ、200回投げて表が121回でた。このとき、このコインは表が出やすいと判断できるか。有意水準5%で片側検定せよ。

７　ある銘柄の牛乳パックを見たところ、内容量1000 ml と表示されていた。任意抽出した100パックの平均内容量の平均を量ったところ、1016 ml であった。母標準偏差が64 ml であるとき、この牛乳パックの内容量表示は正しいといえるか。有意水準1%で次の方法で仮説検定せよ。

(1)　片側検定

(2)　両側検定

【アンケート調査の基本】

以下の内容については、最低限しっかりと決めたうえで、アンケートを実施しよう。

⑴　何のために調査をするか目的を明確にする。

⑵　的確な調査項目の決定。

⑶　調査対象（高校生、女性、医療従事者、etc…）の決定。

⑷　何人に調査すれば、適切なデータを集めたと言えるか。

⑸　いつ（時期）実施するか。

⑹　どのように調査するか。
（紙で回収、アンケートフォーム利用、口頭、etc...）

⑺　どのように解析するか。

⑻　どのような報告方法にするか。
（グラフを利用、数字のみ報告、etc...）

「⑵　的確な調査項目の決定。」については、以下のような内容に気をつけよう。

【質問文について】

①　失礼な語句を使っていないか。

②　１つの質問文に含まれている質問は１つだけか。

③　あいまいな表現は含まれていないか。

④　分かりにくい表現は含まれていないか。

⑤　偏った質問になっていないか。

２　アンケートの質問文として、問題点を抜き出し訂正せよ。

①　学校の制服を廃止しても良いか、良くないか答えやがれ。
良い　　　良くない

②　学校の体育館やプールは、使いやすいですか、使いにくいですか。
使いやすい　　　使いにくい

③　今日、学校に行く前に、きちんと朝食をとりましたか、とりませんでしたか。
とった　　　とらなかった

④　学校が水泳大会を開催することに賛成ですか、反対ですか。
賛成　　　反対

⑤　学校が土曜日に授業することは、反対ですか。
はい　　　いいえ

【選択解答式のアンケートの場合】

①　選択肢の内容は適切か。

②　考えられる選択肢がすべて用意されているか。

３　アンケートの選択肢として、問題のある番号を抜き出し、理由を答えよ。

数学の授業において、
①　強烈な眠気があった。
②　眠気があった。
③　少し眠気があった。
④　眠い時と眠たくない時があった。
⑤　眠気はなかった。

考えられる選択肢が、すべて用意できる場合ばかりでもない。そんな時は、
「その他（具体的にご記入ください　　　　　）」
などの選択肢を入れておくとよい。

「非常に」、「特別に」　などの強調する修飾語を付けた場合、回答者はその選択肢を選びにくい傾向にある。
「どちらともいえない」、「普通」　といった無難な選択肢を用意した場合、それらの選択肢に選択が集まりやすい傾向にあることも頭に入れておきたい。

「(4)　何人に調査すれば、適切なデータを集めたと言えるか。」については、以下のように Excel を利用して求めることができる。

【Excelによる調査すべき人数の決定】

1　調査する対象は、全校生徒 280 人である。ある内容についてアンケートをとき、賛成すると答える生徒が、 50 %前後と考えられるとき、信頼度 99 %として、調査すべき大きさを Excel を利用して求めよ。

B6　=ROUNDUP(B1/((B2/B5)^2*(B1-1)/(B4*(1-B4))+1),0)

	A	B	C	D	E	F	G
1	調査する母集団の大きさ	280					
2	目標精度	0.05	←Yesの比率が50%と予想されるなら、5%前後にとる。				
3	信頼度	0.99					
4	母比率	0.5	←0.5とすると、最も安全な大きさが得られる。				
5	係数	2.58					
6	調査すべき大きさ	198					
7							

2　調査する対象は、全校生徒 840 人である。ある内容についてアンケートをとき、賛成すると答える生徒が、 60 %前後と考えられるとき、信頼度 95 %として、調査すべき大きさを Excel を利用して求めよ。

「(7)　どのように解析するか。」については、以下のように Excel を利用して求めることができる。

【比率に関する検定】

二項分布の性質を利用した、比率に関する検定を Excel を用いて実施してみよう。

1　ある高校では、定期的に生徒に対してアンケートを実施している。その項目の 1 つに、数学の補習の開催時間帯に関するものがあり、開催時間帯に何らかの不満をもつ生徒の比率は、 10 %であった。

今回、このアンケート結果より、開催時間帯を見直し、その後改めて生徒 200 人に調査をしたところ、 12 人が不満をもつことが分かった。不満をもつ人の比率は、今までよりも低くなったといえるか。Excel を用いて考えてみよう。

H21

	A	B	C	D	E	F
1	標本の大きさ	200				
2	出現数	12				
3	基準となる発生確率	0.1				
5	発生確率（=）	0.015256996	←BINOMDIST(B2,B1,B3, 0)			
6	発生確率（＜）	0.016789533	←BINOMDIST(B2-1,B1,B3,1)			
7						
10	有意確率（片側）下	0.032046529	←B5+B6			
11						
12	BINOMDIST(出現数,標本の大きさ,基準となる発生確率,関数形式)					
13						
14	0.032046529＜0.05	より、母集団の不満率は低下したといえる。				
15						

2　ある高校では、定期的に生徒に対してアンケートを実施している。その項目の 1 つに、数学の補習 1 回あたりの時間（長さ）に関するものがあり、時間（長さ）に何らかの不満をもつ生徒の比率は、 15 %であった。

今回、このアンケート結果より、開催時間帯を見直し、その後改めて生徒 300 人に調査をしたところ、 40 人が不満をもつことが分かった。不満をもつ人の比率は、今までよりも低くなったといえるか。Excel を用いて考えてみよう。

【比率の差に関する検定】

　二項分布の正規近似を利用した、比率の差に関する検定を Excel を用いて実施してみよう。

１　ある高校において、校則について、賛成か反対かのアンケートを生徒対象に行った。男子生徒 150 人、女子生徒 130 人からアンケートを回収したところ、男子生徒は 150 人中 120 人が賛成であり、女子生徒は 130 人中 90 人が賛成であった。男子生徒と女子生徒では、賛成率に差があるといえるだろうか。Excel を用いて考えてみよう。

	A	B	C	D
1		男子生徒	女子生徒	
2	標本の大きさ	150	130	
3	出現数	120	90	
4	出現率	0.8	0.692308	
5	平均出現率	0.75		←(B3+C3)/(B2+C2)
6	出現率の差	0.107692		
7	有意水準	0.05		
8	検定統計量	2.075498		←B6/SQRT(B5*(1-B5)*(1/B2+1/C2))
9	棄却値（両側）	1.959964		←NORMSINV(1-B7/2)
10	有意確率（両側）	0.03794		←(1-NORMSDIST(ABS(B8)))*2
11				
12	SQRT（数値）	正の平方根を返す		
13	ABS（数値）	絶対値を求める		
14	NORMSDIST（数値）	標準正規分布の累積分布関数の値を求める。		
15				
16	0.03794＜0.05	より、男子生徒と女子生徒の賛成率に差があるといえる。		
17				

２　ある高校において、校則について、賛成か反対かのアンケートを生徒対象に行った。男子生徒 130 人、女子生徒 150 人からアンケートを回収したところ、男子生徒は 130 人中 120 人が賛成であり、女子生徒は 150 人中 136 人が賛成であった。男子生徒と女子生徒では、賛成率に差があると言えるだろうか。Excel を用いて考えてみよう。

【母平均の差の検定（ｔ検定）】

　比較したい２つの母平均に差があるかどうかを判定するための検定に、 *t* 検定がある。 *t* 検定は、データが正規分布に従っていると仮定できるような状況下で用いられる。

１　ある高校に通う２年生の生徒、男子 20 名、女子 20 名、合計 40 名に１カ月にもらうお小遣いの額について、アンケートを実施した。男子生徒と女子生徒では、もらうお小遣いの額の平均に差があるといえるだろうか。Excel を用いて考えてみよう。

	A	B	F	G	H
1	男子生徒	女子生徒		男子生徒	女子生徒
2	2500	4000	標本の大きさ	20	20
3	3500	4500	平均金額	5700	6050
4	4000	5500			
5	5500	6500			
6	3000	4000	F 検定の結果		
7	3500	5000	有意確率（両側）	0.741547	←FTEST(A2:A21,B2:B21)
8	5000	6000	※　0.741547>0.05	より、	
9	6500	7000	男子生徒のお小遣いの母分散と女子生徒のお小遣いの母分散に違いがあるとはいえない。		
10	10000	10000			
11	9000	9000	t 検定の結果		
12	7500	7000	【等分散】		
13	6500	8000	有意確率（両側）	0.619911	←TTEST(A2:A21,B2:B21,2,2)
14	9500	8500	【非等分散】		
15	8500	9500	有意確率（両側）	0.619928	←TTEST(A2:A21,B2:B21,2,3)
16	7500	7000	※　F検定の結果から、t 検定の結果は等分散の場合を採用		
17	5000	4500	0.619911>0.05	より、	
18	4000	4500	男子生徒のお小遣いと女子生徒のお小遣いの金額に差があるとはいえない。		
19	3500	3000			
20	4500	3500	FTEST(配列 1,配列 2)	F 検定の結果を求める。	
21	5000	4000	TTEST(配列 1,配列 2,尾部,検定の種類)		t 検定の結果を求める。
22					

２　ある高校に通う２年生の生徒、男子 20 名、女子 20 名、合計 40 名に１カ月に食堂で利用する金額について、アンケートを実施した。男子生徒と女子生徒では、利用する金額の平均に差があるといえるだろうか。Excel を用いて考えてみよう。

	A	B
1	男子生徒	女子生徒
2	2500	4000
3	8000	4500
4	6000	5500
5	7500	6500
6	5000	4000
7	3500	4500
8	6500	5500
9	7500	6500
10	12500	9500
11	9000	9000
12	7500	6500
13	6500	8000
14	9500	8500
15	8500	9500
16	7500	8000
17	5000	4500
18	4000	4500
19	3500	3000
20	4500	3500
21	5000	4000
22		