【令和５年度　仙台三高　１学年　SSデータサイエンス×SS数学】

推測統計学の扉

1．仮説検定入門
2．データの標準化
3．偏差値
4．正規分布
5．１群のZ検定（母平均の検定）
6．１群のt検定（母平均の検定）
7．t検定実習
8．二項定理から正規分布への近似
9．２群のｔ検定（２群の平均の差の検定）

年間スケジュール

データ学習の流れ

<table>
<tr><th>No</th><th>週</th><th>SSデータサイエンス</th><th>SS数学Ｉβ</th><th>SS数学Ｉα</th></tr>
<tr><td>1</td><td>入学前春休み</td><td>数学Ｉ「データの分析」を自主学習</td><td></td><td></td></tr>
<tr><td>2</td><td>6月21日　～ 7月10日</td><td>データの表現方法①
平均・中央値・分散と標準偏差</td><td></td><td rowspan="2">場合の数
確率
確率分布
条件付き確率</td></tr>
<tr><td>3</td><td>8月24日　～ 9月14日</td><td>データの表現方法②
散布図・データの変換・
相関係数・回帰分析</td><td></td></tr>
<tr><td>4</td><td>10月24日～ 12月21日</td><td>統計的な推測
仮説検定入門→標準化→偏差値→
正規分布→１群のZ検定（母平均の検定）
→１群のt検定（母平均の検定）→
ｔ検定実習→２群のｔ検定（２群の平均値の検定）</td><td>二項定理・二項分布から正規分布への近似</td><td>期待値</td></tr>
</table>

NO1.

仮説検定入門

仮説検定

与えられたデータをもとに
ある主張が正しいかどうかを判断する手法

仮説検定の証明手順は背理法の証明手順と類似

背理法

ある主張Ａを証明するのに
「Ａでないという主張には矛盾が生ずる」と示すことで
主張Aが正しいと示す証明法

↓

反対の主張の矛盾を示す

仮説検定の例

サイコロが１個あります。真島くんサイコロを振ると、
３回連続で１の目が出ました。

真島君：さすがにこれはおかしい。イカサマなサイコロだろう。

宮本君：そんなことはない。ちゃんとしたサイコロで
たまたまこうなったのだよ。

真島君：いや、そんなことはない。君の言っていることが正しいと仮定するとそんなことが起きる確率はわずか0.48％くらいしかないのだから。そんなことがたまたま起きるのかい？

宮本君：ギャフン。。。（何もいえない。。。）

真島君：ふふふ。何も言えないようだな。やはり私の仮説は正しい。

仮説検定の手順

主張したい仮説$\boxed{1}$が正しいと判断できるかどうかを次の手順で調べる。

仮説$\boxed{1}$：対立仮説
仮説$\boxed{2}$：帰無仮説

① 仮説$\boxed{1}$に反する仮説$\boxed{2}$を立てる。

② 仮説$\boxed{2}$のもとで起きる確率Pを調べる。

③ ②の確率Pが、あらかじめ設定した基準P_0に対して

(1) $P \leqq P_0$のときは、仮説$\boxed{2}$が誤りと判断する。

まれなことが起きたという証明

このとき仮説$\boxed{1}$は正しいと判断できる。

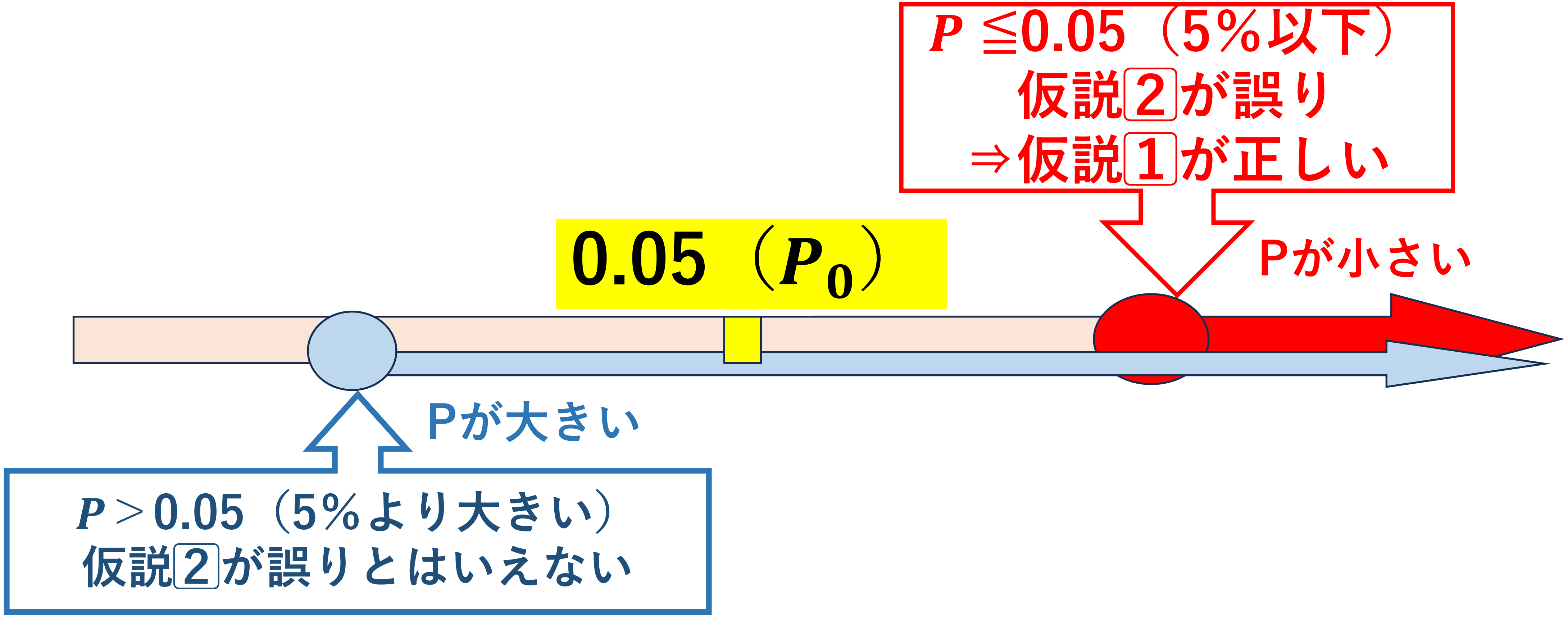
(2) $P > P_0$のときは、仮説$\boxed{2}$が誤りと判断できない。

まれなことが起きなかった証明

このとき、仮説$\boxed{1}$が正しいことは判断できない。

確率Pの基準

例　P_0 ＝0.05　とする。（5％棄却域・P値）



仮説の設定・棄却・採択

サイコロが１個あります。真島くんサイコロを振ると、
３回連続で１の目が出ました。

真島君：さすがにこれはおかしい。イカサマなサイコロだろう

対立仮説の設定

宮本君：そんなことはない。ちゃんとしたサイコロで
たまたまこうなったのだよ。

帰無仮説の設定

真島君：いや、そんなことはない。君の言っていることが正し
いと仮定するとそんなことが起きる確率はわずか
0.48％くらいしかないのだから。そんなことがたまた
ま起きるのかい？

帰無仮説の検証

宮本君：ギャフン。。。（何もいえない。。。）

帰無仮説の棄却

真島君：ふふふ。何も言えないようだな。やはり私の仮説は

対立仮説の採択

問１

実力が同じという評判の卓球部員A、Bが試合を行ったところ、Aが５連勝した。右の度数分布表は、表裏の出方が同様に確からしいコイン１枚を５回投げる操作を1000セット行った結果である。これを用いて、「Aの方がBより実力が上」という仮説が正しいかどうか、基準となる確率を５％として仮説検定せよ。

表の枚数	セット数
5	34
4	153
3	322
2	311
1	150
0	30
計	1000

問１解答

仮説1「Aの方がBより実力が上」に対し、仮説2を（　ア　）とする。

Aが５連勝する確率はコインを５回投げて５回とも表が出る確率に等しい。

度数分布表より、５回とも表が出たときの相対度数は（　イ　）である。

ゆえに、Aが５連勝する確率は（　ウ　）％と考えられ、
基準となる５％（より大きい・以下）である。

よって仮説2（ア）は（誤りである・誤りとはいえない）と判断する。

つまり仮説1は（正しいと判断できる・正しいかどうか判断できない）。

問２

実力が同じという評判のテニス部員A、Bが10回試合を行ったところ、Bが７勝３敗であった。このとき「Bの方がAより実力が上」という仮説が正しいかどうか、基準となる確率を５％として仮説検定せよ。（ただし、引き分けはないものとする）

考え方

「Bの方がAより実力が上」という仮説が正しいかどうかを判断するために
帰無仮説を「BとAの実力は同じ」
対立仮説を「Bの方がAより実力が上」とする。

AとBの実力が同じ（１／２）として、Bが７勝３敗になった事実を参考にして、帰無仮説「BとAの実力が同じ」かそうでないかを考える。

具体的には、Bの方が７勝３敗を基準にBの方が実力が上と判断するならば、Bが８勝２敗、９勝１敗、１０勝０敗の場合もあるので、これらすべて（Bが７勝３敗以上）の確率が５％以下であるのか（BとAの実力が同じ前提で非常にまれなことが起きているのか）、５％より大きいのかを考える

Bから見て	確率
０勝１０敗	
１勝９敗	
２勝８敗	
３勝７敗	
４勝６敗	
５勝５敗	
６勝４敗	
７勝３敗	
８勝２敗	
９勝１敗	
１０勝０敗	

この確率が５％以下なのかどうかを確かめる

問３

５本中３本が当たり（当たる確率が0.6）というくじを８回ひいたところ、そのうち２回で当たりが出た。このとき「このくじの当たりの本数は５本中３本より少ない」という仮説が正しいかどうか基準となる確率を５％として仮説検定せよ。

右の表を完成させ、確かめながら仮説検定をしよう。

当たりの本数	確率	相対度数
0		
1		
2		
3		
4		
5		
6		
7		
8		
計	1	

問３

５本中３本が当たり（当たる確率が0.6）というくじを８回ひいたところ、そのうち２回で当たりが出た。このとき「このくじの当たりの本数は５本中３本より少ない」という仮説が正しいかどうか基準となる確率を５％として仮説検定せよ。ただし、当たる確率が0.6のくじを８回ひいたときの当たりの本数の相対度数は右の表のようになるものとする。

当たりの本数	確率	相対度数
0		0.000655
1		0.007864
2		0.041288
3		0.123863
4		0.232243
5		0.278692
6		0.209019
7		0.08958
8		0.016796
計		1

問４

自分の身近な問題を仮説検定してみよう。

NO２.

データの標準化

データの変換の練習

問１

得点xを、$y=2x+40$　に変換した時の、yの平均、分散、標準偏差を求めなさい。

	A	B	C	D	E	平均	分散	標準偏差
得点x（点）	69	60	81	53	57			
$y=2x+40$に変換								

データの変換の練習結果

問１の結果

	A	B	C	D	E	平均	分散	標準偏差
得点x（点）	69	60	81	53	57	64	100	10
$y=2x+40$ に変換	178	160	202	146	154	168	400	20

yの平均は、xの平均を（　　　　）倍して（　　　　）を加える

yの分散は、 xの分散を（　　　　　）倍する

yの標準偏差は、 xの標準偏差を（　　　　　）倍する

標準化とは　変量 x を平均０、分散１に変換する

Standardization

Z_i ：標準化されたデータ

$$Z_i = \frac{x_i - \bar{x}}{S_x}$$

x_i ：元のデータ

x_i ：変量 x １つ１つのデータ（$i = 1, 2, 3, \cdot\cdot\cdot, n$）

$\bar{x}$ ：変量 x の１つ１つのデータの平均

S_x ：変量 x の標準偏差

Z_i ：変量 x を標準化した１つ１つのデータ
（$i = 1, 2, 3, \cdot\cdot, n$）

標準化の数学的理論

問２

なぜ変量 z は平均が０、分散が１になるのか、考えて見よう。
（ヒント）　データの変換の理論から考えよう。

$$Z_i = \frac{x_i - \bar{x}}{S_x}$$

標準化の練習

問３

標準化した得点zを求め、平均、分散、標準偏差を調べて見よう。

	A	B	C	D	E	平均	分散	標準偏差
得点x（点）	69	60	81	53	57			
標準化した得点z（点）								

$$Z_i = \frac{x_i - \bar{x}}{S_x}$$

標準化したデータで比較する

問４

次の表は、全国の高校１年生男子の握力x（kg）と、立ち幅跳びy（cm）に関するデータと、そのAさん、Bさんの記録である。標準化された記録を算出し、AさんとBさんを比較せよ。

	全国の平均値	全国の標準偏差	A	B
x(kg)	40.0	7.2	43.6	37.1
y(cm)	224.7	22.8	247.5	213.3

標準化したデータで比較する

問 5　空欄を埋め、どのようなことが分かるか考えよう

	A	B	C	D	E	平均	標準偏差
国語の素点	20	40	50	60	80		
標準化した国語の得点							
数学の素点	45	55	70	45	60		
標準化した数学の得点							
英語の素点	40	80	100	70	85		
標準化した英語の得点							

素点の和	105	175	220	175	225		
標準化後の得点の和							

NO３.

偏差値

ウォーミングアップ

- 裕介くんは数学と英語の試験を受けた。
- いずれも得点は75点であった。
- 数学の平均点は51点、英語の平均点は63点であったため、裕介くんは数学の方が良くできたと考えた。
- この考えは正しいだろうか。
- 数学の得点の標準偏差は20点、
英語の標準偏差は6点であったとするとどうか。

↓

標準化をして比べてみよう

偏差値とは

$$偏差値 = 10 \times \frac{x_i - \bar{x}}{s_x} + 50$$

問１

何が分かるか（標準化との違いは何か）

偏差値の計算

問２

偏差値を計算してみよう

	A	B	C	D	E	平均	分散	標準偏差
得点 x （点）	10	45	50	30	100	47	896	29.9
偏差値								

偏差値変換の練習

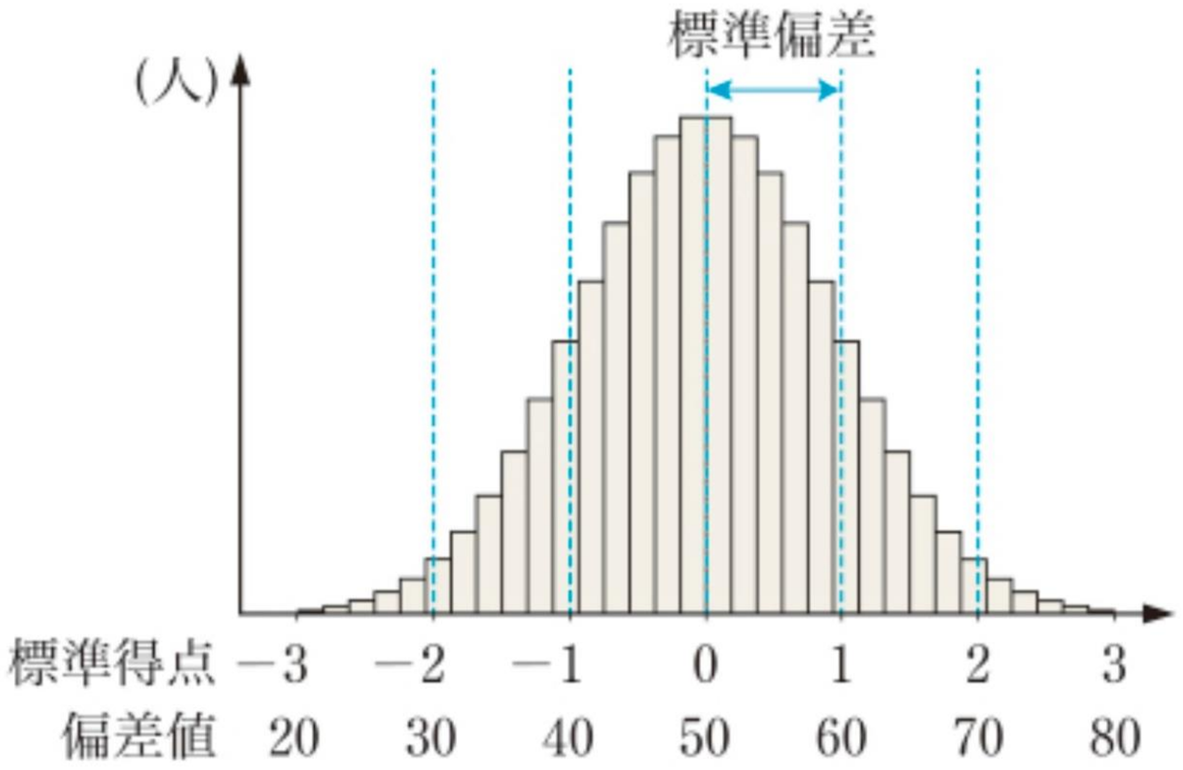
問３　スプレッドシートを利用しよう。

	A	B	C	D	E	平均	標準偏差
国語の素点	20	40	50	60	80		
標準化した国語の得点							
国語の偏差値							
数学の素点	45	55	70	45	60		
標準化した数学の得点							
数学の偏差値							
英語の素点	40	80	100	70	85		
標準化した英語の得点							
英語の偏差値							

偏差値の分布

多くの人が受験した試験の得点のデータを標準化すると、そのヒストグラムは次のような形になることが多い。

右のヒストグラムでは、偏差値60以上の部分にデータ全体のおよそ16%の値が、偏差値70以上の部分におよそ２%の値が含まれている。



<u>見込みの計算</u>（英語の得点の価値から数学の得点を算出）

問４

Ｂさんは数学の試験を受験できなかった。Ｂさんの数学の実力が、英語の実力と同程度であると仮定するとき、Ｂさんの数学の得点は何点であると考えられるか。小数第１位を四捨五入して答えよ。

	Ａさんの得点	Ｂさんの得点	平均点	標準偏差
英語	58	74	53.1	16.2
数学	83	欠席	61.9	19.1

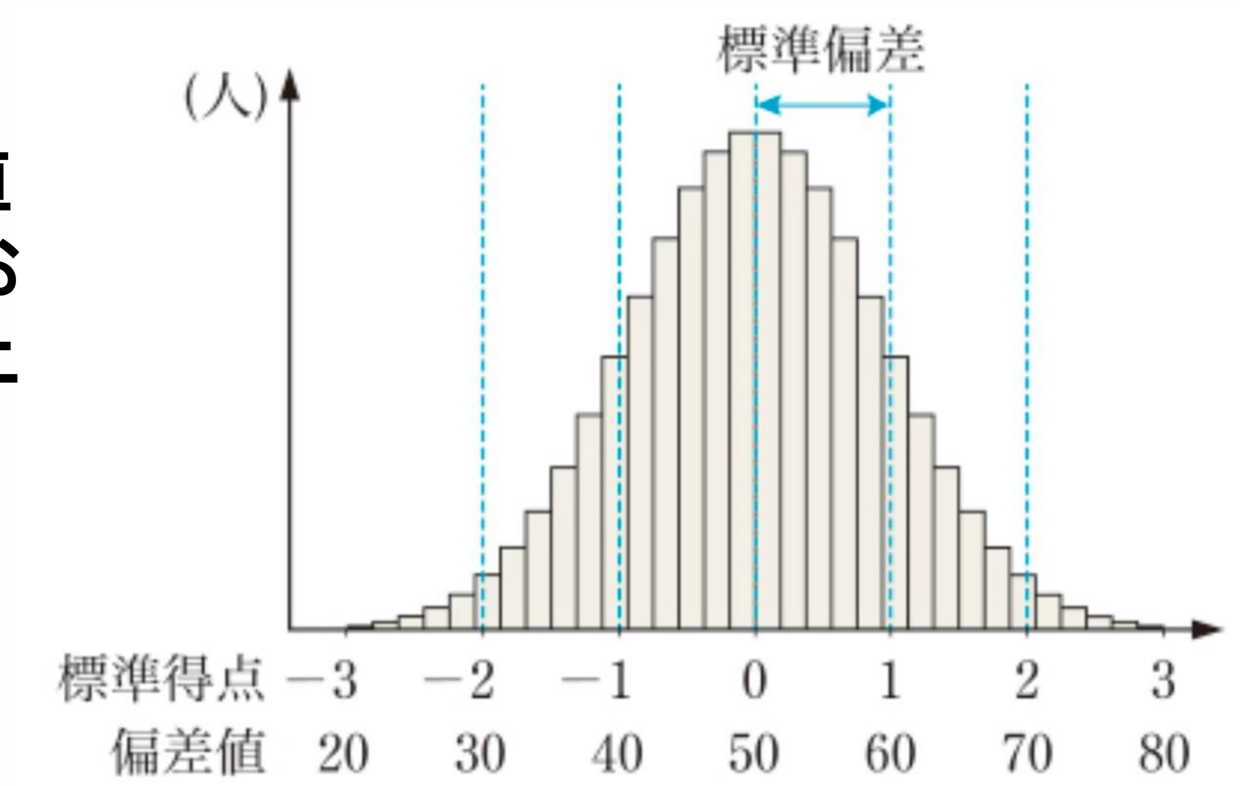
NO４.

正規分布

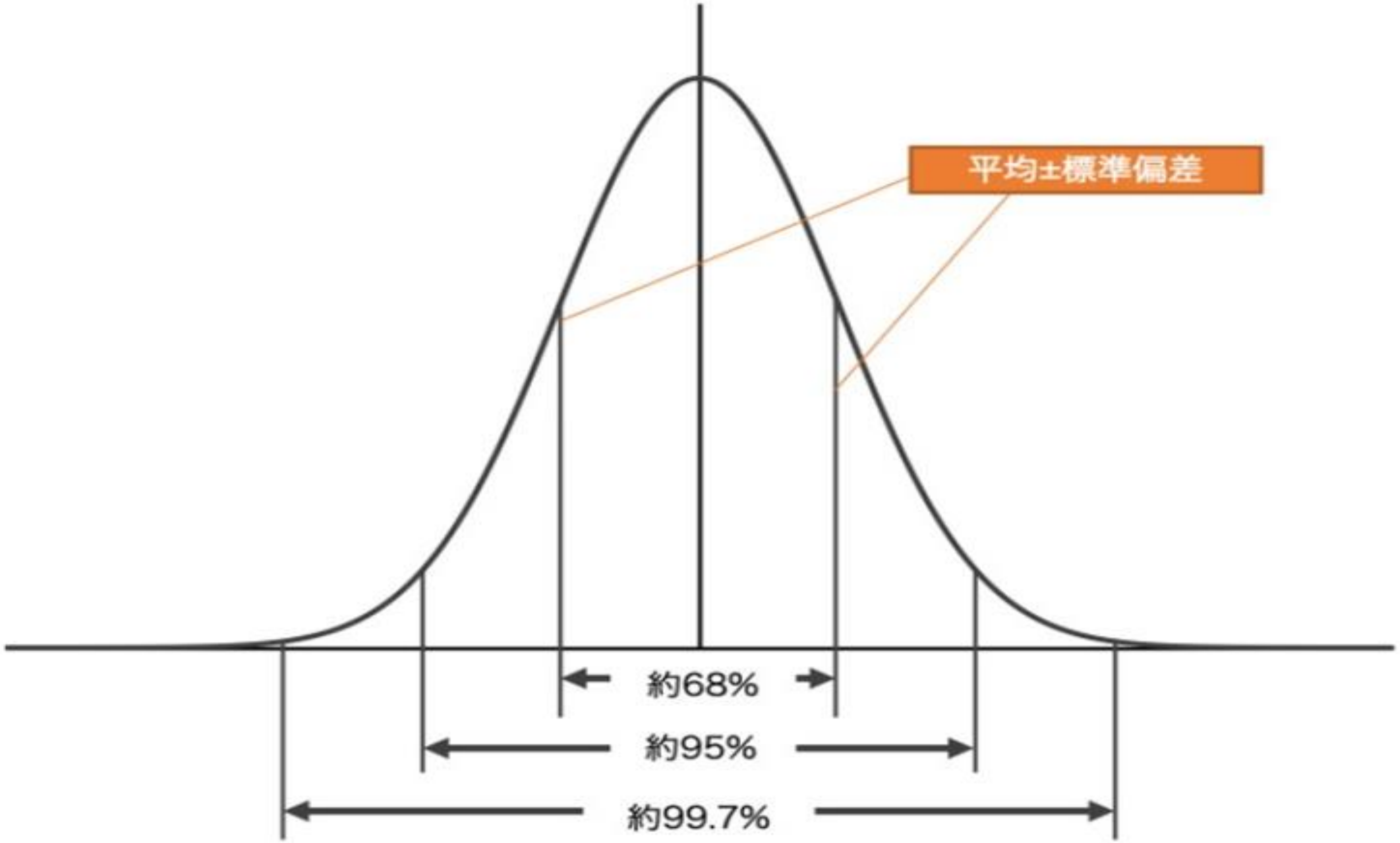
正規分布とは

多くの人が受験した試験の得点のデータを標準化すると、そのヒストグラムは次のような形になることが多い。

右のヒストグラムでは、偏差値60以上の部分にデータ全体のおよそ16%の値が、偏差値70以上の部分におよそ２%の値が含まれている。



正規分布の特徴

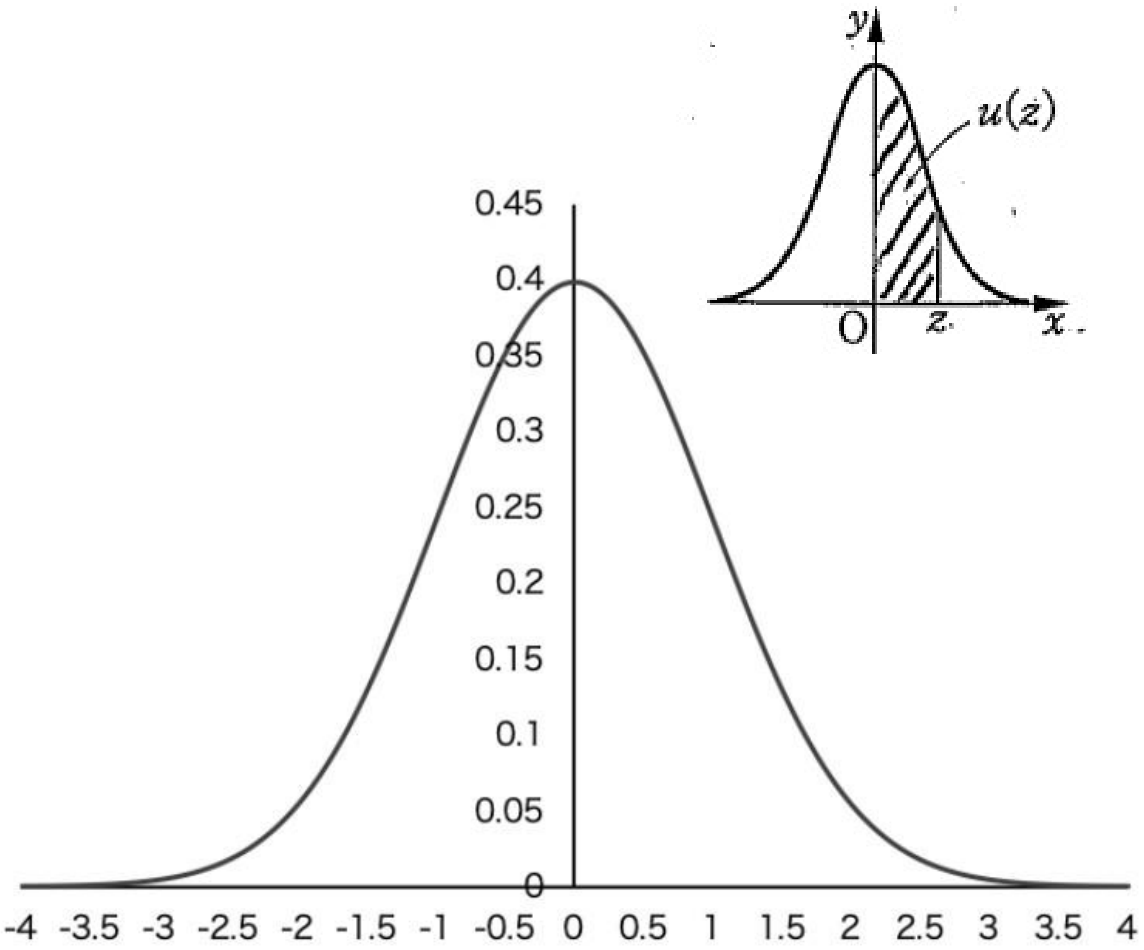


- 人間の身長
- テストの点数
- サイコロの出目の合計数
- 雨粒の大きさ
- 工業製品の規格誤差
- IQスコア
- ある月の平均気温

- 「平均±標準偏差１つ分」の範囲内に全体の約６８%のデータが含まれる。
- 「平均±標準偏差２つ分」の範囲内に全体の約９５%のデータが含まれる。
- 「平均±標準偏差３つ分」の範囲内に全体の約９９．７%のデータが含まれる。

標準正規分布とは

z



標準正規分布表

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34849	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42785	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169

- データを「平均０」「分散１」に標準化（$Z_i = \frac{x_i - \bar{x}}{S_x}$）した正規分布
- グラフ内の面積を「１」としている。

スプレッドシートで正規分布を作ってみよう

※参考 http://kanrikaikei.net/iwami/H29/04-excel4.htm

1

=STANDARDIZE(B2,B83,B84)

標準化

B83、B84
を固定

	A	B	C	D	E
1	番号	数学の得点	標準化得点	偏差値	確率密度関数
2	1	12	-2.603366534	24	0.000841611
3	2	20	-2.103320141	29	0.002730113
4	3	24	-1.853296945	31.5	0.004477058

2

=ROUND(10*C2+50,1)

偏差値＋四捨五入

小数第一位
まで表示

	A	B	C	D	E
1	番号	数学の得点	標準化得点	偏差値	確率密度関数
2	1	12	-2.603366534	24	0.000841611
3	2	20	-2.103320141	29	0.002730113
4	3	24	-1.853296945	31.5	0.004477058

3

=NORM.DIST(B2,B83,B84,FALSE)

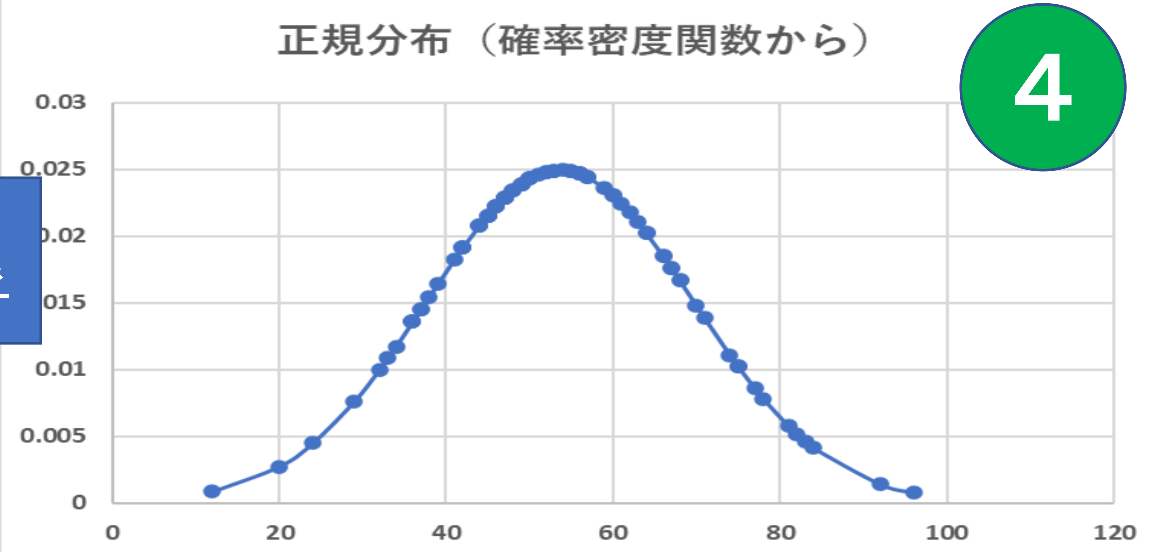
B83、B84を固定

確率密度関数
（正規分布化）

確率密度関数を
小さい順に並び替え

	A	B	C	D	E
1	番号	数学の得点	標準化得点	偏差値	確率密度関数
2	1	12	-2.603366534	24	0.000841611
3	2	20	-2.103320141	29	0.002730113

4



正規分布の演習１　分布から推定

問１

ある高校の２年生男子の身長の分布は平均167㎝、標準偏差7㎝の正規分布とみなせるという。身長が160㎝以上172㎝以下の生徒はおよそ何%いるか。

また、２年生の男子が300人であったとき、185㎝以上の生徒はおよそ何人いると考えられるか。

数学B　P82

正規分布の演習１　解説

問１

ある高校の２年生男子の身長の分布は平均167㎝、標準偏差7㎝の正規分布とみなせるという。身長が160㎝以上172㎝以下の生徒はおよそ何%いるか。
また、２年生の男子が300人であったとき、185㎝以上の生徒はおよそ何人いると考えられるか。

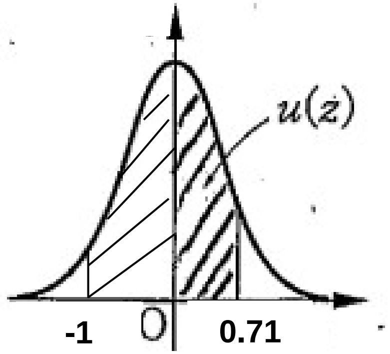
平均167、標準偏差7の正規分布に従う確率変数をXとすると、
求める割合はP(160≦X≦172)である。

$Z=\frac{X-167}{7}$ とすると、Zは平均0、標準偏差1の標準正規分布に従う。

よって、 $P(160\leqq X\leqq 172)=P\left(\frac{160-167}{7}\leqq Z\leqq\frac{172-167}{7}\right)$

$\fallingdotseq P(-1\leqq Z\leqq 0.71)=u(1)+u(0.71)=0.341+0.261=0.602$

したがっておよそ60%である。



また、$P(185\leqq Y)$とすると、$P\left(Z\geqq\frac{185-167}{7}\right)=P(Z\geqq 2.57)=0.5-u(2.57)=0.5-0.4949=0.0051$

これは0.5%より、300人×0.5%＝1.53人
よって185cm以上の男子は300人に１人～２人いると考えられる。

正規分布の演習２　分布から推定

問２

ある工場で生産されるチョコレートの重量は平均209g、標準偏差３gの正規分布に従う。重量が200g未満のものが生産される確率を求めよ。

数学B　P85

正規分布の演習３　分布から推定

問３

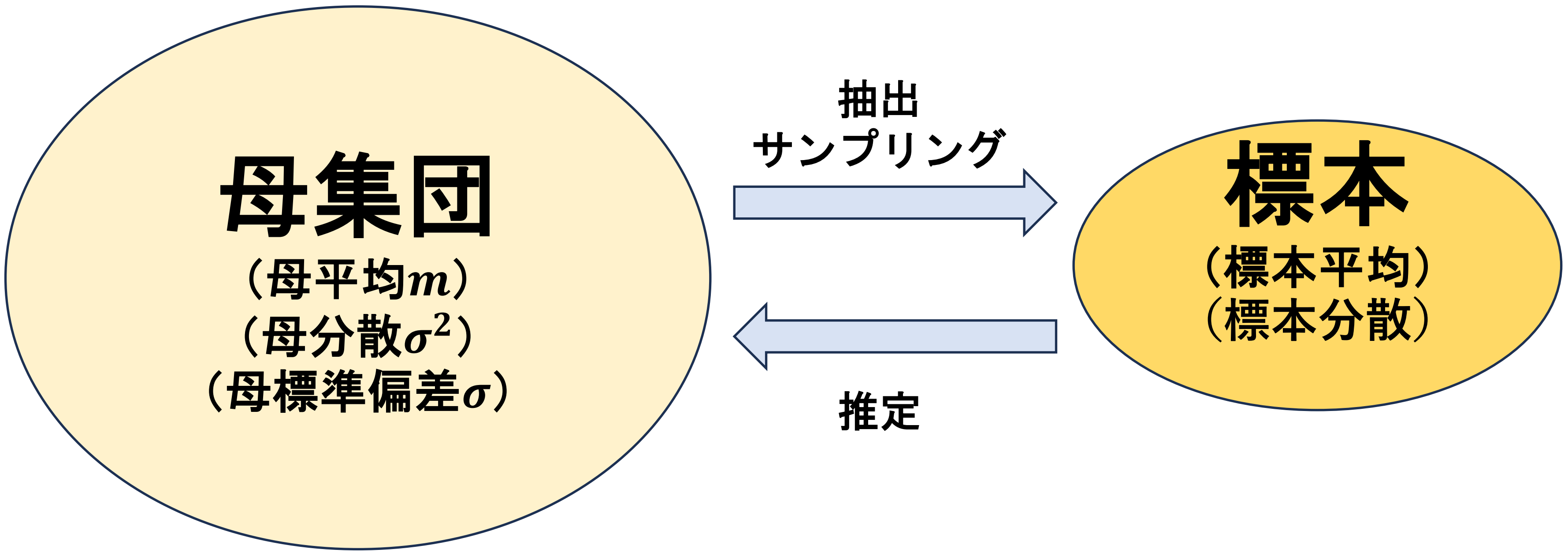
１個のさいころを360回投げるとき、１の目が55回以上でる確率はおよそ何％であると考えられるか。
（平均60、分散$5\sqrt{2}$の正規分布に従っているとします。）

数学B　P84

NO５.

１群のZ検定（母平均の検定）

標本調査　標本から母集団を推定する

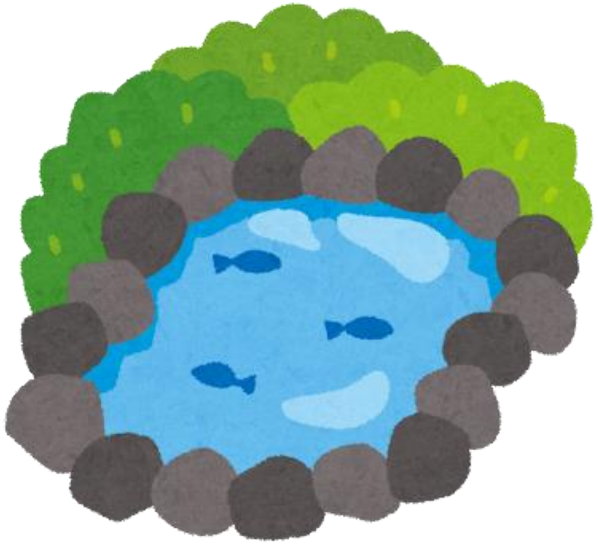


調査の目的	母集団	標本
高校生の学力調査	全ての高校生	学力調査テストを受けた高校生
視聴率	テレビを所有する全ての世帯	視聴率調査対象の一部の世帯
お客様満足度	全てのお客様	アンケートに答えた一部のお客様
内閣支持率	全ての有権者	調査対象の一部の有権者

ウォーミングアップ１　標本調査

ある池で魚の数を推定するために、100匹の魚をつかまえて、目印をつけて池に戻した。そして、１週間後に再び魚を50匹つかまえたところ、目印のついた魚が８匹含まれていた。

この池には、およそ何匹の魚がいると推定できるか。



ウォーミングアップ１　標本調査解答

ある池で魚の数を推定するために、100匹の魚をつかまえて、目印をつけて池に戻した。そして、１週間後に再び魚を50匹つかまえたところ、目印のついた魚が8匹含まれていた。この池には、およそ何匹の魚がいると推定できるか。

標本から　目印なし：目印あり＝42：8

全体の目印なしをｘとすると　　ｘ：100＝42：8

ｘ＝525

全体　　525＋100＝625

母平均の検定（Z検定）とは？

令和４年の全国の高校１年生男子の身長について平均身長は169.8、標準偏差は6であった。

〇全国の各クラスで調査をすると

標本平均たち

クラス	クラスの平均（標本平均$\bar{X}$）
クラス１	170.2
クラス２	168.5
クラス３	172.1
クラス４	167.2
クラス５	169.3
クラス６	169.7
クラス７	170.4
クラス８	171.2
クラス９	171.4
クラス１０	170.5
・・・	
クラス１００	170.6

仙台〇高のとあるクラス40人で調査をしてみると、１年△組40人の平均身長は175㎝であった

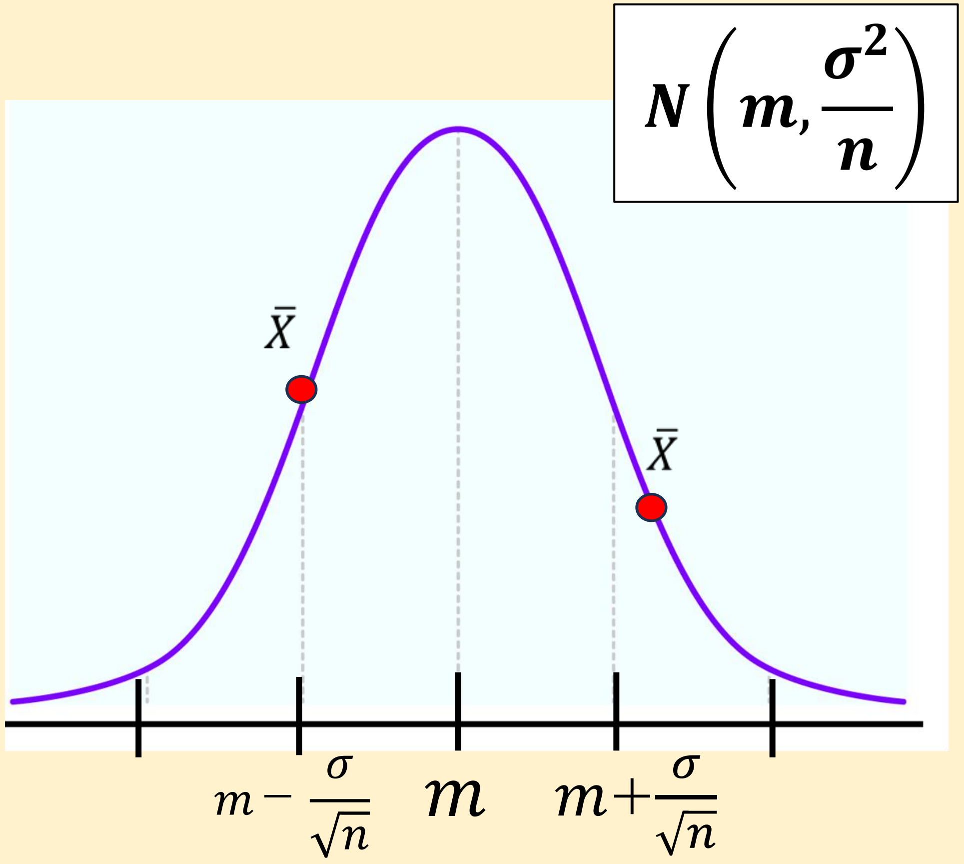
・そんなことはあり得る？
・どういうこと？

標本平均$\bar{X}$は母平均169.8に近いはず

標本平均の分布のルール

標本平均$\bar{X}$の分布

母平均m、母分散σ^2の母集団から無作為に抽出された、大きさnの標本平均$\bar{X}$の分布は、nが大きければ、平均m、標準偏差$\frac{\sigma}{\sqrt{n}}$の正規分布とみなすことができる



標本平均$\bar{X}$の平均はm

標本平均$\bar{X}$の標準偏差は$\frac{\sigma}{\sqrt{n}}$

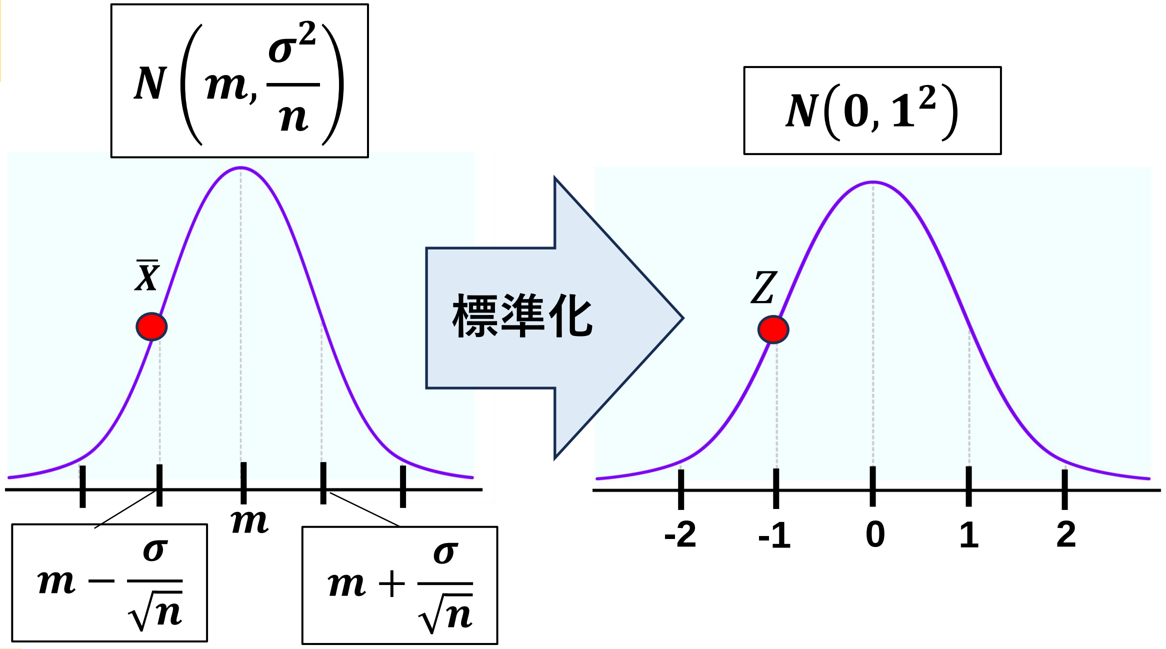
Z検定の手順

1. 帰無仮説、対立仮説、有意水準の設定

2. 検定の前提の確認

- 母平均：m
- 母標準偏差：σ
- 標本平均：$\bar{X}$
- 標本の大きさ：n

nが十分大きいとき、標本平均$\bar{X}$の分布は正規分布$N\left(m, \frac{\sigma^2}{n}\right)$の分布と見なすことができる



3. 検定統計量（Z値）、P値の算出

4. 有意差の判定、帰無仮説の棄却の有無

※P値
帰無仮説が成立するという仮定のもとで
今回の標本観察の結果が起こりうる確率

母平均の検定（Z検定）

（標本と母集団の関係を仮説検定を用いて考えて見よう）

例題

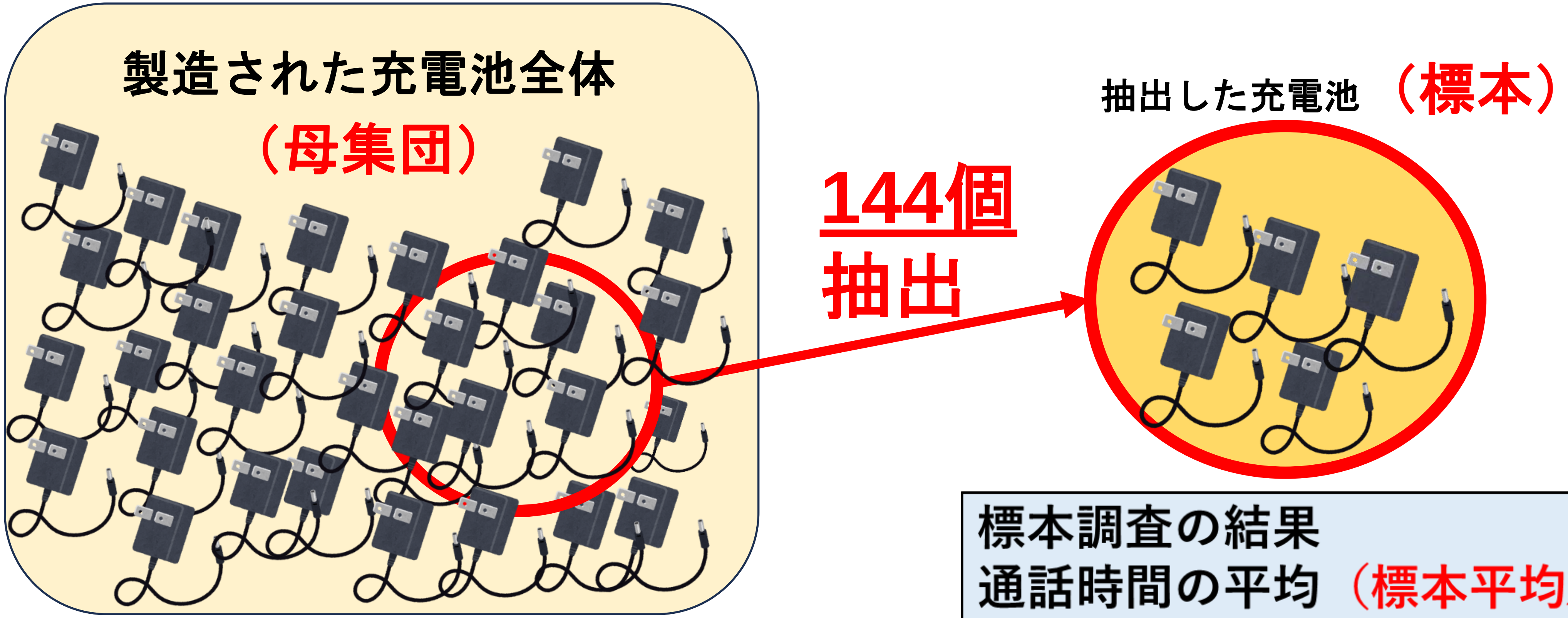
あるスマートフォンは、最大に充電した状態から連続で平均500時間の通話が可能であるという。今回、製造工程の見直しによって充電池の品質が向上し、通話時間が伸びたことが予想された。

そこで、この新しい充電池から144個を無作為に抽出して通話可能な時間を測定したところ、その144個(標本)における平均は510時間であった。

母標準偏差が60時間であるとき、この標本調査の結果から、今回の製造工程の見直しによって、通話可能な時間が伸びたと判断して良いか。

平均で500時間連続通話可能なスマホAがある

B社ではスマホAの連続通話時間を伸ばすために充電池の品質を改良した



通話時間のバラツキ（母標準偏差σ）
は60時間であった。

標本調査の結果
通話時間の平均（標本平均$\bar{X}$）
は510 時間であった。

連続通話時間の長さが伸びたと判断して良いか

母平均の検定（Z検定）手順１

「連続通話時間の長さが伸びたと判断して良いか」を検定する

製造過程見直し後の連続通話可能時間の平均（母平均）を「m」とする

○連続通話可能な時間は伸びた：$m > 500$　対立仮説

○連続通話可能な時間は変わらない：$m = 500$　帰無仮説

統計的に検証

「$P(\bar{X} \geqq 510)$」

標本平均$\bar{X}$が母平均mより10以上大きくなることは、まれなことなのか、確率で確かめる

「$P(\bar{X} - m \geqq 10)$」を計算し、５％以下になるかどうかを確かめる

母平均の検定（Z検定）手順２

母平均：$m=500$と仮定

母標準偏差:　$\sigma=60$

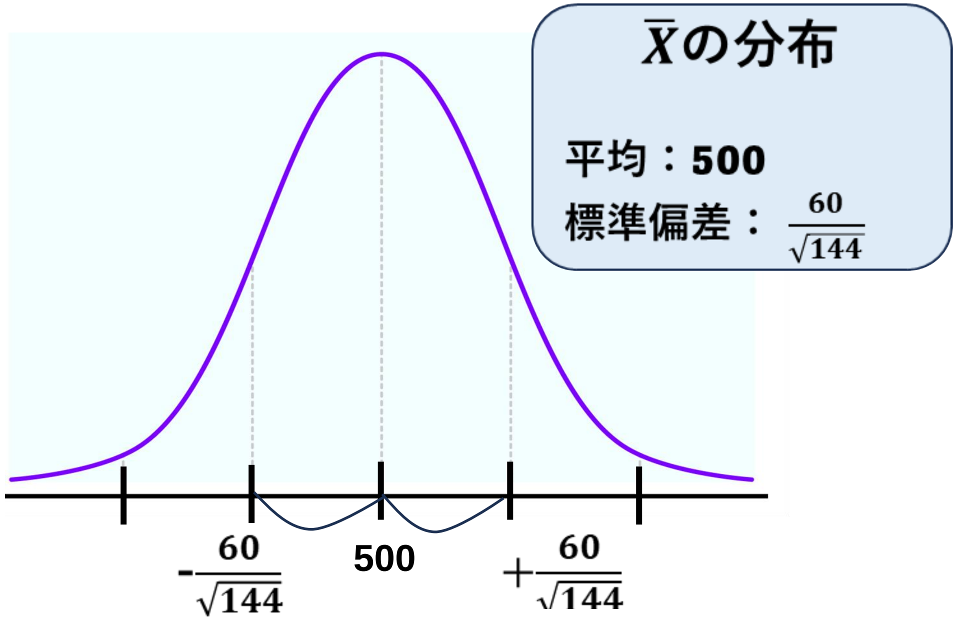
標本の大きさ(サンプルの数):　$\mathrm{n}=144$

標本平均$\bar{X}$の分布は正規分布$N\left(500,\frac{60^2}{144}\right)$とみなすことができる。

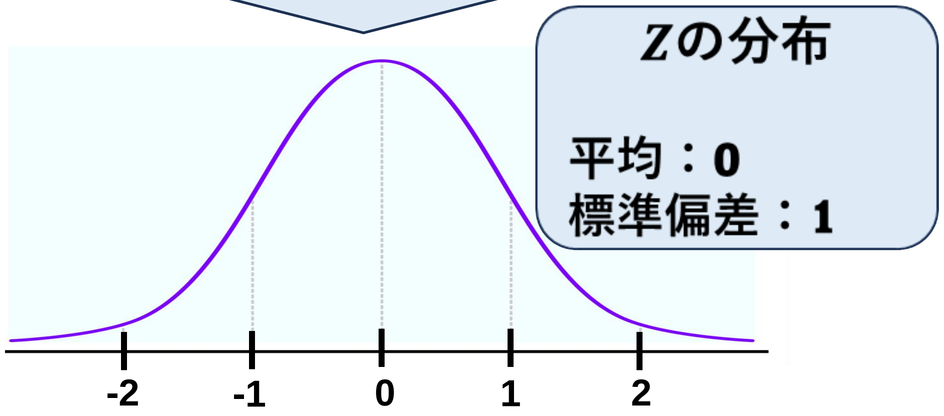
$\bar{X}$を標準化（Zへ変換）

$$Z=\frac{\bar{X}-500}{\frac{60}{\sqrt{144}}}$$

Zの分布は標準正規分布$N(0.1^2)$とみなすことができる。



標準化



母平均の検定（Z検定）手順３・４

$$P(\bar{X}-m \geqq 10) = P\left(\frac{\bar{X}-500}{\frac{60}{\sqrt{144}}} \geqq \frac{10}{\frac{60}{\sqrt{144}}}\right)$$

Z

標本平均が母平均より10以上大きくなる確率

$\dfrac{510-500}{\frac{60}{\sqrt{144}}}=2$ を計算していることと同じ

$$= P(Z \geqq 2) = P(Z \geqq 0) - P(0 \leqq Z \leqq 2)$$

Z値

$$= 0.5 - u(2) = 0.5 - 0.47725 = 0.02275 < 0.05$$

P値0.02275が有意水準５％より小さい

よって、$m=500$という仮定のもとでは、標本平均$\bar{X}$がmより10以上大きくなる確率はおよそ2.3%であり、ほとんど起こりえない

ほとんど起こりえないことが起こったということで、$m=500$という仮定は正しくないと考える

帰無仮説の棄却

すなわち、$m>500$。連続通話可能な時間は伸びたと考える

対立仮説の採択

演習問題（片側検定）

問１

ある全国規模の学力テストの平均点は45.1点で、標準偏差は10点であった。ある県の受験者の中から400人を無作為抽出したところ、その平均点は44.2点であった。この県の受験者の平均点は全国平均より低いと判断できるか。有意水準5%で検定せよ。
また、抽出人数を225人とするとどうか。

数学ⅡB　ハイプライムP161

演習問題（片側検定）

問２

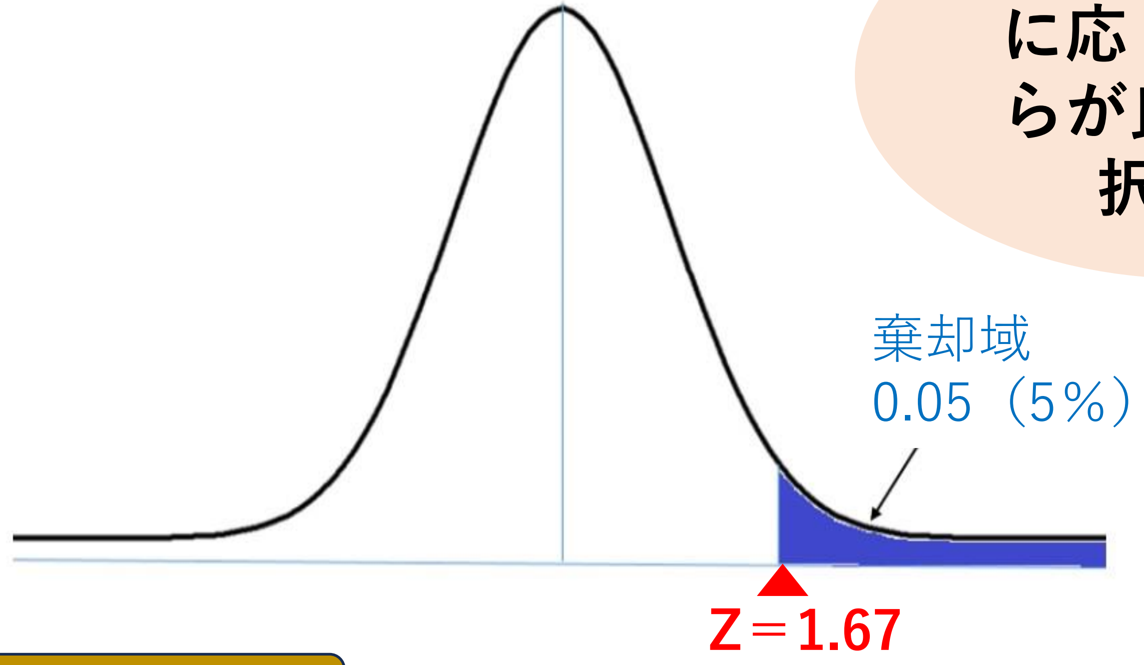
ある種類のねずみは、生まれてから３か月後の体重が平均65g、標準偏差4.8gの正規分布に従うという。今、この種類のねずみ10匹を生まれてから特別な飼料で飼育し、３か月後の体重を測定したところ、次のような結果を得た。この飼料はねずみの体重の増加に与えたと判断できるか。有意水準５％で検定せよ。

６７、７１、６３、７４、６８、６１、６４、８０、７１、７３
（単位はすべてg）

ハイプラP162　旭川医大

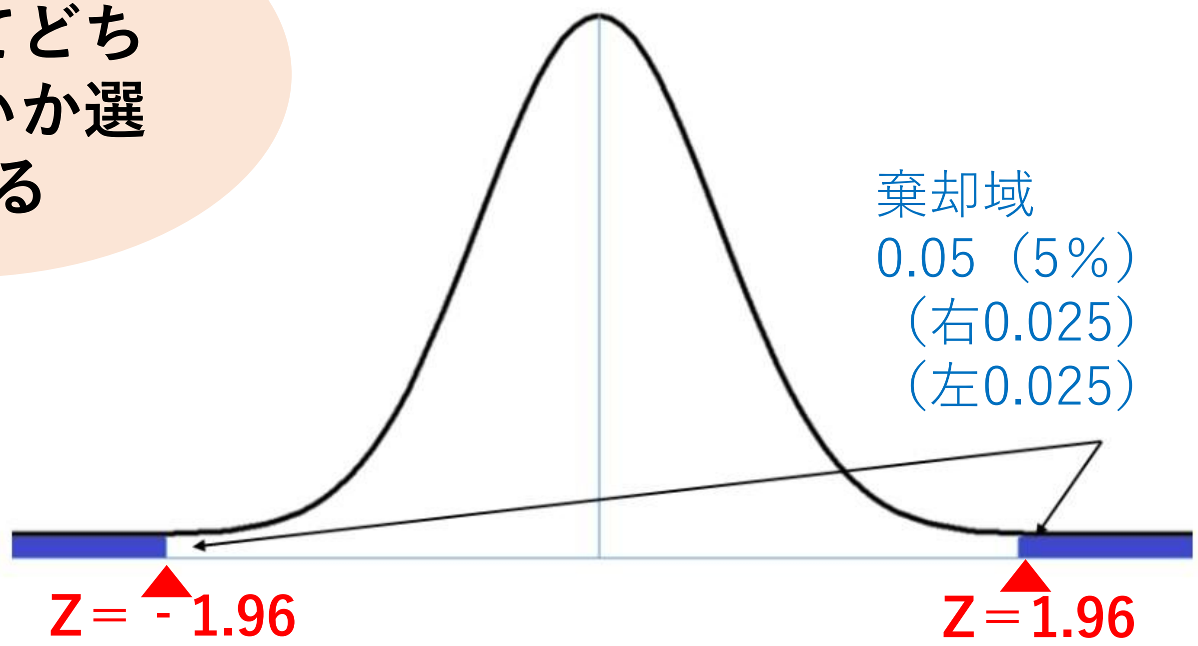
片側検定と両側検定

片側検定



示したい目的に応じてどちらが良いか選択する

両側検定



片側検定の例

対立仮説：薬A中の成分Bの量は100mgより多い
（母平均m＞100）

帰無仮説：薬A中の成分Bの量は100mgである
（母平均m＝100）

両側検定の例

対立仮説：薬A中の成分Bの量は100mgではない
（母平均m≠100）

帰無仮説：薬A中の成分Bの量は100mgである
（母平均m＝100）

演習問題（両側検定にチャレンジ）

問３

ある種類のねずみは、生まれてから３か月後の体重が平均65g、標準偏差4.8gの正規分布に従うという。今、この種類のねずみ10匹を生まれてから特別な飼料で飼育し、３か月後の体重を測定したところ、次のような結果を得た。この飼料はねずみの体重に異常な変化を与えたと判断できるか。有意水準５％で両側検定せよ。

６７、７１、６３、７４、６８、６１、６４、８０、７１、７３
（単位はすべてg）

ハイプラP162　旭川医大

演習問題（両側検定にチャレンジ）

問４

△△会社の食品Aにはある成分Pが20㎎含まれているとパッケージに記載されている。この食品64個を無作為抽出して標本調査をしたところ、成分Pは平均19.2㎎含まれていることが分かった。
△△会社で製造されている食品A全体における成分Pの含まれている量の標準偏差が4㎎であるとき、標本調査の結果から、パッケージの20㎎の記載は誤りであると判断できるか。有意水準５％で両側検定せよ。

ハイプラP162

演習問題（両側検定にチャレンジ）

問５

A社の製品100個を無作為抽出し、1個あたりの重さについて検査した。その結果、１個の重さの平均は表示されている値より0.9ｇ小さく、標本の標準偏差は3.45ｇであった。この製品全体における1個あたりの重さは表示されている値と異なると判断できるかどうか、標本の標準偏差を母標準偏差と見なし、有意水準５％で両側検定せよ。

数学ⅡＢ　教科書　P104

問５ 解答

A社の製品100個を無作為抽出し、1個あたりの重さについて検査した。その結果、１個の重さの平均は表示されている値より0.9ｇ小さく、標本の標準偏差は3.45ｇであった。この製品全体における1個あたりの重さは表示されている値と異なると判断できるかどうか、標本の標準偏差を母標準偏差と見なし、有意水準５%で両側検定せよ。

帰無仮説：１個あたりの重さは表示されている値の通りである

対立仮説：１個あたりの重さは表示されている値と異なる

表示されている重さの値をmとする。

標本の大きさは$n = 100$、母標準偏差は標本の標準偏差$\sigma = 3.45$を用いると、標本平均$\overline{X}$の分布は$N\left(m, \frac{3.45^2}{100}\right)$と見なせるから、$\overline{X}$を標準化した$Z = \frac{\overline{X}-m}{\frac{3.45}{\sqrt{100}}}$の分布は標準正規分布$N(0,1)$と見なせる。

問５ 解答

A社の製品100個を無作為抽出し、1個あたりの重さについて検査した。その結果、１個の重さの平均は表示されている値より0.9ｇ小さく、標本の標準偏差は3.45ｇであった。この製品全体における1個あたりの重さは表示されている値と異なると判断できるかどうか、標本の標準偏差を母標準偏差と見なし、有意水準５%で両側検定せよ。

$$P(|\bar{X}-m| \geqq 0.9) = P\left(\frac{|\bar{X}-m|}{\frac{3.45}{\sqrt{100}}} \geqq \frac{0.9}{\frac{3.45}{\sqrt{100}}}\right)$$

標本平均と母平均の差が0.9以上になる確率

$$= P\left(|Z| \geqq \frac{0.9}{\frac{3.45}{10}}\right) \fallingdotseq P(|Z| \geqq 2.61)$$

$$= 2\left(0.5 - u(2.61)\right) = 2(0.5 - 0.49574)$$

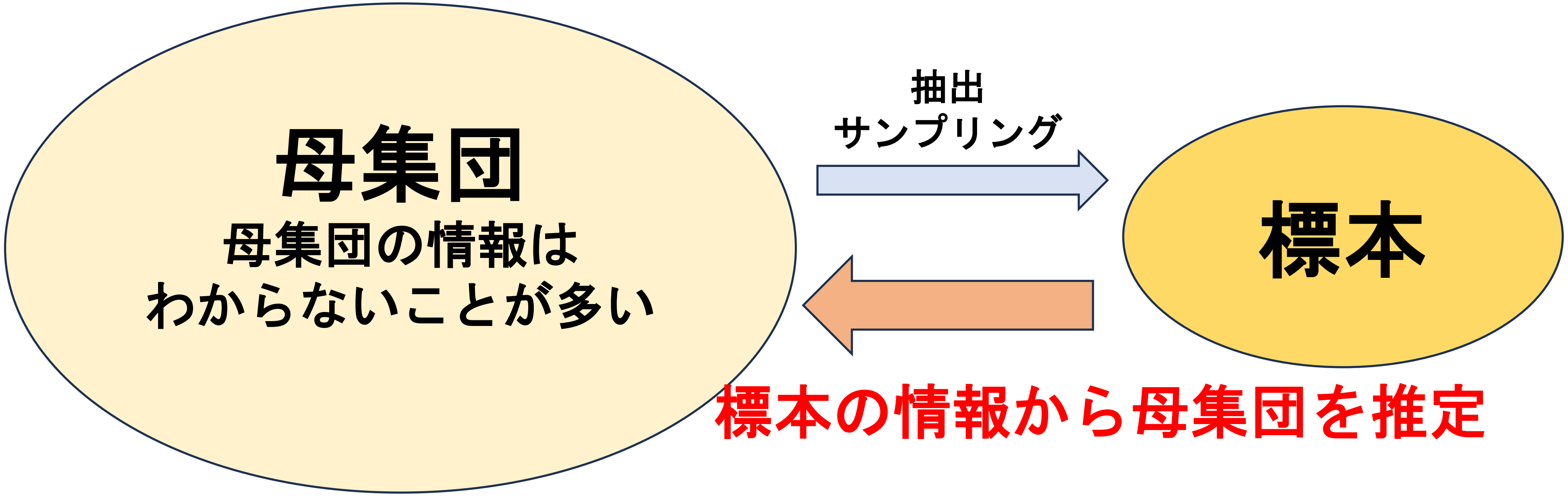
$$= 0.00906 < 0.05$$

有意水準５%より値が小さくなることから、帰無仮説は棄却され、対立仮説を採択する。すなわち、A社の製品全体の１個あたりの重さは表示されている値と異なると判断できる

NO６.

１群のｔ検定
（母平均の検定）

t 検定も標本調査　標本から母集団を推定する



調査の目的	母集団	標本
工場の製品管理	工場の製品すべて	工場の製品の一部
視聴率	テレビを所有する全ての世帯	視聴率調査対象の一部の世帯
お客様満足度	全てのお客様	アンケートに答えた一部のお客様
内閣支持率	全ての有権者	調査対象の一部の有権者

t 検定とは

今回は①をやってみよう（Z検定とほぼ同じ）

①母集団の平均値と標本の平均値が異なるかどうか
（1標本のt検定）
例：全国展開しているラーメンチェーン店の中で
　　A店のラーメンの量が基準に比べて多いかどうかを推定

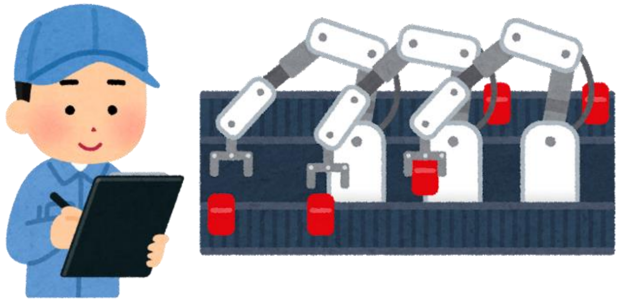
②2つのグループの平均値が互いに異なるかどうか
（独立した2標本のt検定）
例：生徒の数学の平均値が１組と２組で異なるかどうかを推定

③対応のある測定値が有意に異なるかどうか
（対応のある、または関連のある標本のt検定）
例：４０人の生徒の数学Ｉと数学Aの平均値が異なるかどうかを推定

t 検定①の使用例

◎工場の品質管理

サンプルから工業製品の改善が本当に効果的だったか、または品質が向上しているかを t 検定を用いて点検する



◎新薬の効果判定

治験で得られるデータから、t検定で新薬の効果を判定する。

◎アンケート分析（マーケティング）

顧客の生の声を捉え、業務改善ができているかどうかをアンケートデータから t 検定で分析する

t 検定の手順と、Z検定との違い

手順

- 帰無仮説・対立仮説、有意水準を決める
- 検定の前提を確認する
- 検定統計量（t 値）を算出する
- 有意差の判定し、帰無仮説の棄却の有無を決める

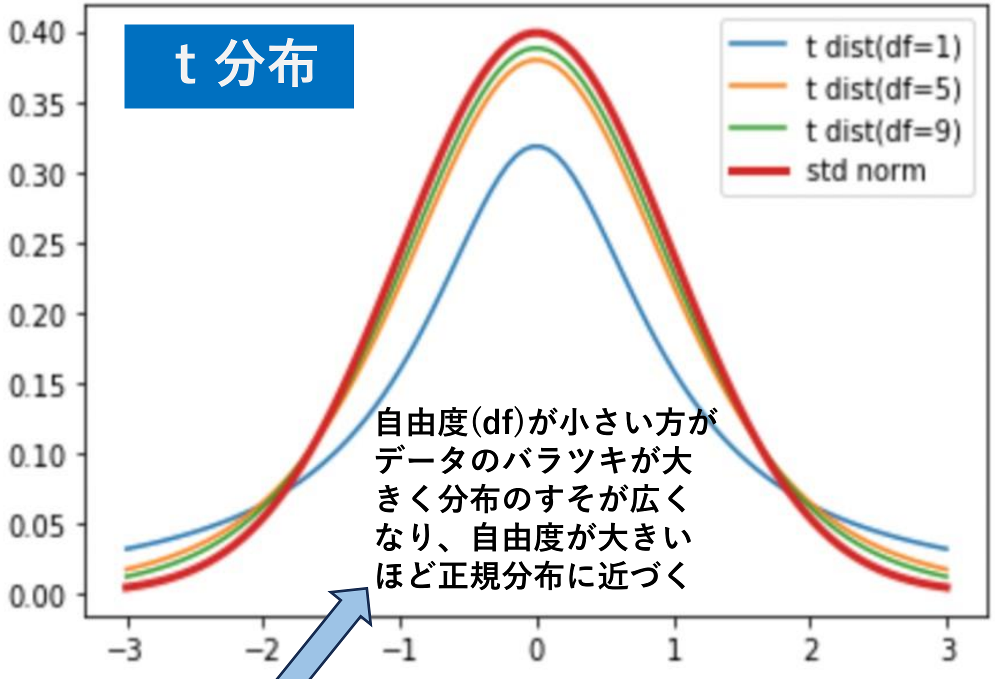
→ **t 検定の検定手順はZ検定の検定手順と同じ**

違い

Z検定	t 検定
○母標準偏差が分かっている	○母標準偏差がわかっていない → 標本の標準偏差から母標準偏差を推定算出する
○標本のサンプルサイズが十分大きい	○標本のサンプルサイズが小さい →「自由度」「不偏分散」で調整する
○ｚ値を正規分布を用いて検定する	○ｔ値をｔ分布を用いて検定する

※Z検定とｔ検定の違いは、①母集団の情報が分かっているか分かっていないか
②標本のサンプルサイズが多いか少ないか

t 分布とは



$$t値 = \frac{\bar{X} - m}{\sqrt{\frac{U}{n}}}$$

t分布の横軸の値。$\bar{X}$をt値に変換して、t分布で検定する

※不偏分散：U

$$U = \frac{n}{n-1} \times (標本分散)$$

nが小さいとき、標本分散を母分散に近づけるために補正した分散

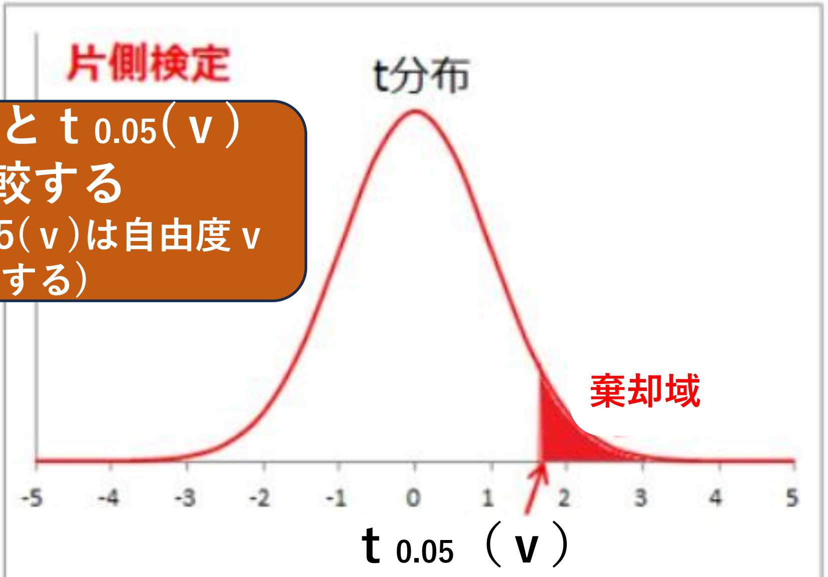
バラツキ補正のためのもの

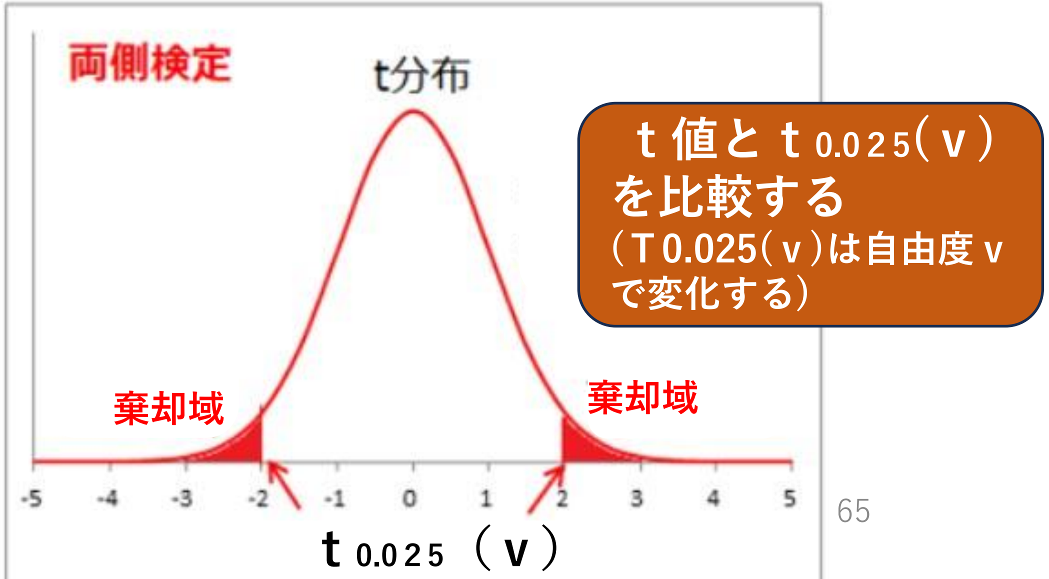
※自由度(df)：$n-1$

標本の大きさ（サンプルサイズ）から１を引いた数

※中心極限定理

母集団から無作為に標本を抽出したときの標本平均の分布は、標本の大きさ(サンプルサイズ)が大きくなるにつれて正規分布に近づく





米国データサイエンティストのブログ
https://datawokagaku.com/t_dist/

統計分析研究所　株式会社アイスタット
https://istat.co.jp/sk_commentary/t-test/

t 分布表

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
1	6.314	12.706
2	2.920	4.303
3	2.353	3.182
4	2.132	2.777
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.813	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
15	1.753	2.132
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
29	1.699	2.045
30	1.697	2.042
31	1.696	2.040
32	1.694	2.037
33	1.692	2.035
34	1.691	2.032
35	1.690	2.030
36	1.688	2.028
37	1.687	2.026
38	1.686	2.024
39	1.685	2.023
40	1.684	2.021
41	1.683	2.020
42	1.682	2.018

例題

お菓子工場で製造しているチョコレート菓子は容量を200ｇとしている。１０個購入して重さを図ってみると次の重さであった。

２０５・１９８・１９７・２０８・２０４
２０２・２０７・１９９・２０７・２０３（単位はg）

この工場で作られているチョコレート菓子の容量は２００ｇより多いといえるか

例題

お菓子工場で製造しているチョコレート菓子は容量を200ｇとしている。１０個購入して重さを図ってみると次の重さであった。

２０５・１９８・１９７・２０８・２０４
２０２・２０７・１９９・２０７・２０３（単位はg）

この工場で作られているチョコレート菓子の容量は２００ｇより多いといえるか

帰無仮説：この工場で作られているチョコレート菓子の容量は２００ｇである
対立仮説：この工場で作られているチョコレート菓子の容量は２００ｇより多い

No.	1	2	3	4	5	6	7	8	9	10	平均	分散
g	205	198	197	208	204	202	207	199	207	203	203	14

自由度：10－1＝9

不偏分散：$\frac{10}{9} \times 14 = 15.56$

t 値：$t = \frac{203-200}{\sqrt{\frac{15.56}{10}}} = 2.41$

めったに起きないことが起きている

→　t 分布表から片側検定　2.41＞ t 0.05(9)＝1.833　より　帰無仮説を棄却し対立仮説を採択
→　結論「この工場で作られているチョコレート菓子の容量は200ｇより多い

演習問題

https://bellcurve.jp/statistics/course/10004.html
統計webより

問１

あるメーカーのビール大瓶14本をサンプリングし、その平均が633mlよりも少ないかどうか検定したい。測定したビール14本の容量が次の表の通りである場合、検定の結果はどのようになるか答えよ。なお、有意水準はとし、ビール大瓶の容量は正規分布に従うものとする。

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
容量 [ml]	633.9	628.1	630.2	634.3	635.1	633	633.6	629.5	632.4	633.6	623.5	632	634.1	626.7

演習問題

https://bellcurve.jp/statistics/course/10004.html
統計webより

問１

あるメーカーのビール大瓶14本をサンプリングし、その平均が633mlよりも少ないかどうか検定したい。測定したビール14本の容量が次の表の通りである場合、検定の結果はどのようになるか答えよ。なお、有意水準はとし、ビール大瓶の容量は正規分布に従うものとする。

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	平均	分散	不偏分散	t 値
容量[ml]	633.9	628.1	630.2	634.3	635.1	633	633.6	629.5	632.4	633.6	623.5	632	634.1	626.7	631.4286	10.69061	11.51297	-1.732

→　t 分布表から片側検定　1.732＜ t 0.05(13)＝1.771　より　帰無仮説を棄却できない。
→　「平均が633mlより少ない」とはいえない

演習問題

問２

ある種類のねずみは、生まれてから３か月後の体重が平均65gであり正規分布に従うという。今、この種類のねずみ10匹を生まれてから特別な飼料で飼育し、３か月後の体重を測定したところ、次のような結果を得た。この飼料はねずみの体重の増加に影響を与えたといえるか

６７、７１、６３、７４、６８、６１、６４、８０、７１、７３ （単位はすべてg）

ハイプラP162　旭川医大　改

NO７.

ｔ検定実習
（実習の手引き）

t 検定実習の手順

①　テーマを決める
②　母平均を決める
③　標本調査内容（項目）を決める
④　調査フォームを作成する
⑤　リンク先をクラスルームにコピーする
⑥　調査に回答する
⑦　スプレッドシートで分析する
⑧　スライドにまとめる
⑨　発表する

ｔ検定実習の手順

①　テーマを決める

・このクラスのこの傾向はまれなのではないか
・この現象は一般的な現象と比べ有意差が生まれるのではないか
（例）三高１年生の１週間の平均学習時間は2時間より多い

②　母平均を決める

・一般的にはこのくらいなはずである
・平均はこのようになると想定できるのだが
・実際の容量は〇〇〇と記載されているのだが
（例）一般的な高校１年生の１週間の平均学習時間は２時間と想定できる

③　標本調査内容（項目）を決める

（例）調査内容（項目）
「あなたの１週間の平均学習時間は（　　　　）時間である」

t 検定実習の手順

④　調査フォームを作成する

別スライドを参照
「発表（アンケート）について」

⑤　リンク先を クラスルームに コピーする

⑥　調査に回答する

全部の班に回答する

ｔ検定実習の手順

⑦　スプレッドシートで分析する

各班で自由に分析

⑧　スライドにまとめる

【ｔ検定テーマ】
三高生はスマホの使用時間を１日１時間までと約束されている。しかし、そんな約束はまったく守れていないのではないか

・母平均m：三高生の生徒のスマホの平均使用時間
・帰無仮説H_0：m＝１
・対立仮説H_1：m＞1
・検定方法：片側検定
・有意水準：5%
・自由度：33

・標本平均：
　1－1生徒のスマホの平均使用時間
　⇒5時間57分
・標本分散：51566
・不偏分散：53128
・t値（検定統計量）：7.52
・t境界値片側（ｔ0.05(33)）：1.692

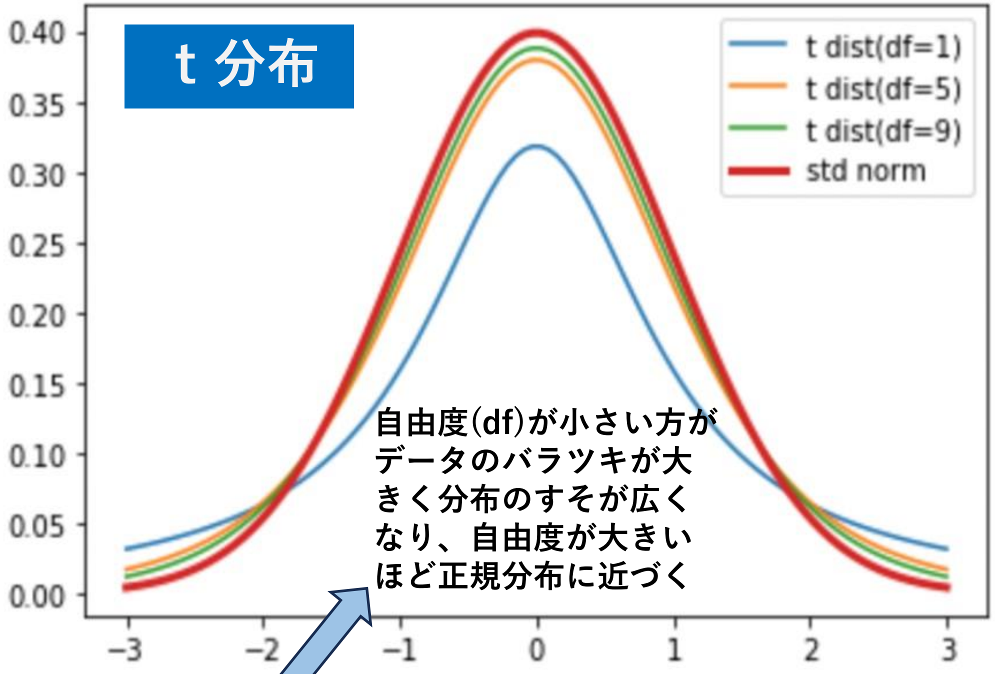
t値＞1.692となり、帰無仮説H_0（m＝１）を棄却し、対立仮説H_1（m＞1）を採択する。
このことより、三高生のスマホの平均使用時間は１時間をはるかに超えていると考えられるからから、三高生のスマホの使用時間はまったく約束を守れていない

⑨　発表する

テーマのネタ（例）

- スマホの利用時間
- 学習時間
- 睡眠時間
- 起床時間
- ゲームの時間
- ラインのスタンプの数
- 風呂の時間
- 通学時間
- メロンパンの重さ
- １つのレーズンパンのレーズンの数

t 分布



$$t値 = \frac{\bar{X}-m}{\sqrt{\frac{U}{n}}}$$

t分布の横軸の値。$\bar{X}$をt値に変換して、t分布で検定する

※不偏分散：U

$$U = \frac{n}{n-1} \times (標本分散)$$

nが小さいとき、標本分散を母分散に近づけるために補正した分散

バラツキ補正のためのもの

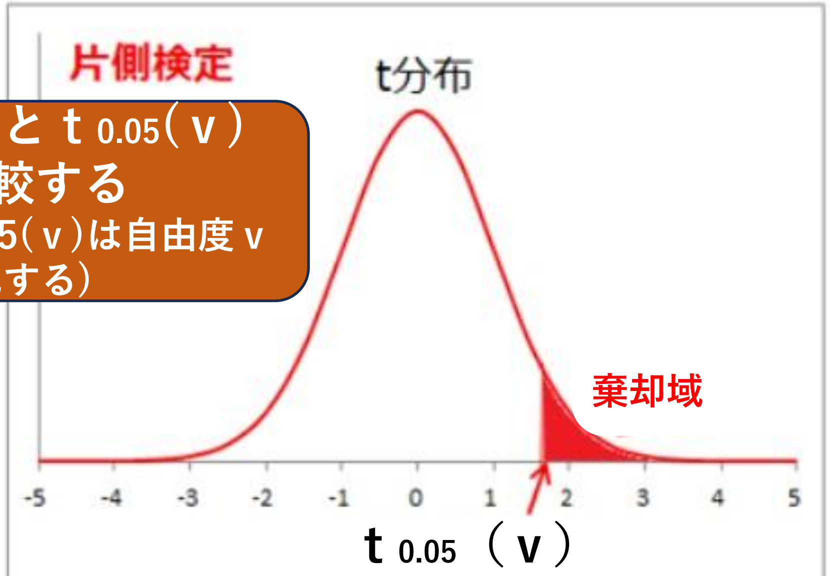
※自由度(df)：$n-1$

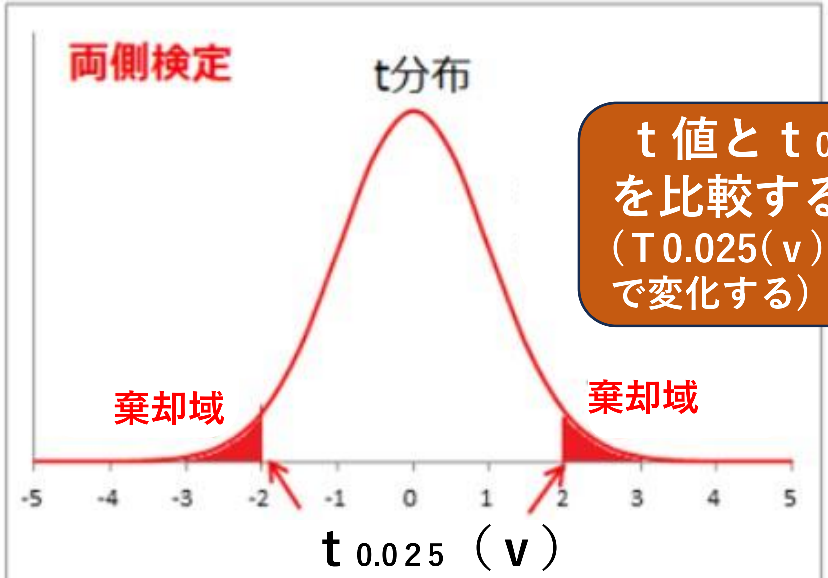
標本の大きさ（サンプルサイズ）から１を引いた数

※中心極限定理

母集団から無作為に標本を抽出したときの標本平均の分布は、標本の大きさ(サンプルサイズ)が大きくなるにつれて正規分布に近づく

ｔ値とｔ0.05(ｖ)を比較する
(Ｔ0.05(ｖ)は自由度ｖで変化する)





ｔ値とｔ0.025(ｖ)を比較する
(Ｔ0.025(ｖ)は自由度ｖで変化する)

米国データサイエンティストのブログ
https://datawokagaku.com/t_dist/

統計分析研究所　株式会社アイスタット
https://istat.co.jp/sk_commentary/t-test/

t 分布表

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
1	6.314	12.706
2	2.920	4.303
3	2.353	3.182
4	2.132	2.777
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.813	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
15	1.753	2.132
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
29	1.699	2.045
30	1.697	2.042
31	1.696	2.040
32	1.694	2.037
33	1.692	2.035
34	1.691	2.032
35	1.690	2.030
36	1.688	2.028
37	1.687	2.026
38	1.686	2.024
39	1.685	2.023
40	1.684	2.021
41	1.683	2.020
42	1.682	2.018

例題

お菓子工場で製造しているチョコレート菓子は容量を200ｇとしている。１０個購入して重さを図ってみると次の重さであった。

２０５・１９８・１９７・２０８・２０４
２０２・２０７・１９９・２０７・２０３（単位はg）

この工場で作られているチョコレート菓子の容量は２００ｇより多いといえるか

帰無仮説：この工場で作られているチョコレート菓子の容量は２００ｇである
対立仮説：この工場で作られているチョコレート菓子の容量は２００ｇより多い

No.	1	2	3	4	5	6	7	8	9	10	平均	分散
g	205	198	197	208	204	202	207	199	207	203	203	14

自由度：10－1＝9　　不偏分散：$\frac{10}{9} \times 14 = 15.56$

t 値：$t = \dfrac{203-200}{\sqrt{\dfrac{15.56}{10}}} = 2.41$

めったに起きないことが起きている

→　ｔ分布表から片側検定　2.41＞ｔ0.05(9)＝1.833　より　帰無仮説を棄却し対立仮説を採択
→　結論「この工場で作られているチョコレート菓子の容量は200ｇより多い

NO８.

二項分布から
正規分布へ

正規分布

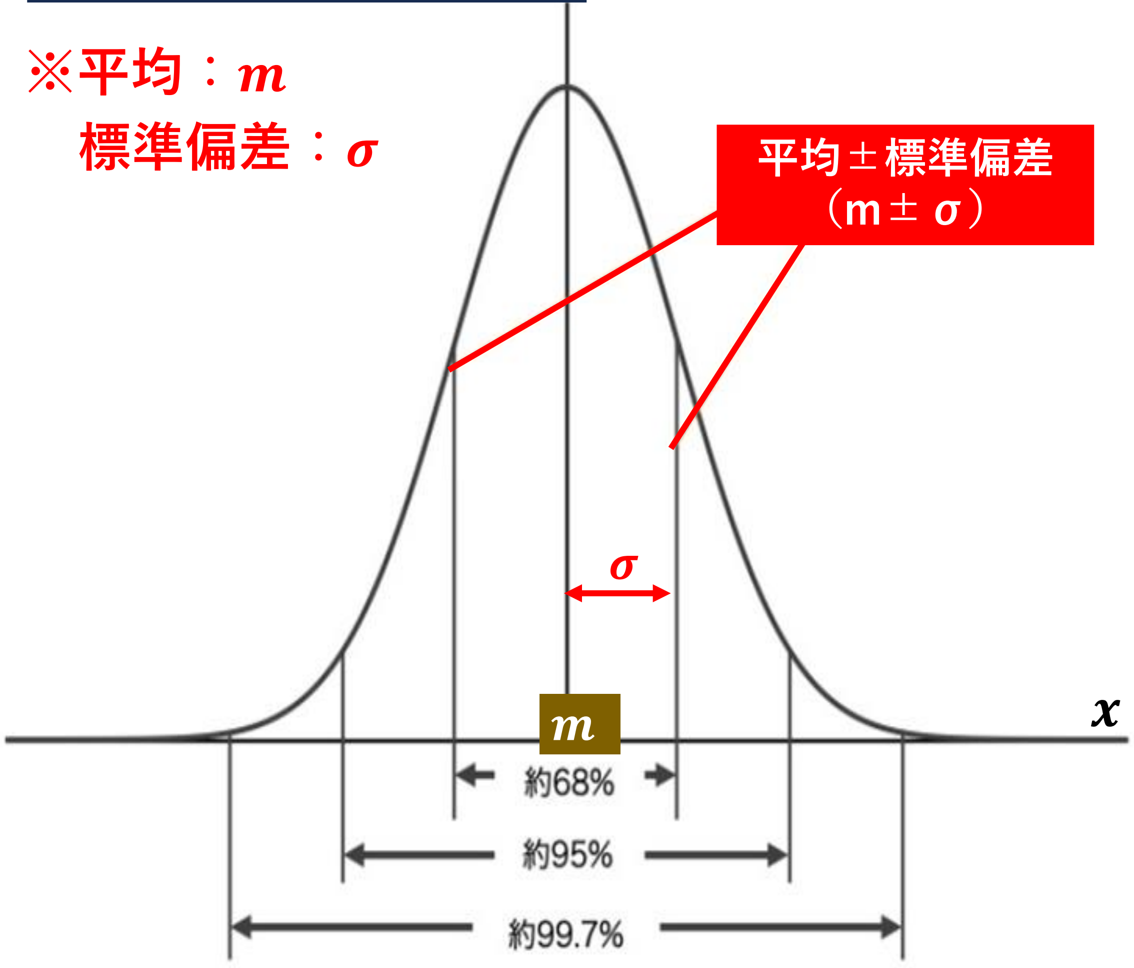
正規分布
$N(m\,,\sigma^2)$

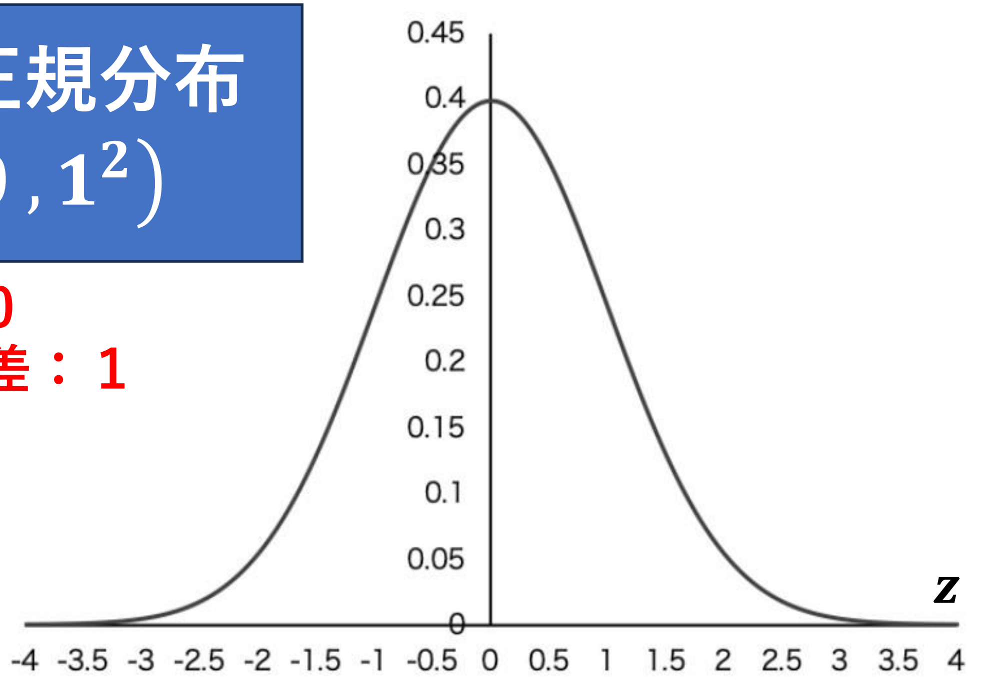
$z=\dfrac{x-m}{\sigma}$ で
標準化変換

標準正規分布
$N(0\,,1^2)$

※平均：0
標準偏差：1

※平均：m
標準偏差：σ





標準正規分布表

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34849	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42785	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327

正規分布演習１（分布から推定）

問１

ある高校の２年生男子の身長の分布は平均167㎝、標準偏差7㎝の正規分布とみなせるという。身長が160㎝以上172㎝以下の生徒はおよそ何%いるか。

また、２年生の男子が300人であったとき、185㎝以上の生徒はおよそ何人いると考えられるか。

数学B　P82

問２

ある工場で生産されるチョコレートの重量は平均209g、標準偏差３gの正規分布に従う。重量が200g未満のものが生産される確率を求めよ。

数学B　P85

正規分布演習１　解説

問１

ある高校の２年生男子の身長の分布は平均167㎝、標準偏差7㎝の正規分布とみなせるという。身長が160㎝以上172㎝以下の生徒はおよそ何%いるか。
また、２年生の男子が300人であったとき、185㎝以上の生徒はおよそ何人いると考えられるか。

平均167㎝、標準偏差7の正規分布に従うとき
求める割合はP(160≦X≦172)となる。
$Z=\frac{X-167}{7}$とすると、Zは平均0、標準偏差1の標準正規分布に従う。
これより、$P(160\leqq X\leqq 172)=P\left(\frac{160-167}{7}\leqq X\leqq\frac{172-167}{7}\right)\fallingdotseq P(-1\leqq X\leqq 0.71)$
$=u(1)+u(0.71)=0.341+0.261=0.602$
したがっておよそ60%となる

また、$P(185\leqq Y)$とすると、$P\left(\frac{185-167}{7}\leqq Z\right)=P(2.57\leqq Z)$
$=0.5-u(2.57)=0.5-0.4949=0.0051$
これは0.5%より、300人であったとすると、300人×0.5%＝1.53人
よって185㎝以上の男子は300人に1人～2人いると考えられる。

正規分布演習２（分布から推定）

問２　ある工場で生産されるチョコレートの重量は平均209g、標準偏差３gの正規分布に従う。重量が200g未満のものが生産される確率を求めよ。

平均209㎝、標準偏差3の正規分布に従うとき
求める割合はP(X＜200)となる。
$Z=\frac{X-209}{3}$とすると、Zは平均0、標準偏差1の標準正規分布に従う。
これより、P(X＜209)=$P\left(X<\frac{200-209}{3}\right)=P(X<-3)$
$=u(1)-u(3)=0.5-0.49865=0.00135$
このことより、200ｇ未満のチョコレートが生産されるのは0.135％である

二項分布 とは　（例）

５本のくじが袋に入っている。あたりくじは３本、はずれくじは２本である。この袋からくじを１本引く動作を n 回繰り返すとき、あたりくじが出た回数の合計の期待値をもとめなさい（ひいたくじはもとに戻すものとする）。

n＝３のとき

r（あたりくじがひかれた回数）	0	1	2	3
P（確率）				
r×P				
E（期待値）				

n＝５のとき

r（あたりくじがひかれた回数）	0	1	2	3	4	5
P（確率）						
r×P						
E（期待値）						

二項分布 とは　（例）

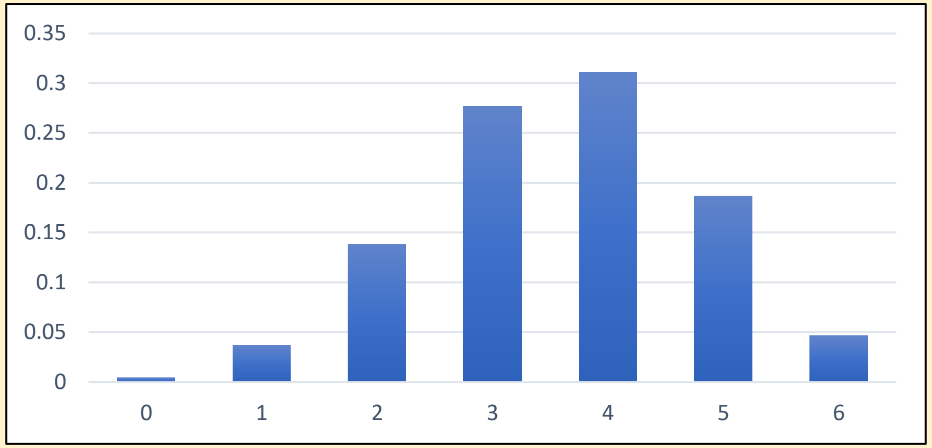
例題２

５本のくじがあり、あたりくじが３本、はずれくじが２本入っている袋がある。この袋からくじを１本引く動作を５回繰り返すとき、あたりくじが引かれた回数の期待値をもとめなさい。（ひいたくじはもとに戻すものとする

二項分布

r（あたりくじがひかれた回数）	0	1	2	3	4	5
P（確率）						

二項分布のグラフ



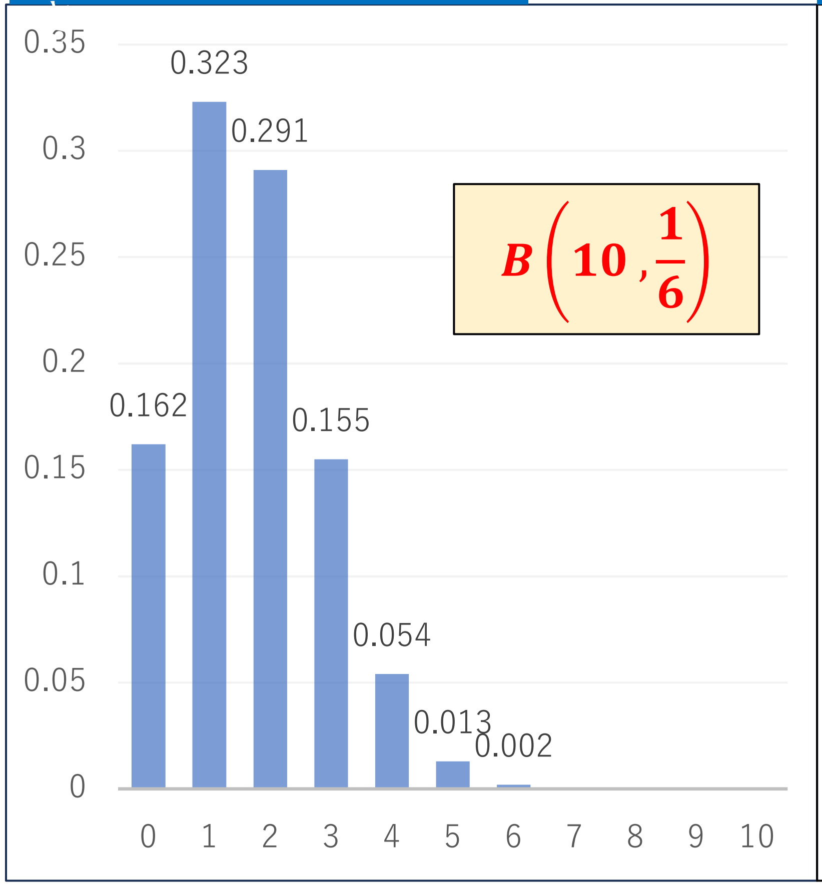
$$\text{期待値E} = 0\times \boxed{5C0\left(\tfrac{2}{5}\right)^5} + 1\times \boxed{5C1\left(\tfrac{3}{5}\right)^1\left(\tfrac{2}{5}\right)^4} + 2\times \boxed{5C2\left(\tfrac{3}{5}\right)^2\left(\tfrac{2}{5}\right)^3} + 3\times \boxed{5C3\left(\tfrac{3}{5}\right)^3\left(\tfrac{2}{5}\right)^2}$$
$$+ 4\times 5C4\left(\tfrac{3}{5}\right)^4\left(\tfrac{2}{5}\right)^1 + 5\times 5C5\left(\tfrac{3}{5}\right)^5$$

二項分布 $B(n,p)$ とは

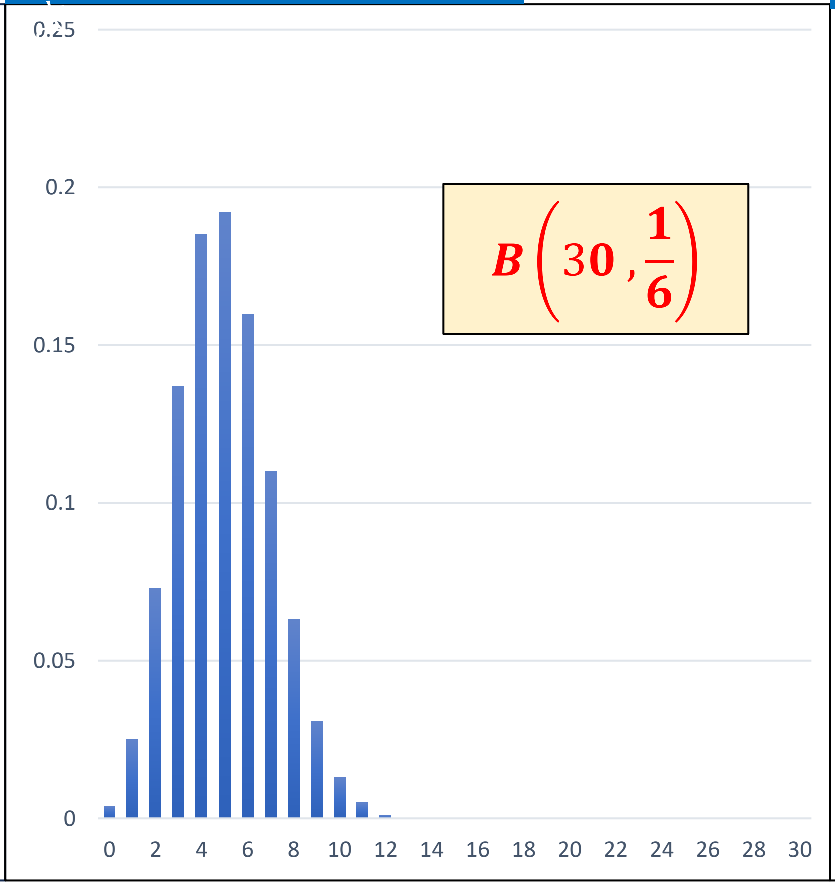
(例)さいころをｎ回振ったときの１の目が出る回数の確率

※　横軸：回数　　　縦軸：確率

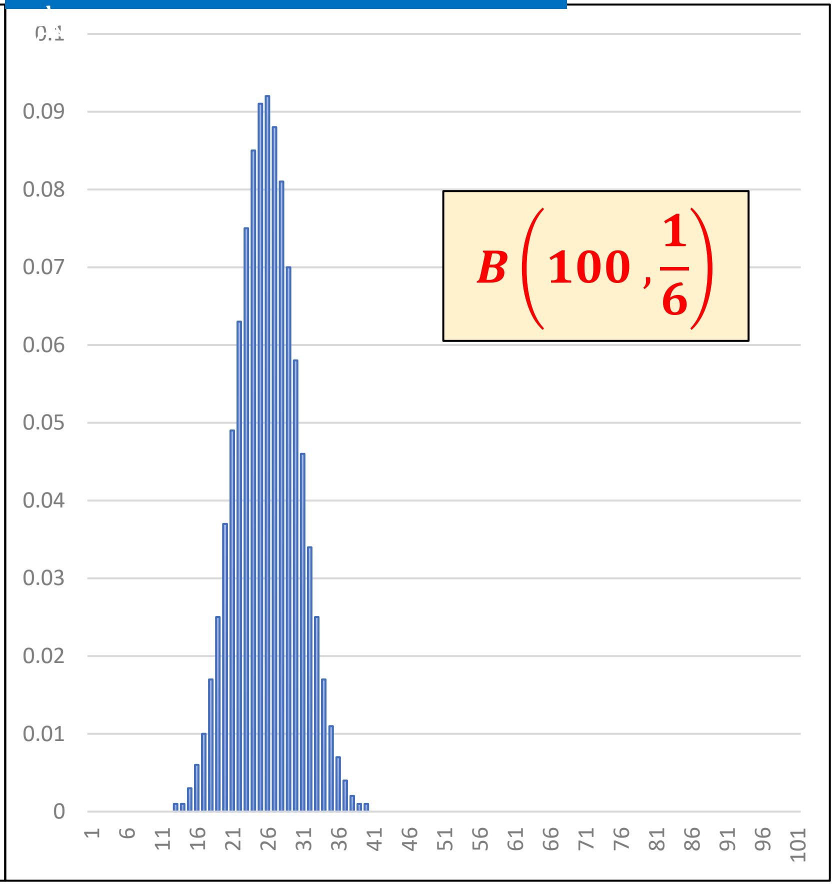
●１０回振ったと



●３０回振ったと



●１００回振ったと



スプレッドシートで二項分布を作ってみよう

５本のくじがあり、あたりくじが３本、はずれくじが２本入っている袋がある。この袋からくじを１本引く動作をｎ回繰り返すとき、あたりくじが引かれた回数とその確率の二項分布をつくろう

1

B2 f_x =COMBIN(3,A2)

	A	B		
1	回数r	3Cr		
2	0	1		
3	1	3		
4	2	3	0.432	0.864
5	3	1	0.216	0.648

標準偏差

組合せの関数「COMBIN」
※CONBIN（ｎ回，あたりの回数）
※B2を作ったらあとは下にドラッグ

2

C2 f_x =B2*((3/5)^A2)*((2/5)^(3-A2))

	A	B	C
1	回数r	3Cr	確率
2	0	1	0.064
3	1	3	0.288
4	2	3	0.432
5	3	1	0.216

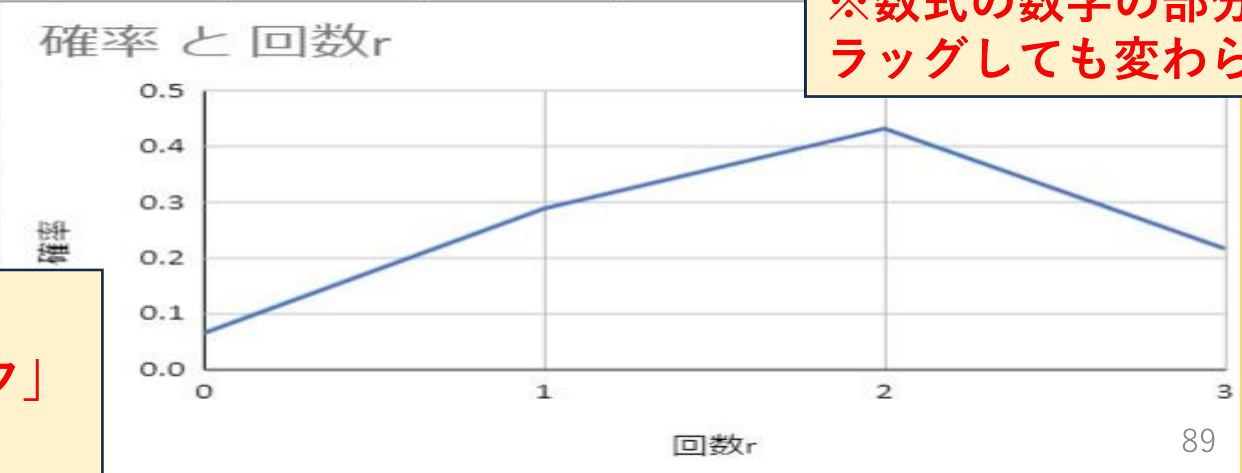
C2は３回中あたり０回の確率
C3は３回中あたり１回の確率

※ $3\mathrm{Cr}\times\left(\frac{3}{5}\right)^{r}\left(\frac{2}{5}\right)^{3-r}$

※C2を作ったらあとは下にドラッグ
※数式の数字の部分はドラッグしても変わらない

3

	A	B	C	D	E
1	回数r	3Cr	確率	確率*r	期待値
2	0	1	0.064	0	1.8
3	1	3	0.288	0.288	
4	2	3	0.432	0.864	
5	3	1	0.216	0.648	標準偏差
6					



①「回数ｒ」と「確率」を共にドラッグ
②「挿入」→「グラフ」→「折れ線グラフ」
③「縦棒グラフ」もあると良い

スプレッドシートで二項分布を作ってみよう

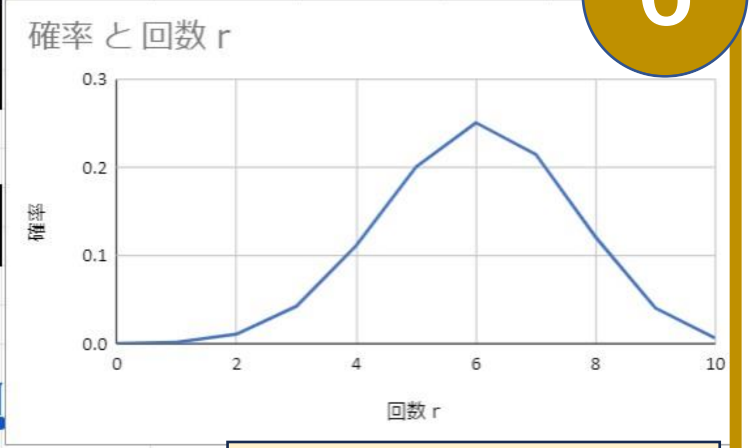
4

	A	B	C	D	E
1	回数r	3Cr	確率	確率*r	期待値
2	0	1	0.064	0	1.8
3	1	3	0.288	0.288	

- 「確率＊ｒ」は「期待値」を求めるために作成
- 「期待値」＝「確率＊ｒ」の和
- 「期待値」はあたりがでる回数の平均値

6

M	N	O	P	Q
回数 r	10C r	確率	確率*r	期待値
0	1	0.0001	0	5.9995
1	10	0.0016	0.0016	
2	45	0.0106	0.0212	
3	120	0.0425	0.1275	標準偏差
4	210	0.1115	0.446	1.549128787
	252	0.2007	1.0035	
	210	0.2508	1.5048	
	120	0.215	1.505	
	45	0.1209	0.9672	
	10	0.0403	0.3627	
10	1	0.006	0.06	



同様に
n=10、n=30、
n=125
を作成する

5

n＝3と同様にn＝6を作成する

I2 fx =H2*((3/5)^G2)*((2/5)^(6-G2))

	A	B	C	D	E	F	G	H	I	J	K
1	回数r	3Cr	確率	確率*r	期待値		回数 r	6C r	確率	確率*r	期待値
2	0	1	0.064	0	1.8		0	1	0.004096	0	3.6
3	1	3								36864	
4	2	3								7648	
5	3	1								2944	標準偏差

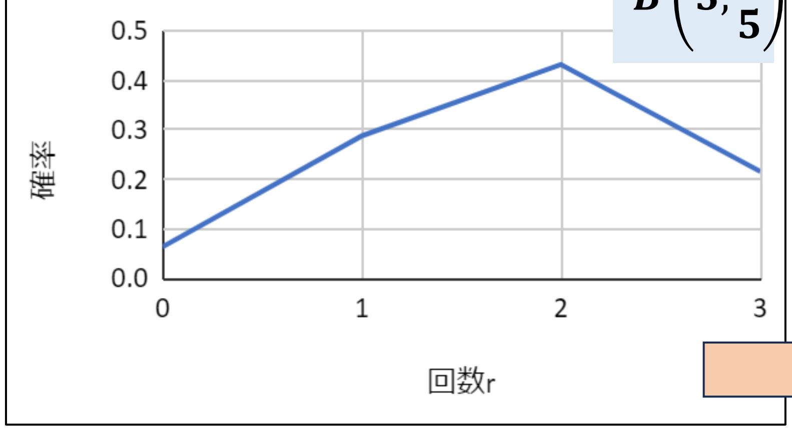
①「C2」をコピー
②「I2」に、「特殊貼り付け」→「数式のみ貼り付け」
（B2→H2、A２→G2、に自動で変わる）
③数式の中の「３」を「６」に変える

スプレッドシートから

回数：n=3　平均：1.8

標準偏差　0.849（$\sqrt{3\times\frac{3}{5}\times\frac{2}{5}}$）

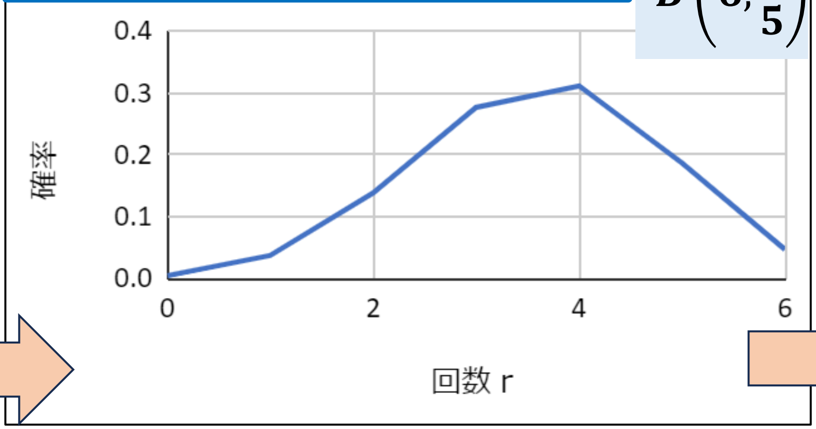
$B\left(3,\frac{3}{5}\right)$



回数：n=6　平均：3.6

標準偏差　1.2（$\sqrt{6\times\frac{3}{5}\times\frac{2}{5}}$）

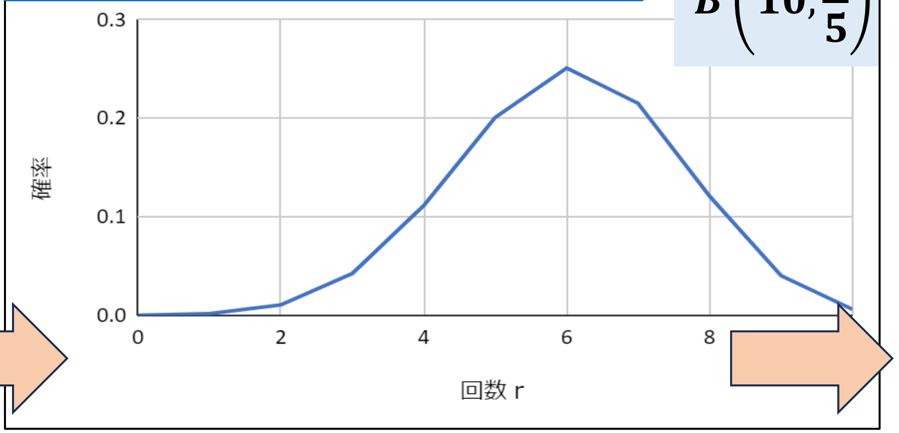
$B\left(6,\frac{3}{5}\right)$



回数：n=10　平均：6.0

標準偏差　1.55（$\sqrt{10\times\frac{3}{5}\times\frac{2}{5}}$）

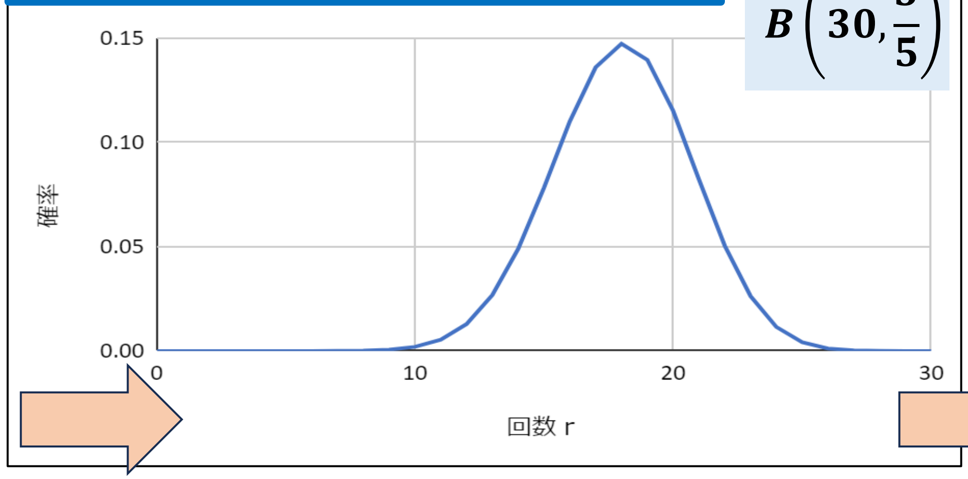
$B\left(10,\frac{3}{5}\right)$



回数：n=30　平均：18

標準偏差　2.68（$\sqrt{30\times\frac{3}{5}\times\frac{2}{5}}$）

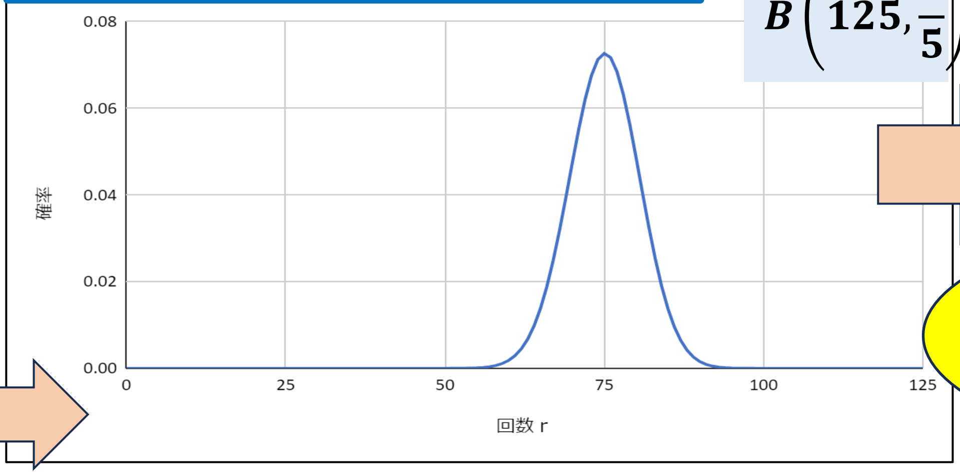
$B\left(30,\frac{3}{5}\right)$



回数：n=125　平均：75

標準偏差　5.48（$\sqrt{125\times\frac{3}{5}\times\frac{2}{5}}$）

$B\left(125,\frac{3}{5}\right)$



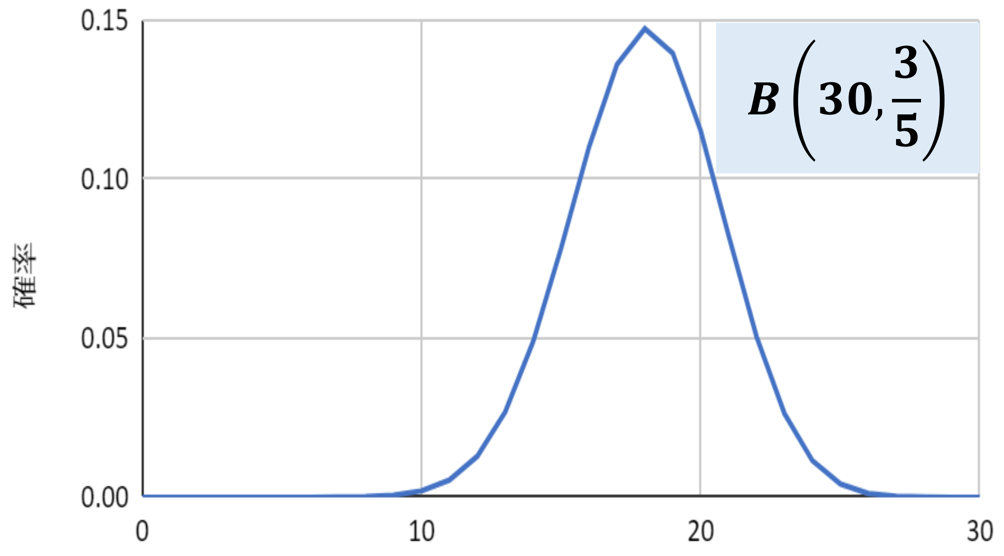
徐々に整った正規分布に近づいている

中心極限定理

スプレッドシートから

回数：n=30　平均：18

標準偏差　2.68（$\sqrt{30\times\frac{3}{5}\times\frac{2}{5}}$）

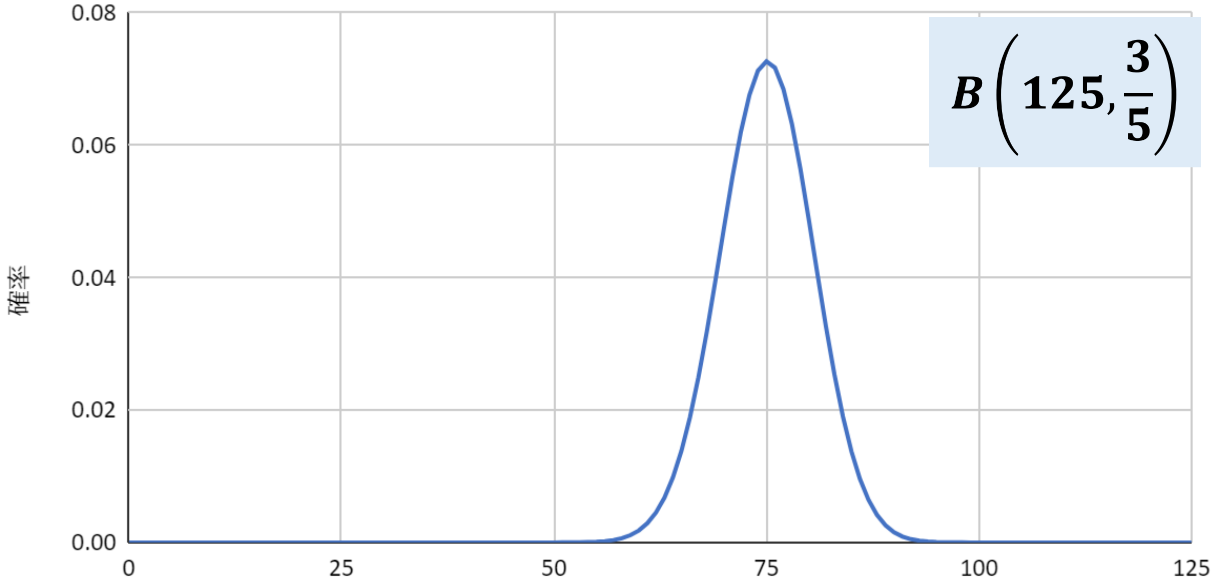


$N\left(18\,,30\times\frac{3}{5}\times\frac{2}{5}\right)$

平均18、分散$30\times\frac{3}{5}\times\frac{2}{5}$　の正規分布と見なせる

回数：n=125　平均：75

標準偏差　5.48（$\sqrt{125\times\frac{3}{5}\times\frac{2}{5}}$）

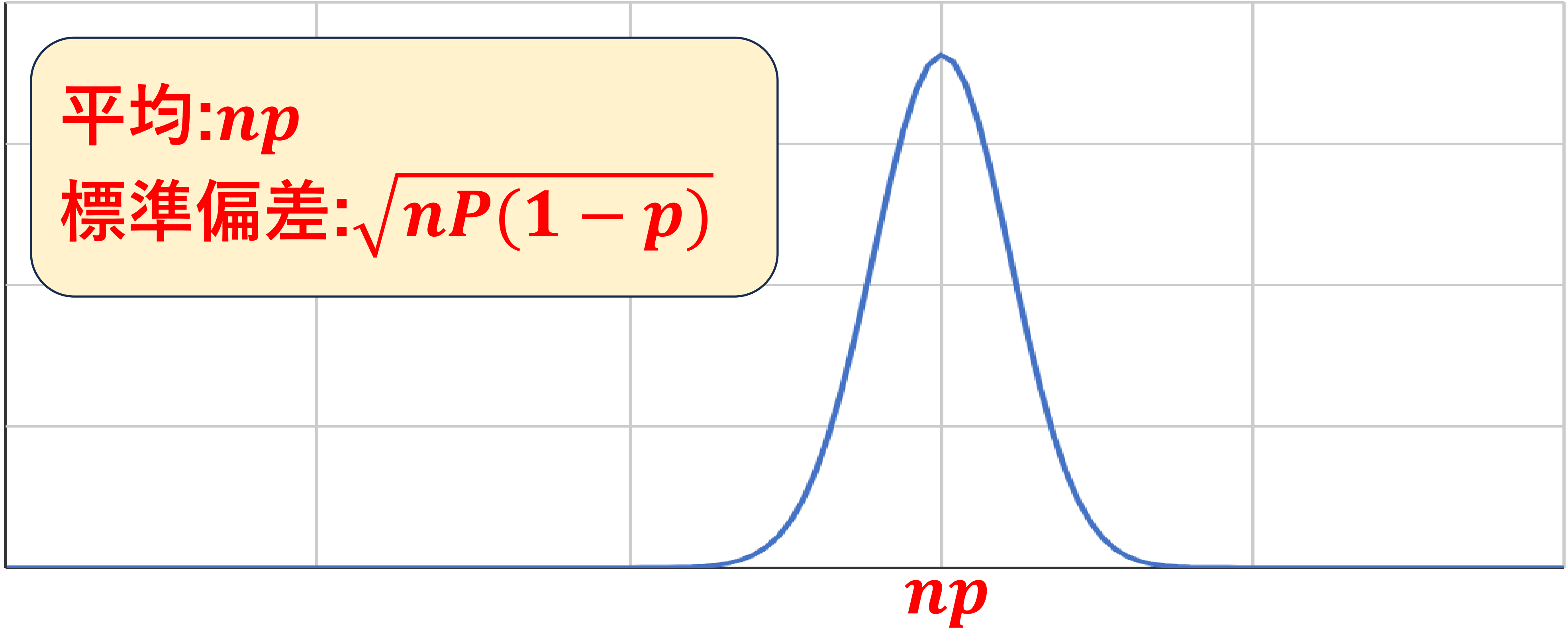


$N\left(75\,,125\times\frac{3}{5}\times\frac{2}{5}\right)$

平均75、分散$125\times\frac{3}{5}\times\frac{2}{5}$　の正規分布と見なせる

二項分布の正規分布近似

確率pの事象をn回行う二項分布$B(n, p)$は、nが十分大きいとき、
平均np、分散$np(1-p)$の正規分布$N(np, np(1-p)$）に従う。



平均:np

標準偏差:$\sqrt{nP(1-p)}$

演習をしてみよう（二項定理から正規分布近似）

問１

１枚のコインを600回投げて、表が出る合計回数をXとする。ただし、コインを１回投げたときの表と裏が出る確率は等しいとして、Xの値が180≦X≦220となる確率を正規分布を利用して求めよ。

演習をしてみよう（二項定理から正規分布近似）

問２

１個のさいころを720回投げるとき、6の目の出る回数が100回以上150回以下である確率を求めよ。

数学B　ハイプライム　P157

演習をしてみよう（仮説検定）

問３

１枚の硬貨を800回投げたところ、表が出た回数は430回であった。この硬貨は表が出やすいと判断できるか。有意水準５％と１％でそれぞれ片側検定せよ。

数学B　ハイプライム　P172

NO９．

２群のｔ検定
（２群の平均の差の検定）

※数式（ｔ値）の意味を厳密に理解するのは非常に難しいです。
　今回は、ｔ検定は「何を調べられるのか」「どのように調べられるのか」を
　理解することを目的として下さいね。
（統計は、まずやってみることが非常に大切）

２群の平均の差のｔ検定とは

２つの群の母平均に統計的有意差があるかを検証する
（２群の特徴を調べる手法）

①　対応のある２群のｔ検定

②　ステューデントのｔ検定（対応のない２群で分散が等しいとき）

③　ウェルチのｔ検定　（対応のない２群で分散が異なるとき）

対応のある２群・対応のない２群の違い

対応のある２群

A組20人の生徒の数学の得点	A組20人の生徒の物理の得点

A組の中で、数学と物理の得点に差があるかを調べる

特定の患者50人の投薬前の血圧	特定の患者50人の投薬後の血圧

血圧を下げる薬があり、投薬前と投薬後で変化したかを調べる

対応のない２群

A組20人の生徒の数学の得点	B組18人の生徒の数学の得点

A組とB組の生徒の数学の得点に差があるかを調べる

畑Aから取れたアスパラガス50本の長さ	畑Bから取れたアスパラガス50本の長さ

畑Aと畑Bのアスパラガスの長さに差があるかを調べる

ｔ検定の手順

・帰無仮説・対立仮説、有意水準を決める
・検定の前提を確認する
・検定統計量（ｔ値）を算出する
・有意差の判定し、帰無仮説の棄却の有無を決める

※　帰無仮説・対立仮説の立て方
帰無仮説H_0：２群の平均値に差がない
対立仮説H_1：２群の平均値に差がある

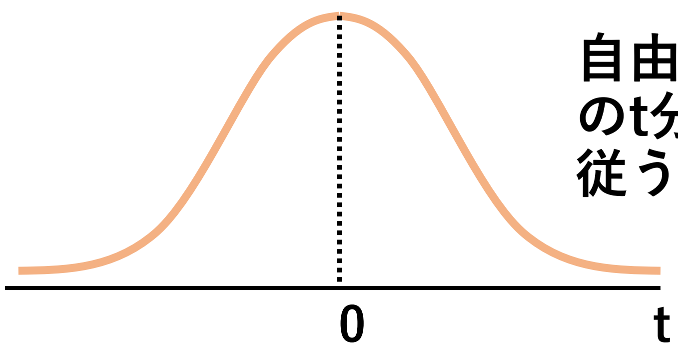
① 対応のある２群の t 検定

検定統計量

$$t = \frac{\bar{d} - m}{\frac{S_u}{\sqrt{n}}}$$

$\bar{d}$:差の平均値
m:差の設定値や理論値
S_u:差の不偏標準偏差
n: 標本の大きさ

t分布



自由度n-1のt分布に従う

例題

高校生10人を対象に、あるストレッチ前後での長座体前屈の記録を測定した結果がある。あるストレッチの効果があるかどうかを、有意水準5％で、片側の t 検定を用いて調べる。

$H_0: m = 0$ （あるストレッチの効果はない。差がない）
$H_1: m > 0$ （あるストレッチの効果はある。差はある）

t分布表より自由度10－1＝9のときは

$t_{0.05}(9) = 1.833$　　$t = \frac{3.60-0}{\frac{3.21}{\sqrt{10}}}$ ≒3.546>1.833

よって有意水準５％で、あるストレッチに効果があるといえる。

	前（cm）	後（cm）	差
A	46.1	48.2	2.1
B	32.5	37.7	5.2
・・・	・・・	・・・	・・・
平均値			3.6
不偏標準偏差			3.21

② ステューデントの t 検定（対応のない２群で分散が等しいとき）

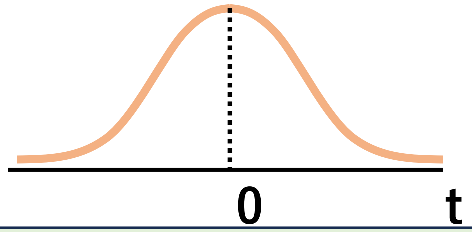
検定統計量

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{(n_1-1){S_{u1}}^2+(n_2-1){S_{u2}}^2}{n_1+n_2-2}}}$$

$\overline{X_1}, \overline{X_2}$: ２群それぞれの平均値
S_{u1}, S_{u2}: ２群それぞれの不偏標準偏差
n_1, n_2: ２群それぞれの標本の大きさ

t分布

自由度n1+n2-2のt分布に従う



例題

地域A、地域Bから切り出した同じ樹種の木材の密度(単位はg/cm3で測定)をそれぞれ10ずつ調べたところ、標本平均と不偏標準偏差は表のようになった。有意水準５％で、地域A、地域Bの母平均m_1、m_2に差があるかを両側の t 検定を用いて調べる。

$H_0: m_1 = m_2$ （木材の密度に地域による差はない）
$H_1: m_1 \neq m_2$ （木材の密度に地域による差はある）
t分布表より自由度10＋10－2＝18のときは、$t_{\frac{0.05}{2}}(18) = 2.101$

$$t = \frac{0.38 - 0.41}{\sqrt{\frac{1}{10}+\frac{1}{10}}\sqrt{\frac{(10-1)0.028^2+(10-1)0.032^2}{10+10-2}}} = -2.231\cdots < -2.101$$

よって有意水準５％でこの木材の密度は地域によって差があるといえる。

	地域A	地域B
標本平均	0.38	0.41
不偏標準偏差	0.028	0.032

③ ウェルチの t 検定（対応のない２群で分散が異なるとき）

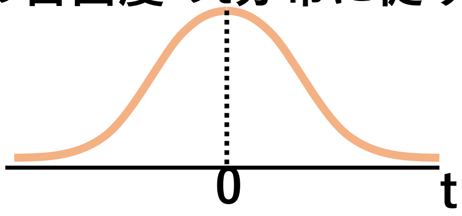
検定統計量

$$t=\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\frac{{S_{u1}}^2}{n_1}+\frac{{S_{u2}}^2}{n_2}}}$$

$\overline{X_1},\overline{X_2}$: ２群それぞれの平均値
S_{u1},S_{u2}: ２群それぞれの不偏標準偏差
n_1,n_2: ２群それぞれの標本の大きさ

t分布

t値はfに最も近い整数値の自由度のt分布に従う



$$f=\frac{\left(\frac{{S_{u1}}^2}{n_1}+\frac{{S_{u2}}^2}{n_2}\right)^2}{\frac{\left(\frac{{S_{u1}}^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{{S_{u2}}^2}{n_1}\right)^2}{n_2-1}}$$

例題

ある種類の魚を観察するために、２つの池A、Bから捕獲してきた。捕獲した池によって差がないかを確かめるために体重を（単位はgで）測定したところ、表のような結果であった。有意水準５％で池A、Bの魚の体重の母平均m_1、m_2について、m_1の方が大きいことを検証するために、母標準偏差は異なるとして片側の t 検定を行う。

$H_0: m_1=m_2$　（魚の体重に池による差はない）
$H_1: m_1>m_2$　（池Aの魚の体重の方が重い。池による差はある）
$f\fallingdotseq 12$なので、t分布表より自由度12のときは、

$t_{0.05}(12)=1.782$

$$t=\frac{412-380}{\sqrt{\frac{15.3^2}{6}+\frac{26.1^2}{8}}}=2.871\cdots>1.782$$

よって有意水準５％で、池Aの魚の体重の方が重いといえる。

	池A	池B
魚の数（匹）	6	8
標本平均	412	380
不偏標準偏差	15.3	26.1

Excel・スプレッドシートで演習してみよう

超簡単

①　対応のある２群のｔ検定

＝T.TEST(X1の範囲,X2の範囲,片側・両側,1)

②　ステューデントのｔ検定（対応のない２群で分散が等しいとき）

＝T.TEST(X1の範囲,X2の範囲,片側・両側,2)

今回は標本平均から判断する

③　ウェルチのｔ検定　（対応のない２群で分散が異なるとき）

＝T.TEST(X1の範囲,X2の範囲,片側・両側,3)

※　片側なら1、両側なら2を入れる

Excel・スプレッドシートで演習してみよう

① 対応のある２群の t 検定

② ステューデントの t 検定（分散が等しい）

③ ウェルチの t 検定（分散が異なる）

例題１

標本を特定の患者とし、血圧が投薬前と後で変化したかを調べ、投薬の効果を検証する

F2 =T.TEST(B2:B27

	A	B	C	D	E	F
1	被験者	投薬前	投薬後	差		
2	A	180	130	50		3.03245
3	B	135	135	0		
4	C	150	132	18		
5	D	138	136	2		
6	E	143	133	10		
7	F	170	132	38		
8	G	122	123	-1		
9	H	168	142	26		
10	I	159	137	22		
11	J	126	123	3		
12	K	149	136	13		

例題２

農家Aと農家Bでそれぞれいくつか取れたアスパラガスの長さに違いがあるかどうかを調べ、農家Aと農家Bのアスパラガスの成育を検証する

E2 =T.TEST(B2:B8,C2:C8,1,2)

	A	B	C	D	E
1	標本数	農家A	農家B		
2	1	20	21		0.17548
3	2	23	24		
4	3	24	25		
5	4	19	21		
6	5	24	24		
7	6	25	28		
8	7	25	26		
9	8		24		
10					
11	標本平均	22.8571	24.125		
12	標本分散	4.97959	4.85938		
13	標本標準偏差	2.2315	2.2044		

例題３

ビーカーとメスシリンダーで水50mℓがどれだけ正確に測れるかを調べ、ビーカーとメスシリンダーの用途を検証する

F3 =T.TEST(B3:B78,C3:C78,1,3)

	A	B	C	D	E
1		ビーカー	メスシリンダー		
2	標本数	100mL	100mL	50mL	
3	1	49.85	49.38	48.68	
4	2	51.06	48.89	48.95	
5	3	51.38	49.38	48.99	
6	4	49.51	48.68	49.4	
7	5	46.93	48.26	49.58	
8					
75	73	43.91	49.08	49.86	
76	74	43.93	48.94	49	
77	75	42.06	49.36	49.58	
78	76	43.93	48.81	49.91	
79					
80	内径	50.45	27.5	20.55	
81	理論値	49.8815	49.8815	49.8815	
82	平均	46.6	48.85	49.3	
83	分散	8.735761	0.596096	0.291751	
84	標準偏差	2.955632	0.772073	0.540139	
85	理論値－平均値	3.2815	1.0315	0.5815	

計算結果の見方

① 例題１で

被験者	投薬前	投薬後	差		
A	180	130	50		0.0000303245359
B	135	135	0		
C	150	132	18		
D	138	136	2		
E	143	133	10		
F	170	132	38		

＝T.TEST(B2:B27,C2,C27,1,1)

X1の範囲

X2の範囲

片側検定

対応のある

＝T.TEST(B2:B27,C2,C27,1,1)
で計算をした結果
0.0000303…と返ってきます。
この値はｔ値ではなく、ｔ値から自動的にｐ値（ｔ分布のｔ値より外側の面積）が計算されます。有意水準５％での検定の場合、ｐ値＜0.05であれば帰無仮説を棄却します。有意水準１％での場合、ｐ値＜0.01であれば帰無仮説を棄却します。

付録　t 分布表

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
1	6.314	12.706
2	2.920	4.303
3	2.353	3.182
4	2.132	2.777
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.813	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
15	1.753	2.132
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048

	片側検定	両側検定
有意水準	0.05	0.025
自由度 v		
29	1.699	2.045
30	1.697	2.042
31	1.696	2.040
32	1.694	2.037
33	1.692	2.035
34	1.691	2.032
35	1.690	2.030
36	1.688	2.028
37	1.687	2.026
38	1.686	2.024
39	1.685	2.023
40	1.684	2.021
41	1.683	2.020
42	1.682	2.018