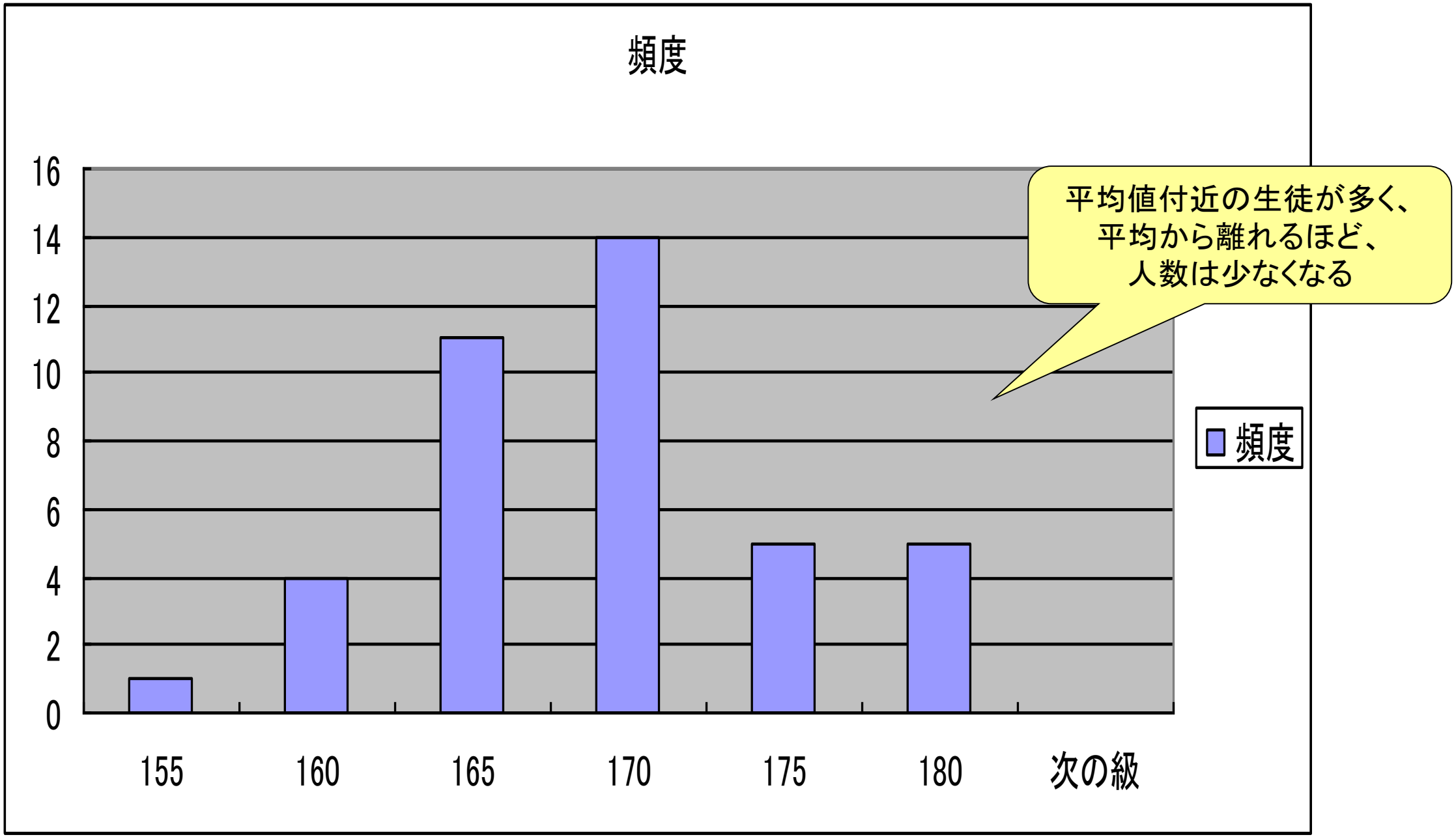
データの統計的な分析
～　仮説検定と相関　～

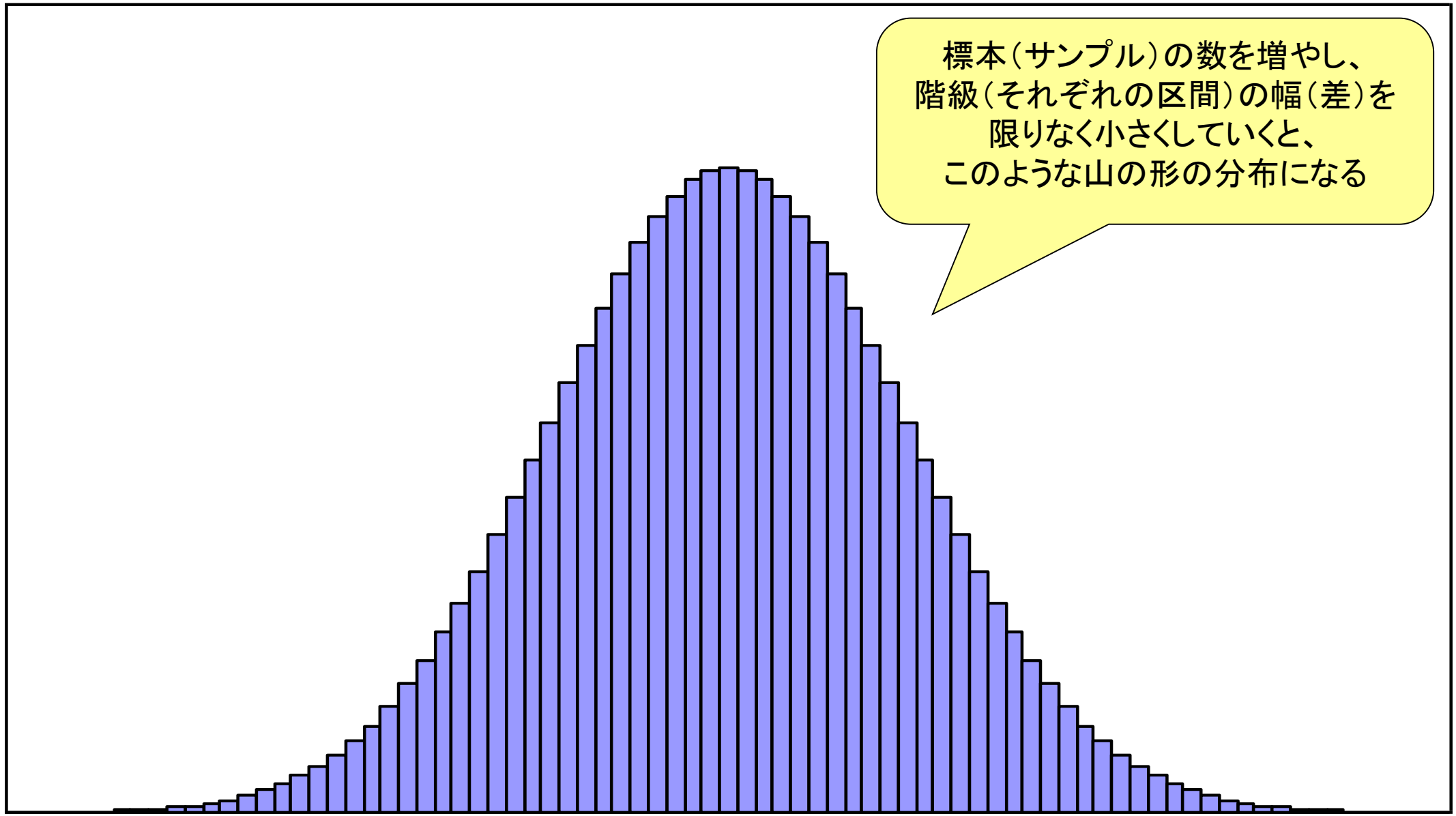
情報の科学　第23回授業

04情報の蓄積と管理

対応データ　21exp23.xls

1 「仮説検定」について

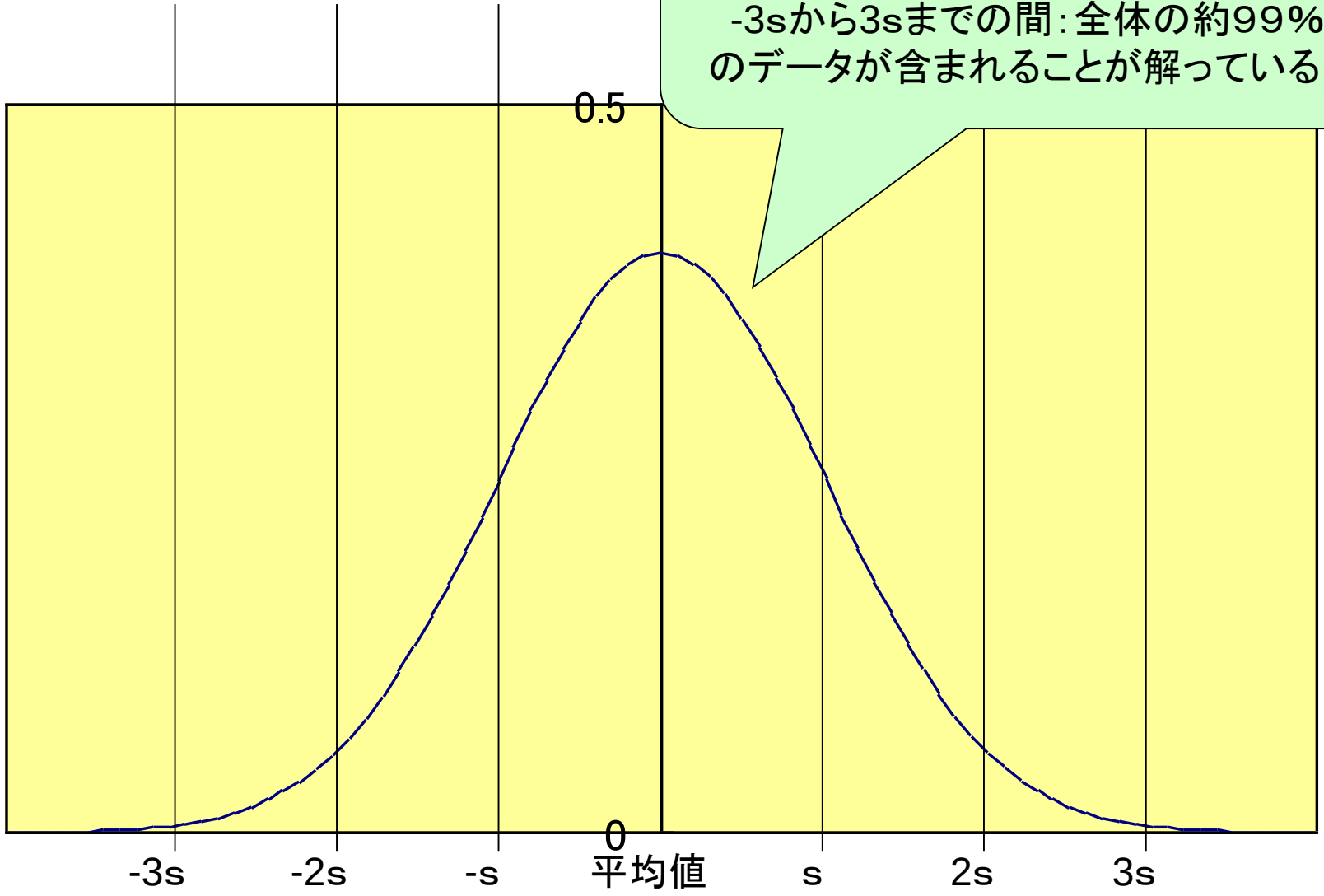

頻度
16
14
12
10
8
6
4
2
0
155
160
165
170
175
180
次の級
平均値付近の生徒が多く、
平均から離れるほど、
人数は少なくなる
頻度


標本（サンプル）の数を増やし、
階級（それぞれの区間）の幅（差）を
限りなく小さくしていくと、
このような山の形の分布になる

正規分布

- 平均値の周辺が最も度数が多く、
- 平均値から離れるに従って、度数が少なくなっていくような分布（山のような形）
- 世の中の多くの分布が、ほぼ正規分布のような形になると見なすことができる。
  - 模試の結果、体重、身長、・・・
- 正規分布で「分かっていること」を活用できる

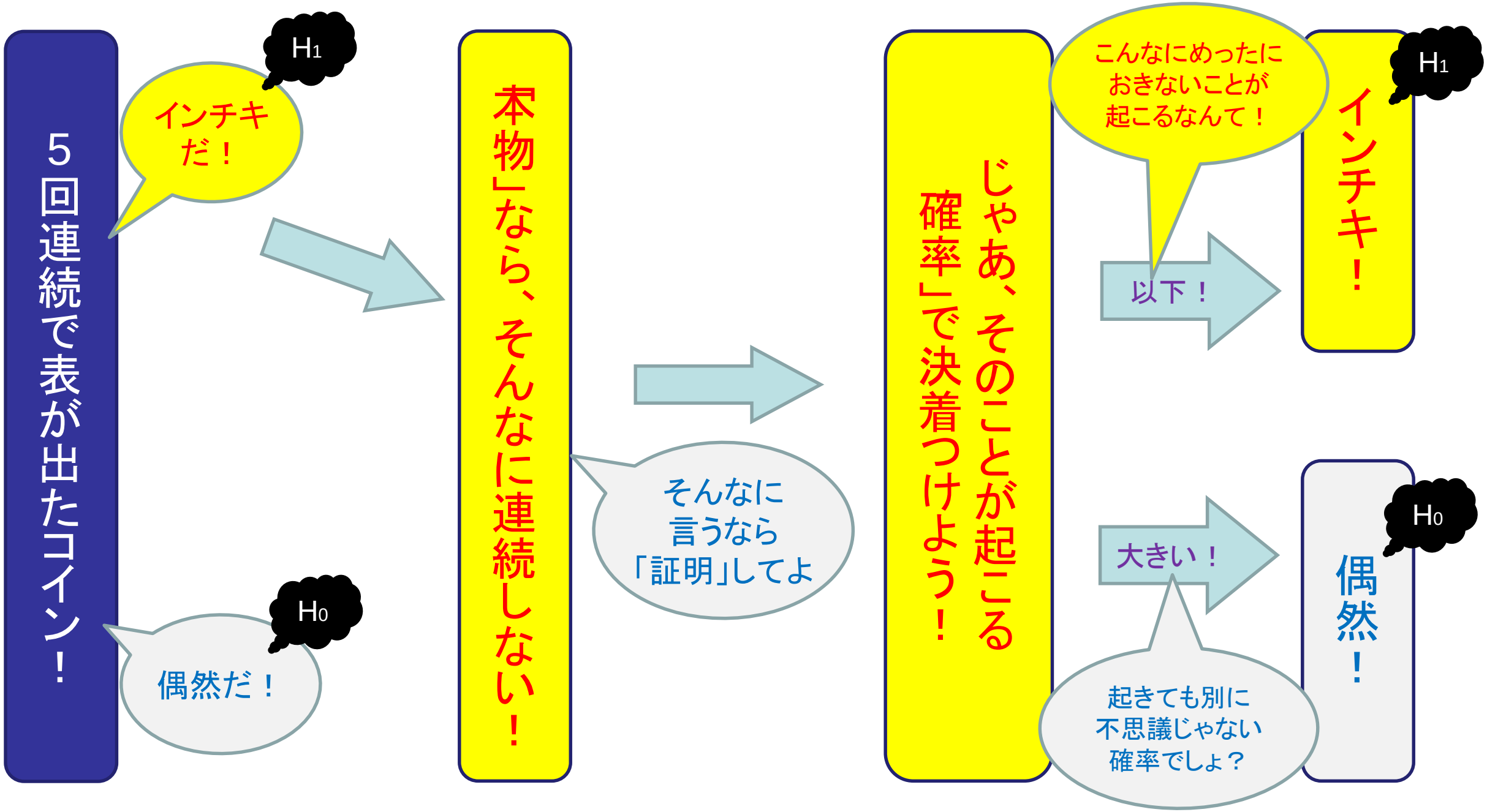
標準偏差がsの正規分布



-sからsまでの間：全体の約68％
-2sから2sまでの間：全体の約95％
-3sから3sまでの間：全体の約99％
のデータが含まれることが解っている。

統計的仮説検定

ある出来事に着目した場合、
その起こった出来事の確率が、
一定[有意水準といいます]
（0.05　あるいは　0.01）以下の場合に、
「違いがあるのでは」とする考え方


5回連続で表が出たコイン！
H1
インチキだ！
H0
偶然だ！
「本物」なら、そんなに連続しない！
そんなに言うなら「証明」してよ
じゃあ、そのことが起こる確率で決着つけよう！
こんなにめったにおきないことが起こるなんて！
以下！
H1
インチキ！
大きい！
起きても別に不思議じゃない確率でしょ？
H0
偶然！

例）5回連続で表が出たコイン。「表が出やすい」といえるか？

H_0：コイン表裏の出やすさは同様

→「守り」（＝帰無仮説）

H_1：コインは表が出やすい

→「示したいこと」（＝対立仮説）

確率を計算し、「基準値」以下ならば、H_0は棄却

→「示したいこと」が示される（「有意差がある」という）

「検定」と「過誤」

$(0.5)^5 = 0.03125$

有意水準　5%

H_0を棄却　→　このコインは表が出やすい

※もちろん「本当は表裏同じ」かもしれない！　←　第1種の過誤

有意水準　1%

H_1を棄却　→　このコインは表が出やすい

とは言い切れない

※もちろん「本当は表が出やすい」かもしれない！　←　第2種の過誤

＜注意！＞この場合、決して「表裏が同じ」と言い切れるわけではない！

＜練習２＞

１の目が3回連続して出たサイコロがある。
このサイコロは1の目が出やすいと言えるか。
有意水準１％で検定せよ。

H_0：このサイコロの目の出やすさは同様
H_1：このサイコロは1の目が出やすい

１の目が3回連続・・・（１／６）×３　≒０．００463
＜０．０１
よって、H_0を棄却　→　このサイコロは、有意水準1％で
１の目が出やすい！

（発展・参考）

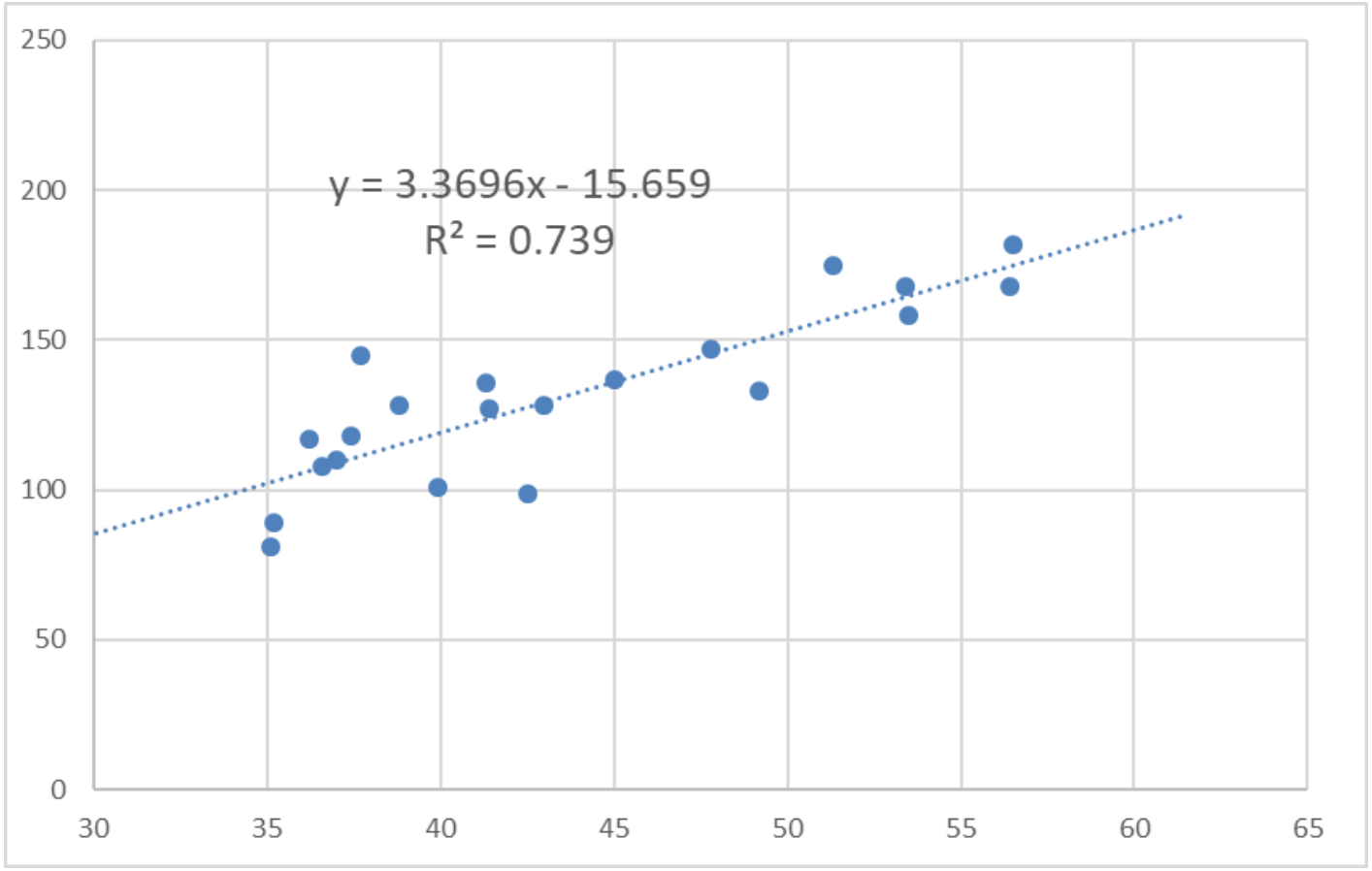
既に知られている分布を活用する

- 分散に「違い」があるかを確かめる
  - 「F分布」の活用　→　F検定
- 平均に「違い」があるかを確かめる
  - 「正規分布」の活用　→　Z検定（n≧30）
  - 「t分布」の活用　→　t検定（n＜30）
- クロス集計表に「違い」があるかを確かめる
  - 「χ^2（カイ二乗）分布の活用」　→　χ^2検定

→　これらの「表」から得られた確率に基づき、
有意差があるかを判断している。

２　相関について

相関を調べる



※この直線を「回帰直線」という

＜練習３＞

- ワークシートにある「握力と背筋力」のデータから、
  - 散布図を作成する
  - 回帰直線を表示させる
  - 回帰直線の方程式を表示させる

相関行列

	身長	体重	座高	握力	上体起こし	長座体前屈	反復横跳び	シャトルラン	50m走	立ち幅跳び	ハンドボール投
身長	1.000										
体重	0.382	1.000									
座高	0.756	0.497	1.000								
握力	0.250	0.559	0.315	1.000							
上体起こし	0.066	0.092	-0.029	0.360	1.000						
長座体前屈	0.257	0.235	0.235	0.317	0.309	1.000					
反復横跳び	0.149	0.110	0.093	0.386	0.457	0.477	1.000				
シャトルラン	0.142	-0.090	0.029	0.175	0.341	0.277	0.372	1.000			
50m走	-0.211	-0.098	-0.215	-0.454	-0.329	-0.294	-0.544	-0.553	1.000		
立ち幅跳び	0.359	0.063	0.273	0.412	0.256	0.361	0.483	0.341	-0.674	1.000	
ハンドボール投げ	0.292	0.315	0.278	0.470	0.457	0.408	0.519	0.400	-0.490	0.419	1.000

＜練習４＞

- 「科学の道具箱」を開き、「高等学校体力測定データ」を確認する
- すでに整形されたデータを元に、相関行列を作ってみる
- 相関行列から相関の高い2系列を選び、散布図と回帰直線を作成する