仮説検定

１　カイ二乗検定

２　ｔ検定

３　回帰分析

「ｔ検定」のパターン

１　１つのデータしかないが

（母集団の平均はわかってる）

２　２つのデータがある

Excel　シート

「【例題】A病院」

「検定　A病院」

これを見ながら

「ｔ検定」のパターン

１　１つのデータしかないが
（母集団の平均はわかってる）

２　２つのデータがある

ある病院では、放射線技師の日給は

「平均30000円」と言ってます

１０人に聞いてみた結果がこれです。

「日給30000円」と言える？　言えない？

	日給
Aくん	27149
Bくん	35800
Cさん	26382
Dくん	28329
Eさん	21988
Fさん	27441
Gくん	25195
Hくん	20475
Iさん	23173
Jさん	31090

１０人に聞いた結果

「日給30000円」

と言える？　言えない？

（等分散と仮定）

ｔ検定の流れ

１　「両側」か「片側」かを決める

２　仮説を立てる（H_0 H_1）

３　有意水準を決める（普通は0.05）

４　ｔ値（絶対値）を求める

５　ｔ値の境界値を求める

６　ｔ値とｔ値の境界値から判断する

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「両側検定」？

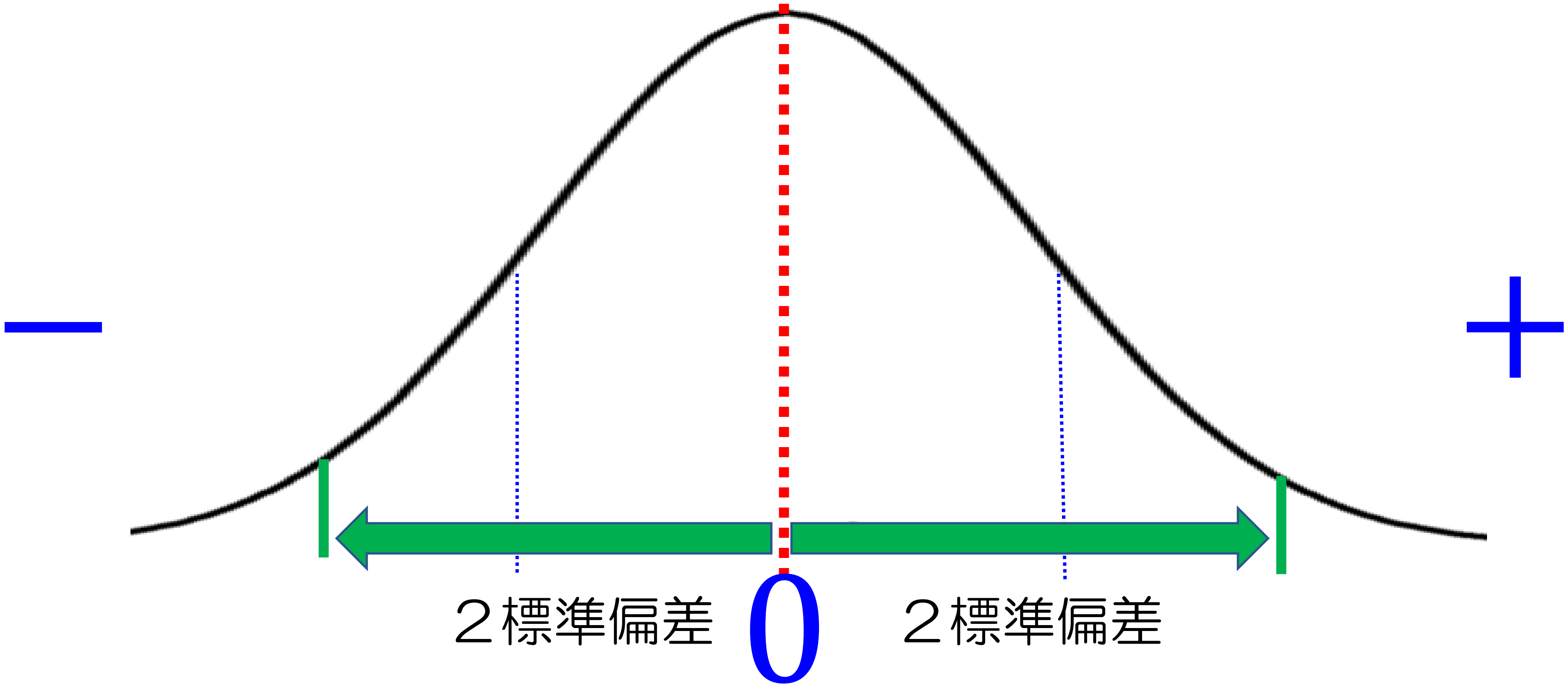
「片側検定」？

思いだしてみて

「正規分布」と「標準偏差」

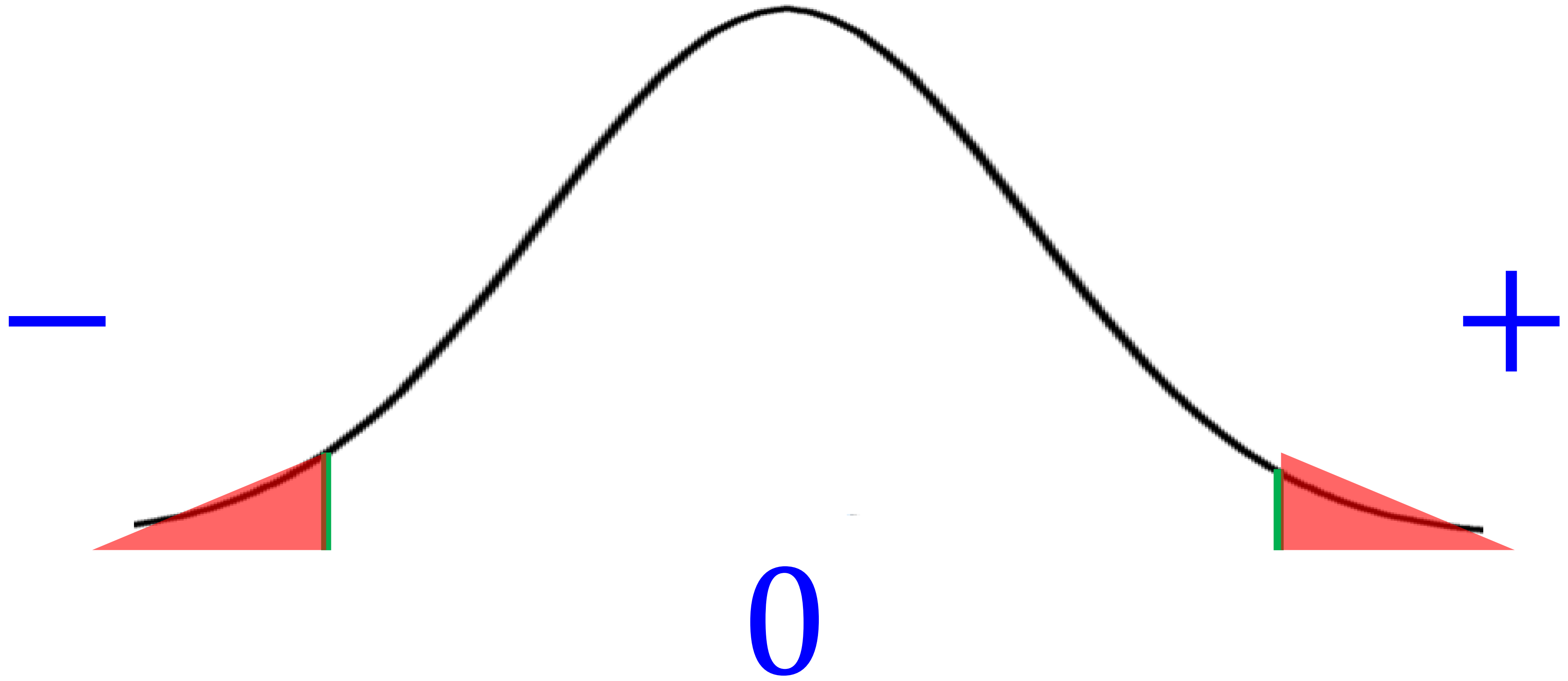
「正規分布 と 標準偏差」

- 平均値 ± ２標準偏差（面積）の中に
  データの「９５％」がある



「正規分布 と 標準偏差」

逆に言うと、ここにデータの「5%」がある



「正規分布 と 標準偏差」

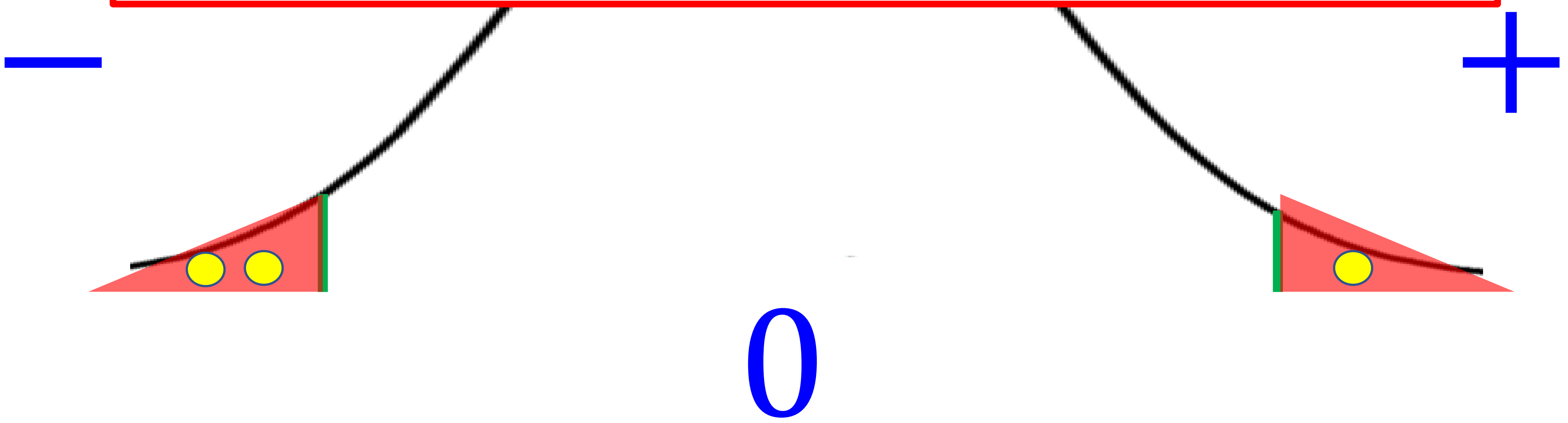
ここにデータが入るってことは

分　い…

ほとんどない！と言い切れる

明らかに間違ってると言っていい

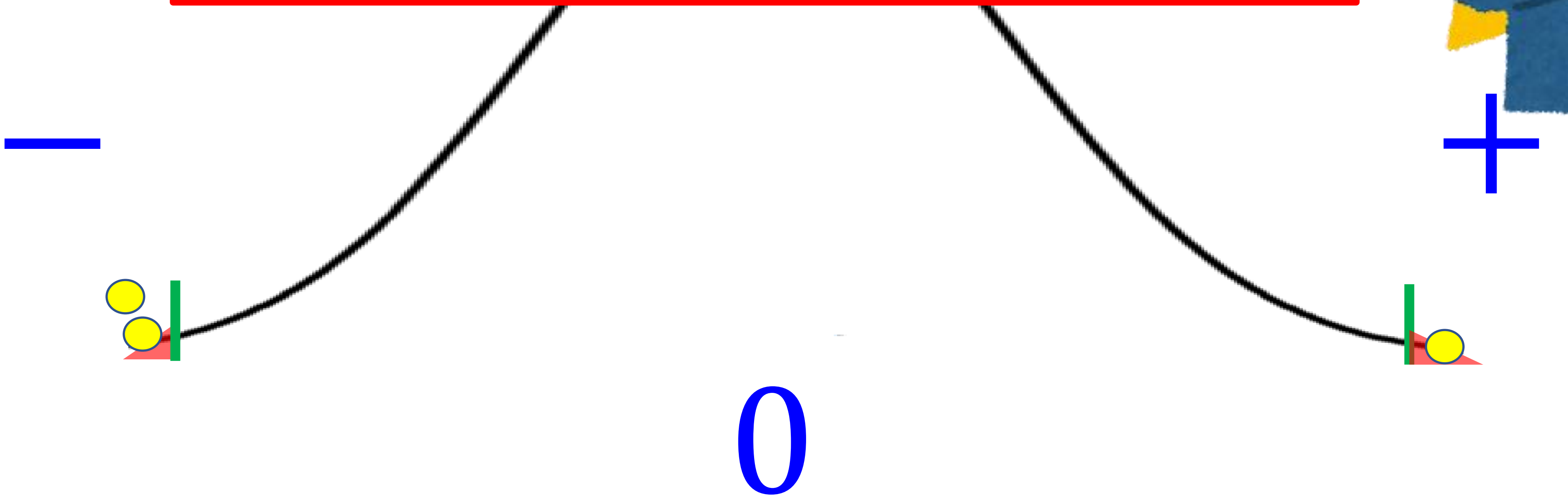
「棄却できる！」



「正規分布 と 標準偏差」

「１％」ならなおさら！

ほぼ絶対あり得ない！



話は「両側」「片側」に戻って

例えば

「日給が30000円と仮定」(H_0)すると

対立仮説(H_1)は３通りできる

１　日給は30000円でない

２　日給は30000円より少ない

３　日給は30000円より多い

１　日給は30000円でない

２　日給は30000円より少ない

３　日給は30000円より多い

日給が「30000円」かどうかだけで

30000円より多いか少ないかは、

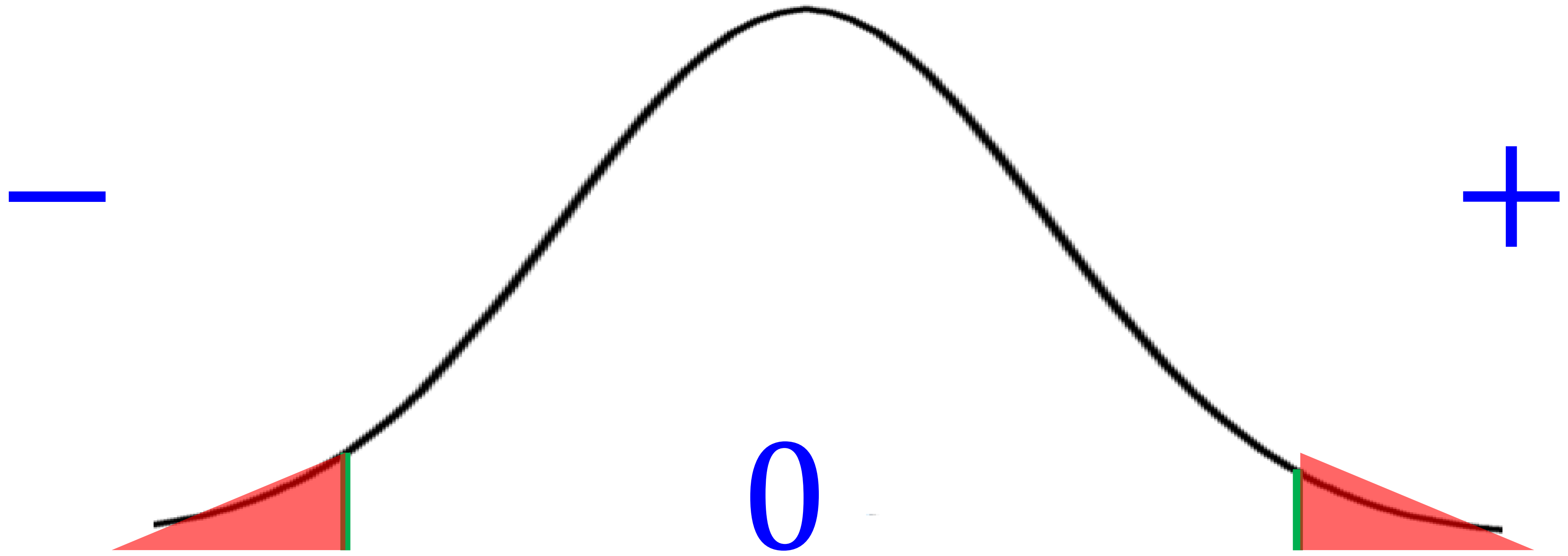
全く関係ない！

「両側検定」　　　　　1　日給は30000円でない

これが「両側検定」

どっちに入ってもいい

（多い少ないは全く考えない）



1　日給は30000円でない

2　日給は30000円より少ない

3　日給は30000円より多い

「30000円」より少ないかを調べるだけで

30000円より多いかは、

全く考えない

1　日給は30000円でない

2　日給は30000円より少ない

3　日給は30000円より多い

「30000円」より多いかを調べるだけで

30000円より少ないかは、

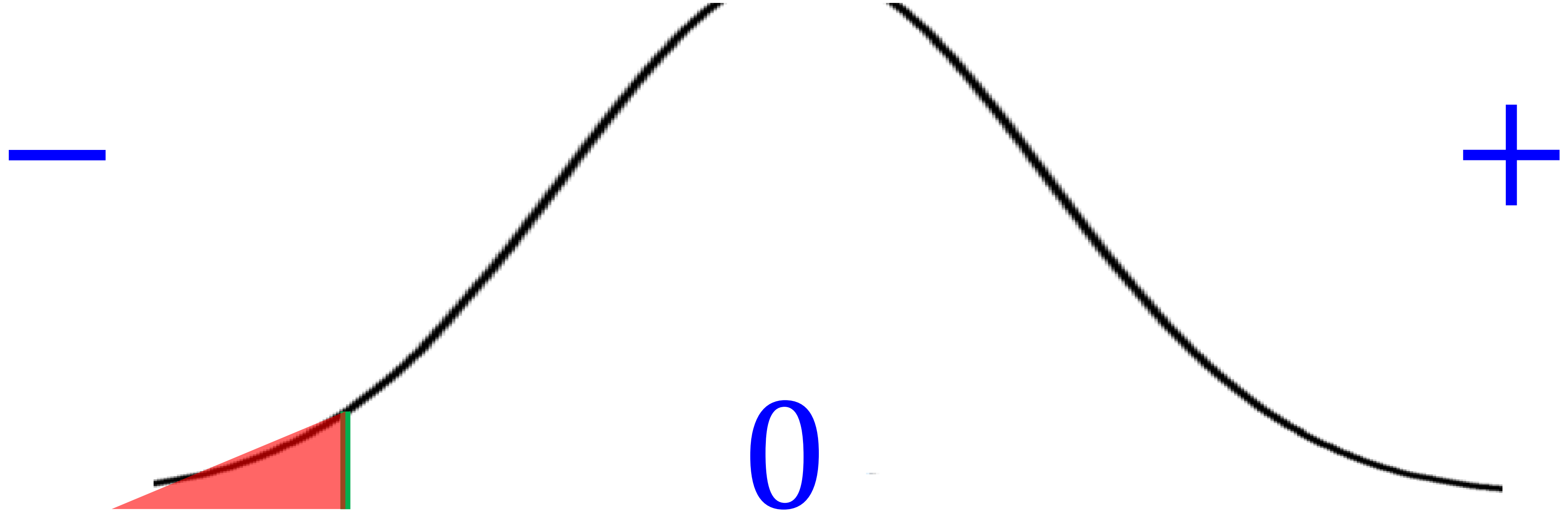
全く考えない

「片側検定」　　２　日給は30000円より少ない

これが「片側検定」

マイナスの領域に入ればいい

（大小の関係があることを検定したい）

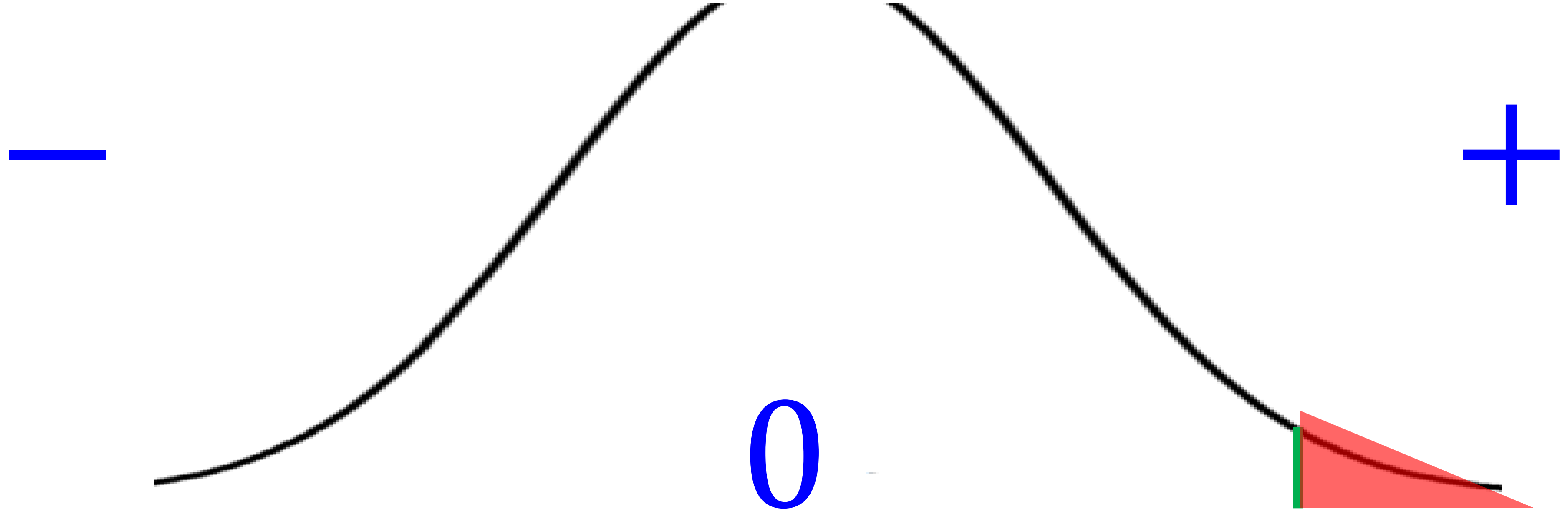


「片側検定」　　３　日給は30000円より多い

これが「片側検定」

プラスの領域に入ればいい

（大小の関係があることを検定したい）



今回は「日給30000円」

と言える？

「より少なくても」　「より多くても」

どっちでもいい！

だから、両側検定！

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「仮説」をたてる

H_0：「日給が30000円である」

H_1：「日給は30000円でない」

何回も言うけど

自分が言いたいことが「H_1」

「H_0」：証明したくないなって事

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

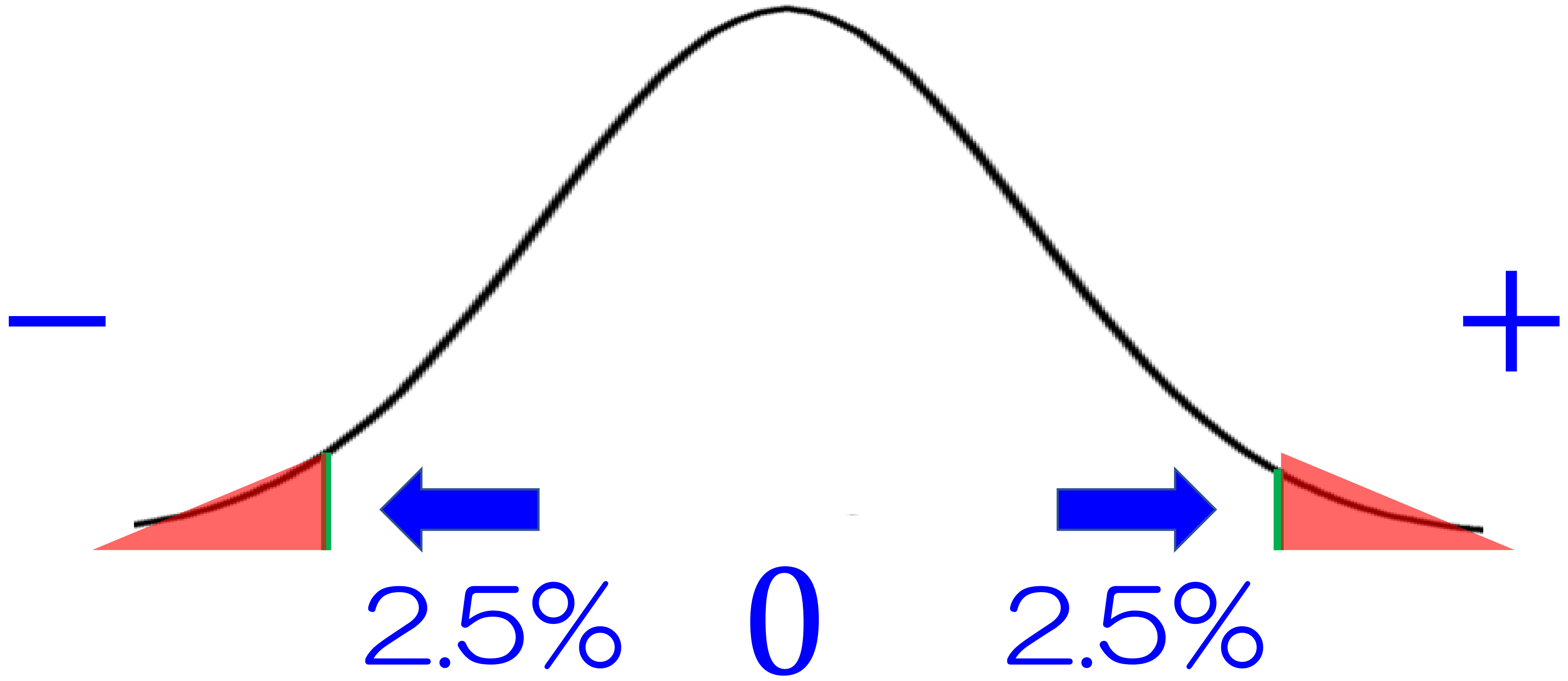
5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「両側検定」と「片側検定」

ここの面積（両側）が　「５%」

ということは片側は「2.5%」



「両側検定」と「片側検定」

ここの面積（両側）が　「5％」

ということは片側は「2.5％」

片側検定では　有意水準を２倍にしないとだめ

（そのままやると有意水準2.5％になってしまう）

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「ｔ値の求め方」

深く考えずにこれ！

$$\boxed{t = \frac{(\text{標本平均}) - (\text{仮説平均})}{\sqrt{\dfrac{\text{不偏分散}}{\text{データの個数}}}}}$$

不偏分散：＝VAR.S(データの範囲)

やってみよう

	日給	平均（仮説）	30000
Aくん	27149	標本平均	
Bくん	35800	平均の差	
Cさん	26382	有意水準	0.05
Dくん	28329	データ個数	10
Eさん	21988	境界値	
Fさん	27441		
Gくん	25195	不偏分散	
Hくん	20475	分母√の中	
Iさん	23173	分母の√	
Jさん	31090	ｔ値	

やってみよう

	日給	平均（仮説）	30000
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差	-3297.8
Cさん	26382	有意水準	0.05
Dくん	28329	データ個数	10
Eさん	21988	境界値	
Fさん	27441		
Gくん	25195	不偏分散	20138424.62
Hくん	20475	分母√の中	2013842.462
Iさん	23173	分母の√	1419.099173
Jさん	31090	ｔ値	2.323868594

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

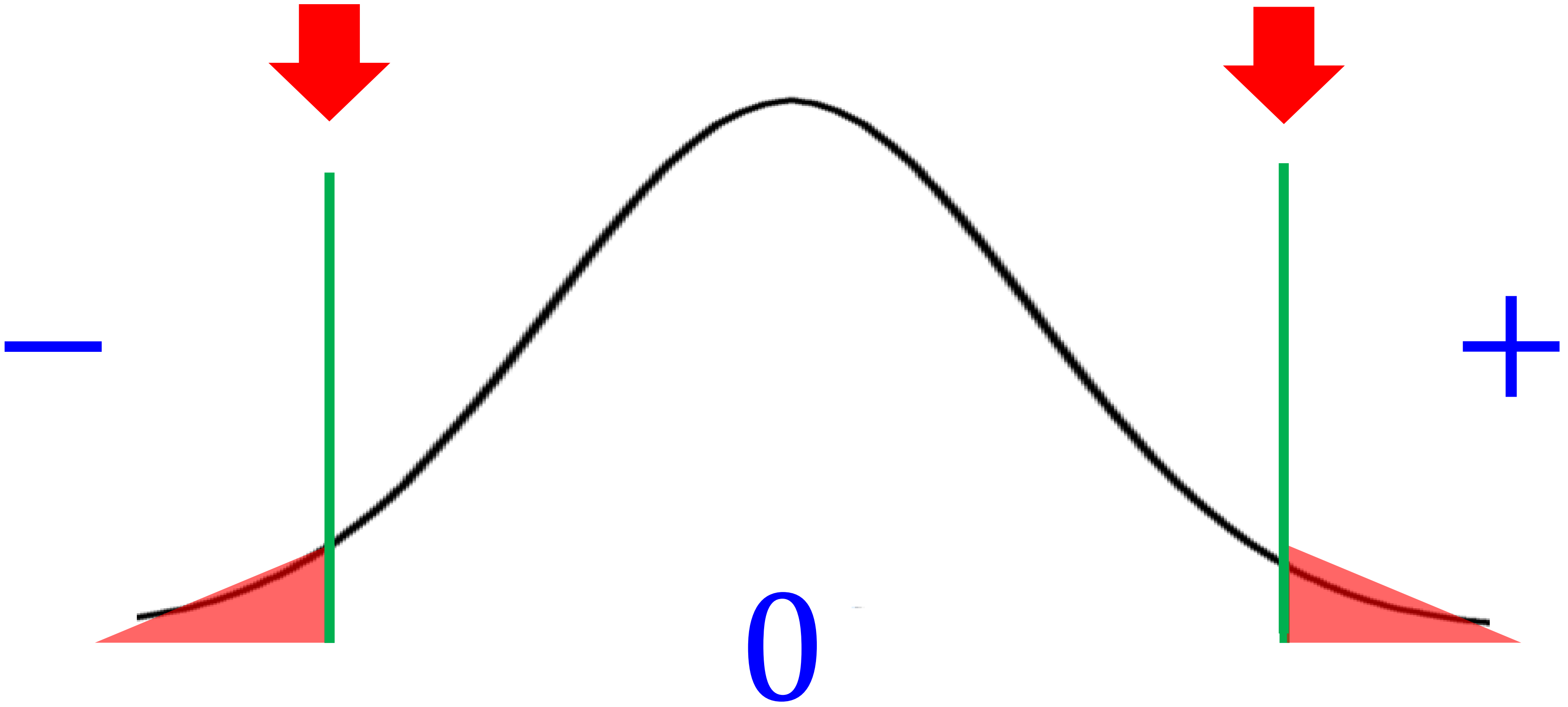
5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「ｔ値の境界値」

データの「5%」が入る

このラインのこと

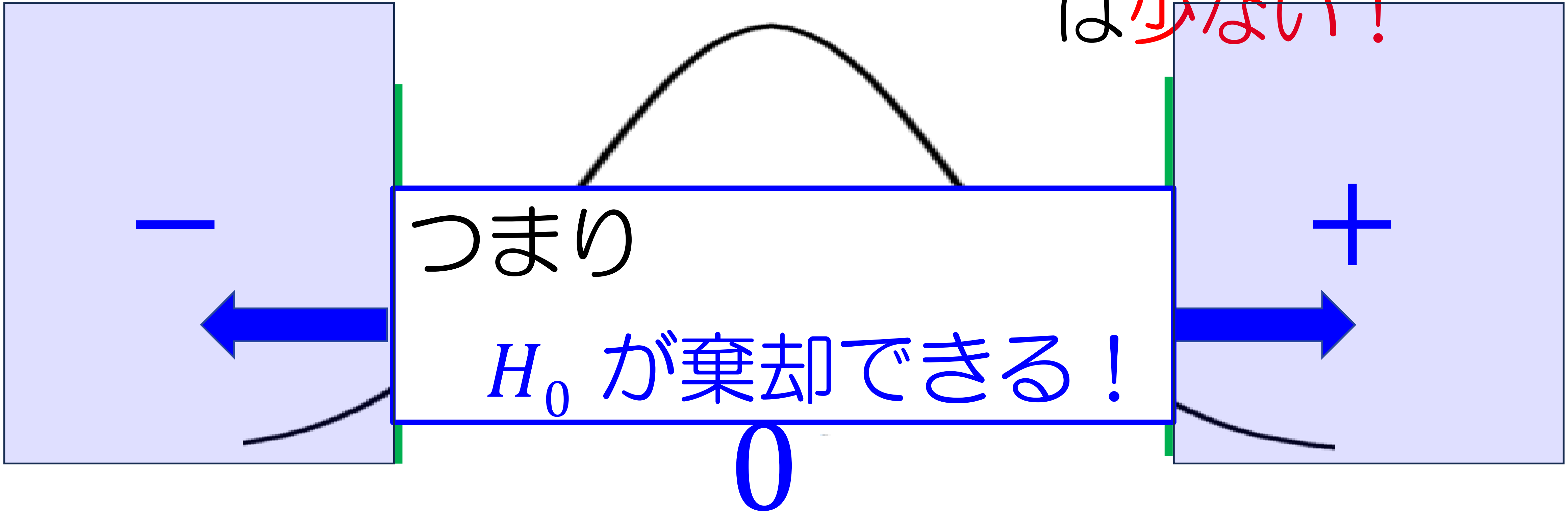


「ｔ値の境界値」

これより外側（値が大きい）と

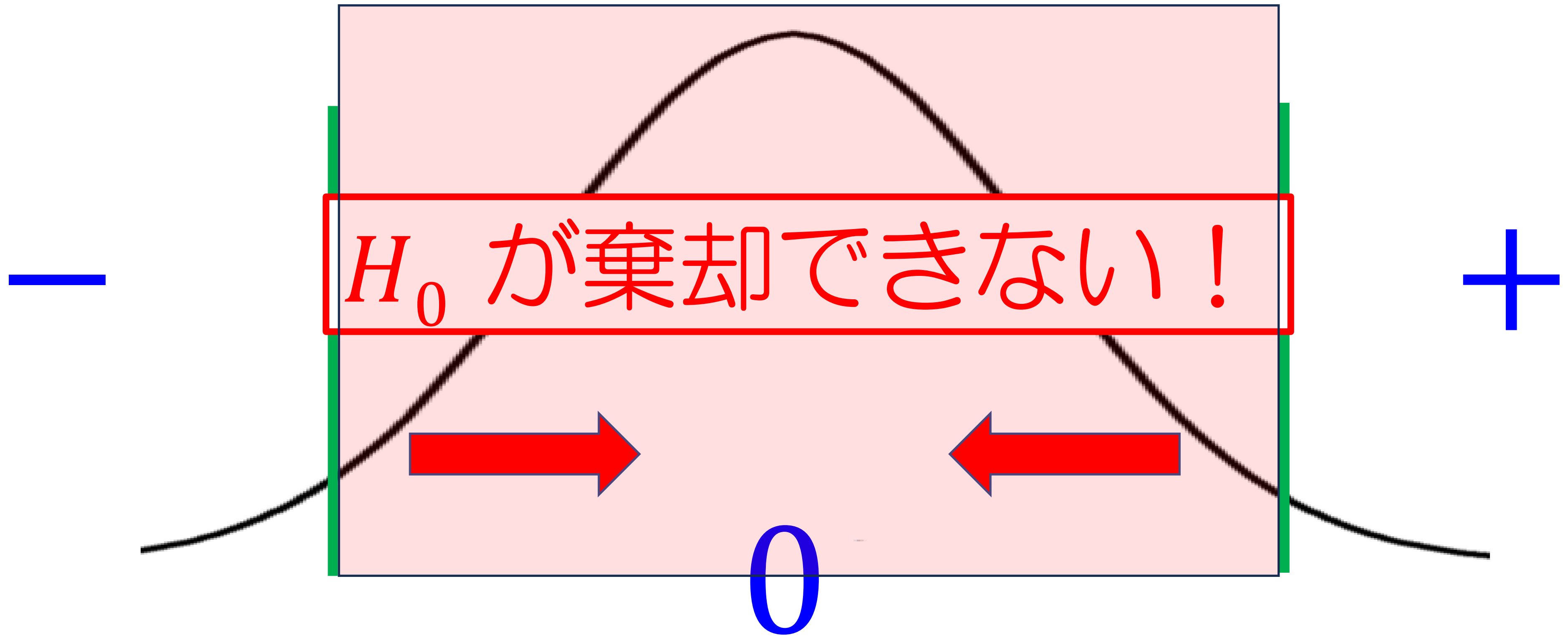
間違ってデータがここに入る確率

は少ない！



「ｔ値の境界値」

逆に内側（値が小さい）だと…



「ｔ値の境界値」

これも深く考えずに、

Excelに任せせましょう

＝ T.INV.2T(有意水準,自由度)

有意水準：0.05

自由度：データ個数－１

注意！

やってみよう

	日給	平均（仮説）	30000
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差	-3297.8
Cさん	26382	有意水準	0.05
Dくん	28329	データ個数	10
Eさん	21988	境界値	
Fさん	27441		
Gくん	25195	不偏分散	20138424.62
Hくん	20475	分母√の中	2013842.462
Iさん	23173	分母の√	1419.099173
Jさん	31090	t値	2.323868594

やってみよう

	日給	平均（仮説）	30000
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差	-3297.8
Cさん	26382	有意水準	0.05
Dくん	28329	データ個数	10
Eさん	21988	境界値	2.262157163
Fさん	27441		
Gくん	25195	不偏分散	20138424.62
Hくん	20475	分母√の中	2013842.462
Iさん	23173	分母の√	1419.099173
Jさん	31090	t値	2.323868594

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「ｔ値とｔ値の境界値」の比較

ｔ値の境界値と比べて

ｔ値（絶対値）が

大きい：H_0 を棄却し、H_1 を採用

小さい：H_0 を棄却できない

やってみよう

	日給	平均（仮説）	30000
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差	-3297.8
Cさん	26382	有意水準	0.05
Dくん	28329	データ個数	10
Eさん	21988	境界値	2.262157163
Fさん	27441		
Gくん	25195	不偏分散	20138424.62
Hくん	20475	分母√の中	2013842.462
Iさん	23173	分母の√	1419.099173
Jさん	31090	ｔ値	2.323868594

ｔ値の境界値 ≦ ｔ値

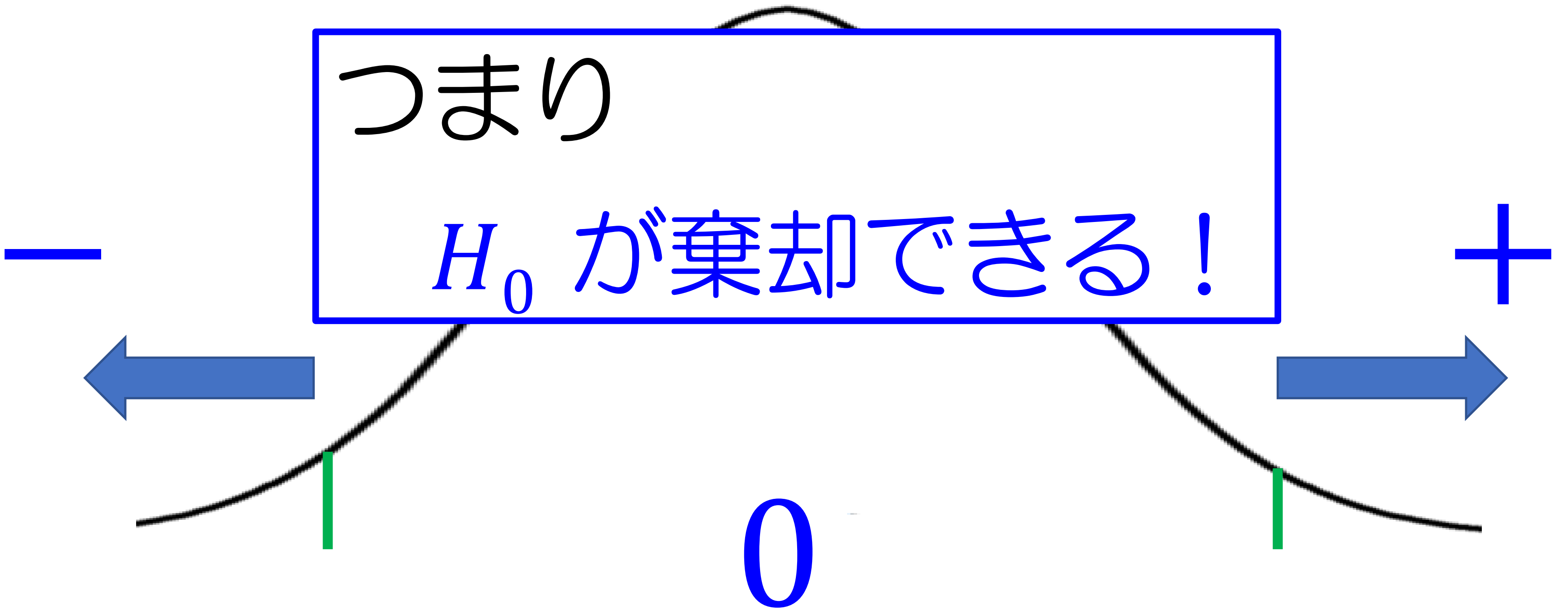
「ｔ値とｔ値の境界値」の比較

これより外側（値が大きい）と

間違ってデータがここに入る確率は少ない！

つまり

H_0が棄却できる！



今回の検定結果

「日給が30000円である」という仮定では

日給の平均が26702.2円という

データが得られる可能性は5%以下

（これが得られる可能性は低い）

「日給が30000円」という仮説が

間違っていると言っていい。

つまり「日給は30000円でない」と言える

Excel　シート

「居酒屋」

「B病院」

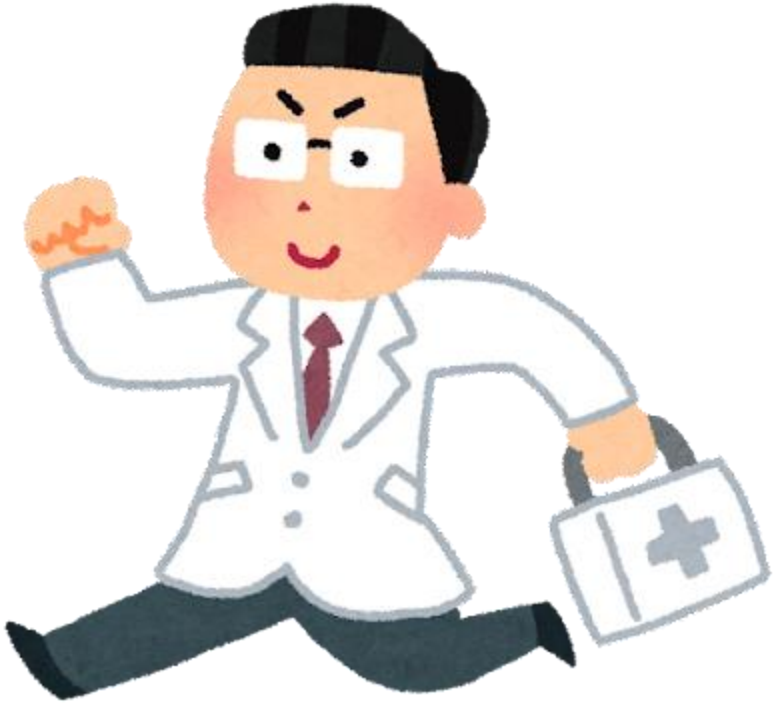
「C病院」

これは確実に
やっといて

隣の人と協力してやってみよう！

今日は、やることがいっぱいあるから、

ノンビリやらずにパパっと終わらせて！



終わってなくても

次へ行くよ！

そろそろ慣れてきたと思うので

「小テスト」やってみよう

いつものところの

「小テスト　１」から

- 時間：３０分
- 提出はいつものところへ

問題文をよく読んで、

H_0仮説　H_1仮説をたてて

「ｔ検定」のパターン

１　１つのデータしかないが

（母集団の平均はわかってる）

２　２つのデータがある

前回は「母集団の平均がわかってた」

（日給30000円とか、容量350ｍｌとか）

母集団の平均との比較

今度は「２つのデータの平均の比較」

やる流れは全く一緒

ｔ検定の流れ

１　「両側」か「片側」かを決める

２　仮説を立てる（H_0 H_1）

３　有意水準を決める（普通は0.05）

４　ｔ値（絶対値）を求める

５　ｔ値の境界値を求める

６　ｔ値とｔ値の境界値から判断する

前回の流れに追加の<u>新しい考え</u>

１　「対応のあるデータ」か

　　「対応のないデータ」か

２　片側検定の大小の判断

前回の流れに追加の新しい考え

１　「対応のあるデータ」か
　　「対応のないデータ」か

２　片側検定の大小の判断

「対応のあるデータ」

同じ対象で２回計測したデータの比較など

・同じ物を条件を変えて撮像

・同じ人で投薬前後の血圧

「対応のないデータ」

全く関係のない対象で２つの比較など

・２つの集団からの標本の平均

・JRと地下鉄の遅延時間

「対応のあるデータ」

同じ対象で２回計測したデータの比較など

- 同じ物を条件を変えて撮像
- 同じ人で投薬前後の血圧

今までと

同じやりかたで進めてOK！

「対応のあるデータ」

同じ対象で２回

計測したデータの比較

同じ人で

投薬前後の血圧

血圧は下がった

と言える？

患者No	投与前	投与後
1	150	129
2	146	135
3	152	149
4	139	125
5	149	141
6	137	135
7	141	137
8	152	140
9	141	126
10	138	134

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

今回は「血圧の大小について検定」だから

「片側検定」

H_0：投薬前後で差はなかった

H_1：投薬前後で血圧が下がった

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「対応のあるデータ」

「ｔ検定：一対の標本による平均の検定」

t-検定: 一対の標本による平均の検定ツール		
	変数 1	変数 2
平均	144.5	135.1
分散	35.38889	53.21111
観測数	10	10
ピアソン相関	0.569709	
仮説平均との差異	0	
自由度	9	
t	4.750411	
P(T<=t) 片側	0.000522	
t 境界値 片側	1.833113	
P(T<=t) 両側	0.001044	
t 境界値 両側	2.262157	

ｔ検定の流れ

１　「両側」か「片側」かを決める

２　仮説を立てる（H_0 H_1）

３　有意水準を決める（普通は0.05）

４　ｔ値（絶対値）を求める

５　ｔ値の境界値を求める

６　ｔ値とｔ値の境界値から判断する

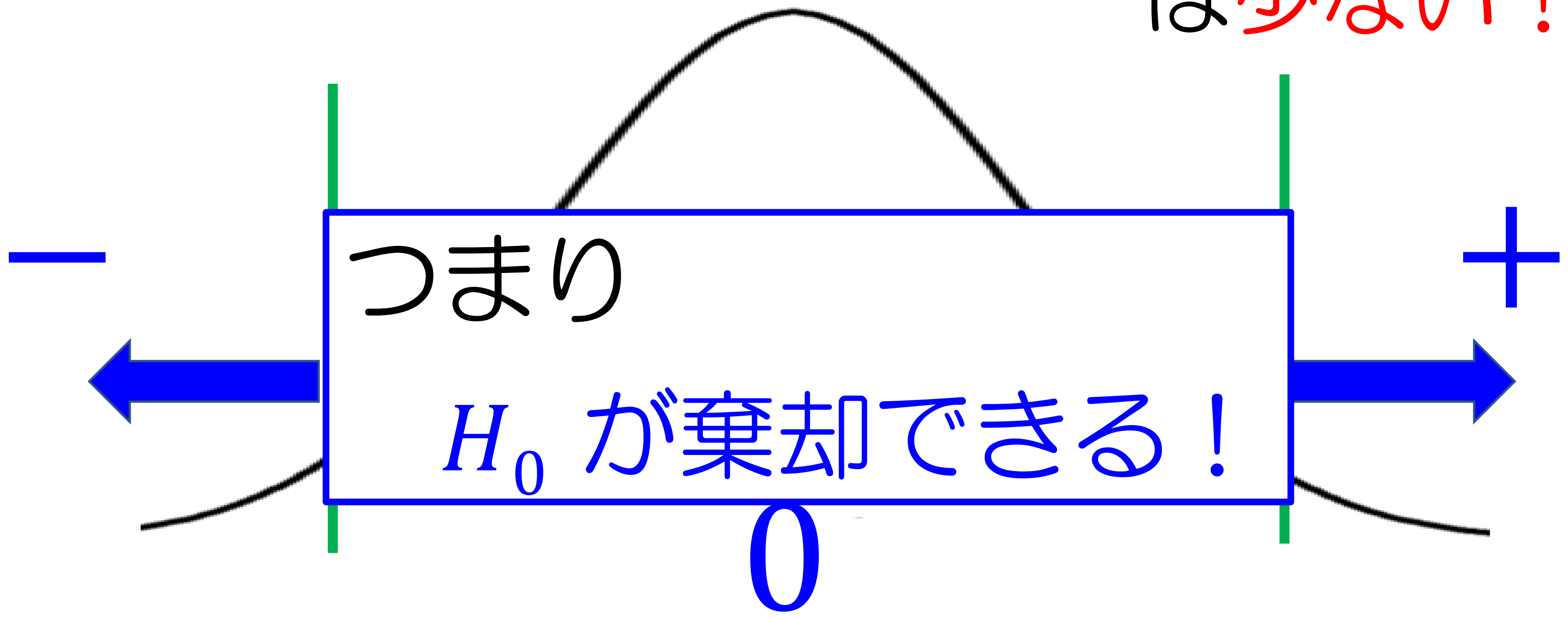
t-検定: 一対の標本による平均の検定ツール		
	変数 1	変数 2
平均	144.5	135.1
分散	35.38889	53.21111
観測数	10	10
ピアソン相関	0.569709	
仮説平均との差異	0	
自由度	9	
t	4.750411	
P(T<=t) 片側	0.000522	
t 境界値 片側	1.833113	
P(T<=t) 両側	0.001044	
t 境界値 両側	2.262157	

「ｔ値の境界値」

境界値より外側（ｔ値が大きい）と

間違ってデータがここに入る確率

は少ない！



「ｔ値とｔ値の境界値」の比較

ｔ値の境界値と比べて

ｔ値（絶対値）が大きい

H_0：投薬前後で差はなかった

H_1：投薬前後で血圧が下がった

もうひとつの見方

t-検定: 一対の標本による平均の検定ツール		
	変数 1	変数 2
平均	144.5	135.1
分散	35.38889	53.21111
観測数	10	10
ピアソン相関	0.569709	
仮説平均との差異	0	
自由度	9	
t	4.750411	
P(T<=t) 片側	0.000522	
t 境界値 片側	1.833113	
P(T<=t) 両側	0.001044	
t 境界値 両側	2.262157	

「P(T<=t)片側とα値」の比較

「P(T<=t)片側」とα値(有意水準)

を比べる

「P(T<=t)片側」の方が

大きい：H_0が棄却できない

小さい：H_0が棄却できる

t-検定: 一対の標本による平均の検定ツール

	変数 1	変数 2
平均	144.5	135.1
分散	35.38889	53.21111
観測数	10	10
ピアソン相関	0.569709	
仮説平均との差異	0	
自由度	9	
t	4.750411	
P(T<=t) 片側	0.000522	
t 境界値 片側	1.833113	
P(T<=t) 両側	0.001044	
t 境界値 両側	2.262157	

α＝0.05

「P(T<=t)片側とα値」の比較

「P(T<=t)片側」のほうが

α値(有意水準)より小さい

H_0：投薬前後で差はなかった

H_1：投薬前後で血圧が下がった

問題やってみよう！

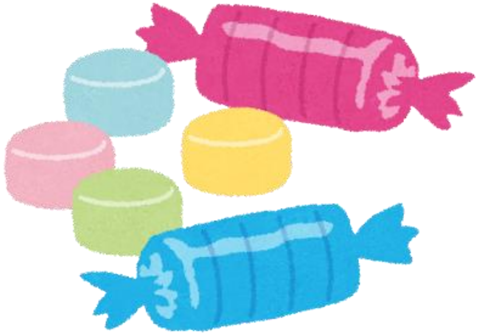
「医療統計　７回目　Excel」

問題１

あるダイエット法の効果を調べるために
10人の被験者で調査した。
このダイエット法は効果がある？
この結果を有意水準5%のｔ検定によっ
て検定して下さい。

問題２

「お菓子を食べたほうが良いのか」



「水を飲んだほうが良いのか」

「ストレッチをしたほうがよいのか」

テスト前にどうするべきかｔ検定！

仮説検定

１　カイ二乗検定

２　ｔ検定２

３　回帰分析

「対応のあるデータ」

同じ対象で２回計測したデータの比較など

- 同じ物を条件を変えて撮像
- 同じ人で投薬前後の血圧

「対応のないデータ」

全く関係のない対象で２つの比較など

- ２つの集団からの標本の平均
- JRと地下鉄の遅延時間

今までは

等分散：ばらつきが等しい

「母集団が等分散である」と仮定

して検定してた

「対応のないデータ」

全く関係のない対象で２つの比較など

・２つの集団からの標本の平均

・JRと地下鉄の遅延時間

「対応のないデータ」

全く関係のない対象で２つの比較

・２つの集団からの標本の平均

関係のない対象だから

「等分散である」かわからない！

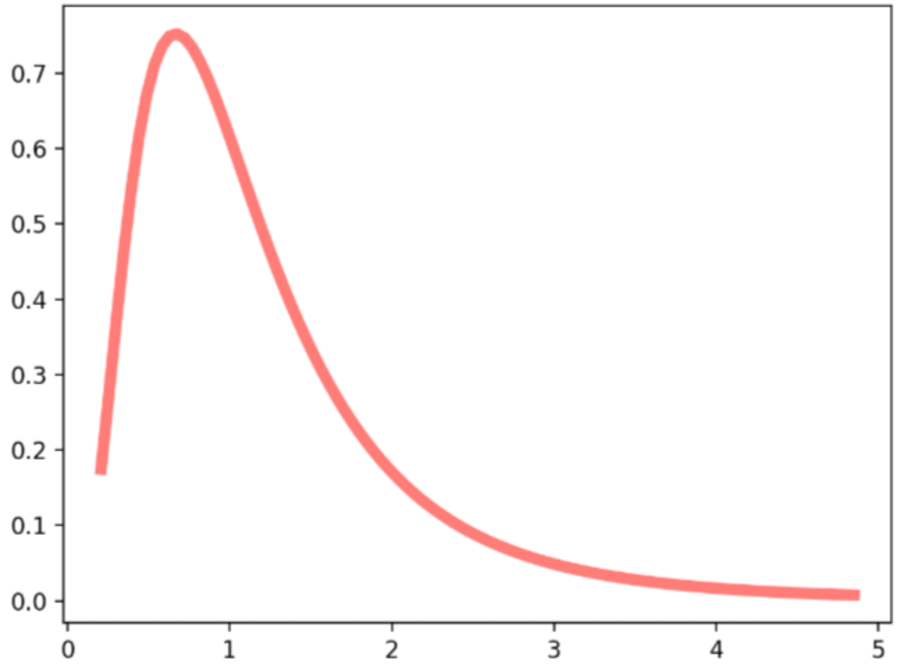
まず、等分散かどうか調べる必要がある！

「Ｆ検定」

全く関係のない母集団が

等分散かどうか調べる検定

等分散：ばらつきが等しい



こんな分布になるけど

別に知らなくていい！

「Ｆ検定」

JR	地下鉄
19.3	23.4
25.1	22.6
15.9	17.4
21.5	15.7
20.5	20.9
18.8	18.1
16.9	16.2

ＪＲと地下鉄の
遅延時間に
差はあるのか

「データ分析」
「Ｆ検定」から

こんなのが出るはず

F-検定: 2 標本を使った分散の検定		
	変数 1	変数 2
平均	19.71428571	19.18571429
分散	9.381428571	9.631428571
観測数	7	7
自由度	6	6
観測された分散比	0.974043311	
P(F<=f) 片側	0.487674236	
F 境界値 片側	0.233434021	

見るのはココ！

F-検定: 2 標本を使った分散の検定		
	変数 1	変数 2
平均	19.71428571	19.18571429
分散	9.381428571	9.631428571
観測数	7	7
自由度	6	6
観測された分散比	0.974043311	
P(F<=f) 片側	0.487674236	
F 境界値 片側	0.233434021	

「Ｆ検定」

「P(F<=f)片側」とα値(有意水準)

を比べる

「P(F<=f)片側」の方が

大きい：分散が等しい

小さい：分散は等しくない

F-検定: 2 標本を使った分散の検定		
	変数 1	変数 2
平均	19.71428571	19.18571429
分散	9.381428571	9.631428571
観測数	7	7
自由度	6	6
観測された分散比	0.974043311	
P(F<=f) 片側	0.487674236	
F 境界値 片側	0.233434021	

$\boxed{\alpha = 0.05}$

この２つは等分散であると言える

「Ｆ検定」

F検定の結果

データに対応がなく、等分散と仮定できる

⇒　スチューデントのｔ検定

データに対応がなく、等分散と仮定できない

⇒　ウェルチのｔ検定

最近は、どっちにしろ「ウェルチのｔ検定」
でいいんじゃない？って説もある

ここから「ｔ検定」をスタート

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

今回は「遅延時間に差があるのか検定」

「両側検定」

H_0：遅延時間に差がない

H_1：遅延時間に差がある

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「対応のあるデータ」

「ｔ検定：<u>等分散を仮定</u>した平均の検定」

t-検定: 等分散を仮定した 2 標本による検定		
	変数 1	変数 2
平均	19.71428571	19.18571429
分散	9.381428571	9.631428571
観測数	7	7
プールされた分散	9.506428571	
仮説平均との差異	0	
自由度	12	
t	0.320722192	
P(T<=t) 片側	0.376968679	
t 境界値 片側	1.782287556	
P(T<=t) 両側	0.753937358	
t 境界値 両側	2.17881283	

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

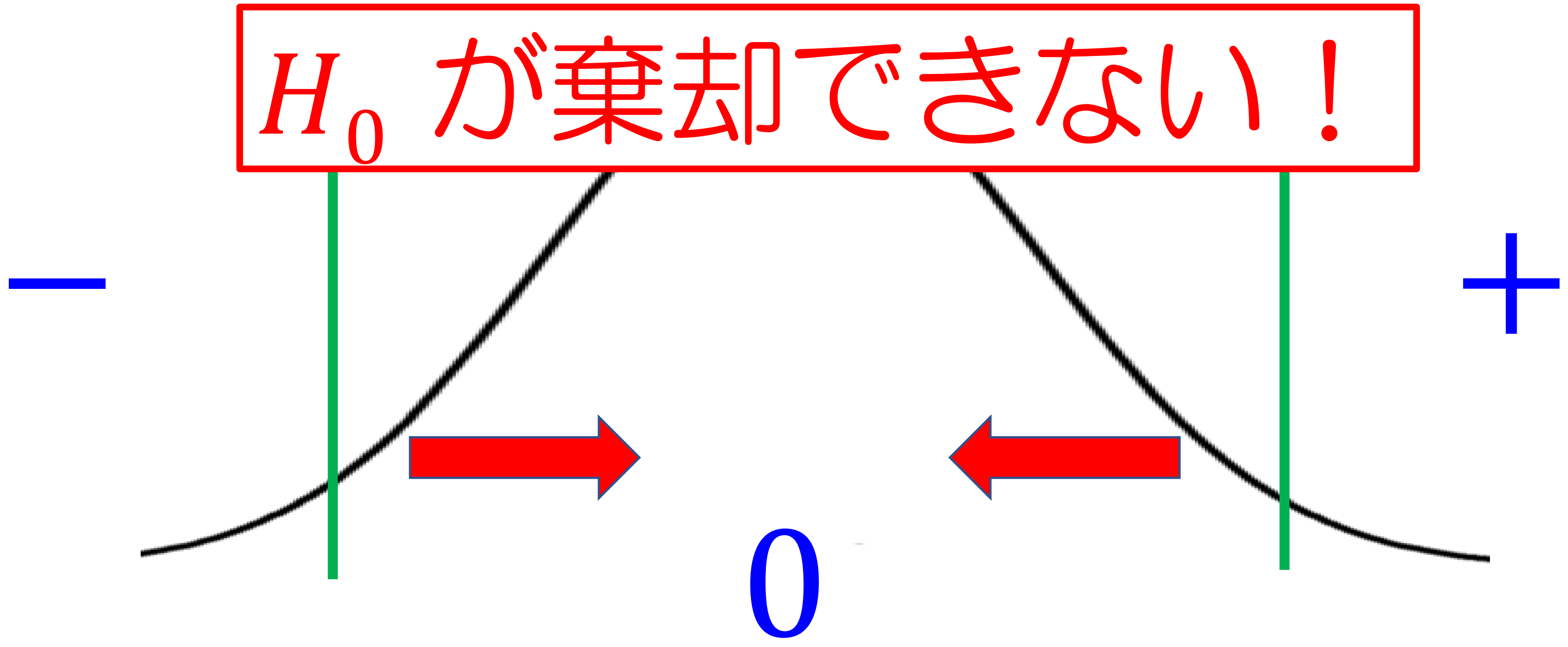
5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

t-検定: 等分散を仮定した 2 標本による検定		
	変数 1	変数 2
平均	19.71428571	19.18571429
分散	9.381428571	9.631428571
観測数	7	7
プールされた分散	9.506428571	
仮説平均との差異	0	
自由度	12	
t	0.320722192	
P(T<=t) 片側	0.376968679	
t 境界値 片側	1.782287556	
P(T<=t) 両側	0.753937358	
t 境界値 両側	2.17881283	

「ｔ値の境界値」

ｔ値が、境界値より小さい（内側）



「ｔ値とｔ値の境界値」の比較

ｔ値の境界値と比べて

ｔ値（絶対値）が小さい

H_0：遅延時間に差はない ←棄却できない！

H_1：遅延時間に差がある

遅延時間に差があるとは言い切れない！

t-検定: 等分散を仮定した 2 標本による検定

	変数 1	変数 2
平均	19.71428571	19.18571429
分散	9.381428571	9.631428571
観測数	7	7
プールされた分散	9.506428571	
仮説平均との差異	0	
自由度	12	
t	0.320722192	
P(T<=t) 片側	0.376968679	
t 境界値 片側	1.782287556	
P(T<=t) 両側	0.753937358	
t 境界値 両側	2.17881283	

$\alpha = 0.05$

「P(T<=t)片側とα値」の比較

「P(T<=t)片側」のほうが

α値(有意水準)より大きい

H_0：遅延時間に差はない　←棄却できない！

H_1：遅延時間に差がある

<u>遅延時間に差があるとは言い切れない！</u>

もし、F検定の結果が

「等分散でない」だったら

どうしよう？

そんなの簡単！

「対応のあるデータ」

「ｔ検定：分散が等しくないと仮定～」

をやればいいだけ！！

分散が等しくないパターン

問題やってみよう！！

「並ぶ人数の調査」を行いました。

アトラクションの人気に差があると

言えるのは？

タワー・オブ・テラー

73 94 89 92 49 28 99 75 9

79 31

スチールドラゴン

33 22 43 30 33 63 46 57 25

ハリウッド・ドリーム

61 60 67 79 50 60 79 75 55

67 69 41

前回の流れに追加の新しい考え

１　「対応のあるデータ」か
　　「対応のないデータ」か

２　片側検定の大小の判断

「片側検定の大小の判断」

H_0：ＡとＢの平均は等しい

H_1：Ａの方がＢより平均が大きい

H_0：ＡとＢの平均は等しい

H_1：Ａの方がＢより平均が小さい

の２つの仮説パターンが立てられる

例えば

「日給が30000円と仮定」(H_0)すると

対立仮説(H_1)は３通りできる

１　日給は30000円でない

２　日給は30000円より少ない

３　日給は30000円より多い

「片側検定の大小の判断」

H_0：日給は30000円である
H_1：日給は30000円より大きい

H_0：日給は30000円である
H_1：日給は30000円より小さい

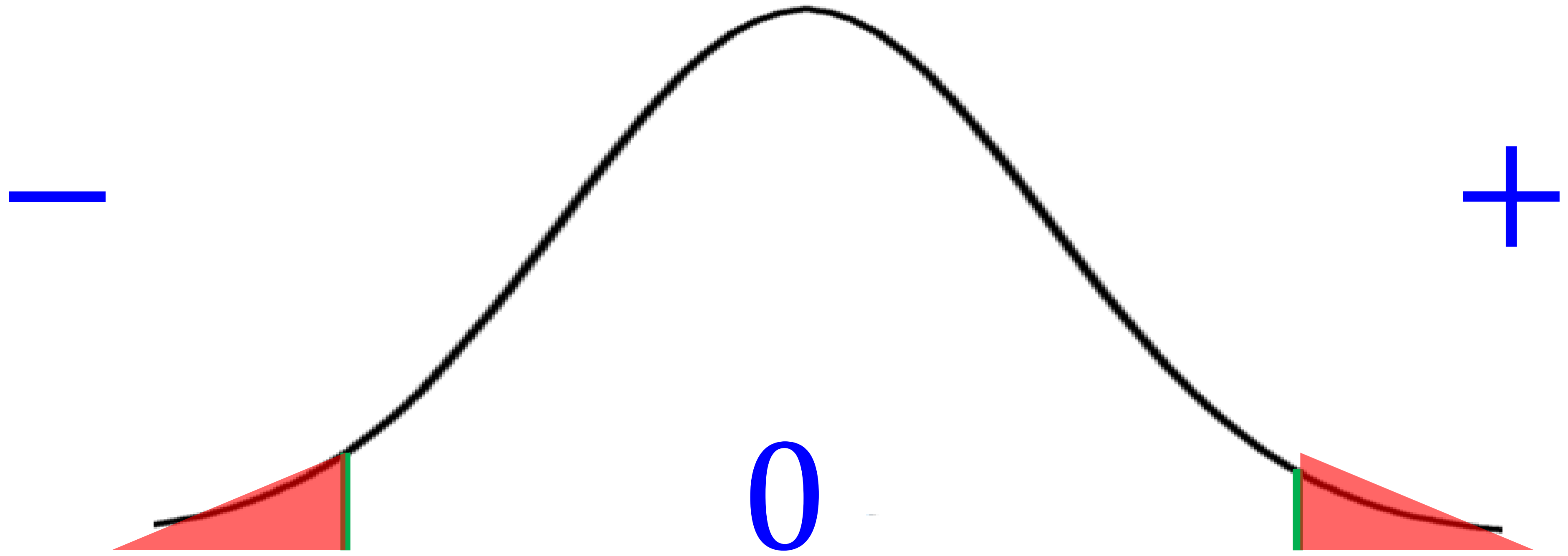
の２つの仮説パターンが立てられる

「両側検定」　　　　　1　日給は30000円でない

これが「両側検定」

どっちに入ってもいい

（多い少ないは全く考えない）

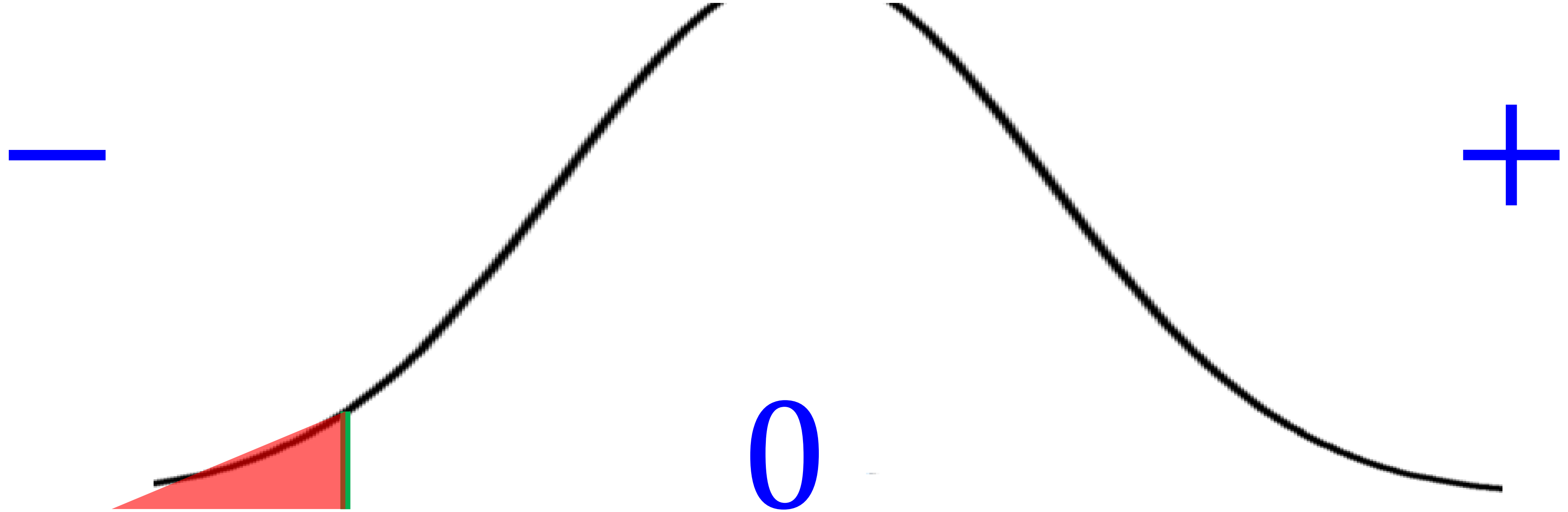


「片側検定」　　２　日給は30000円より少ない

これが「片側検定」

マイナスの領域に入ればいい

（大小の関係があることを検定したい）

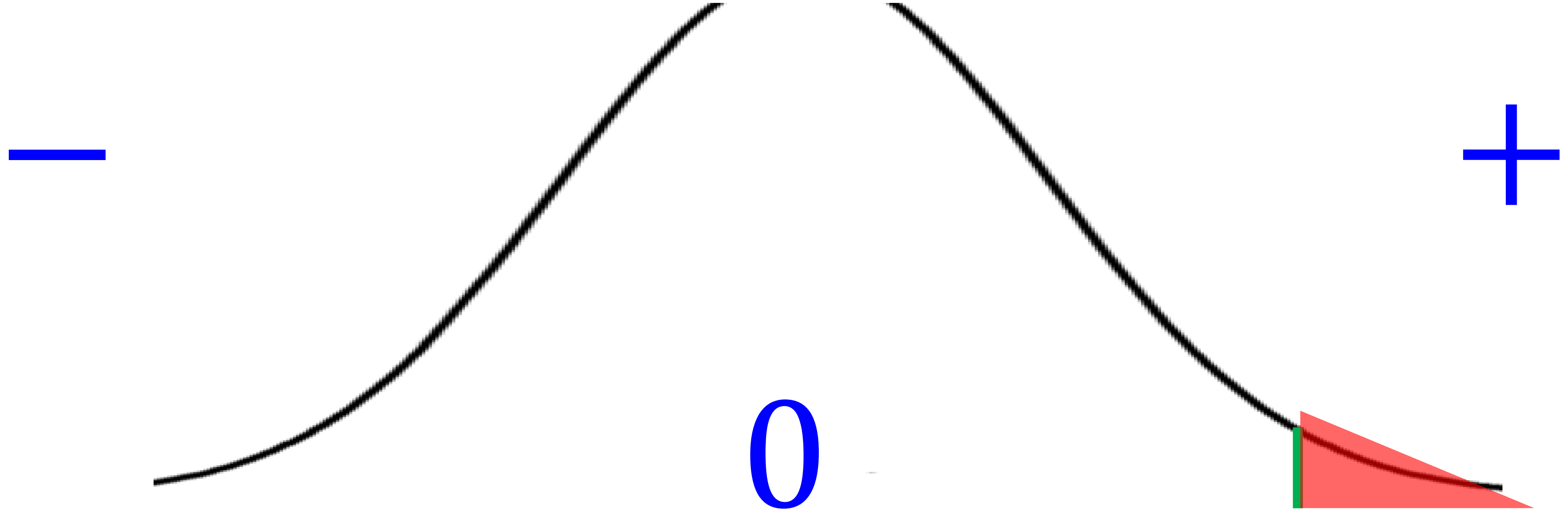


「片側検定」　　　３　日給は30000円より多い

これが「片側検定」

プラスの領域に入ればいい

（大小の関係があることを検定したい）



一緒にやってみよう

- 「性別」と「自宅か下宿」
- 「性別」と「本」

有意にどんな差があるのか？

一緒にやってみよう

- 「性別」と「自宅か下宿」
- 「性別」と「本」

有意にどんな差があるのか？

まずは、Ｆ検定

「性別」と「自宅か下宿」のＦ検定

F-検定: 2 標本を使った分散の検定		
	変数 1	変数 2
平均	1.515	1.52
分散	0.251030151	0.250854
観測数	200	200
自由度	199	199
観測された分散比	1.000701122	
P(F<=f) 片側	0.498030298	
F 境界値 片側	1.263340341	

$\boxed{\alpha = 0.05}$

この２つは等分散であると言える

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

今回は「性別と住居はどんな関係かの検定」

「片側検定」

ｔ検定の流れ

１　「両側」か「片側」かを決める

２　仮説を立てる（H_0 H_1）

３　有意水準を決める（普通は0.05）

４　ｔ値（絶対値）を求める

５　ｔ値の境界値を求める

６　ｔ値とｔ値の境界値から判断する

H_0：男性と女性の住居の平均は等しい

H_1：男性の方が女性より平均が大きい

つまり、男性の方が「下宿が多い」

H_0：男性と女性の住居の平均は等しい

H_1：男性の方が女性より平均が小さい

つまり、男性の方が「実家が多い」

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「性別」と「自宅か下宿」の t 検定

t-検定: 等分散を仮定した2標本による検定		
	変数 1	変数 2
平均	1.608247423	1.436893204
分散	0.240764605	0.248429469
観測数	97	103
プールされた分散	0.244713171	
仮説平均との差異	0	
自由度	198	
t	2.44824914	
P(T<=t) 片側	0.007612973	
t 境界値 片側	1.652585784	
P(T<=t) 両側	0.015225946	
t 境界値 両側	1.972017478	

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「性別」と「自宅か下宿」の t 検定

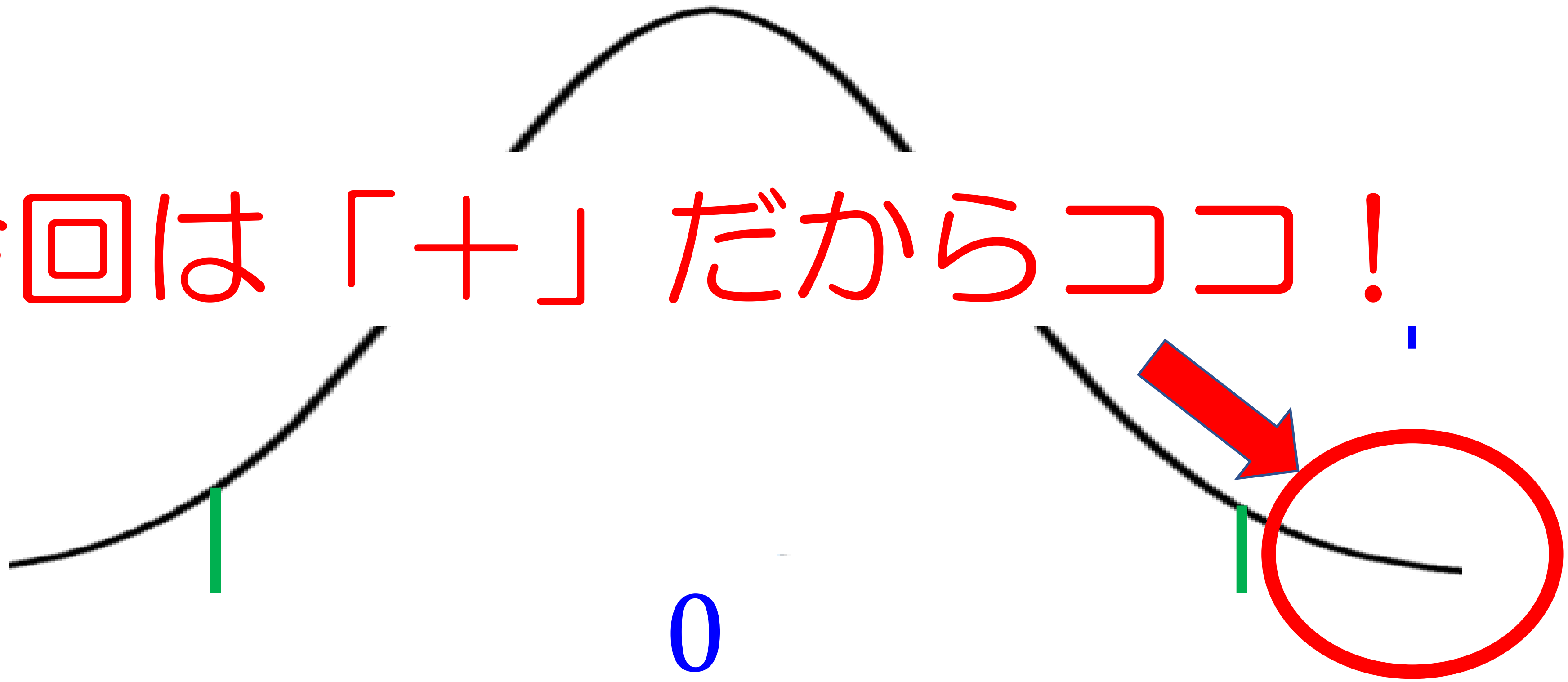
t-検定: 等分散を仮定した 2 標本による検定		
	変数 1	変数 2
平均	1.608247423	1.436893204
分散	0.240764605	0.248429469
観測数	97	103
プールされた分散	0.244713171	
仮説平均との差異	0	
自由度	198	
t	2.44824914	
P(T<=t) 片側	0.007612973	
t 境界値 片側	1.652585784	
P(T<=t) 両側	0.015225946	
t 境界値 両側	1.972017478	

「ｔ値の境界値」

これより外側（絶対値が大きい）

H_0が棄却できる！

今回は「＋」だからココ！



H_0：男性と女性の住居の平均は等しい

H_1：男性の方が女性より平均が大きい

つまり、男性の方が「下宿が多い」

H_0：男性と女性の住居の平均は等しい

H_1：男性の方が女性より平均が小さい

つまり、男性の方が「実家が多い」

H_1：男性の方が女性より平均が大きい

「検定結果」

男性の方が「下宿が多い」

女性の方が「実家が多い」

一緒にやってみよう

- 「性別」と「自宅か下宿」
- 「性別」と「本」

有意にどんな差があるのか？

まずは、Ｆ検定

「性別」と「本」のＦ検定

F-検定: 2 標本を使った分散の検定		
	変数 1	変数 2
平均	1.515	1.565
分散	0.251030151	0.24701
観測数	200	200
自由度	199	199
観測された分散比	1.016275048	
P(F<=f) 片側	0.454727421	
F 境界値 片側	1.263340341	

$\alpha = 0.05$

この２つは等分散であると言える

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

今回は「性別と本はどんな関係かの検定」

「片側検定」

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

H_0：男性と女性の本の平均は等しい

H_1：男性の方が女性より平均が大きい

つまり、男性の方が「本を読まない」

H_0：男性と女性の本の平均は等しい

H_1：男性の方が女性より平均が小さい

つまり、男性の方が「本を読む」

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「性別」と「本」のｔ検定

t-検定: 等分散を仮定した２標本による検定		
	変数 1	変数 2
平均	1.484536082	1.640776699
分散	0.252362543	0.232438607
観測数	97	103
プールされた分散	0.242098697	
仮説平均との差異	0	
自由度	198	
t	-2.244332507	
P(T<=t) 片側	0.01295881	
t 境界値 片側	1.652585784	
P(T<=t) 両側	0.02591762	
t 境界値 両側	1.972017478	

ｔ検定の流れ

1　「両側」か「片側」かを決める

2　仮説を立てる（H_0 H_1）

3　有意水準を決める（普通は0.05）

4　ｔ値（絶対値）を求める

5　ｔ値の境界値を求める

6　ｔ値とｔ値の境界値から判断する

「性別」と「本」のｔ検定

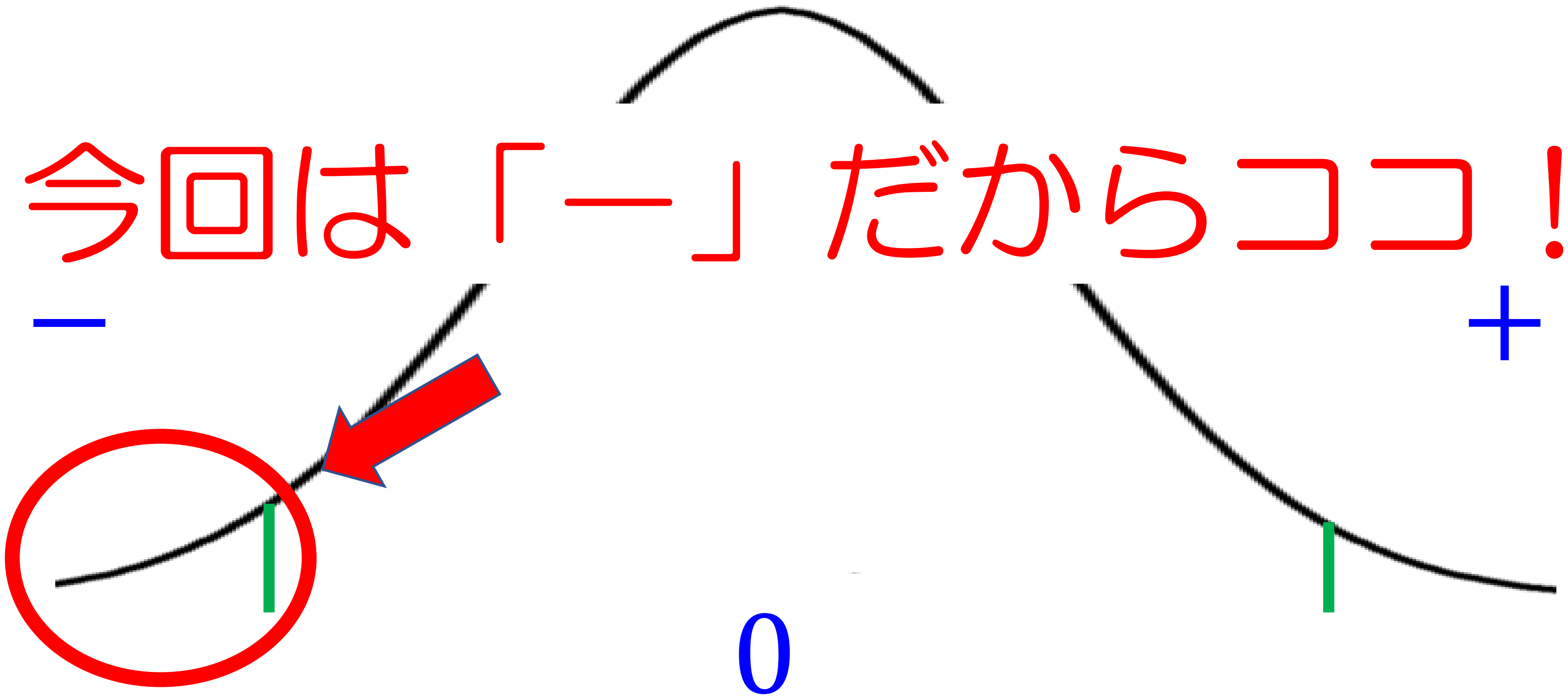
t-検定: 等分散を仮定した２標本による検定		
	変数 1	変数 2
平均	1.484536082	1.640776699
分散	0.252362543	0.232438607
観測数	97	103
プールされた分散	0.242098697	
仮説平均との差異	0	
自由度	198	
t	-2.244332507	
P(T<=t) 片側	0.01295881	
t 境界値 片側	1.652585784	
P(T<=t) 両側	0.02591762	
t 境界値 両側	1.972017478	

「ｔ値の境界値」

これより外側（絶対値が大きい）

H_0 が棄却できる！

今回は「ー」だからココ！



H_0：男性と女性の本の平均は等しい

H_1：男性の方が女性より平均が大きい

つまり、男性の方が「本を読まない」

H_0：男性と女性の本の平均は等しい

H_1：男性の方が女性より平均が小さい

つまり、男性の方が「本を読む」

「検定結果」

つまり、

男性の方が「本を３冊以上読む」

女性の方が「本を３冊以上読まない」

H_1：男性の方が女性より平均が小さい