

仮説検定（統計的検定）

1 カイ二乗検定

2 t 検定

3 回帰分析



「仮説検定」の基本的な考え方



「たぶん、こうなるはずや！」

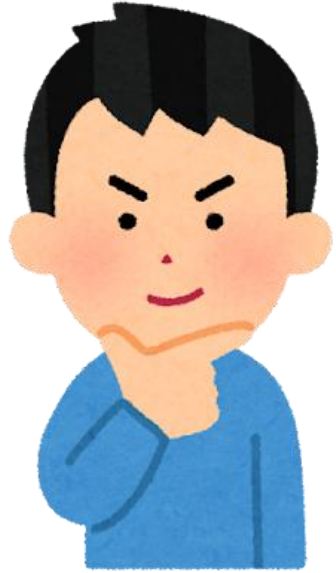
「そんなはずはない！」



これを判断するには、自分の中に**判断基準**が必要

判断基準があるから「ありえない」と感じる！

「仮説検定」の基本的な考え方



「たぶん、こうなるはずや！」

「そんなはずはない！」



その判断基準こそが「有意水準」

「仮説検定」の基本的な考え方

その判断基準こそが「有意水準」

「偶然による誤判定を

どの程度まで許す？」

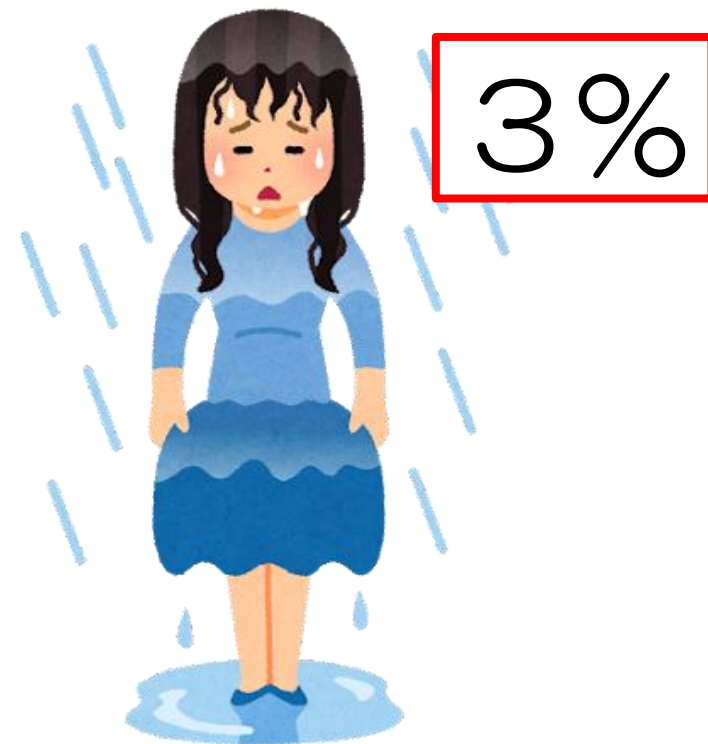
をあらかじめ決める数値です。

降水確率が3%だから、傘を持ってこなかった
つまり、「雨は降らない」と判断した

逆に言うと

3%の可能性で

雨に濡れる**危険**がある！



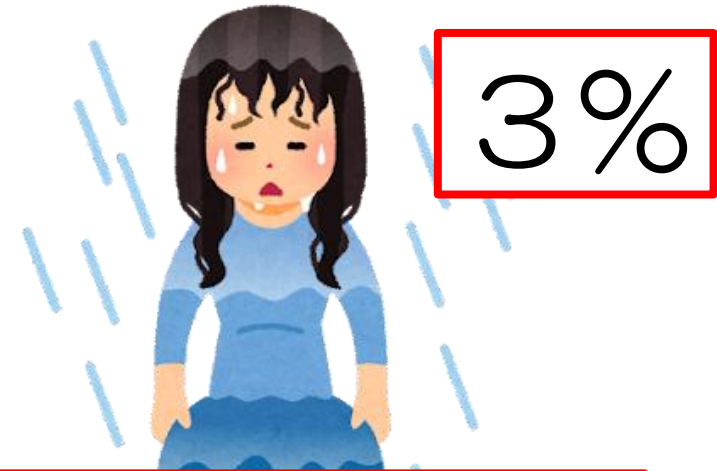
「今日は傘を持って出かけたほうがいい！」

という仮説に対して

「3%」の危険をおかしても、

仮説が間違っていると思うから

傘を持ってこなかった



ある仮説が有意に（明らかに）間違っていると

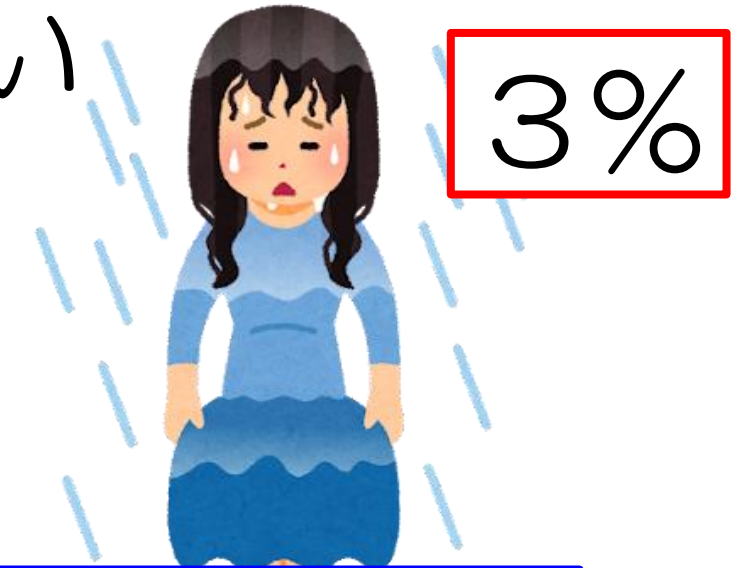
判断する基準を確率で表す！

天気予報の降水確率が「3%」のとき

自分の判断の基準になる降水確率が

「5%」なら 傘を持っていかない

「1%」なら 傘をもっていく



「有意水準」(α) 普通「5%」か「1%」

仮説検定

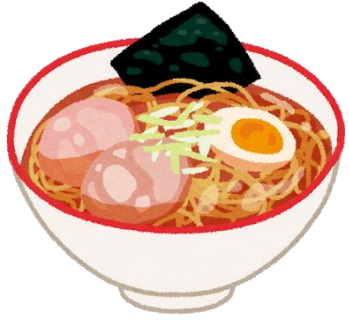
1 カイ二乗検定

2 t 検定

3 回帰分析



今回の「t検定」



「美味しいか」

「美味しくくない」

の2択

じゃなくて



「**点数**」をつけてもらって

その平均を比較する

今回の「t検定」

結果は複数

t検定

条件
2択

男性？ 女性？

何点つけた？

性別で点数に差があるのか？

今回の「t検定」

例えば、

朝食を「食べる人」と「食べない人」で

テストの点数に差があるのか？



結果がバラバラだから、割合は無理！

結果の**平均を比較する**（検定）

第3の要因を見つけるための統計手法（一部）

多変量解析：複数の変数を同時分析して、ある変数がある他の変数に与える影響を評価する

回帰分析：ある変数を目的変数とし、他の変数を説明変数として、目的変数が説明変数によってどう説明できるかモデル化（数式化）

「t検定」のパターン

- 1 一つのデータしかないが
(母集団の平均はわかっている)
- 2 二つのデータがある



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 一つのデータしかないが
(母集団の平均はわかっている)

2 二つのデータがある

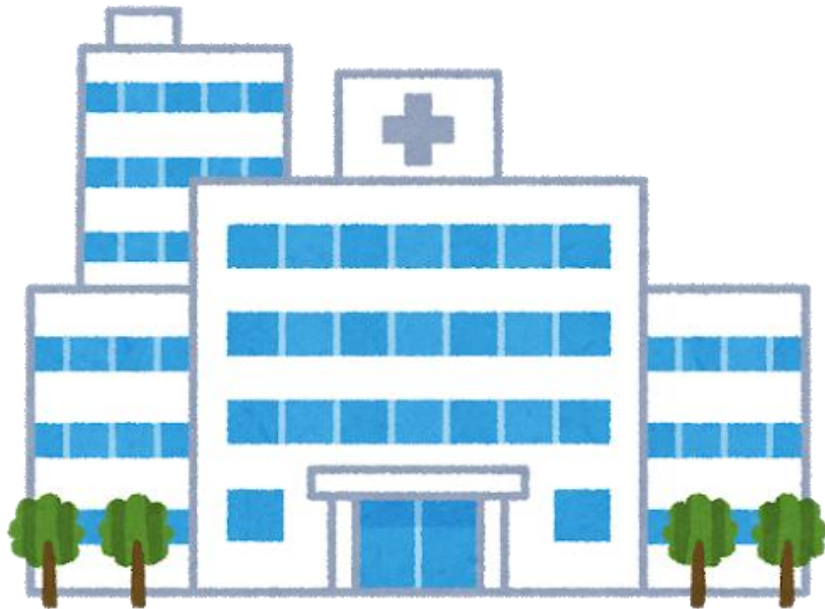


ある病院では、放射線技師の日給は

「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」かを定める
- 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」かを定める
- 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」？

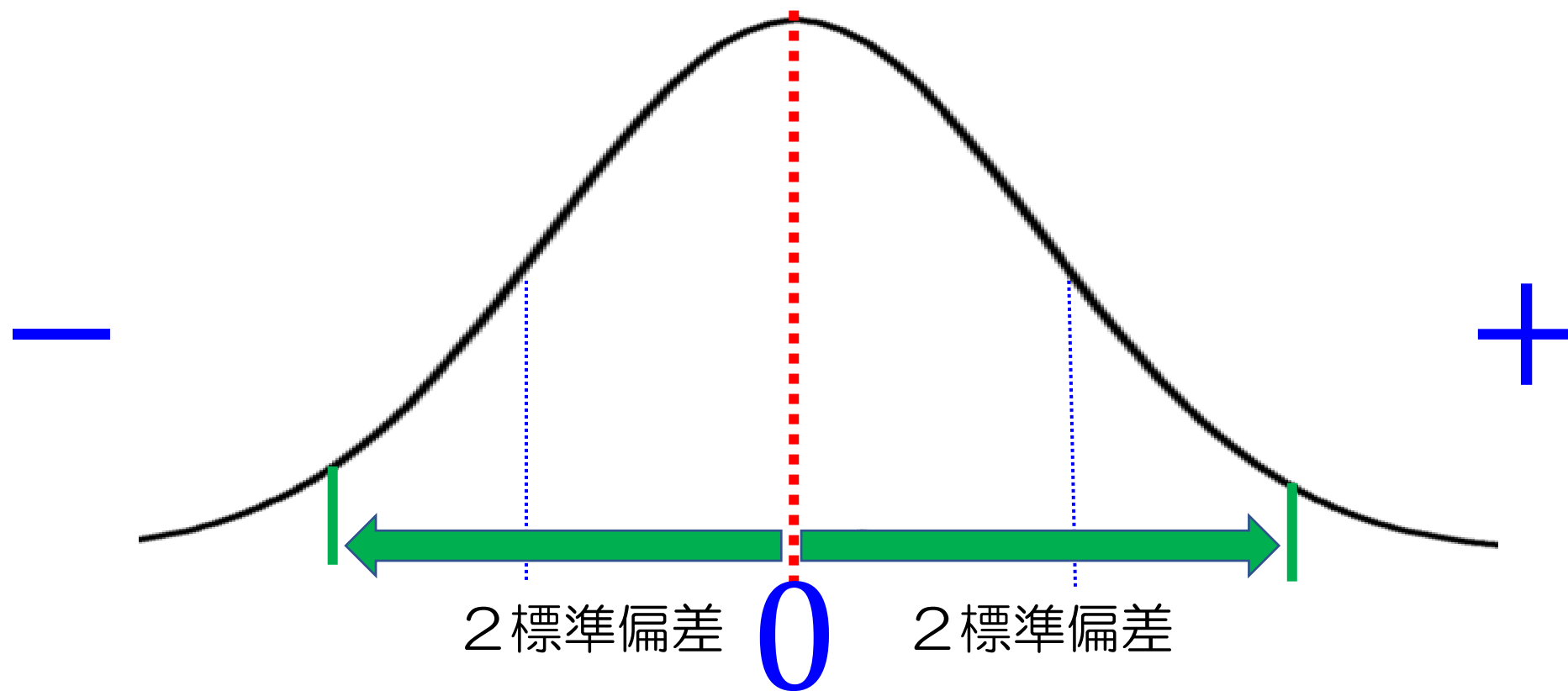


思いだしてみて

「正規分布」と「標準偏差」

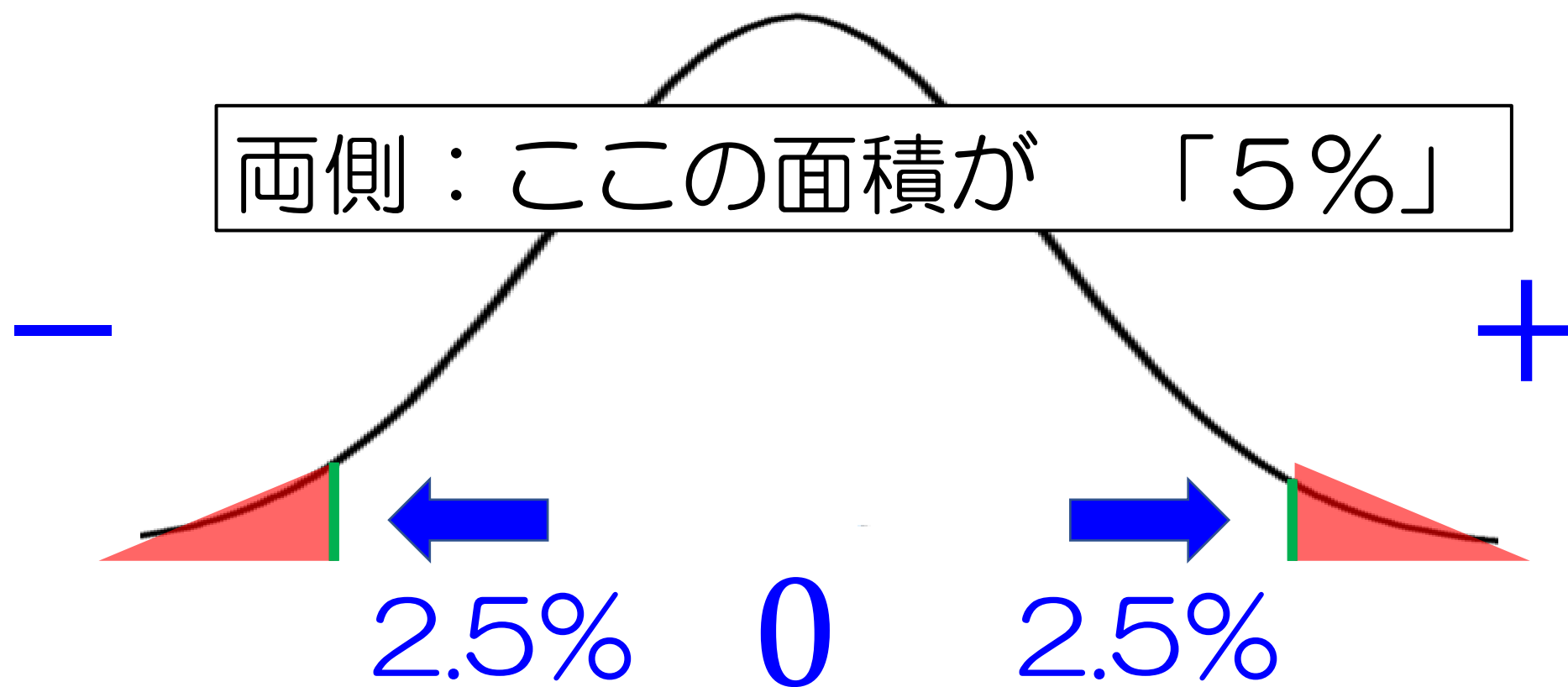
「正規分布 と 標準偏差」

- 平均値 \pm 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



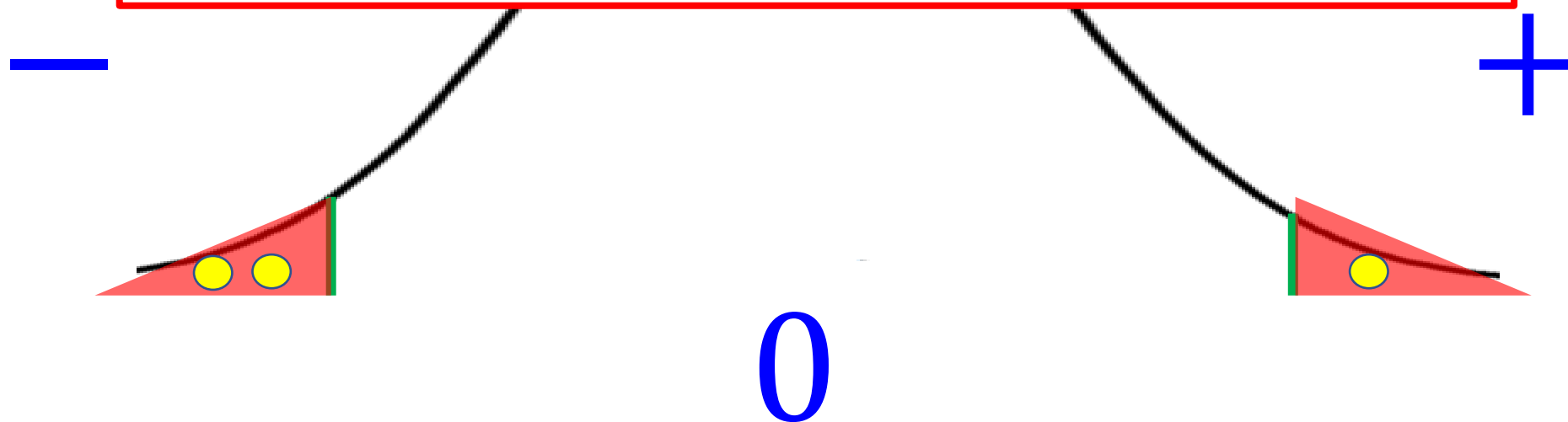
「正規分布 と 標準偏差」

ここにデータが入るってことは

分ほとんどない！と言いきれる

明らかに間違ってると言っている

「棄却できる！」



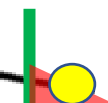
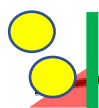
「正規分布 と 標準偏差」

「1%」ならなおさら！
ほぼ絶対あり得ない！



—

+



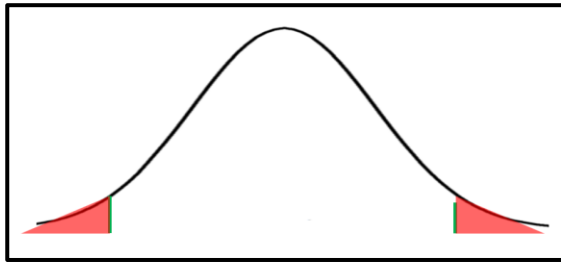
0

「両側検定」と「片側検定」

「両側検定」

ある値が基準値と異なるか

を調べたいとき



どちらの方向に値がずれているか、

両方の可能性を考えたいとき！

t 検定の流れ

- 1 「両側」か「片側」かを定める
- 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 」

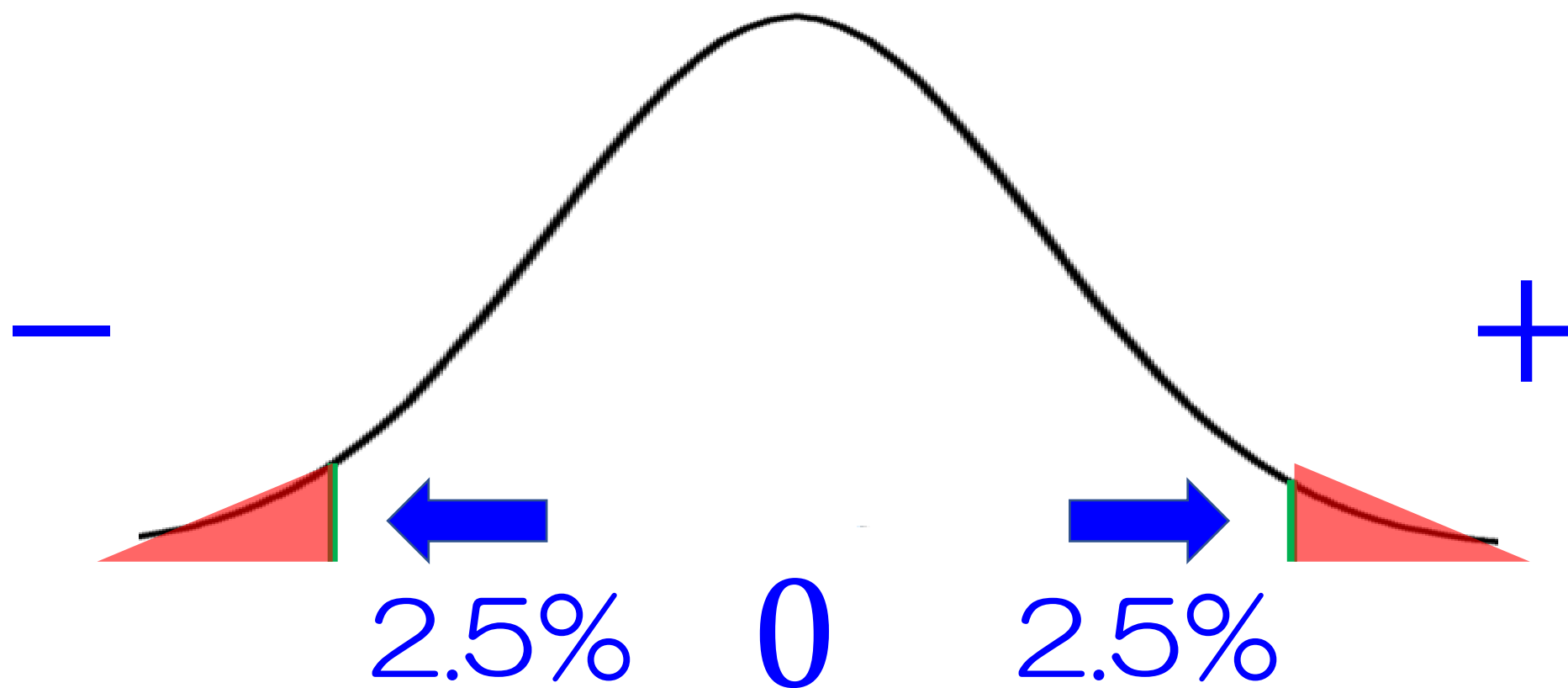
「 H_0 」 : 証明したくないなって事

t 検定の流れ

- 1 「両側」か「片側」かを定める
- 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」

両側：ここの面積が「5%」



t 検定の流れ

- 1 「両側」か「片側」かを定める
- 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！

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

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	データ个数	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 $\sqrt{\quad}$ の中	
Iさん	23173	分母の $\sqrt{\quad}$	
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	データ个数	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 $\sqrt{\quad}$ の中	2013842.462
Iさん	23173	分母の $\sqrt{\quad}$	1419.099173
Jさん	31090	t 値	2.323868594

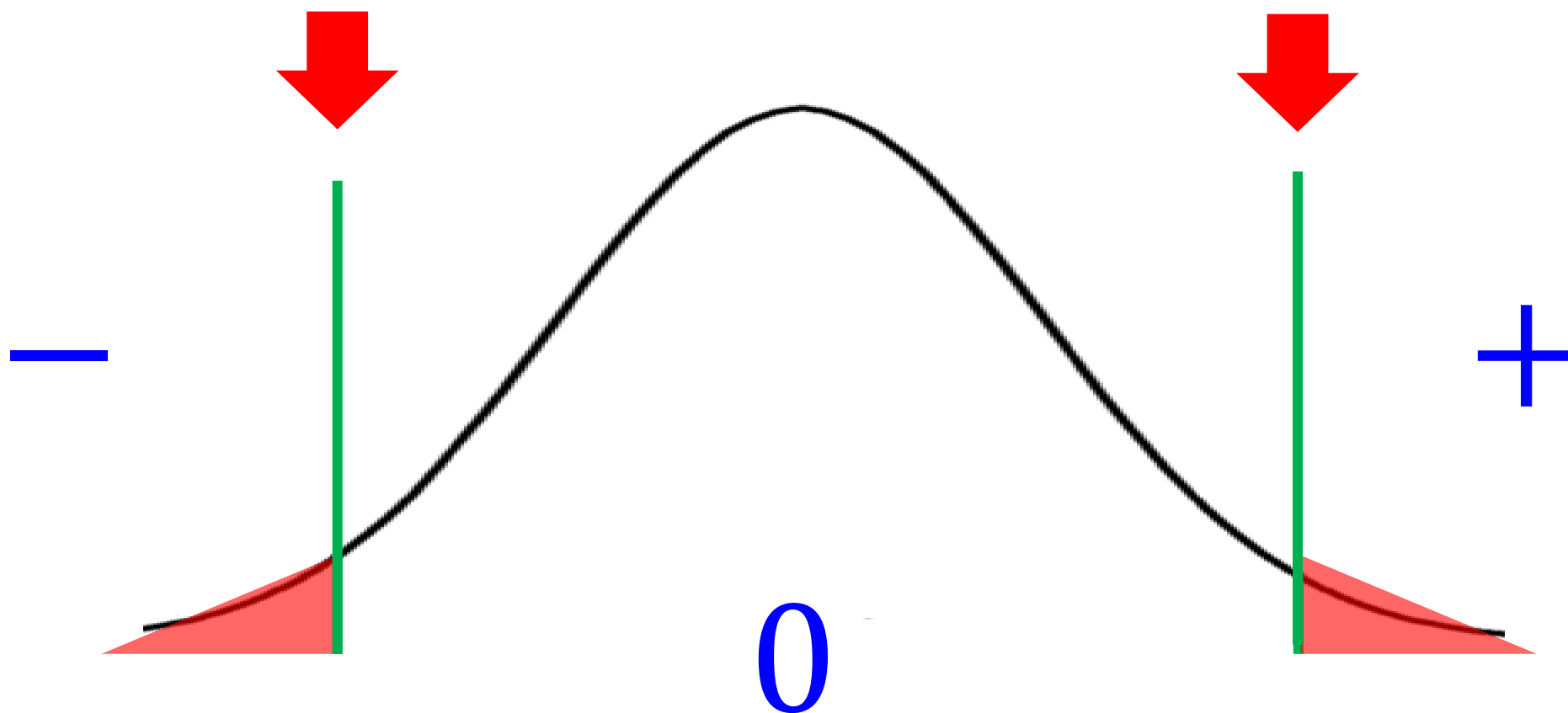
t 検定の流れ

- 1 「両側」か「片側」かを定める
- 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

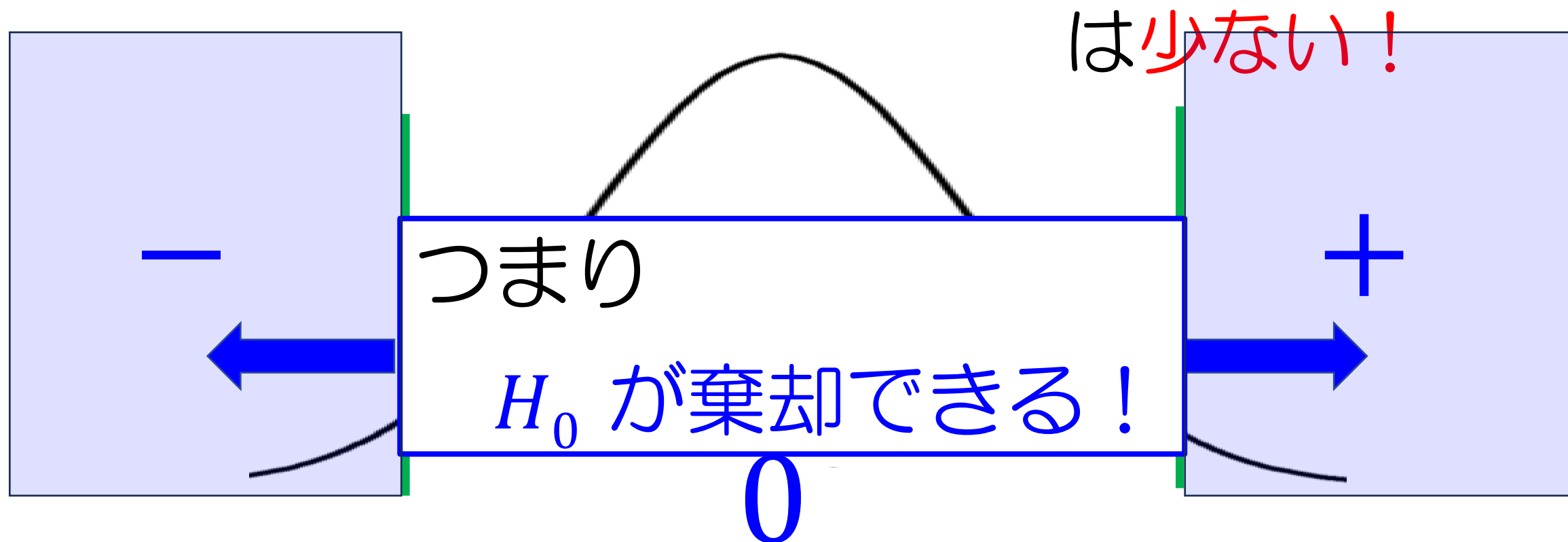
このラインのこと



「t 値の境界値」

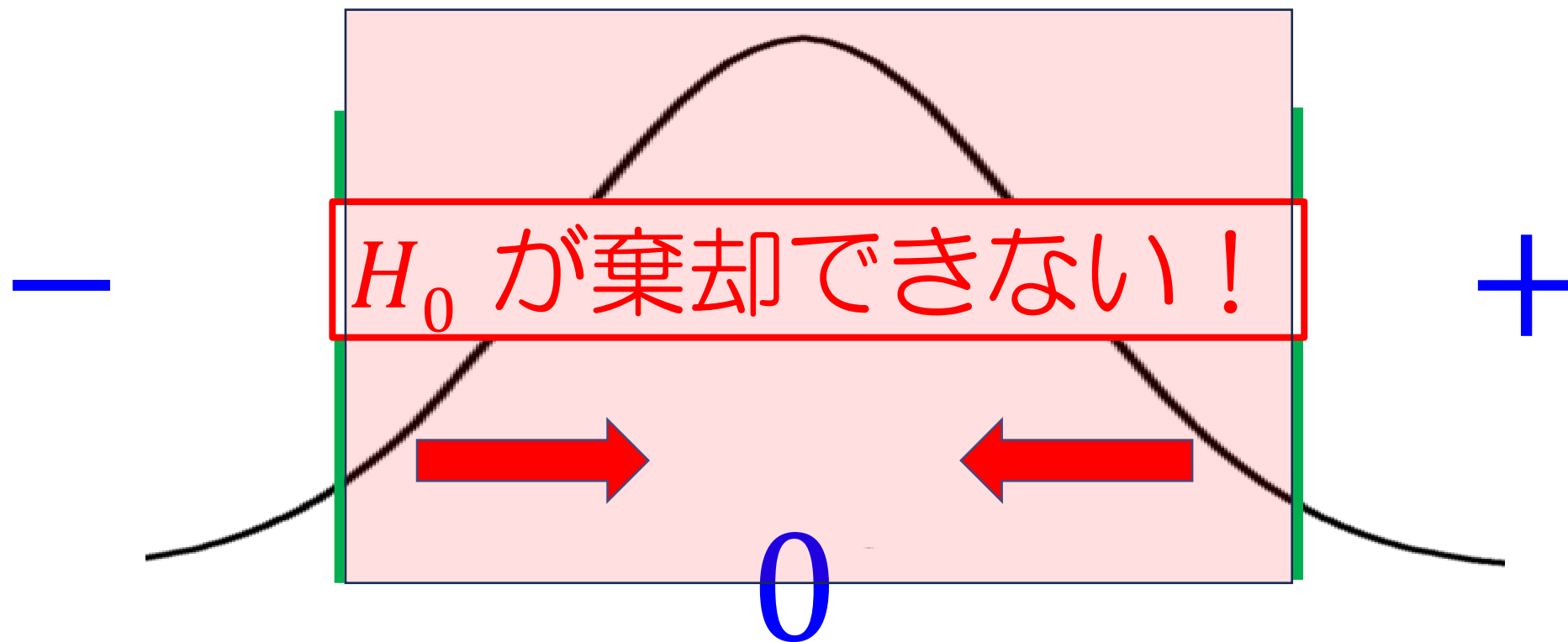
これより外側（値が大きい）と

間違っってデータがここに入る確率



「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	データ个数	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 $\sqrt{\quad}$ の中	2013842.462
Iさん	23173	分母の $\sqrt{\quad}$	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	データ个数	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 $\sqrt{\quad}$ の中	2013842.462
Iさん	23173	分母の $\sqrt{\quad}$	1419.099173
Jさん	31090	t 値	2.323868594

t 検定の流れ

- 1 「両側」か「片側」かを定める
- 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

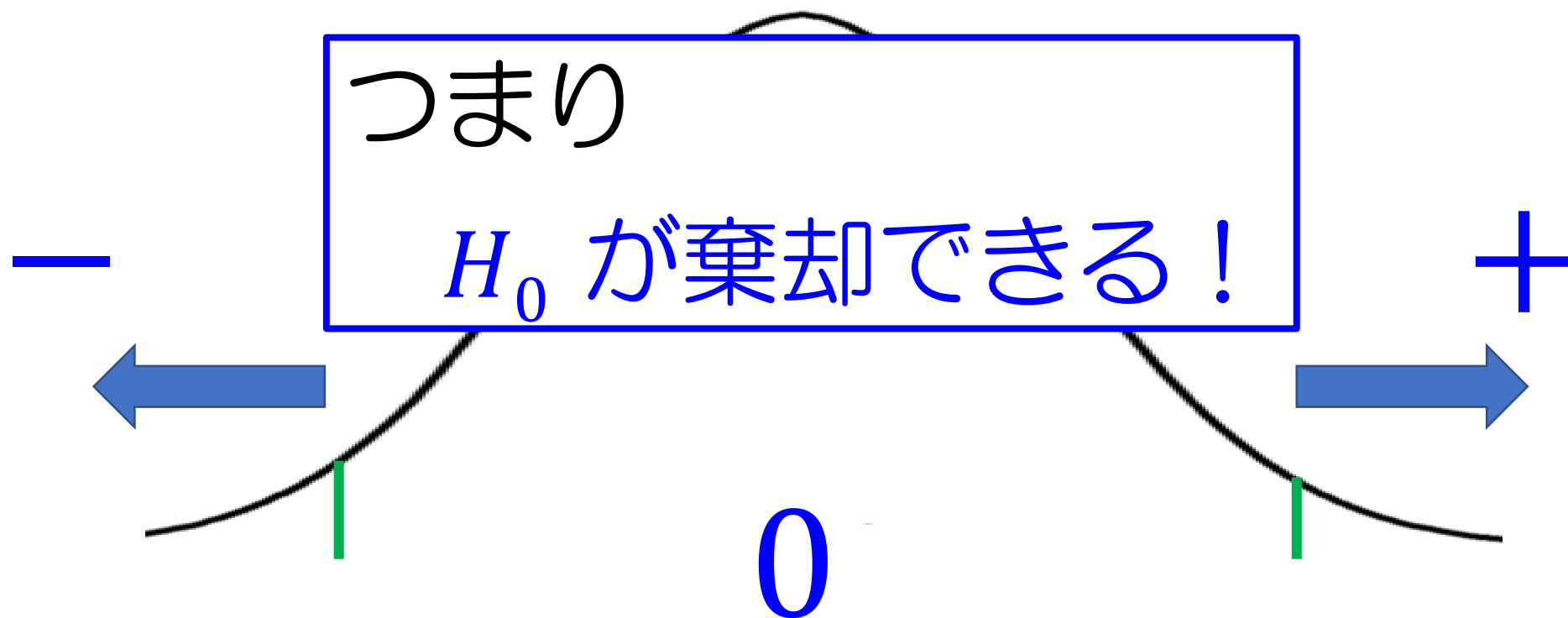
	日給	平均 (仮説)	30000
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差	-3297.8
Cさん	26382	有意水準	0.05
Dくん	28329	データ个数	10
Eさん	21988	境界値	2.262157163
Fさん	27441		
Gくん	25195	不偏分散	20138424.62
Hくん	20475	分母√の中	2013842.462
Iさん	23173	分母の√	1419.099173
Jさん	31090	t 値	2.323868594

t 値の境界値 \leq 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

間違っってデータがここに入る確率は少ない！



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る。

つまり「日給は30000円でない」と言える