

「仮説検定の有意水準」

統計学仮説検定では

帰無仮説 H_0 を棄却するかどうかの判断基準

判断する基準を確率で表す！

それが「有意水準」(α)

普通は「0.05」か「0.01」

先に決めておく

有意水準

H_0 が正しいと仮定したとき、

どれくらい珍しい結果なら H_0 を棄却するかの基準

0.05 → p 値が0.05未満なら
「珍しい」と判断

0.01 → p 値が0.01未満なら
「かなり珍しい」と判断

有意水準

どっちにするかは、

どれだけ厳しい基準値で判定するかで

変えればいい！

有意水準

「0.05 (5%)」



有意水準

「0.01 (1%)」

0.01の方が厳しい！

仮説検定

1 カイ二乗検定

2 t 検定

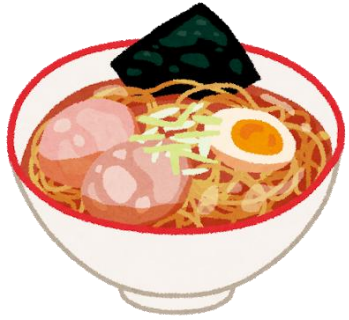
3 回帰分析



前回の「カイ二乗検定」

	結果も2択	
条件 2択	×二乗検定	
	赤と青のお皿で、食べるかどうかの 割合を検定	

前回の「カイ二乗検定」



「美味しいか」

「美味しくない」
の2択

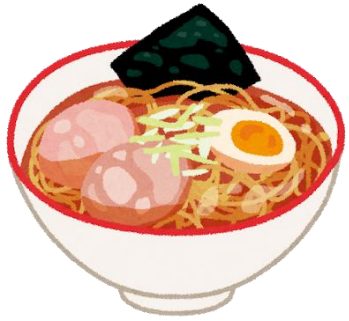
全体（1000人）では



美味しい : 800 人

美味しくない : 200 人

前回の「カイ二乗検定」



「美味しいか」

「美味しくない」
の2択

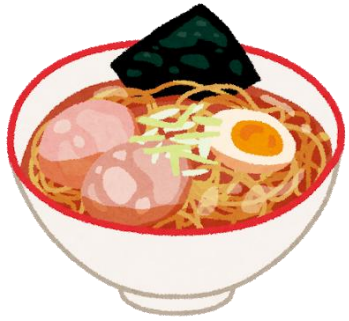
全体（1000人）では



美味しい	美味しくない	合計
800	200	1000

80%が美味しい

前回の「カイ二乗検定」



「美味しいか」

「美味しくない」

の2択

ということは

「男性」と「女性」でわけても

それぞれの 80%が美味しい

ってことなんやろうから…



前回の「カイ二乗検定」

男性300、女性700の結果



	美味しい	美味しくない	合計
男	240	60	300
女	560	140	700
合計	800	200	1000

ってなるはず！？

前回の「カイ二乗検定」

男性300、女性700の結果



	美味しい	美味しくない	合計
男	100	200	300
女	700	0	700
合計	800	200	1000

予想と違って、実際はこうなった！

前回の「カイ二乗検定」

このデータをもとにカイ二乗検定の結果

性別と、このラーメンの味覚の間には

関連性があると言える

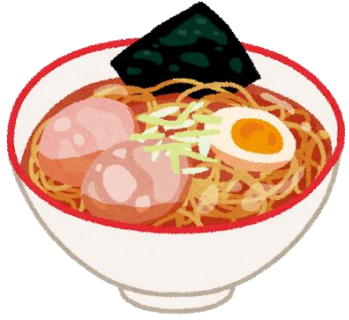
これも研究（卒研）

今回は年齢とかは考えてないから

考えてみてもいいかも！



今回の「t 検定」



「美味しいか」

「美味しくない」
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）

「t 検定」

朝食を 「食べる人」 の平均点

「食べない人」 の平均点

を比較してみるとどんな関係がある？

わかるようで
わからんな



CMとかでこんなの良くきくことない？



「緑茶を飲む人は、
飲まない人より免疫力が高い」



「ヨーグルトで
がんリスクが半分に！」

まさにコレが「t 検定！」



緑茶を飲む人の
免疫力の平均

VS

飲まない人
免疫力の平均

この2つの平均を比較するんやけど

この情報だけだと、

緑茶を飲んだ人の平均が高かったら

「緑茶を飲むと免疫力が上がる」

と見えてしまう





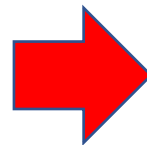
緑茶を飲む人の
免疫力の平均

VS

飲まない人
免疫力の平均

大事なこと

こんなことは
全くわからない



緑茶を飲まない人が

- 別のものを飲む
- 全く飲まない

あくまで「t 検定」は平均の比較だけ



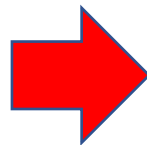
緑茶を飲む人の
免疫力の平均

VS

飲まない人
免疫力の平均

大事なこと

こんなことは
全くわからない



緑茶を飲まない人が

- 別のものを飲む
- 全く飲まない

「t 検定」を始める時に、
この話したと思うねん

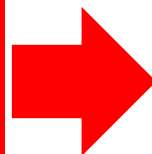


緑茶を飲む人の
免疫力の平均

VS

飲まない人
免疫力の平均

これを
「隠れた要因」
「第3の要因」
とかって言う



大事なこと

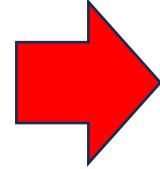
緑茶を飲まない人が

- 別のものを飲む
- 全く飲まない

ちょっと前にやったのを思い出して！



コンビニが
多いと



検挙数が
多い

第3の要因：人口密度

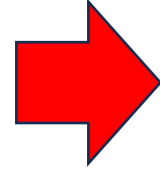


コンビニが犯罪を引き起こしてるんじゃないくて
人口が多いからコンビニが多く、犯罪数も多い！

こんな例もある



アイスが
売れる日は



溺死事故が
増える



第3の要因：気温

呪われたアイス屋さんとかじゃなくて

暑くなったから、泳ぎに行って溺れる人が増える

問題 みんなが働く病院には



救急病院



救急病院の当直は、
大晦日は意外と暇、
元旦はくそ忙しいと
言われています。

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 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」

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

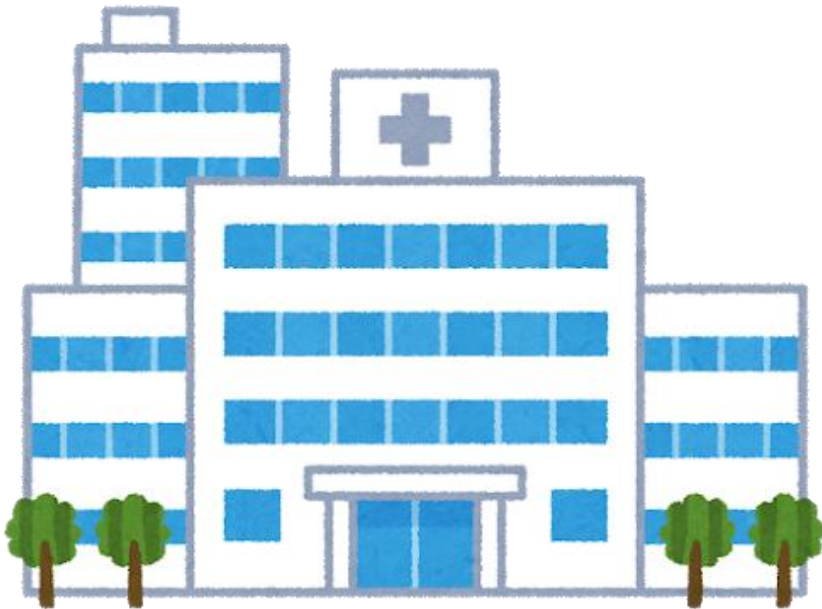


ある病院では、放射線技師の日給は

「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？



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

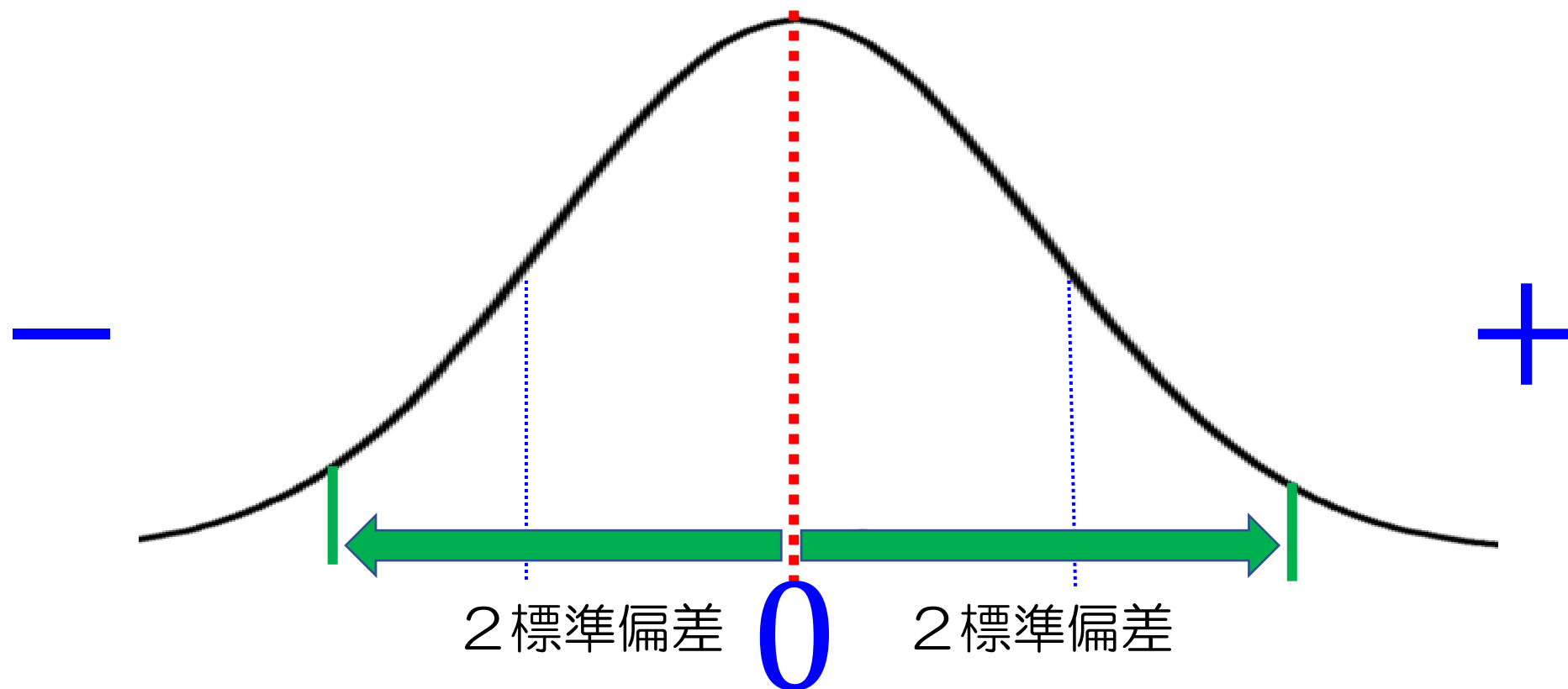
「両側検定」と「片側検定」
何が違うの？



思いだしてみて
「正規分布」と「標準偏差」

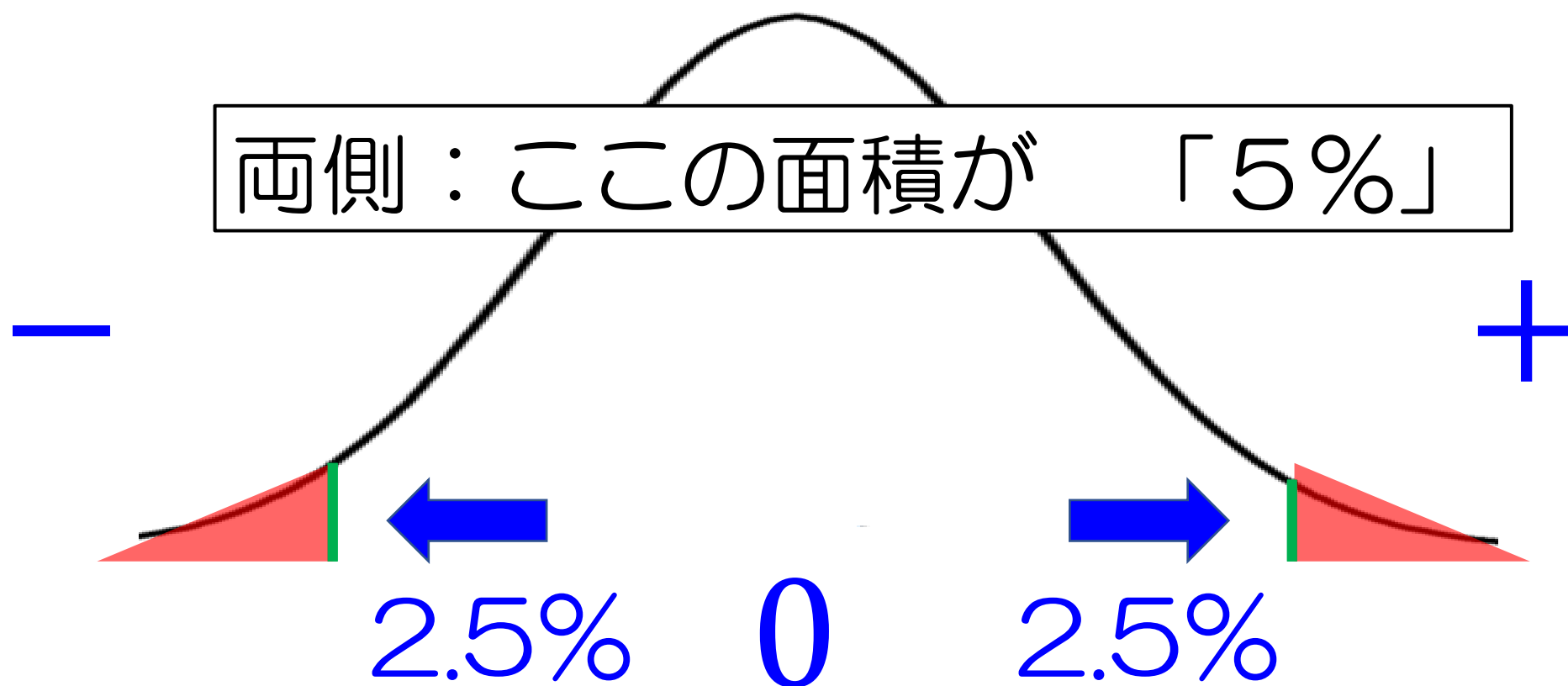
「正規分布 と 標準偏差」

- 平均値 $\pm$ 2標準偏差（面積）の中にデータの「約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

分布を正確に反映してるとは言えない…

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い

「 H_0 を棄却できる！」

0

「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！



—

+

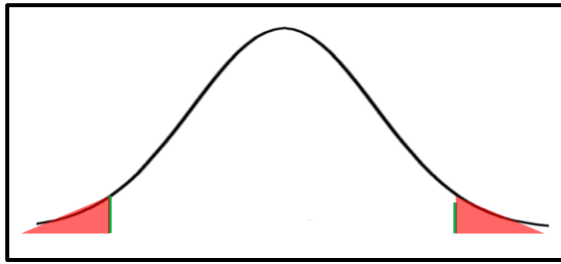


0

「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！

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

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

「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！**

「両側検定」と「片側検定」 何が違うの？

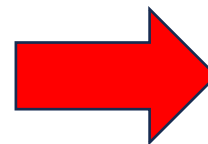


平均と比べて「差がある？」

両側検定

平均より

「大きい？」



片側検定

平均より

「小さい？」

片側検定

とりあえず、まずは

基本の「両側検定」で！



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 」

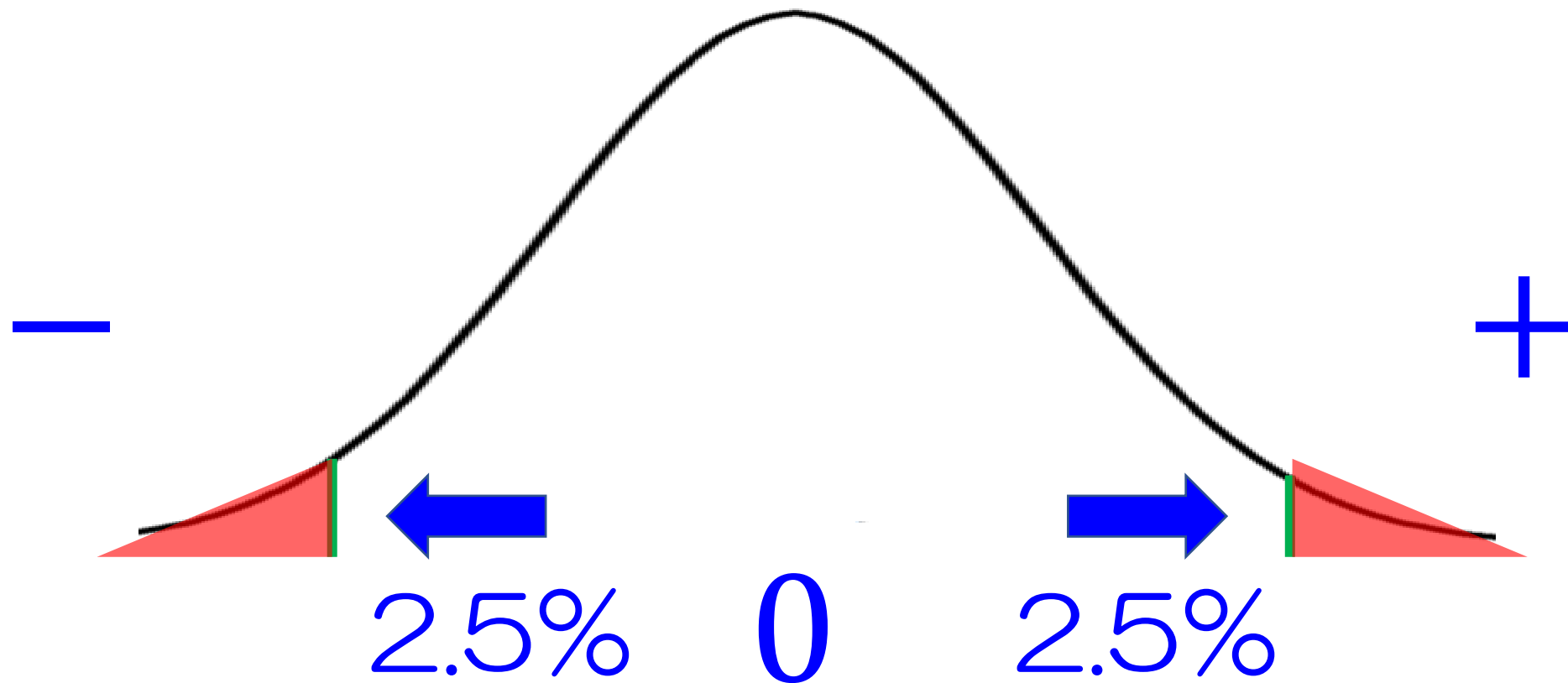
「 H_0 」 : 証明したくないなって事

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」

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)

今はここを求めている！

	日給	平均（仮説平均）	30000.0
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差（標本－仮説）	-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	（ t 値の）境界値	
Fさん	27441		
Gくん	25195	不偏分散	20138424.6222
Hくん	20475	分母√の中	2013842.4622
Iさん	23173	分母の√	1419.0992
Jさん	31090	t 値	2.3239

t 検定の流れ

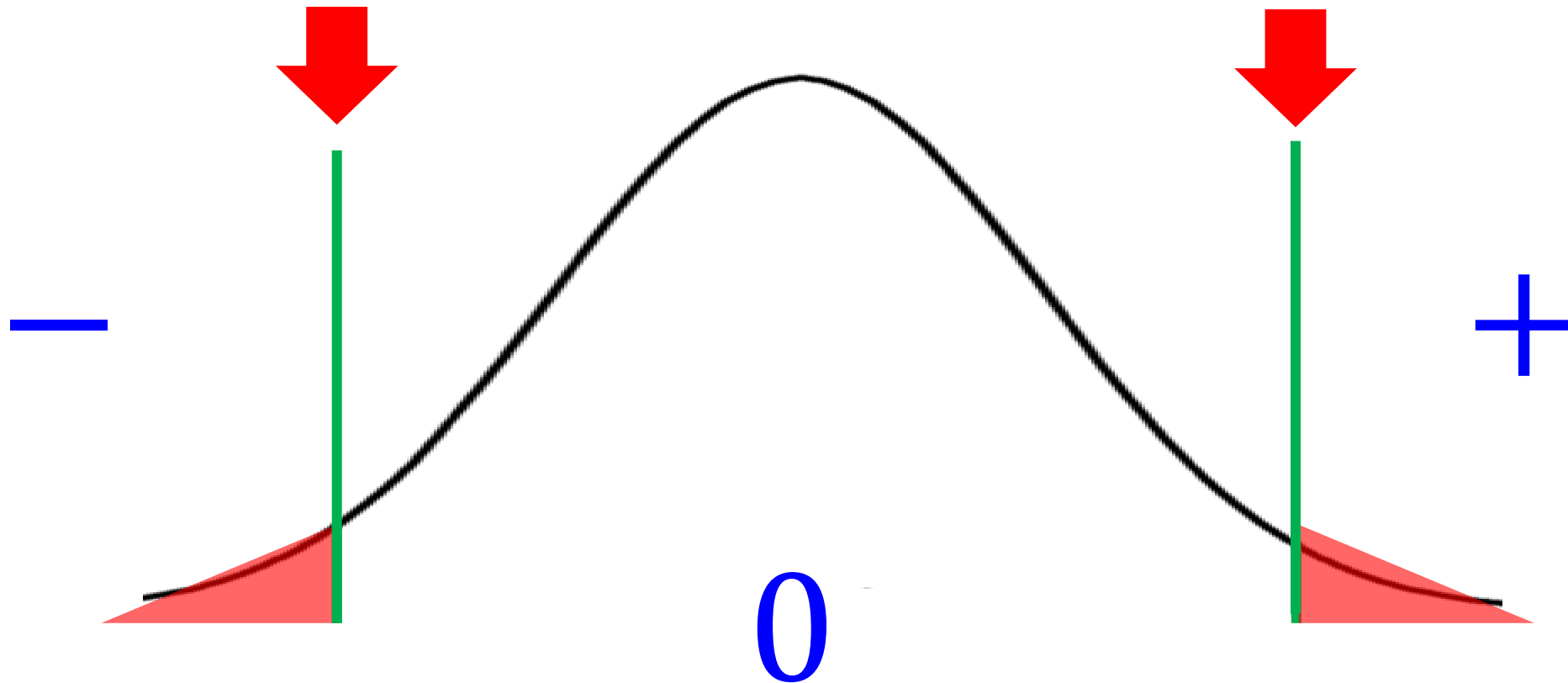
- 1 「両側」か「片側」かを定める
- 2 仮説を立てる (H_0 H_1)
- 3 有意水準を決める (普通は0.05)
- 4 t 値 (絶対値) を求める
- 5 t 値の境界値を求める
- 6 t 値と t 値の境界値から判断する

「t 値の境界値」

H_0 が正しいとき、t値が入る確率が5%の領域

このラインのこと

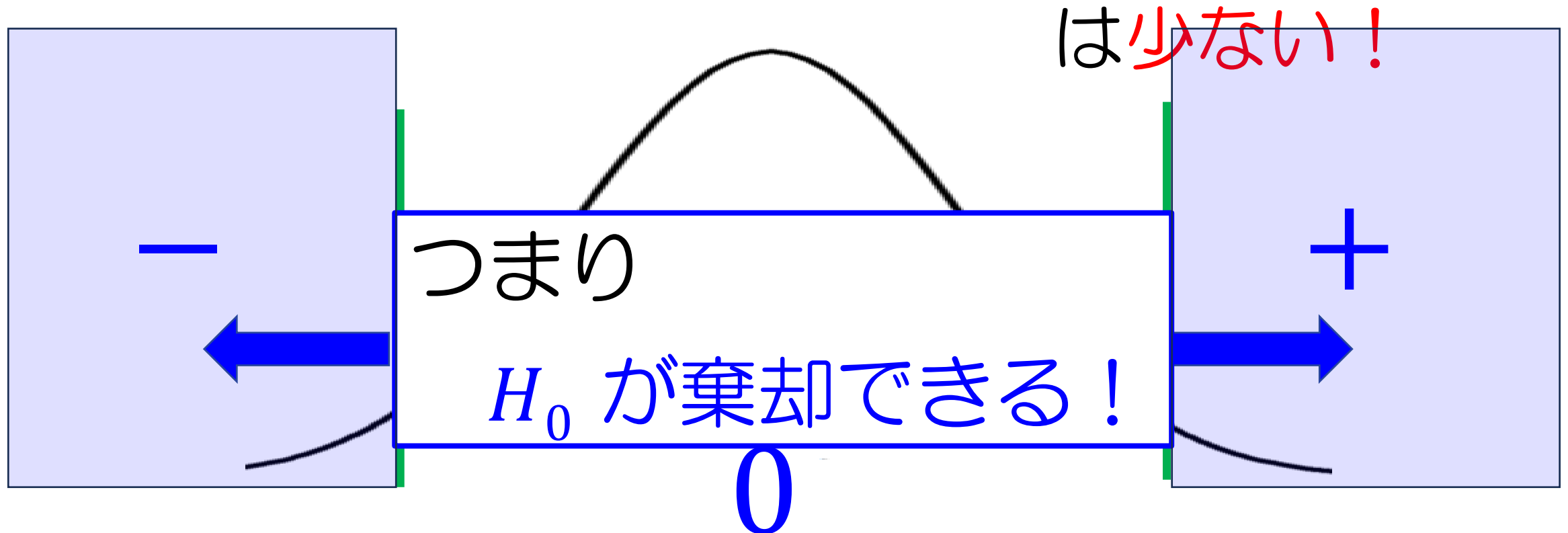
このラインのこと



「t 値の境界値」

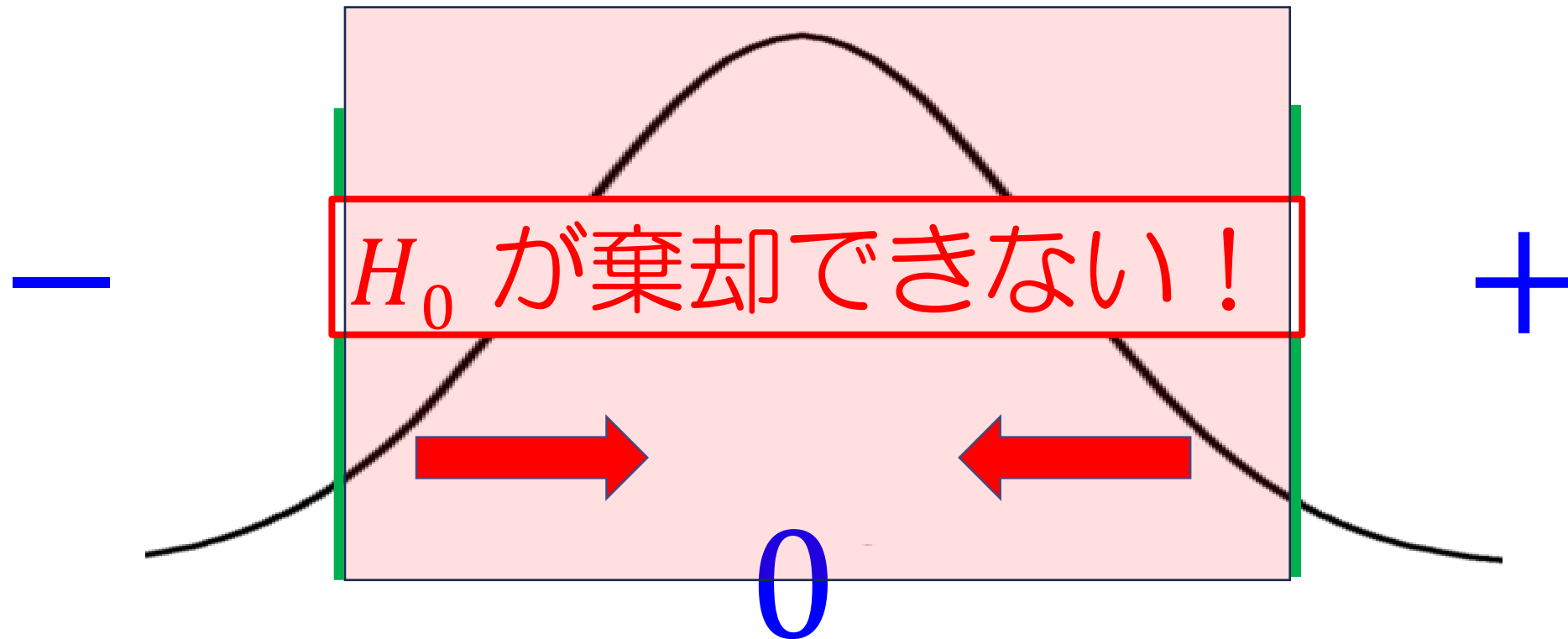
これより外側（値が大きい）と

間違ってデータがここに入る確率



「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！

今はここを求めている！

	日給	平均（仮説平均）	30000.0
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差（標本－仮説）	-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	（ t 値の）境界値	2.26216
Fさん	27441		
Gくん	25195	不偏分散	20138424.6222
Hくん	20475	分母√の中	2013842.4622
Iさん	23173	分母の√	1419.0992
Jさん	31090	t 値	2.3239

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 を支持！

小さい： H_0 を棄却できない

比べてみよう

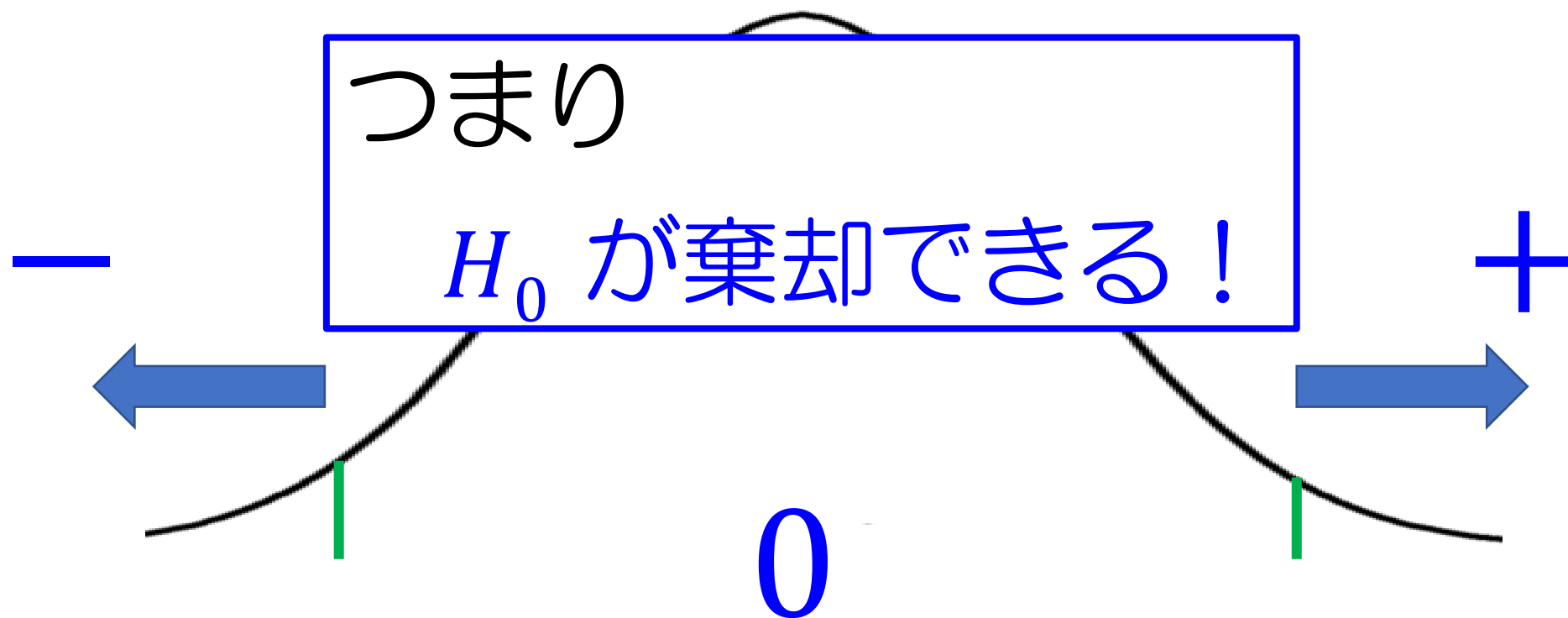
	日給	平均（仮説平均）	30000.0
Aくん	27149	標本平均	26702.2
Bくん	35800	平均の差（標本－仮説）	-3297.8
Cさん	26382	有意水準	0.05
Dくん	28329	データ個数	10
Eさん	21988	（ t 値の）境界値	2.26216
Fさん	27441		
Gくん	25195	不偏分散	20138424.6222
Hくん	20475	分母√の中	2013842.4622
Iさん	23173	分母の√	1419.0992
Jさん	31090	t 値	2.3239

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！



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円と同じかそれ以上の

極端な結果が出る確率は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

とりあえず、まずは

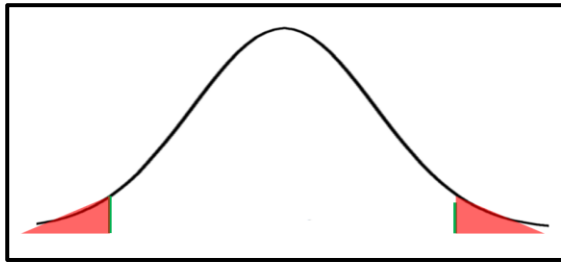
基本の「両側検定」で！

「課題1」やってみよう



「前回」は

「両側検定」



ある値が基準値と異なるか
を調べたいとき

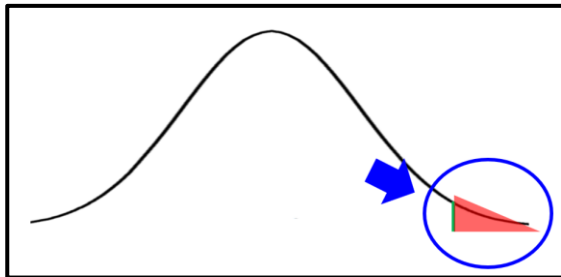
どちらの方向に値がずれているか、

両方の可能性を考えたいとき！

「両側検定」と「片側検定」

今回はこのパターン！

「片側検定」

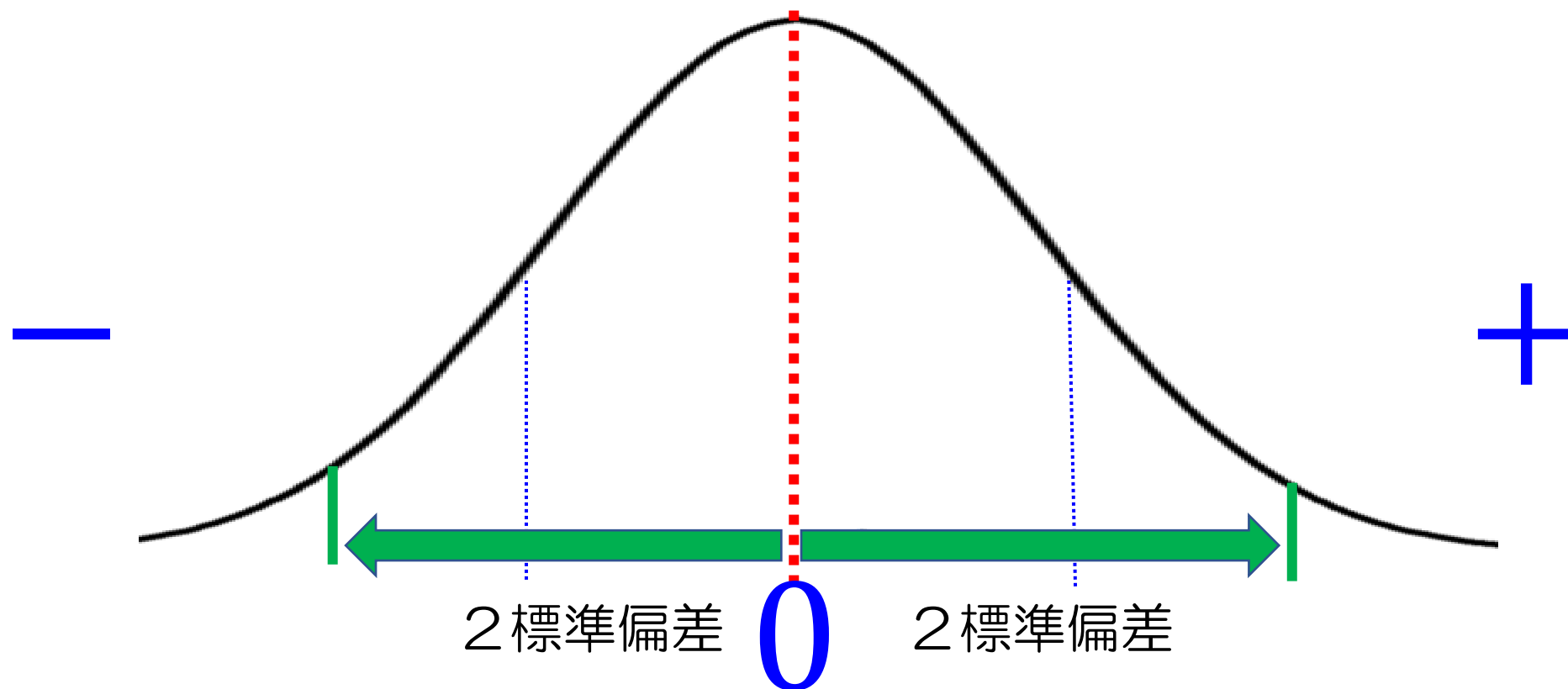


ある値が基準値よりも大きいかを調べたいとき

「正規分布 と 標準偏差」

思いだしてー！

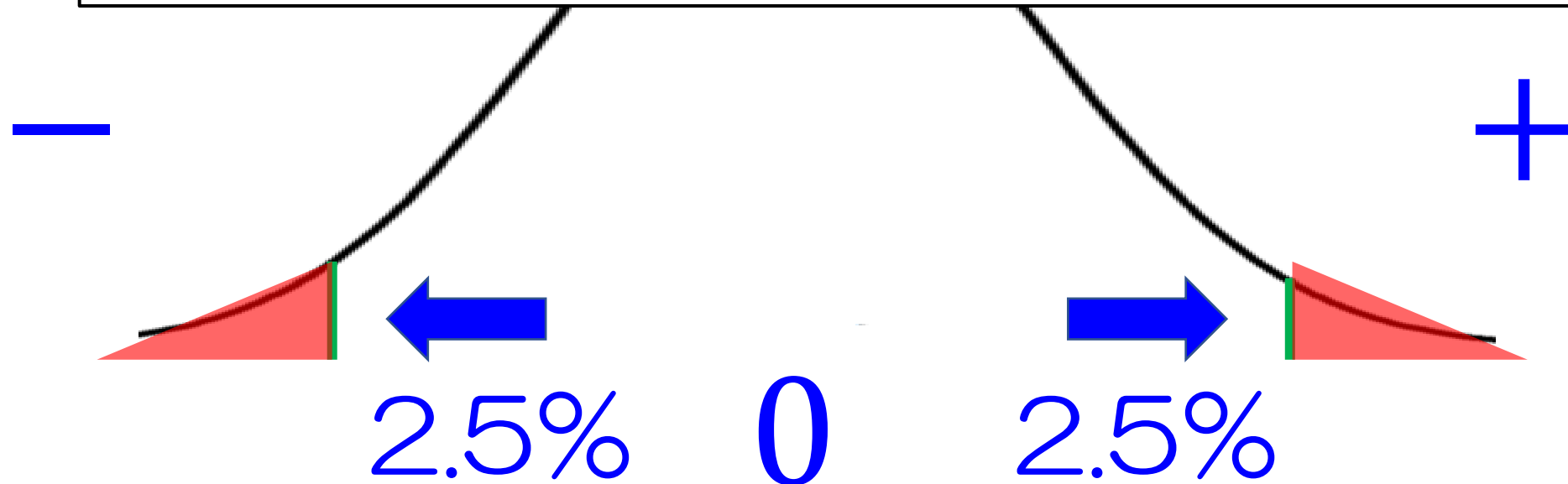
- 平均値 $\pm$ 2標準偏差（面積）の中にデータの「95%」がある



「両側検定」と「片側検定」

両側：左右合わせて面積が「5%」

片側：ここの面積が半分やから
「2.5%」ずつになるん？



「両側検定」と「片側検定」



両側：ここの面積が「5%」

じゃあ、片側検定では「5%」にするために
有意水準を2倍にしないとだめ？

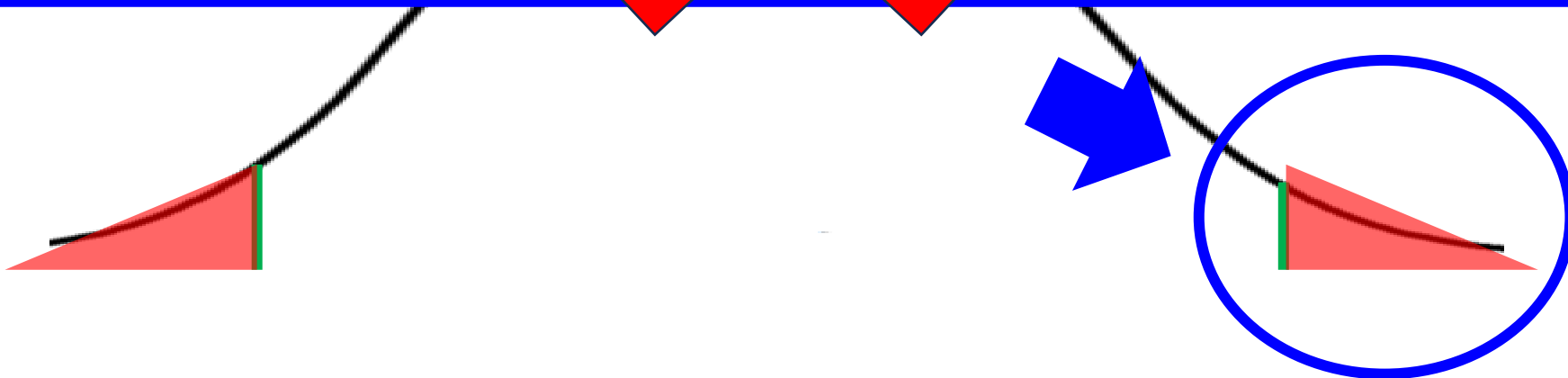


「両側検定」と「片側検定」



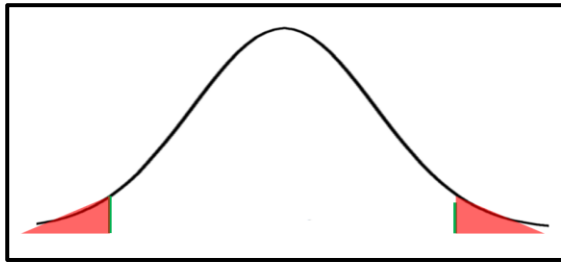
この考え方は間違い！

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が2.5%になってしまう)



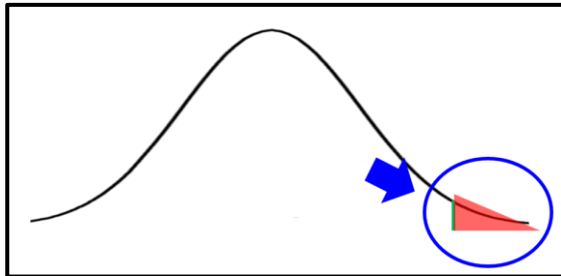
「両側検定」と「片側検定」の有意水準

「両側検定」



左右に2.5%ずつ

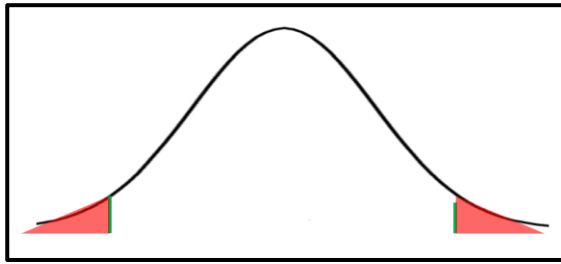
「片側検定」



片側に5%全部

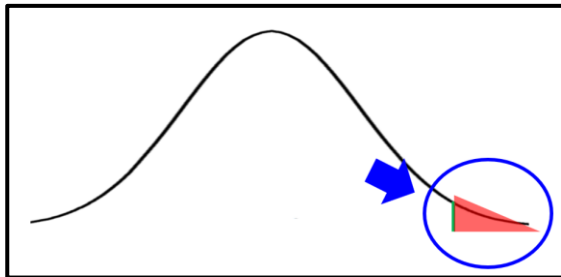
「両側検定」と「片側検定」

「両側検定」



ある値が基準値と異なるか
を調べたいとき

「片側検定」

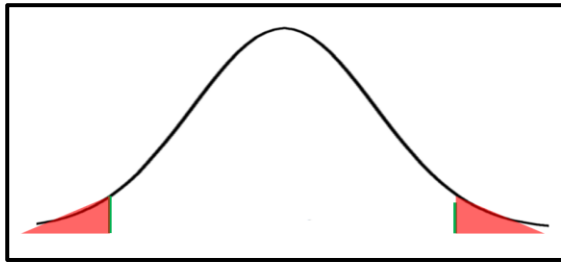


ある値が基準値よりも大きいか
を調べたいとき

「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



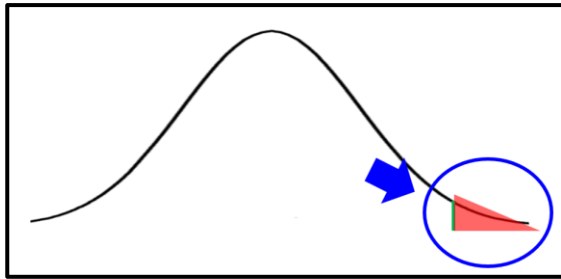
どちらの方向に値がずれているか、

両方の可能性を考えたいとき！

「両側検定」と「片側検定」

「片側検定」

ある値が基準値よりも大きいかを調べたいとき



（「小さいか」もありやで）

特定の方向に値がずれているか、

一方向だけの可能性を考えたいとき！

※方向はデータを見る前に決める！

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！

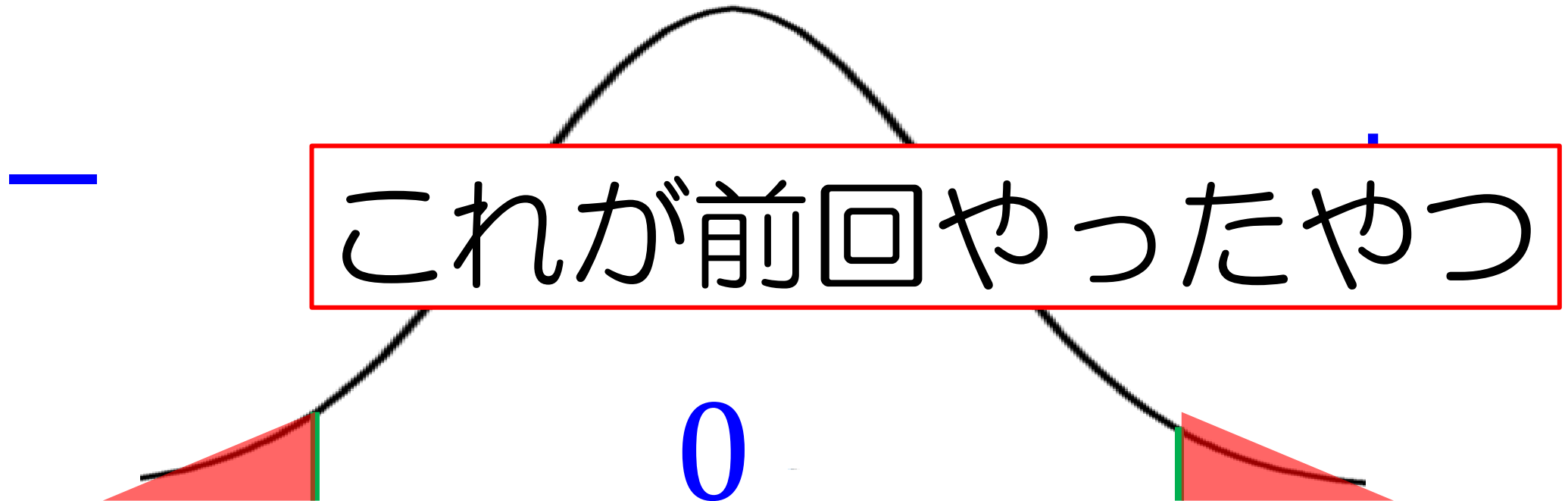
「両側検定」

1 日給は300000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300000円でない
- 2 日給は300000円より少ない
- 3 日給は300000円より多い

「300000円」より多いかを調べるだけで
300000円より少ないかは、
全く考えない

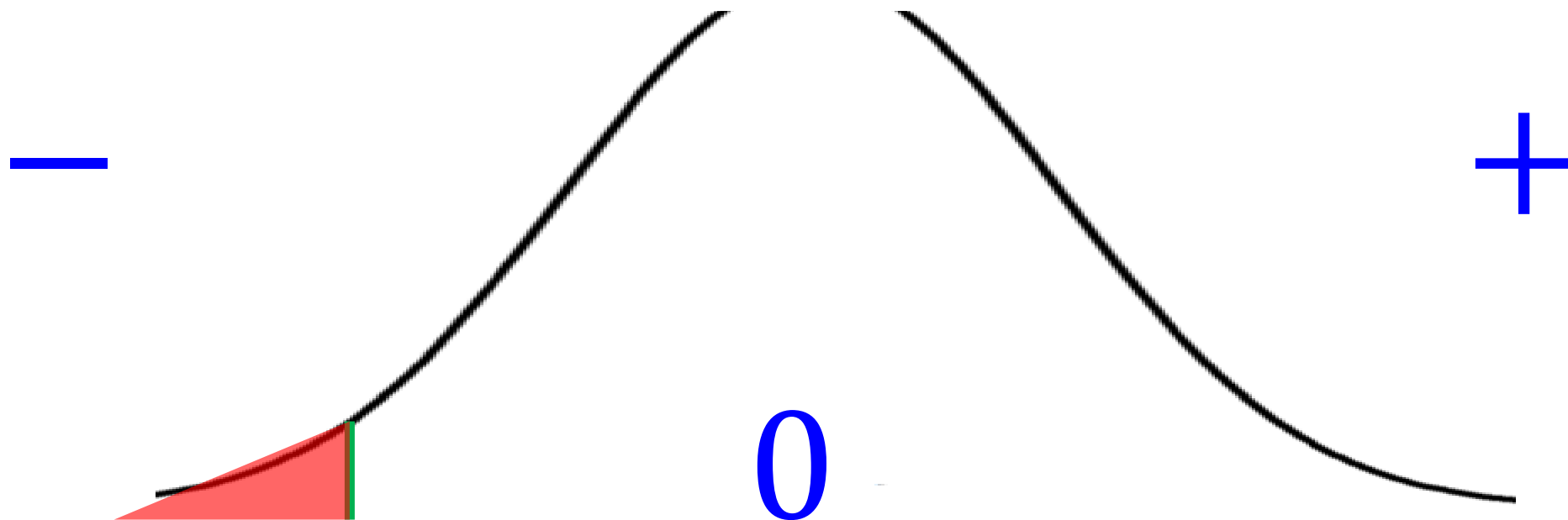
「片側検定」

2 日給は300000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)



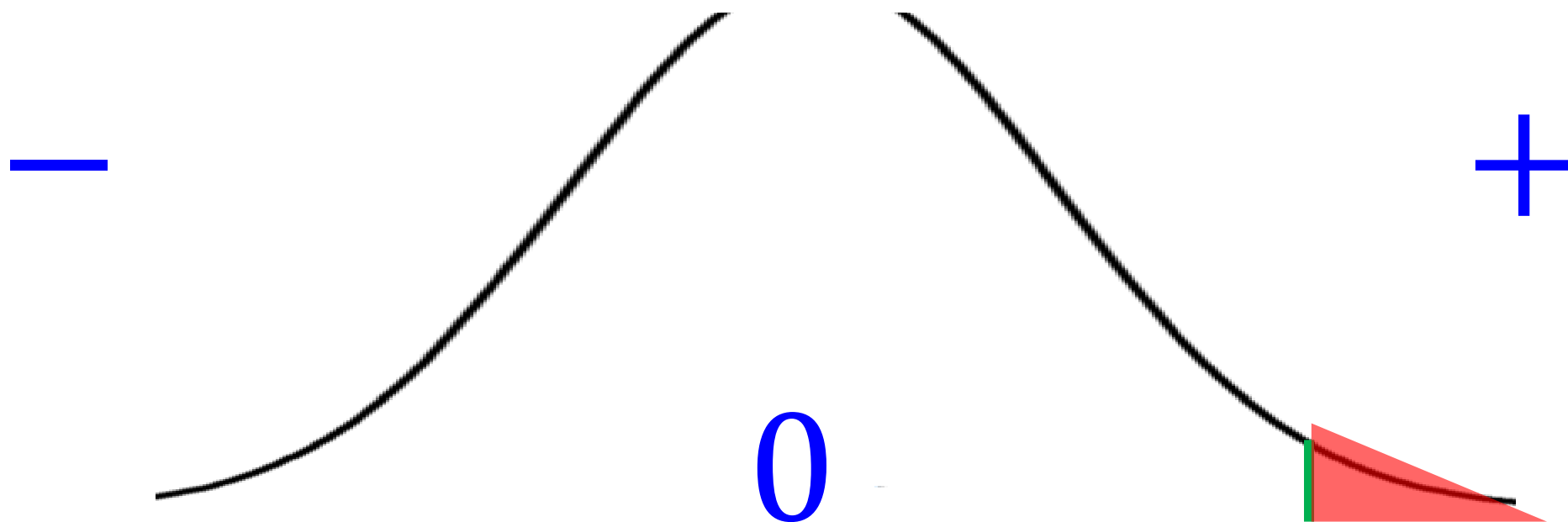
「片側検定」

3 日給は300000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)



次は「片側検定」やってみよう！

やる流れは全く一緒！

残りのシート
全部やってみよう！

