2019年度春学期　統計学　第１４回

分布についての仮説を検証する
— 検定

浅野　晃
関西大学総合情報学部

仮説検定の考え方は，単純

くじのあたり確率

「夏祭り，夜店のくじに当たりなし
　露天商の男を逮捕」
（朝日新聞大阪版2013年7月29日）

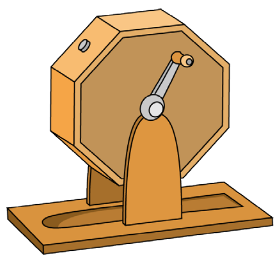
「１万円以上をつぎ込んだ男性が不審に思い、府警に相談。２８日に露店を家宅捜索し、 当たりがないことを確認した」

半分当たるというくじへの疑問

「半分の確率で当たる」というくじを
10回ひいても，１回も当たらなかった

運が悪いのか？

それとも
「半分の確率で当たる」
というのがウソか？



https://illpop.com/png_season/dec01_a07.htm

こう考える

警察みたいに全部のくじを調べられないなら，

仮に，本当に「確率1/2で当たる」とする

そのとき，
10回ひいて1回も当たらない確率は，
$(1/2)^{10}=1/1024$

こう考える

本当に「確率1/2で当たる」なら，

10回ひいて1回も当たらない確率は
1/1024（約0.001）

それでも「確率1/2で当たる」を信じるのは，

確率0.001でしか起きないことが，
いま目の前で起きていると信じるのと同じ

こう考える

確率0.001でしか起きないことが，
いま目の前で起きていると信じる

そりゃちょっと無理がありませんか？

というわけで，
「確率1/2で当たる」はウソ，と
考えるほうが自然　　これが［仮説検定］

復習：t分布と区間推定

正規分布の場合の区間推定

標本X1, ... Xnをとりだす
サイズn

標本平均 $\overline{X}$

母集団

母平均μ

正規分布
と仮定する

母平均μの95%信頼区間が
知りたい

(説明の都合です)

母分散σ²がわかっているものとする

正規分布の場合の区間推定

考え方

標本は，母集団分布と同じ確率分布にしたがう

正規分布N(μ, σ²)

標本平均 $\overline{X}$ は，やはり正規分布にしたがうが，分散が1/nになる　正規分布N(μ, σ²/n)　［性質２］

正規分布の［性質１］により

$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$$

は標準正規分布にしたがう
N(0, 1)

この例題は

標本X1, ... Xnをとりだす
サイズn

標本平均 $\overline{X}$

母集団
(受験者全体)

母平均μ

正規分布
と仮定する

母平均μの95%信頼区間が
知りたい

不偏分散s²で代用

母分散σ²がわからないので，

t分布

t統計量　$t = \dfrac{\bar{X} - \mu}{\sqrt{s^2/n}}$　は

自由度(n-1)のt分布にしたがう

t(n-1)

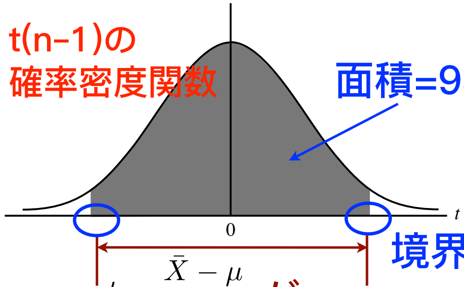
(「スチューデントのt分布」という)

発見者ウィリアム・ゴセットのペンネーム

t分布を用いた区間推定

$t = \dfrac{\bar{X} - \mu}{\sqrt{s^2/n}}$ は自由度(n−1)のt分布にしたがう

t(n−1)

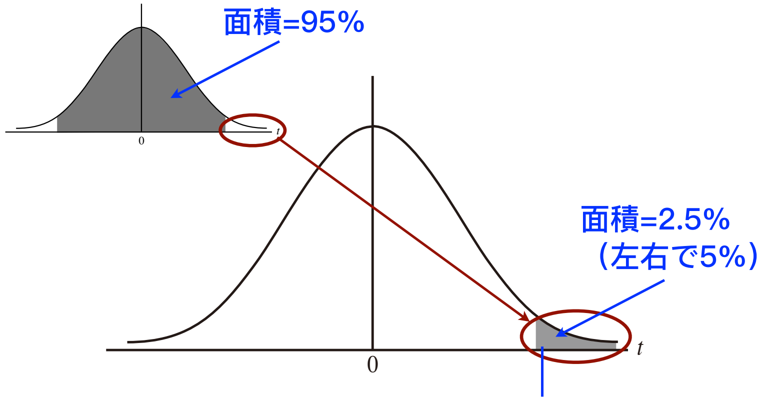


境界の値はいくら？

$t = \dfrac{\bar{X} - \mu}{\sqrt{s^2/n}}$ が

この区間に入っている確率=95%とすると

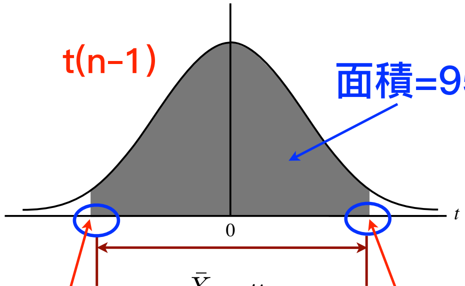
t分布を用いた区間推定



境界の値は自由度によってちがうので

$t_{0.025}$(n−1) としておく［上側2.5%点］という

t分布を用いた区間推定



$t = \dfrac{\bar{X} - \mu}{\sqrt{s^2/n}}$ が

この区間に入っている確率=95%

$-t_{0.025}$(n−1)　　$t_{0.025}$(n−1)

t分布を用いた区間推定

$t = \dfrac{\bar{X} - \mu}{\sqrt{s^2/n}}$ が $-t_{0.025}$(n−1) と $t_{0.025}$(n−1) の間に入っている確率が95%

式で書くと

$$P\left(-t_{0.025}(n-1) \leqq \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \leqq t_{0.025}(n-1)\right) = 0.95$$

μ の式に直すと

$$P\left(\bar{X} - t_{0.025}(n-1)\sqrt{\frac{s^2}{n}} \leqq \mu \leqq \bar{X} + t_{0.025}(n-1)\sqrt{\frac{s^2}{n}}\right) = 0.95$$

前回のプリントの例題

$t_{0.025}(10-1)=2.262$

標本平均=50　不偏分散=25　標本サイズ=10

$$P\left(\bar{X} - t_{0.025}(n-1)\sqrt{\frac{s^2}{n}} \leqq \mu \leqq \bar{X} + t_{0.025}(n-1)\sqrt{\frac{s^2}{n}}\right) = 0.95$$

μの95%
信頼区間の
下限

μの95%
信頼区間の
上限

で，信頼区間を求めるのは，
今日の本題ではありません。

t分布と検定

t分布と検定：例題

10人の被験者に，
薬Aを与えた場合と薬Bを与えた場合とで，それぞれある検査を行うと，その結果の数値は次の表の通りとなりました。このとき，

薬Bは，薬Aよりも，検査の数値を高くする働きがあるといえるでしょうか。

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66

t分布と検定：例題

問題は，
それぞれの被験者について，
薬Aと薬Bで数値がどう変化しているか。

各被験者について，
（薬Bでの数値）－（薬Aでの数値）を求める

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	−2	−2	5	3	−2	2	3	3	6

t分布と検定：例題

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	-2	-2	5	3	-2	2	3	3	6

薬Bでの数値のほうが高い（+）
薬Aでの数値のほうが高い　(-)
どちらの被験者もいる

差の平均値について
「薬Bでの数値のほうが高い」か？

A. Asano, Kansai Univ.

「本質的な差」

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	-2	-2	5	3	-2	2	3	3	6

10人の被験者について，差の平均値は +2
薬Bでの数値のほうが高い

その差は，
偶然生じたものではなく
「本質的な」差なのか？

「本質的」とは？

A. Asano, Kansai Univ.

検定で考える

1.
母集団（ここでは，世界のすべての患者）については

「薬Aと薬Bでの差」の平均は0

と仮説を設定する。

つまり，「本質的な差はない」という仮説を設定する。

A. Asano, Kansai Univ.

検定で考える

1.「母集団（ここでは，世界のすべての患者）については
『薬Aと薬Bでの差』の平均は0」と仮説を設定する。
つまり，「本質的な差はない」という仮説を設定する。

2.
被験者は，母集団から無作為抽出された，10人からなる標本と考える。

A. Asano, Kansai Univ.

検定で考える

1.「母集団（ここでは，世界のすべての患者）については
『薬Aと薬Bでの差』の平均は0」と仮説を設定する。
つまり，「本質的な差はない」という仮説を設定する。

2. 被験者は，母集団から無作為抽出された，10人からなる標本と考える。

3.
このとき，
被験者10人での「薬Aと薬Bでの差」の
平均値が，わずかな確率でしか生じないほどの
大きな差であるなら，この差は
「偶然によって生じたものではない」と考える。

A. Asano, Kansai Univ.

検定で考える

1.「母集団（ここでは，世界のすべての患者）については
『薬Aと薬Bでの差』の平均は0」と仮説を設定する。
つまり，「本質的な差はない」という仮説を設定する。

2. 被験者は，母集団から無作為抽出された，10人からなる標本と考える。

3. このとき，被験者10人での「薬Aと薬Bでの差」の
平均値が，わずかな確率でしか生じないほどの
大きな差であるなら，この差は
「偶然によって生じたものではない」と考える。

4.
すなわち，「本質的な差はない」という
当初の仮説は誤り，と結論する。

A. Asano, Kansai Univ.

検定で考える

1.「母集団（ここでは，世界のすべての患者）については
『薬Aと薬Bでの差』の平均は0」と仮説を設定する。
つまり，「本質的な差はない」という仮説を設定する。

2. 被験者は，母集団から無作為抽出された，10人からなる標本と考える。

3. このとき，被験者10人での「薬Aと薬Bでの差」の
平均値が，わずかな確率でしか生じないほどの
大きな差であるなら，この差は
「偶然によって生じたものではない」と考える。

4. すなわち，「本質的な差はない」という当初の仮説は誤り，と結論する。

この論理を仮説検定（検定）という

A. Asano, Kansai Univ.

例題に検定で答える

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	−2	−2	5	3	−2	2	3	3	6

薬Bでの数値のほうが
「本質的に」高いか？

母集団全体での「薬Aと薬Bでの差」は，
平均 μ の正規分布にしたがうと考える

標本サイズを n（例題では10）
標本平均を $\overline{X}$（例題では，10人の被験者における差の平均値で，+2）
不偏分散を s^2（例題では，10人の被験者についての不偏分散で，8.89）

t統計量 $t = \dfrac{\overline{X} - \mu}{\sqrt{\dfrac{s^2}{n}}}$ は，自由度(n−1)のt分布にしたがう

A. Asano, Kansai Univ.

例題に検定で答える

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	−2	−2	5	3	−2	2	3	3	6

薬Bでの数値のほうが
「本質的に」高いか？

標本サイズを n（例題では10）
標本平均を $\overline{X}$（例題では，10人の被験者における差の平均値で，+2）
不偏分散を s^2（例題では，10人の被験者についての不偏分散で，8.89）

t統計量 $t = \dfrac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n}}}$ は，自由度(n-1)のt分布にしたがう

「母集団については『薬Aと薬Bでの差』の平均は0」
という仮説

$$\to \mu = 0$$

A. Asano, Kansai Univ.

例題に検定で答える

標本サイズを n（例題では10）
標本平均を $\overline{X}$（例題では，10人の被験者における差の平均値で，+2）
不偏分散を s^2（例題では，10人の被験者についての不偏分散で，8.89）

仮説より，μ = 0

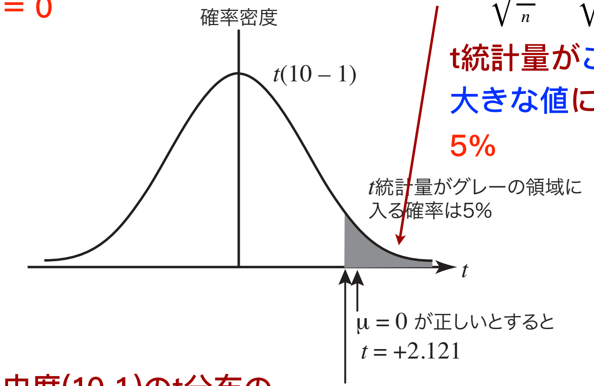
このとき，t統計量は

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$$

A. Asano, Kansai Univ.

t統計量= +2.121 の意味

仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$
μ = 0



t統計量がこんなに
大きな値になる確率は
5%

自由度(10-1)のt分布の
上側5%点

$t_{0.05}(10-1) = +1.8331$

A. Asano, Kansai Univ.

仮説は間違っている，と考える

仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$
μ = 0

t統計量がこんなに大きな値になる確率は5%

そんな小さな確率でしか起きないはずのこと
が起きているのは不自然

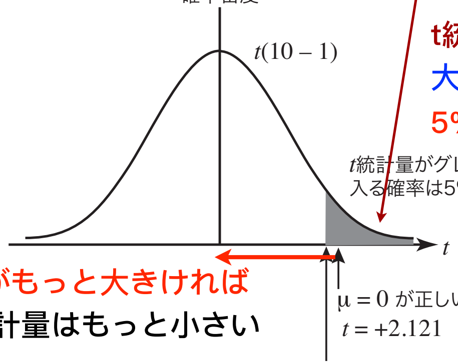
仮説が間違っていると考える

A. Asano, Kansai Univ.

では，どういう結論なら

仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0



t統計量がこんなに
大きな値になる確率は
5%

μがもっと大きければ
t統計量はもっと小さい

それなら起きる確率は $t_{0.05}(10-1) = +1.8331$
5%より大きい

仮説は間違っている，と考える

仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0

t統計量がこんなに大きな値になる確率は5%

仮説が間違っていると考える

本当は，μはもっと大きいと考える

μ > 0

薬Bでの数値のほうが高い，と考える

検定の言葉

［帰無仮説］ H_0: μ = 0

仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0

［有意水準］

t統計量がこんなに大きな値になる確率は5%

仮説が間違っていると考える　帰無仮説を［棄却］する

［対立仮説］ H_1: μ > 0　　対立仮説を［採択］する

本当は，μはもっと大きいと考える

μ > 0

偶然とは思わない
［有意］である

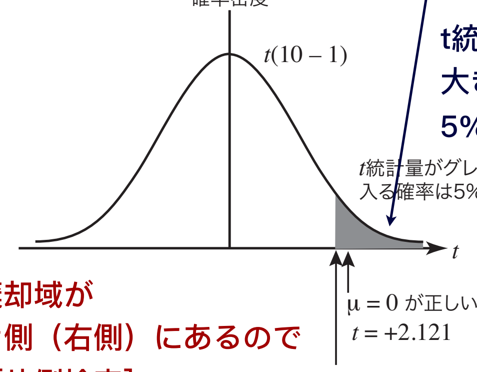
薬Bでの数値のほうが高い，と考える

検定の言葉

［検定統計量］

仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0



t統計量がこんなに
大きな値になる確率は
5%［棄却域に落ちる］

［棄却域］

棄却域が
片側（右側）にあるので
［片側検定］

$t_{0.05}(10-1) = +1.8331$

両側検定

例題を少し変更

10人の被験者に，
薬Aを与えた場合と薬Bを与えた場合とで，それぞれ
ある検査を行うと，その結果の数値は次の表の通りと
なりました。このとき，

~~薬Bは，薬Aよりも，検査の数値を高くする働きがあると いえるでしょうか。~~ 薬Aと薬Bで，検査の数値に違いがあるといえるでしょうか。

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66

前の例題とどう違うのか

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	−2	−2	5	3	−2	2	3	3	6

薬Bは，薬Aよりも，検査の数値を高くする働きがあるといえるでしょうか。

差の平均値について
「薬Bでの数値のほうが高い」か？

薬Aと薬Bで，検査の数値に違いがあるといえるでしょうか。

どちらの数値が高いにしても，
本質的に差があるか？

この場合，対立仮説は

帰無仮説が正しいとするとき，t統計量 $t = \dfrac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} = \dfrac{2 - 0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0

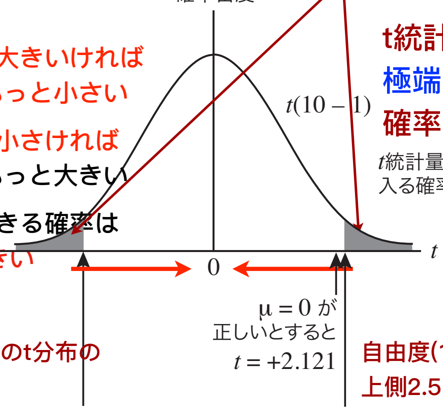
μがもっと大きいければ
t統計量はもっと小さい

μがもっと小さければ
t統計量はもっと大きい

それなら起きる確率は
5%より大きい

t統計量がこんなに
極端な値になる
確率は5%

t統計量がグレーの領域に
入る確率は, 左右合わせて5%



対立仮説は
μ ≠ 0

$\mu = 0$ が
正しいとすると
$t = +2.121$

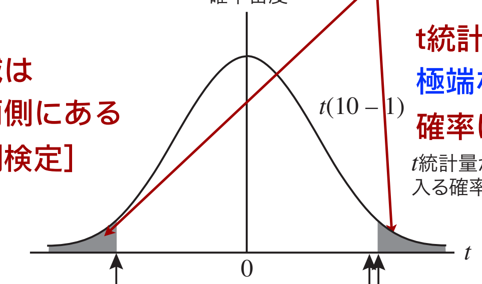
自由度(10-1)のt分布の
下側2.5%点
$-t_{0.025}(10-1) = -2.2622$

自由度(10-1)のt分布の
上側2.5%点
$t_{0.025}(10-1) = +2.2622$

この場合，棄却域は

帰無仮説が正しいとするとき，t統計量 $t = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}} = \frac{2-0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0



棄却域は
左右両側にある
［両側検定］

t統計量がこんなに
極端な値になる
確率は5%

t統計量がグレーの領域に
入る確率は，左右合わせて5%

μ = 0 が
正しいとすると
$t = +2.121$

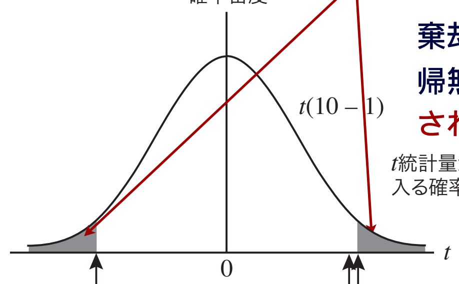
自由度(10-1)のt分布の
下側2.5%点
$-t_{0.025}(10-1) = -2.2622$

自由度(10-1)のt分布の
上側2.5%点
$t_{0.025}(10-1) = +2.2622$

例題では

帰無仮説が正しいとするとき，t統計量 $t = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}} = \frac{2-0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0



棄却域に落ちない
帰無仮説は棄却
されない

t統計量がグレーの領域に
入る確率は，左右合わせて5%

μ = 0 が
正しいとすると
$t = +2.121$

自由度(10-1)のt分布の
下側2.5%点
$-t_{0.025}(10-1) = -2.2622$

自由度(10-1)のt分布の
上側2.5%点
$t_{0.025}(10-1) = +2.2622$

棄却されないときは

仮説が正しいとするとき，t統計量 $t = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}} = \frac{2-0}{\sqrt{\frac{8.89}{10}}} = +2.121$

μ = 0

t統計量がこんなに極端な値になる確率は5%
とはいえない

本当は，μ は0ではないとはいえない

薬Bでの数値と薬Aでの数値に
本質的な違いがある，とはいえない

棄却されないときは

帰無仮説が棄却されるのは
帰無仮説が正しいとすると，
とても小さな確率でしか起きないはずのこ
とが，いま起きていることになるから

帰無仮説が棄却されないときは
「いま起きていることがおきる確率は
とても小さい，とまではいえない」

棄却されないときは

帰無仮説が棄却されないときは
「いま起きていることがおきる確率は
とても小さい，とまではいえない」

だから
「帰無仮説が棄却されない」とは

~~帰無仮説が正しい~~
帰無仮説が間違っているとはいいきれない

有意水準について

検定とはそういうものです

有意水準は
物言いの慎重さを表す

有意水準が
大きい（5%）
大胆だが，蛮勇

確率5%でおきることでも
「こんなことがおきるのは偶然
とは思えない」として棄却

小さい（1%）
慎重だが，臆病

確率1%より大きいことなら
「偶然でないと言い切れない」
として棄却しない

なんかおかしくない？

なんかおかしくない？

被験者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	−2	−2	5	3	−2	2	3	3	6

薬Bは，薬Aよりも，検査の数値を高くする
働きがあるといえるでしょうか。　片側検定

「薬Bでの数値のほうが高い」といえる

薬Aと薬Bで，検査の数値に違いが
あるといえるでしょうか。　両側検定

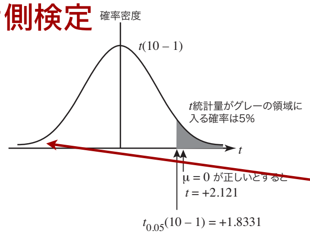
「本質的な差がある」とはいえない

???

A.Asano, Kansai Univ.

実はおかしくない

片側検定

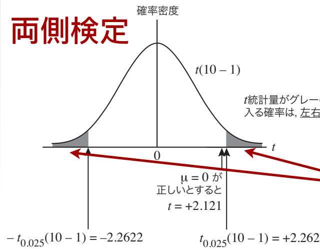


薬Bは薬Aよりも，
検査の数値を高くする働きが
あるだろう，という目論見がある

標本を調べると逆の結果に
なっていたとしても，見逃す

同じ有意水準でも，
大胆な検定を行う

両側検定



薬Aと薬Bのどちらの数値が
高いかという目論見はない

どちらが高い場合でも棄却

同じ有意水準でも慎重な検定を行う

A.Asano, Kansai Univ.

くじびきの例でいうと

帰無仮説：「当たり確率は50%である」

くじをひく立場なら

10回中1回も当たらなかったら
→ 帰無仮説が正しいとするとき，
そんなことが起きる確率は小さいし，
しかも結果に不満だから棄却したい

10回中10回当たったら
→ 帰無仮説が正しいとするとき，
そんなことが起きる確率はやはり小さい が，
結果に不満はないから棄却しない

片側検定

A.Asano, Kansai Univ.

くじびきの例でいうと

帰無仮説：「当たり確率は50%である」

賞品を出す立場なら

10回中1回も当たらなかったら
→ 帰無仮説が正しいとするとき，
そんなことが起きる確率は小さい が，
とくに損はしないから棄却しない

10回中10回当たったら
→ 帰無仮説が正しいとするとき，
そんなことが起きる確率はやはり小さい
それでは破産してしまうので棄却したい

やはり
片側検定

A.Asano, Kansai Univ.

くじびきの例でいうと

帰無仮説：「当たり確率は50%である」

中立の立場（商店会長？）なら

10回中1回も当たらなかったときも
10回中10回当たったときも
→ 帰無仮説が正しいとするとき，
そんなことが起きる確率はどちらも小さいし，
どちらにしても信用にかかわるので棄却したい

これが
両側検定

どの検定を用いるかは，
「立場」にもとづいて先に決めて
おかなければならない

検定は
どんなときにするものなのか

有意水準と第１種の誤り

有意水準5%のときは

確率5%でしかおきないことがおきているなら，それは偶然ではないと考える

［有意］

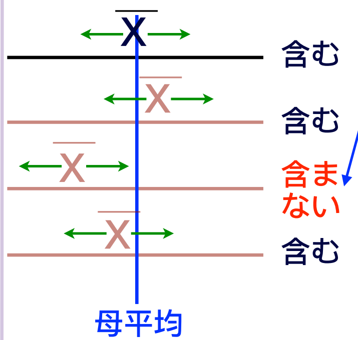
でも，確率5%でしかおきないことは，

言い換えれば，確率5%でおきるのでは？

有意水準と第１種の誤り

確率5%でしかおきないことは，
確率5%でおきる

区間推定において
「95%信頼区間」が
確率5%ではずれるように



帰無仮説が正しくても，
偶然標本が偏っていたために，
棄却してしまう確率が5%

［第１種の誤り］
その確率=有意水準

有意水準と第１種の誤り

つまり

帰無仮説が本当に正しいとしても，

有意水準5%の仮説検定を何度も行うと，
そのうちの5%では第1種の誤りを犯す

正しいはずの帰無仮説を棄却し，
採択すべきでない対立仮説を採択してしまう

検定の結論が言っていること

検定で「帰無仮説を棄却する」とは

私は，帰無仮説は間違いだ，と判断する。

ただし

私は100回中5回はウソを言う
（第1種の誤りを犯す）。

私が今回，本当のことを言っているのか，
ウソを言っているのか，
それは誰にもわからない。

検定はどんなときに

何度でも標本をとりだして検定できるようなら，
検定などする必要はない

小さな標本を１回しかとりだせないときに，
それでも十分にいえる結論を導く

何度も検定をすれば，
棄却されないはずの帰無仮説も
たまには棄却される

「血液型と性格に関係がない」という帰無仮説も
たまに棄却されることがある

最近はこんな研究も

血液型と性格「関連なし」　読売新聞 2014.7.19

血液型と性格の関連性に科学的根拠はないとする統計学的な解析結果を、九州大の縄田健悟講師（社会心理学）が発表した。

日米の１万人以上を対象にした意識調査のデータを分析した。

質問に対する回答のうち、血液型によって差があったのは３項目だけで、その差もごくわずかだったため「無関連であることを強く示した」と結論づけた。

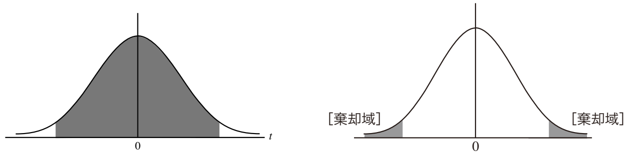
検定で，標本サイズが大きいと

標本サイズが大きいと

推測がよりはっきりする

区間推定では　信頼区間が狭くなる

検定では　棄却域が広くなる



検定で，標本サイズが大きいと

標本サイズが大きいと

区間推定では　信頼区間が狭くなる

検定では　棄却域が広くなる

標本サイズが非常に大きいと

帰無仮説が
ちょっとでも疑わしいと棄却される

それでも「血液型と性格に関係がない」
という帰無仮説が棄却されないなら
「無関連であることを強く示し」ている