◎統計的仮説検定について

・よく使われる仮説検定の種類

- 母平均の検定
  - 母分散既知 → z検定
  - 母分散未知 → t検定
- 平均値の差の検定（データ間の対応なし）
  - 母分散既知 → z検定
  - 母分散未知
    - 等分散が仮定できる → t検定
    - 等分散でないか未知 → Welchのt検定
- 平均値の差の検定（データ間の対応あり） → t検定

母分散の差の検定 → F検定

①母分散がわかる → z検定
②　　　　わからない → t検定
③分散の検定は → F検定

ただし、初歩の段階でF検定したことは…
QC検定の試験の時くらい。

② 母平均の検定

・サンプリングしてきたデータの平均が母集団の平均と等しいかどうか？

母集団

平均 μ

標準偏差 σ

抜き取り →

サンプル

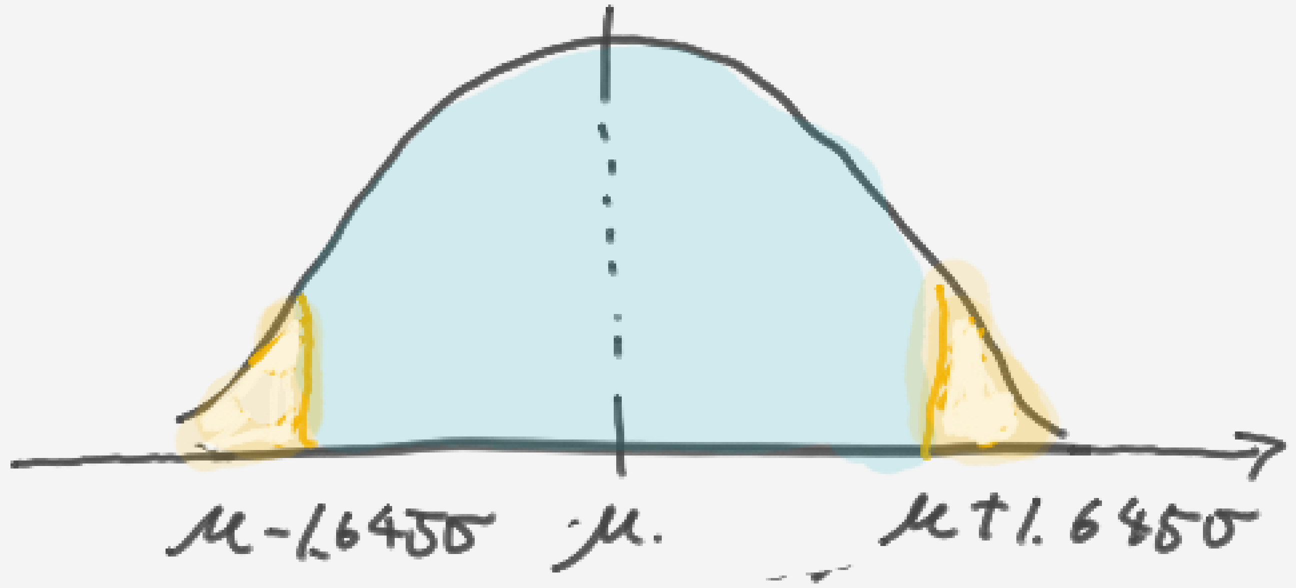
・平均 $\bar{x}$

・不偏標準偏差 s

・帰無仮説 $H_0 : \mu = \bar{x}$

・対立仮説 $H_1 : \mu \neq \bar{x}$

このときに $\mu = \bar{x}$ と言っても問題ないか？を確認できる。



$\bar{x} = \mu,\ s = \sigma$ を仮定したときに、$\bar{x}$ は青色の領域に90%入る。

入らないときは「仮定がおかしい」と判断して対立仮説を選ぶ

◎ 平均値の差の検定（対応なし）

・2つの独立した母集団の間に平均値の差はあるか？

母集団A
Avg: μ_A
st.dev: σ_A

↓

サンプルA
Avg: $\bar{x}_A$
st.dev: $\bar{s}_A$

母集団B
Avg: μ_B
st.dev: σ_B

↓

サンプルB
Avg: $\bar{x}_B$
st.dev: s_B

$\mu_A = \mu_B$ であるかを確認する。

帰無仮説 H_0: $\mu_A = \mu_B$
対立仮説 H_1: $\mu_A \neq \mu_B$

プールした分散 s

$$s^2 = \frac{(n_1-1)\times s_1^2 + (n_2-1)\times s_2^2}{n_1+n_2-2}$$

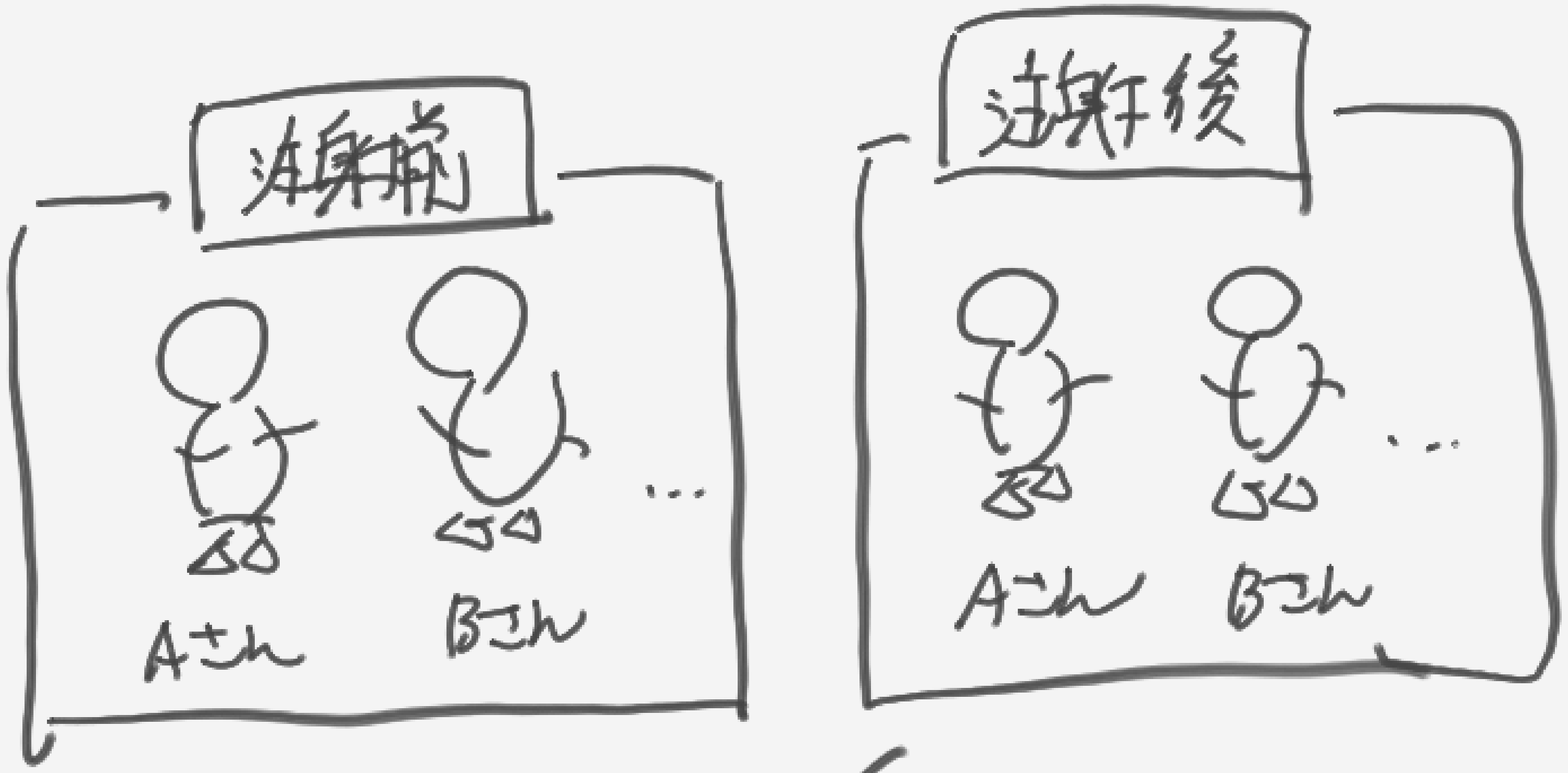
（分散の推定精度が上がるらしい）

◎ 平均値の差の検定

・同一の対象から抽出された
対になる2群の標本に対して
平均値の差はあるか？

$H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 \neq \bar{x}_2$



標本が同一 ＝ 対応あり.

Q 統計的仮説検定の実用例

・普通に計量値で仮説検定の例を出しても
たぶん、他のサイトの方がわかりやすいので…
計数値を題材にします。

・サンプル数がとっても多ければ、「正規分布で近似できる」
大数の法則を利用すれば、二項分布でも同じように検定できる

去年と今年で、不良率は下がった？

不良率 = 不良数／生産数 なので

計数値，1つ 2つと数える

不良率も計数値

不良は、Yes/Noの二項で表されるため
二項分布に従う = 正規分布ではない

→  これを使った検定ができない！

・不良率は下がったのか？の検定

母集団

不良率：$\mu = P$

st.dev：$\sigma = \sqrt{P(1-P)}$

理由はまだ理解してないけど そういうものらしい。

サンプル

不良率：$\bar{x} = p$

st.dev：$s = \sqrt{\dfrac{p(1-p)}{n}}$

仮説の設定

$H_0 : P = p$

$H_1 : P \neq p$

検定統計量の計算

$$u = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

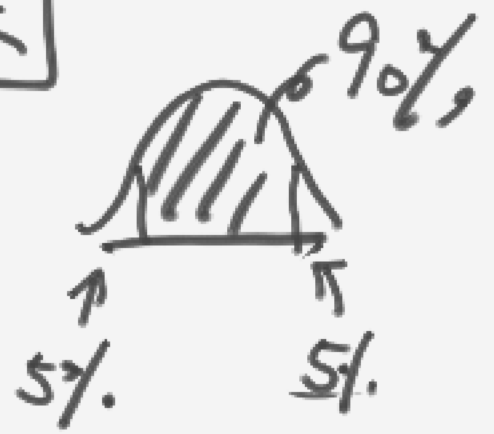
$$= \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

← $N(0, 1^2)$の標準正規分布に近似する

∵ QC2級の範囲で取扱うことができる

棄却域の設定

$u(0.05) = 1.96$



判定

u値がUmaxを超えたら、H1を採択

◎数値を仮定して使ってみる

	確率p	サンプル数n
期間A	$p_A = \frac{0.44}{100}$	61919
期間B	$p_B = \frac{0.70}{100}$	253104

① 帰無仮説を立てる

H_0 : 期待値 $E(p_A) = E(p_B)$

H_1 : $E(p_A) \neq E(p_B)$

② $\bar{p}$ の計算

$$\bar{p} = \frac{r_A + r_B}{n_A + n_B}$$

③ u の計算

$$u = \frac{p_A - p_B - E(p_A - p_B)}{\sigma(p_A - p_B)}$$

$$= \frac{(p_A - p_B) - E(p_A - p_B)}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$$

期待値は平均の足し算引き算

分散は各分散の足し算

帰無仮説が成立する前提だと

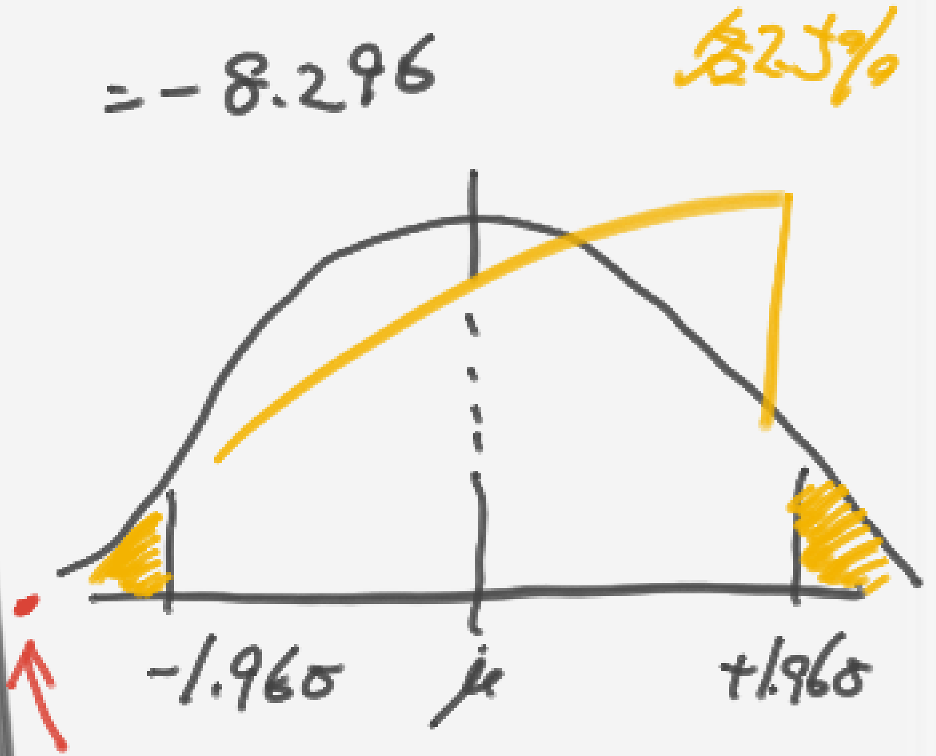
$E(p_A) = E(p_B) \Leftrightarrow E(p_A - p_B) = 0$

$$\therefore u = \frac{p_A - p_B}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$$

④ 数値を代入して

$$u = \frac{(0.44 - 0.70) \times 1/100}{\sqrt{\frac{\frac{0.44}{100}\left(1-\frac{0.44}{100}\right)}{61919} + \frac{\frac{0.70}{100}\left(1-\frac{0.70}{100}\right)}{253104}}}$$

$= -8.296$



−8.296ということは

2.5%よりもかなり低い確率でしか起こらない

→前提が違う ⇒ $p_A \neq p_B$ だ!!

＠統計的仮説検定 まとめ

- 統計的仮説検定とは?
- どんなときに使う?
- 統計的仮説検定の手順
- 検定の種類の選び方
- 統計的誤りについて

◎統計的仮説検定とは

・仮説が正しいかどうかを、データから判定する手法。

主張したいこと　vs　主張とは逆の仮説＝帰無仮説

↑これを「真だ」と言うために　　↑これを設定して

95%とか99%とかの
信頼性のもとに、帰無仮説を否定。

「だから、対立仮説が真なんじゃないの？」と
強く主張する

帰無仮説が正しければ：
データは 90% の割合で
95%
99%
など

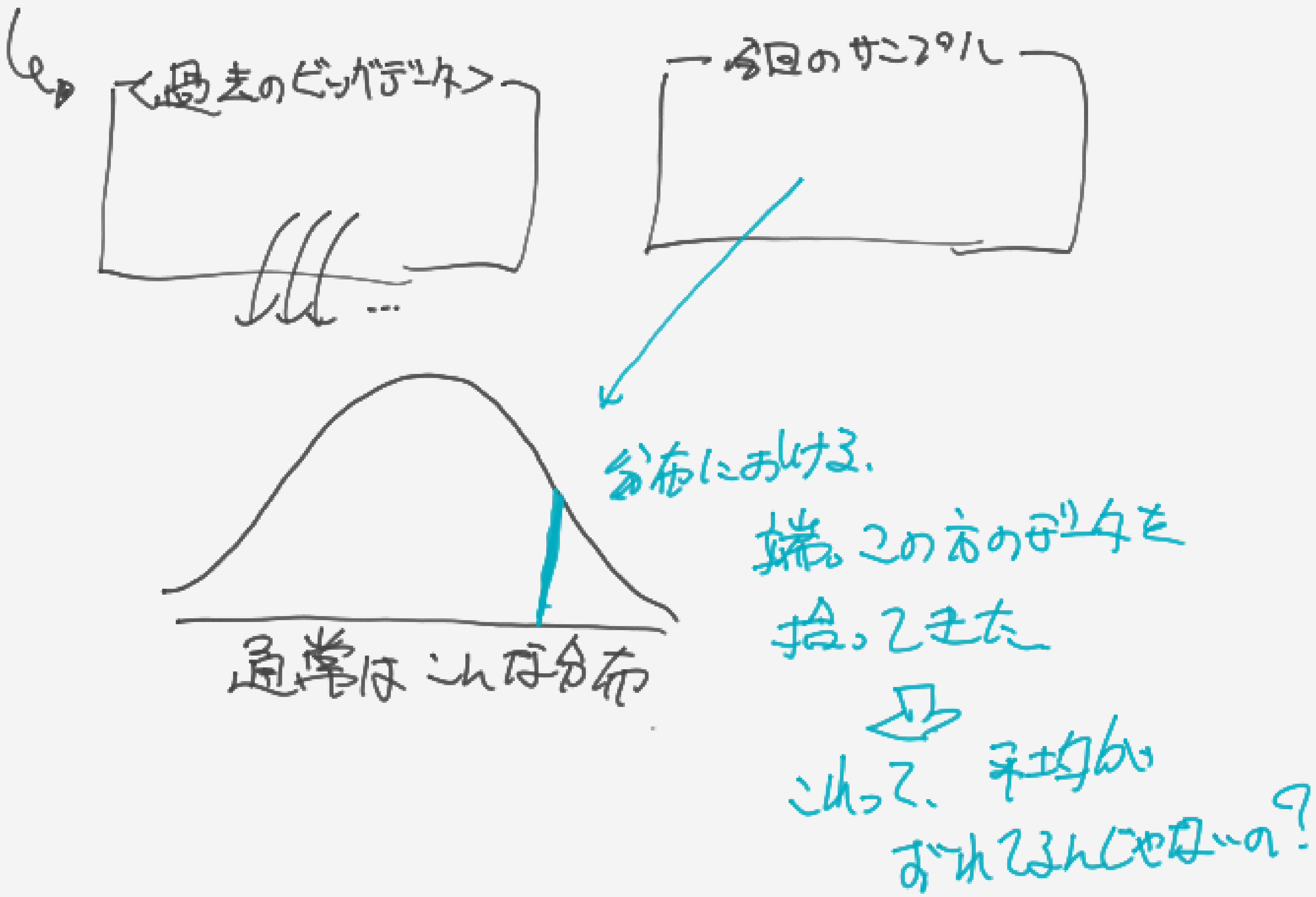
緑の領域に
収まるはず

なのに、今回とってきたサンプルは
黄色の中に入った

これは…前提条件（帰無仮説）は
おかしいと、統計的に判断して
対立仮説が正しいと結論づける。

統計的仮説検定ってどう使うの？

・母集団（いつもの状態）と変わった？



・製造業なら、たとえば管理図を作っているとき
得られたサンプルデータが「いつもと違うな」と思ったら

平均が変わったんじゃないか？
ばらつきが変わったんじゃないか？

なんてことを確認できる。

◎統計的仮説検定の手順

step1. 帰無仮説・対立仮説を立てる.

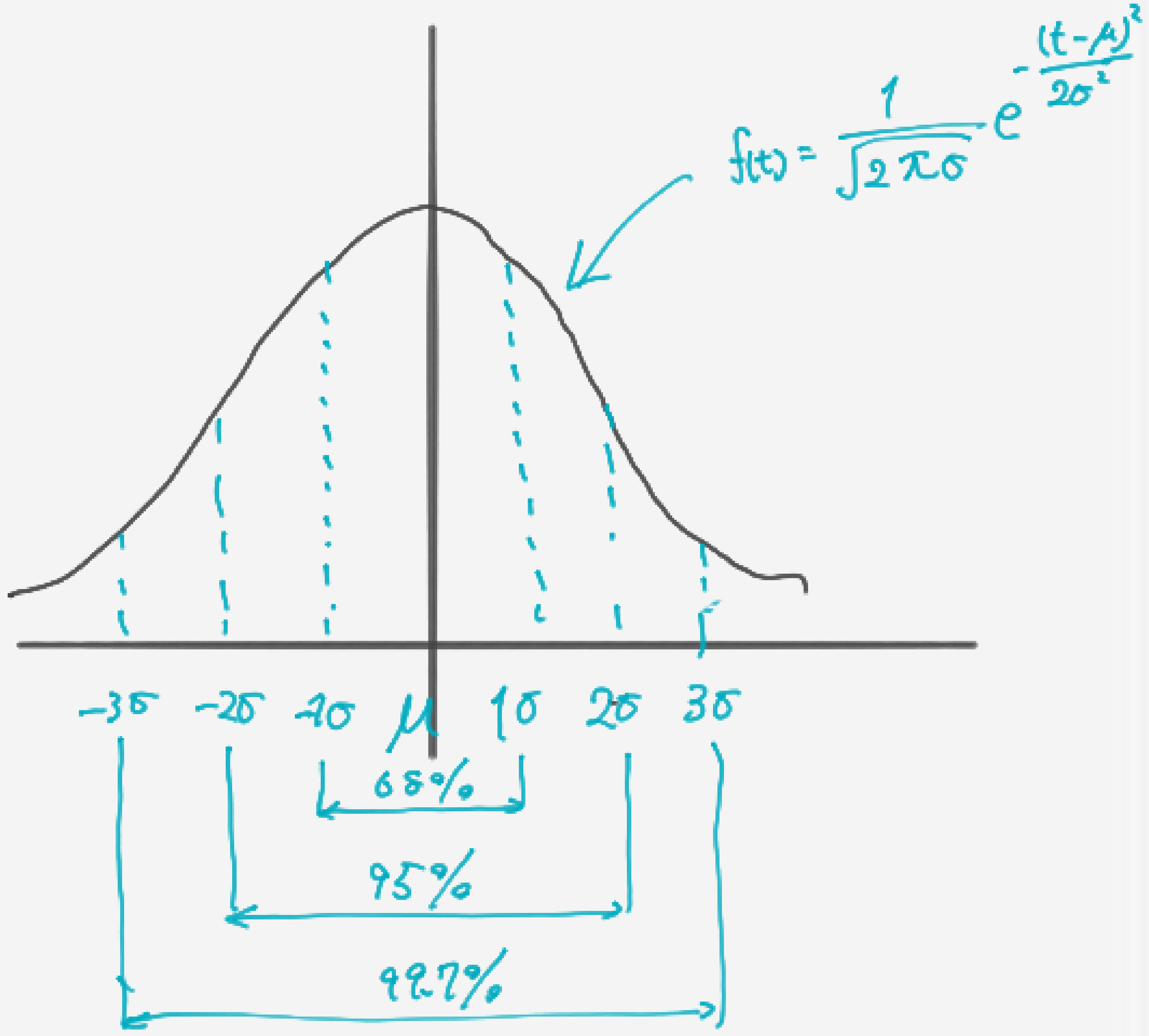
Step2. 有意水準αを決める.

Step3. 帰無仮説のもとで、現実におこったことの確率を計算する.
検定統計量.

step4. 検定統計量の確率分布をみる.

Step5. 検定統計量が有意水準以内であるかを確認する

Step6. 帰無仮説を棄却または採択する



◎検定の種類の選び方

- 母平均μの検定（母分散σ^2が既知） → z検定
- 母平均μの検定（母分散σ^2が未知） → t検定
- 母平均μ_1, μ_2の検定（母分散σ_1^2, σ_2^2が既知，かつ$\sigma_1^2 = \sigma_2^2$） → z検定
- 母平均μ_1, μ_2の検定（母分散σ_1^2, σ_2^2が未知，かつ$\sigma_1^2 = \sigma_2^2$） → t検定
- 母平均μ_1, μ_2の検定（母分散既知，かつ$\sigma_1^2 \neq \sigma_2^2$） → z検定
- 母平均μ_1, μ_2の検定（母分散未知，かつ$\sigma_1^2 \neq \sigma_2^2$） → ウェルチの検定
- 母分散σ_1^2, σ_2^2の検定 → F検定

$$\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \to N(0,1)$$

$$\frac{\bar{X} - \mu_0}{\sqrt{\frac{s^2}{n}}} \to t_{n-1}$$

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{m+n}{mn}\sigma^2}} \to N(0,1)$$

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{m+n}{mn}\hat{\sigma}^2}} \sim t_{m+n-2}$$

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \to N(0,1)$$

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}} \to t_f$$

$$f := \frac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{s_X^2}{m(m-1)} + \frac{s_Y^2}{n(n-1)}}$$

$$\frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{s_X^2}{s_Y^2} \sim F_{m-1,n-1}$$

（∵ $\sigma_X^2 = \sigma_Y^2$：仮定より）

○ 統計的誤りについて

・第1種の誤り

本当は帰無仮説が正しいのに、棄却してしまう。

本当はセーフなのに、アウトにしてしまう。

・第2種の誤り

本当は対立仮説が正しいのに、帰無仮説を棄却しない

本当はアウトなのに、セーフにしてしまう。