入門 統計学　第5章

χ^2分布とF分布

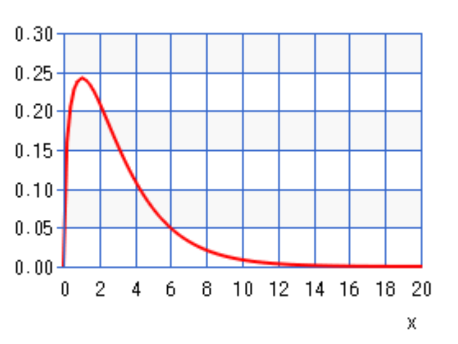
『入門 統計学 第2版　―検定から多変量解析・実験計画法・ベイズ統計学まで―』（オーム社）

※注：本書を購入された方へのサービスですので，教科書指定（参考図書は不可）していない授業での使用はお控えください。

本章で学ぶ2つの確率分布

χ^2分布 と F分布

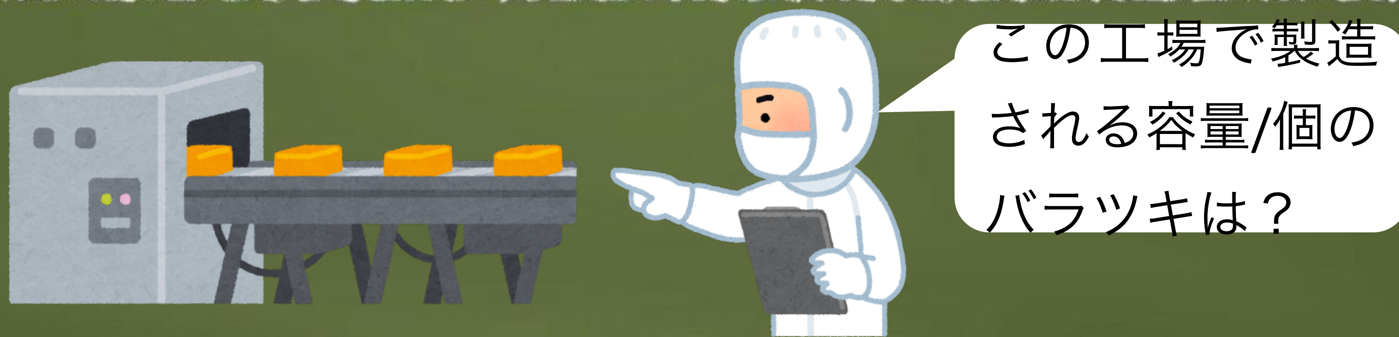
- カイジジョウ
  χ^2分布：データの平方和である統計量（χ^2）が従う連続型の確率分布
  →母分散の区間推定や独立性の検定に使用
- F分布：2つのχ^2の比が従う連続型の確率分布
  →等 分散分析に使用



χ^2分布
（$\nu=3$）

F分布
（$\nu_1=3, \nu_2=5$）

5.1　χ^2分布（カイジジョウ）



- 母分散の区間推定をしたい
  →標本分散が従う確~~率~~分布を使えよい？
  残念ながらありません
- でも，標本分散と比例する統計量（χ^2）が従う確率分布はある…それがχ^2分布

χ²はデータの平方和

- χ²分布に従うχ²は，複数のデータx_iを2乗して足し合わせた統計量（各データは，正規母集団からの独立抽出）χ²統計量

$$\chi^2 = x_1^2 + x_2^2 + \cdots + x_n^2 = \sum x_i^2$$

独立抽出

正規分布した母集団

特徴1：データの平方和なので負にはならない

特徴2：データの数（自由度）が増えると値が大きくなる

つまり，χ²分布の母数は自由度

実際のχ²は標準化変量の平方和

- 尺度や単位が異なる対象でも同じ分布表（後掲）を使えるように，標準化（節2.4）したデータz_iの平方和を考える

$$\chi^2_{(n)} = z_1^2 + z_2^2 + \cdots + z_n^2 = \frac{(x_1-\mu)^2}{\sigma^2} + \frac{(x_2-\mu)^2}{\sigma^2} + \cdots + \frac{(x_n-\mu)^2}{\sigma^2} = \frac{\sum (x_i-\mu)^2}{\sigma^2}$$

自由度

未だ標本平均を使っていないので標本サイズが自由度（ν=n）

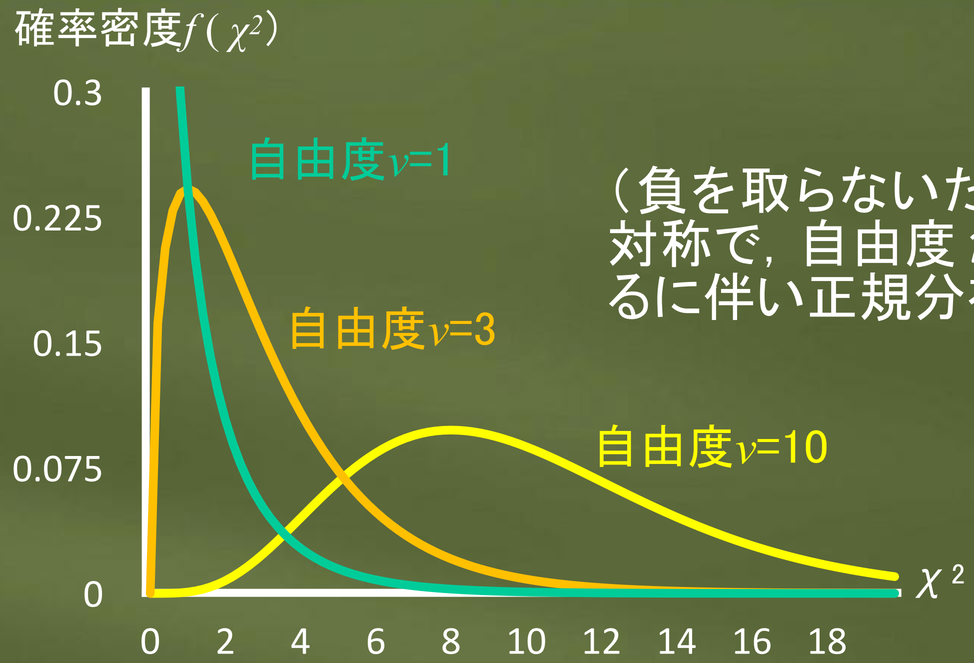
同じ母集団から抽出しているので母平均μと母分散σ^2は全て同一

χ^2分布の確率密度関数 $$f(\chi^2) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}(\chi^2)^{\frac{\nu}{2}-1}e^{-\frac{\chi^2}{2}}$$

自由度νのみが分布形に影響する母数であることが確認できる（図を次掲）

任意のχ^2値よりも右側の確率を求めるExcel関数
=CHISQ.DIST.RT（χ^2値，自由度）

χ²分布の形

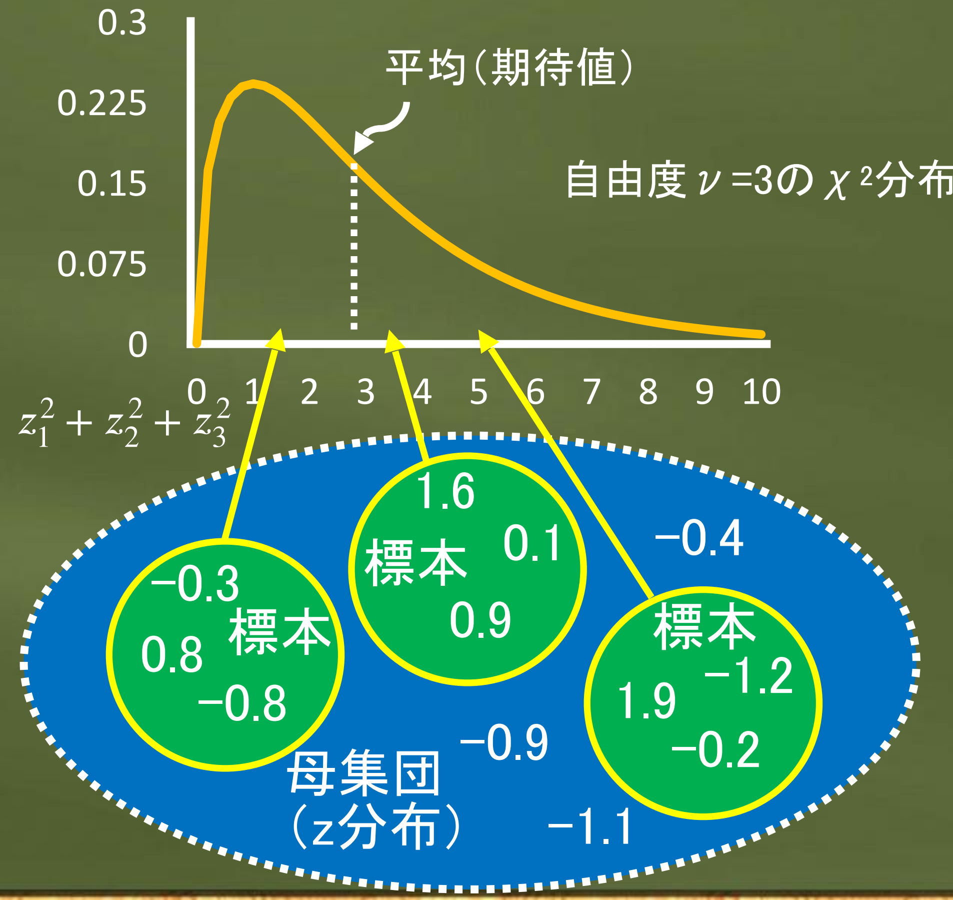


（負を取らないため）左右非対称で，自由度νが大きくなるに伴い正規分布に近づく

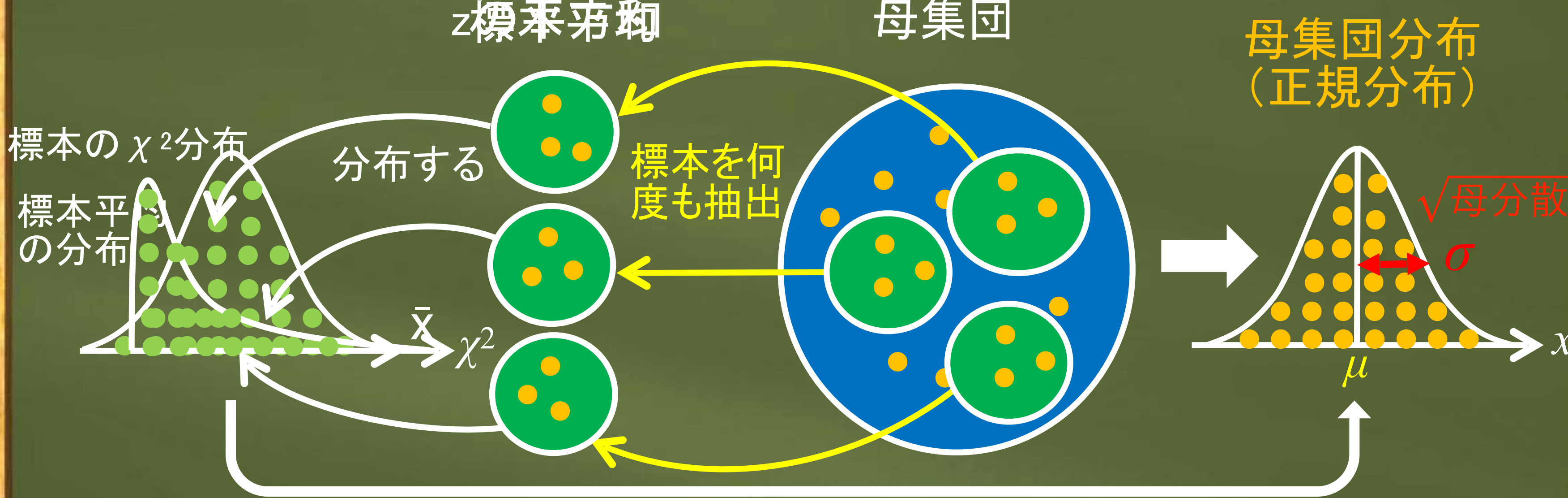
平均（期待値）＝自由度ν
分散＝自由度の2倍（2ν）

χ²分布の再確認

（母分散が既知で自由度が3の場合）



5.2　母分散の区間推定

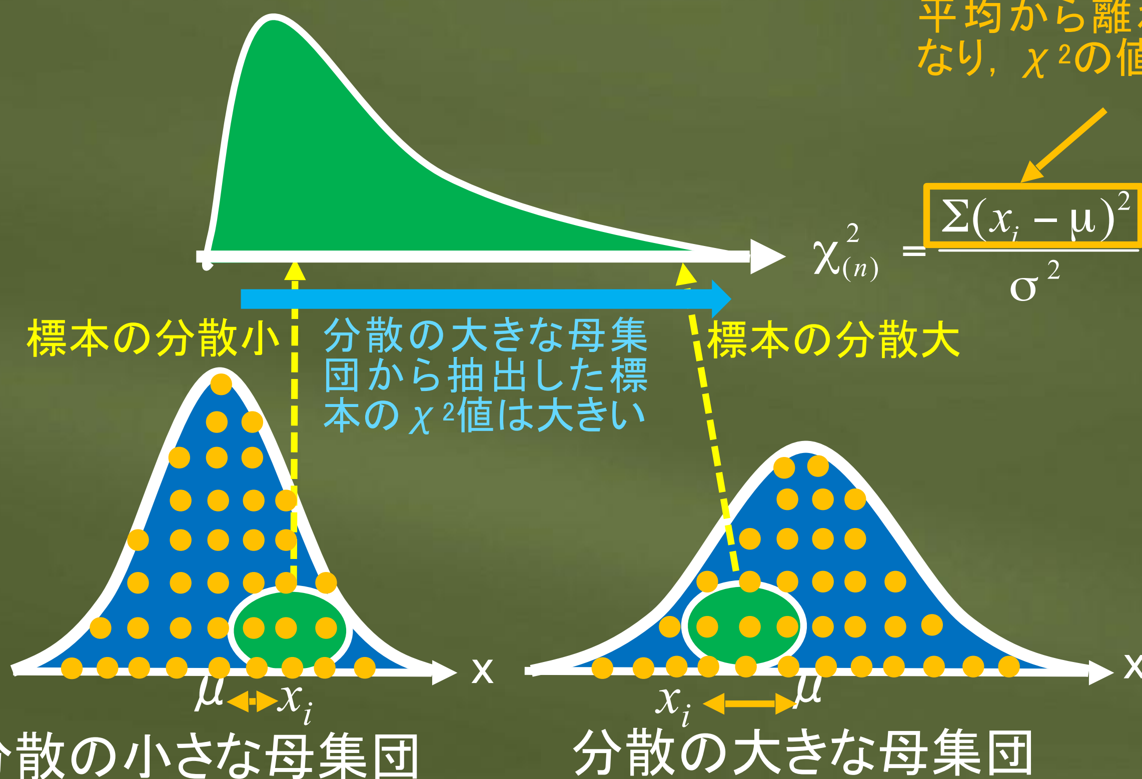


母集団の分散を区間推定（第4章）

なぜχ²分布から母分散の区間推定ができるのか？

異なる分散の母集団から抽出した標本でχ^2を計算したとする…（本来は母分散は不変）

データのバラツキが大きいと，平均から離れたデータも多くなり，χ^2の値は大きくなる



母標準偏差を基準とした「平均からのズレ具合」（つまり分散）

χ^2は標本の分散と連動している
→この性質を利用して母集団の分散を推定
（この後，式で展開）

（χ²と）不偏分散との比例関係

未知の母平均の代わりに標本平均

不偏分散の分子と同じ

母平均が未知の $\chi^2_{(n-1)} = \dfrac{\boxed{\sum (x_i - \bar{x})^2}}{\sigma^2}$

$\hat{\sigma}^2 = \dfrac{\boxed{\sum (x_i - \bar{x})^2}}{n-1}$

標本平均を1つ使ったので
自由度も1つ減る

両式の分子を整理する

$\sigma^2 \times \chi^2_{(n-1)} = (n-1) \times \hat{\sigma}^2$

↓ χ²の式にする

$\chi^2_{(n-1)} = \dfrac{(n-1) \times \hat{\sigma}^2}{\sigma^2}$

定数

χ²は母集団の
不偏分散と比例

母分散の式にする

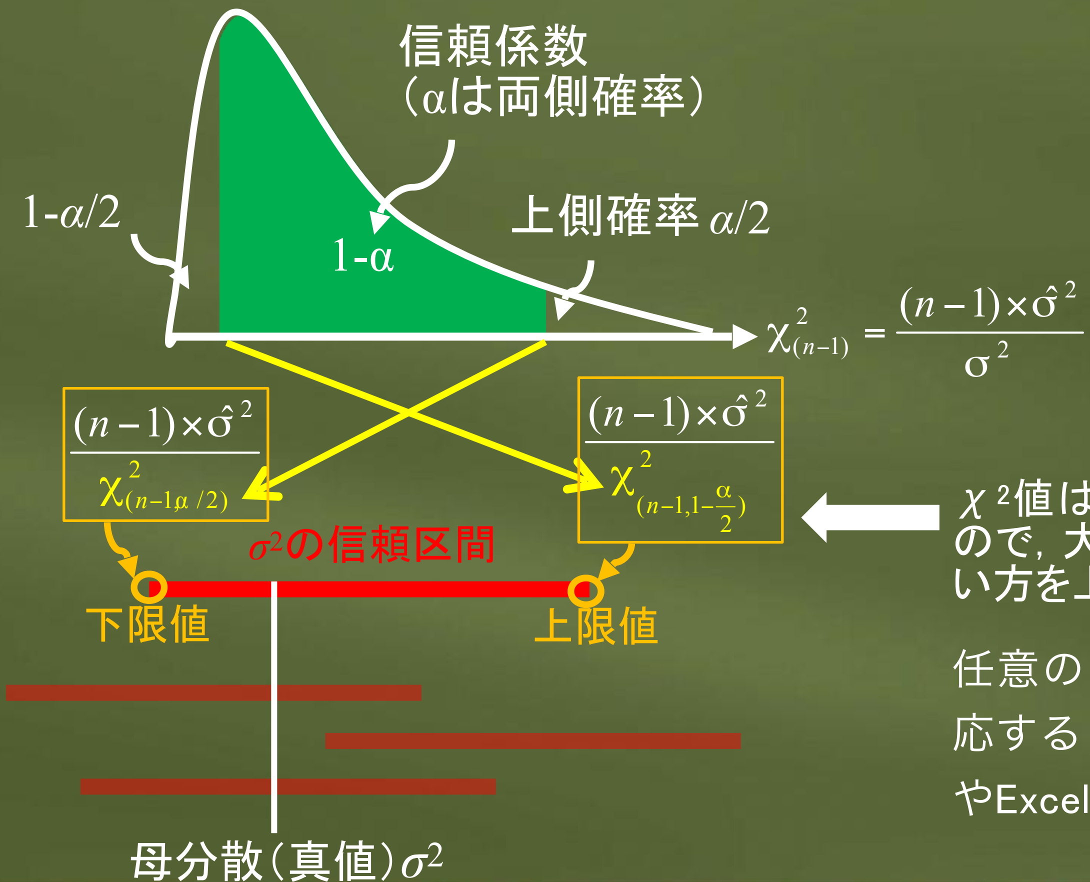
標本データ
から計算可

母分散の推定式

$\sigma^2 = \dfrac{(n-1) \times \hat{\sigma}^2}{\chi^2_{(n-1)}}$

分布するので幅を持つ

母分散の区間推定



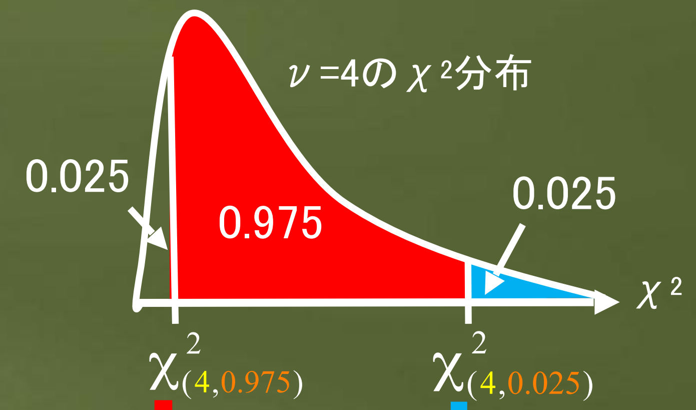
χ^2値は信頼限界の分母になるので，大きい方を下限値に，小さい方を上限値の計算に用いる

任意の自由度と上側確率に対応するχ^2値は分布表（次掲）やExcel関数から得る

χ^2 分布表の読み方(付録Ⅲ)

上側確率p

自由度 ν	0.995	0.990	0.975	0.025	0.010
1	0.000	0.000	0.001	5.024	6.635
2	0.010	0.020	0.051	7.378	9.210
3	0.072	0.115	0.216	9.348	11.345
4	0.207	0.297	0.484	11.143	13.277
5	0.412	0.554	0.831	1[illegible].833	15.086
6	0.676	0.872	1.237	14.4[illegible]	[illegible]



信頼係数95%の区間を求めるためのχ²値

パソコンがあるならば，例えば下記のExcel関数で上側確率に対応するχ²値が得られる

Excel関数=CHISQ.INV.RT(上側確率，自由度)

例：=CHISQ.INV.RT(0.975,4)→0.484

例　題

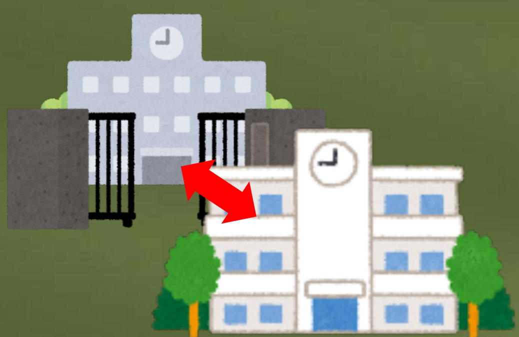
テントウムシを5匹採集しました。それらの体長は5，8,10，11，15でした（㎜）。このテントウムシの体長の母分散に対する信頼区間を信頼係数95％で推定せよ。

解：まず，不偏分散$\hat{\sigma}2$を計算すると13.7㎜²となる。それを母分散の信頼区間を求める下の式に代入すれば良い。ただし，n=5で，χ²値は先ほどの分布表から読み取る（α=0.05）。すると，信頼係数95％の母分散σ²に対する信頼区間は（4.92㎜², 113.22㎜²）となる。これの平方根を取った（2.22㎜, 10.64㎜）は母標準偏差の信頼区間となる。

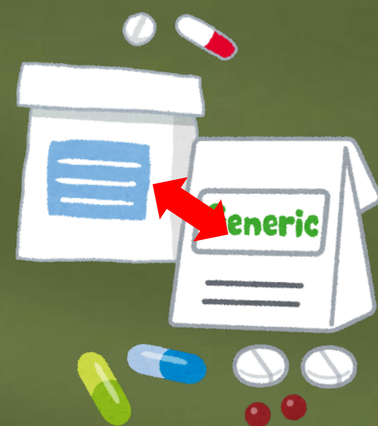
$$\frac{(n-1)\times\hat{\sigma}^2}{\chi^2_{(n-1,\alpha/2)}} < \sigma^2 < \frac{(n-1)\times\hat{\sigma}^2}{\chi^2_{(n-1,1-\alpha/2)}} \rightarrow \frac{(5-1)\times 13.7}{11.143} < \sigma^2 < \frac{(5-1)\times 13.7}{0.484}$$

5.3　バラツキを比較できる確率分布

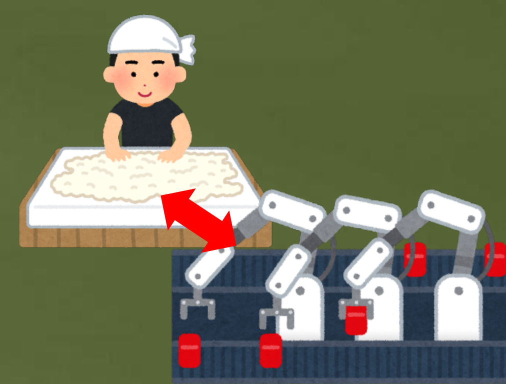
中学校Aと中学校Bとでは，数学の成績のバラツキに差はあるか？



先発薬とジェネリックでは，効果のバラツキに差はあるか？

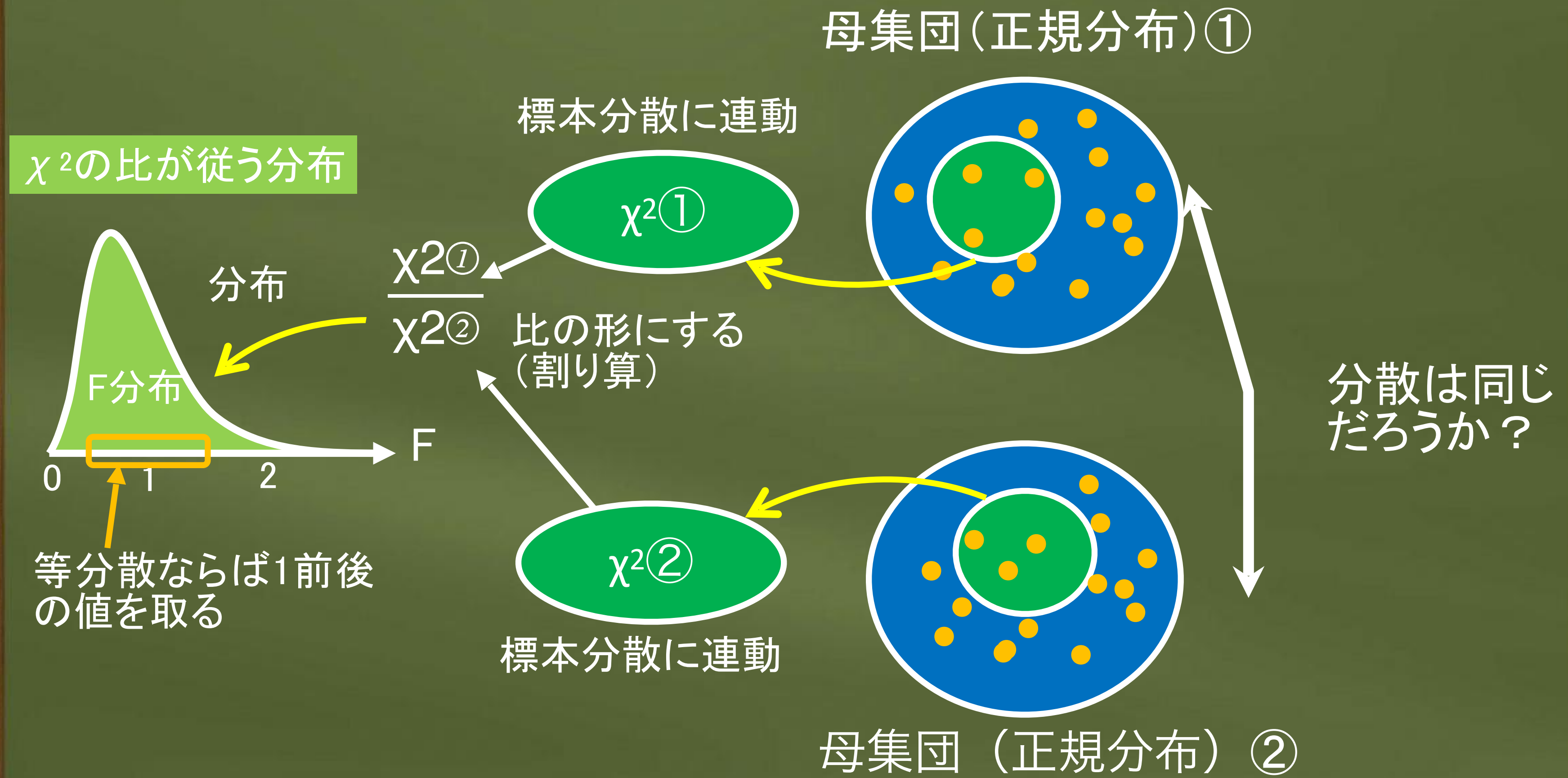


ロボット導入の前後では品質のバラツキに差はあるか？



こうした2群の母分散の差の検定（第7章）のほか，処理効果の有無の検定（分散分析：第8章）に使える確率分布がF分布

F分布


母集団（正規分布）①
標本分散に連動
χ2の比が従う分布
χ²①
分布
χ2①
χ2②
比の形にする
（割り算）
F分布
F
0
1
2
分散は同じ
だろうか？
等分散ならば1前後
の値を取る
χ²②
標本分散に連動
母集団（正規分布）②

F分布の確率密度関数

F分布に従う統計量 $F_{(\nu_1,\nu_2)} = \dfrac{\dfrac{\chi^2_{(\nu_1)}}{\nu_1}}{\dfrac{\chi^2_{(\nu_2)}}{\nu_2}}$

第1自由度

第2自由度

自由度は2つ

χ^2は自由度に伴い大きくなるため，それぞれの自由度で割って自由度の影響を取り除く

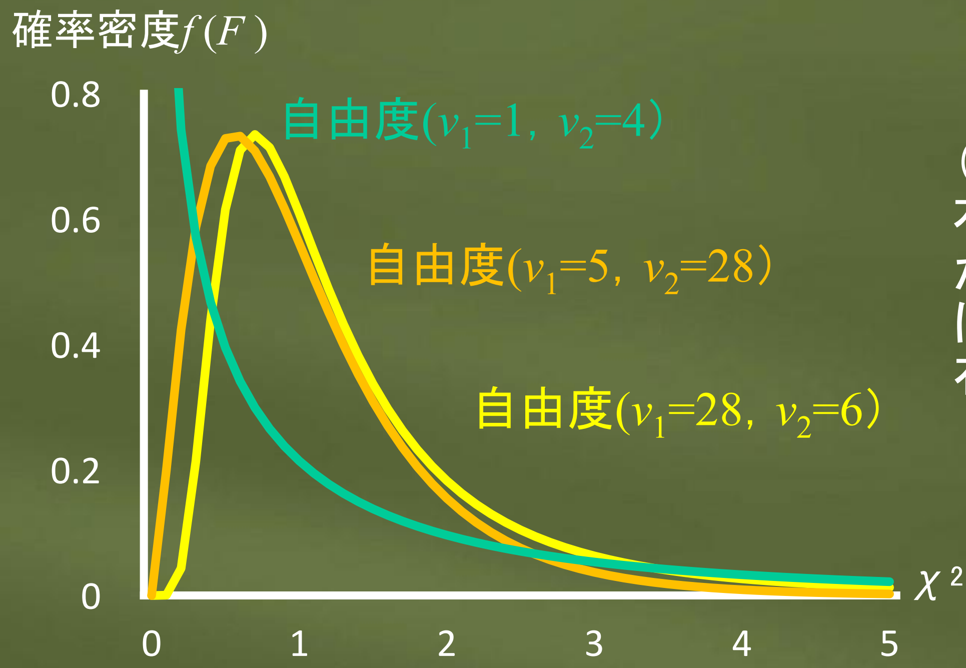
F分布の確率密度関数

$$f(F) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})F^{\frac{\nu_1-2}{2}}}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})(1+\frac{\nu_1}{\nu_2}F)^{\frac{\nu_1+\nu_2}{2}}}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}$$

2つの自由度が母数（形に影響する）ことが確認できる（図を次掲）

任意のF値よりも右側の確率を求めるExcel関数
=F.DIST.RT(F値，第1自由度，第2自由度）

F分布の形



（負を取らないため）左右非対称で，両自由度が大きくなるに伴い右に移動するが，正規分布に近づくことはない

平均（期待値）= $\dfrac{\nu_2}{\nu_2 - 2}$ (ただし $\nu_2 > 2$

分散= $\dfrac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$ (ただし $\nu_2 > 4$)

5.4　特別なF

2つの不偏分散の比になる場合

不偏分散と比例していることを示したχ^2の式
（8枚前のスライド）

$$\chi^2_{(n-1)} = \frac{(n-1)\times\hat{\sigma}^2}{\sigma^2} \xrightarrow{\text{自　由（度1)を}} \chi^2_{(\nu)} = \frac{\nu\hat{\sigma}^2}{\sigma^2}$$

Fの式に代入

2つの不偏分散/母分散の比となるF

$$F_{(\nu_1,\nu_2)} = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} = \frac{\dfrac{\cancel{\nu_1}\hat{\sigma}_1^2}{\sigma_1^2}/\cancel{\nu_1}}{\dfrac{\cancel{\nu_2}\hat{\sigma}_2^2}{\sigma_2^2}/\cancel{\nu_2}} = \frac{\hat{\sigma}_1^2/\sigma_1^2}{\hat{\sigma}_2^2/\sigma_2^2}$$

抽出元の母集団が同じなら母分散も同じ

同じ母集団から抽出したときのF（等分散の検定や分散分析の仮説）

$$F_{(\nu_1,\nu_2)} = \frac{\hat{\sigma}_1^2/\cancel{\sigma^2}}{\hat{\sigma}_2^2/\cancel{\sigma^2}} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

2つの不偏分散の比になる

1前後の値ならば同じぐらいの分散なので仮定は正しそう…

Fとtの関係

分母と分子を母標準誤差で割る

$$t_{\bar{x}} = \frac{\bar{x}-\mu}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}} = \frac{\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\dfrac{\hat{\sigma}/\cancel{\sqrt{n}}}{\sigma/\cancel{\sqrt{n}}}} = \frac{z_{\bar{x}}}{\dfrac{\hat{\sigma}}{\sigma}} = \frac{z_{\bar{x}}}{\sqrt{\dfrac{\chi^2_{(\nu)}}{\nu}}}$$

←これを2乗すると…

$$\chi^2_{(\nu)} = \frac{\nu\hat{\sigma}^2}{\sigma^2}$$

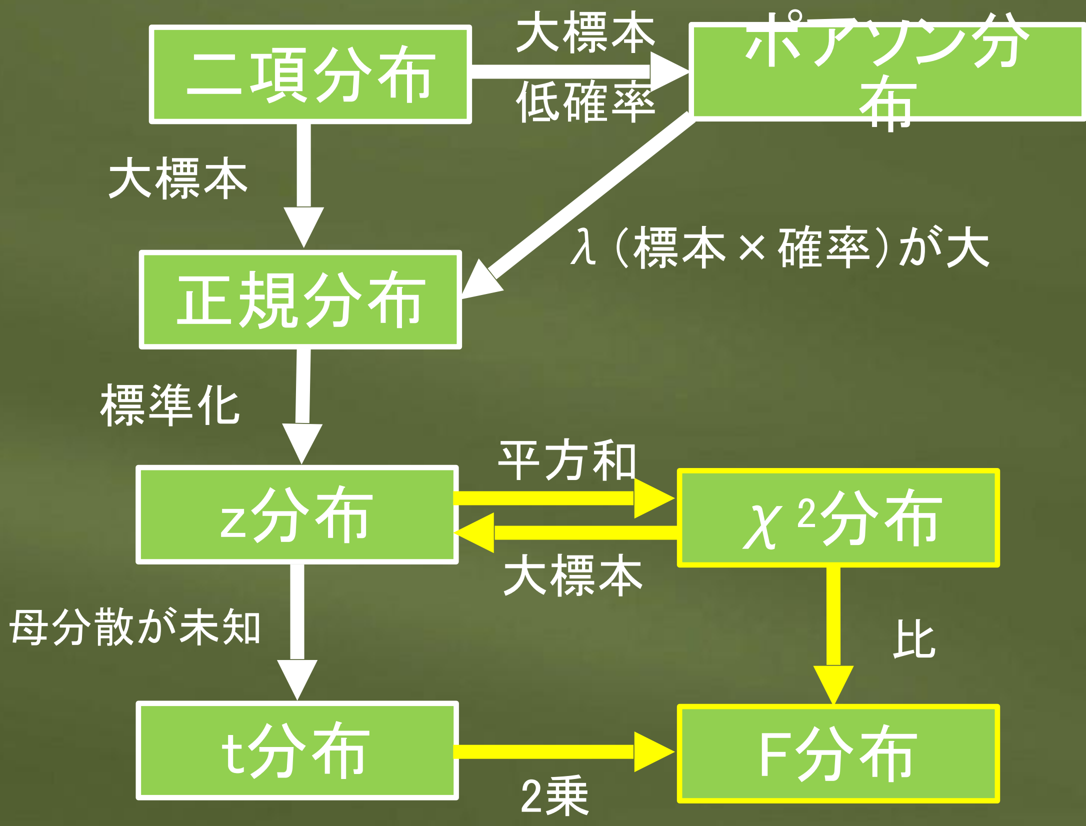
z^2は自由度1のχ^2

$$t^2_{\bar{x}} = \frac{z^2_{\bar{x}}}{\chi^2_{(\nu)}/\nu} = \frac{\chi^2_{(1)}/1}{\chi^2_{(\nu)}/\nu} = F_{(1,\nu)}$$

tの2乗は分子の第1自由度が1の“F”

2群の分散分析（F検定）とt検定は同じになる

いろいろな確率分布の関係



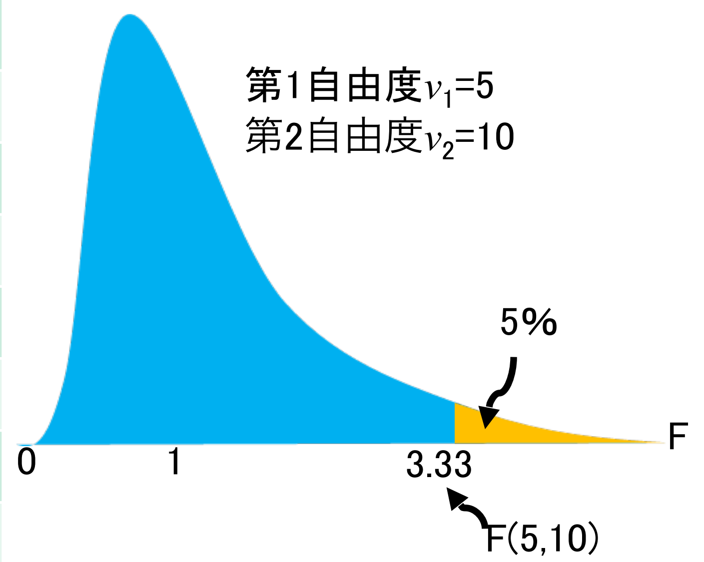
F分布表と読み方（付録IV）

上側確率5%で1つの表を使う

第1自由度（分子）

第2自由度（分母）

5%	1	2	3	4	5	6	7	8	9	10	15	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94						
4	7.71	6.94	6.59	6.39	6.26	6.16						
5	6.61	5.79	5.41	5.19	5.05	4.95						
6	5.99	5.14	4.76	4.53	4.39	4.28						
7	5.59	4.74	4.35	4.12	3.97	3.87						
8	5.32	4.46	4.07	3.84	3.69	3.58						
9	5.12	4.26	3.86	3.63	3.48	3.37						
10	4.96	4.10	3.71	3.48	3.33	3.22						



=F.INT.RT(0.05,5,10)

以上で第５章は終了です。