エクセルを用いた区間推定のシミュレーション

# シミュレーションの目的

* 母集団平均の95％信頼区間の構成を100回繰り返し，100回のうち95回前後は母集団平均をとらえられているということを観察します．これにより，95％という意味を頻度的に理解します．
* ＜オプション＞特定の標本における測定値の分布のグラフと，100回計算された標本平均の分布（理論的には，標本分布）のグラフを描きます．これにより，標本分布の概念を理解します．
＃グラフを描くためには，適当に階級を設定して，出現した値を分類する必要があります．

# シミュレーションの方法（概要）

1. 平均50，標準偏差10の正規分布に従う母集団から，大きさ100の標本を抽出して標本平均と標準偏差を計算します．標準偏差の計算では，不偏推定値が得られるように，偏差平方和を99で割ります（エクセルのSTDEV.S関数を使います）．
2. 標本から計算された平均と標準偏差を用いて，母集団平均の区間推定を行います．母集団平均の95％信頼区間を構成します．
3. 以上の作業（標本抽出と平均の計算）を100回繰り返します．構成された100個の信頼区間のうち，母集団平均をとらえているものはいくつあるか数えます．おおよそ95回前後は母集団平均をとらえていることがわかります．

# シミュレーション（Microsoft Excel 2010を使用）の実行手順

A2 セルに標本001と入力します．この行（第２行）で１回目の標本抽出を行います．



A3セルに「標本002」，A4セルに標本「003」，・・・，A101セルに「標本100」と入力します．ひとつひとつ入力していく必要はありません．A2セルを選択した状態で，枠の右下にマウスを合わせます．マウスのポインタが「+」になったら，左ボタンを押したままA101セルまでドラッグします．ボタンを離すと，下図のように，「標本100」までが入力されます．



B1セルに x\_01 と入力します．



C1セルに「x\_002」，D1セルに標本「x\_003」，・・・，CW1セルに「x\_100」と入力します．ひとつひとつ入力していく必要はありません．B1セルを選択した状態で，枠の右下にマウスを合わせます．マウスのポインタが「+」になったら，左ボタンを押したままCW1セルまでドラッグします．ボタンを離すと，下図のように，「x\_100」までが入力されます．



１回目の標本抽出を行います．B2 セルに，=NORM.INV(RAND(), 50, 10) と入力します．これは，平均50，標準偏差10 の正規分布に従う乱数を発生させます．

注意：Microsoft Excel 2007 以前のエクセルでは，NORMINV関数を用います（NORMとINVの間にドットがない）．



平均50，標準偏差10 の正規分布から乱数がひとつ抽出されます．乱数ですから，みなさんのシミュレーションでは，下図とは異なった値が出現します．



B2 セルに入力した式を，「標本001」の行の，x\_02からx\_100までコピーします．上で説明したようにマウスのドラッグ操作でコピーできますが，もっと簡単にコピーを行う方法を示します．B2セルを選択し，「ホーム」タブにある「コピー」のアイコンをマウスで左クリックします．Ctrl キーとCキーを同時に押してもいいです．「コピー」のアイコンをクリックしたことと同じになります．



B2セルの周囲を破線が回転している状態で（下図では示されていませんが），C2セルをマウスで左クリックします．



エクセルのウィンドウ内部下側にあるスライダーで，シート内を右に移動します．CW2セルが見えたら，Shiftキーを押しながら，CW2セルをマウスで左クリックします．すると，B2セルをコピー元として，C2セルからCW2セルがコピー先として選択された状態になります．



「ホーム」タブにある「貼り付け」のアイコンをマウスで左クリックして，コピーを実行します．Ctrl キーとVキーを同時に押してもいいです．「貼り付け」のアイコンをクリックしたことと同じになります．



これで，平均50，標準偏差10 の正規分布をしている母集団から，大きさ100の標本抽出ができたことになります．B2セルの乱数は再計算されて値が変わりますが，気にしなくていいです．以下，何か操作を実行するたびに乱数の再計算がされて値が変わりますが，気にする必要はありません．



データは得られているけれども，母集団の平均と標準偏差は未知であるとしましょう．われわれはもちろんこれらの値を知っています．どのようにデータを作成したかは秘密にして，誰かにデータを渡すことを想像します．データを渡された人は，この標本での平均と標準偏差を計算し，信頼区間を構成します．下図のように，CXセルからDAセルまでに，「標本平均」，「標準偏差」，「95%下限」，「95%上限」とそれぞれ入力してください．



標本平均を計算します．CX2セルに，=AVERAGE(B2:CW2) と入力します．



標準偏差（不偏分散の正の平方根）を計算します．CY2セルに，=STDEV.S(B2:CW2) と入力します．

注意：Microsoft Excel 2007 以前のエクセルでは，STDEV関数を用います．



計算された標本平均と標準偏差を使って，母集団平均の95%信頼区間を構成します．

下限は，平均から標準誤差の1.96倍を引いた値です．ここでの標準誤差とは，平均値の標本分布での標準偏差です．母集団での標準偏差をとすると，大きさ*n*の標本では，標準誤差はになります．母集団での標準偏差が未知の場合は，代わりに標本から計算される不偏分散を使います．

CZ2セルに =CX2 – 1.96\*(CY2/10) と入力します．



信頼区間の上限は，下限の計算での引き算を足し算に変えるだけです．

DA2セルに =CX2 + 1.96\*(CY2/10) と入力します．



構成した信頼区間が，母集団平均（50）をとらえているかどうかテストします．とらえられていた場合に限り，OK と表示することにしましょう．

DB2 セルに，=IF(AND(CZ2<=50, DA2>=50), “OK”, “ ”) と入力します．信頼区間の下限が50以下であり，かつ，上限が50以上であれば，この信頼区間は母集団平均をとらえていることになります．



概念的には，この作業を100回繰り返し，区間推定を100回行います．実際の作業はコピーだけです．ワークシートの２行目（１回目の標本抽出）で入力した数式（B2セルからDB2セル）を，他の行（３行目から101行目）にコピーしましょう．コピー先が広範囲なので，上で説明した方法をうまく使ってください．

1. B2セルからDB2セルを選択して，「コピー」のアイコンをクリックする．
2. B3セルをマウスで左クリックする．
3. スライダーを使ってシート内を移動し，DB101セルが見えるようにする．
4. Shiftキーを押しながら，DB101セルをマウスで左クリックする．これにより，B2セルからDB2セルをコピー元として，B3セルからDB101セルがコピー先として選択された状態になる．
5. 「貼り付け」のアイコンをクリックして，コピーを実行する．



区間推定を100回繰り返しました．DB列を見て，OKとなった数を数えてください．実際には，OKとならなかった数を数えて，それを100から引きます．95%の信頼区間ですから，おおよそ95回前後はOKとなるはずです．