

サンプリング (1/2)

2012年11月27日

後 保範

1

0. 分析ツールの使用

- Excelの分析ツールの使用準備
 - (1) Excel左上の「[ファイル](#)」をクリック
 - (2) 「[Excelオプション](#)」をクリック
 - (3) 「[アドイン](#)」をクリック
 - (4) 管理に「[Excelアドイン](#)」を選定
 - (5) 「[設定](#)」をクリック
 - (6) 「[分析ツール](#)」をオンにし「[OK](#)」をクリック

2

0.1 分析ツールの使用方法

- Excel2010を立ち上げる
 - (1) Excel画面の「[データ](#)」を選定
 - (2) 画面右上の「[データ分析](#)」をクリック
 - (3) 「[分析ツール](#)」から必要ツールを選択
 - (4) 「[OK](#)」をクリック
→ 各分析ツールにより処理が異なる

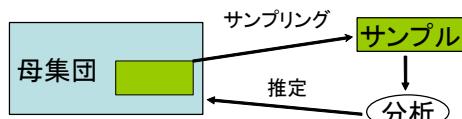
3

0.2 結果の判定方法

- 2章(検定)、3章(分散分析)及び4章(サンプリング)の結果判定
 - (1) 評価値が有意水準(通常5%)以下なら
→ 「差がある」あるいは「異なる」と判定
 - (2) 有意水準以上なら
→ 「差がない」あるいは「同じ」と判定

4

1. サンプリング



目的:ある集団(母集団)の状態を知る
→ 母集団全体の把握は困難
→ 母集団を代表するもの(サンプリング)
→ 傾りをなくす(ランダム)

5

1.1 サンプリングの方法

- (1) 単純ランダムサンプリング
- (2) 2段サンプリング
- (3) 層別サンプリング
- (4) 集落サンプリング
- (5) 系統サンプリング

6

2. 単純ランダムサンプリング

- 母集団を代表するサンプルをランダムに抽出する。
- ランダムとは「規則性や人為性でなく、出現確率がほぼ均等な状態を示す」、代表例はサイコロを振ったときの目。
- Excelはランダムに値をつくる機能がある。
- 単純ランダムサンプリングでは、全体(母集団)からランダムに抽出する。

7

2.1 単純サンプリングの方法

- F4-1.xlsxのシート1の図4-2 1ロットの製品から「単純ランダムサンプリング」で10個の製品をランダムに抜き出す。
- データ → データ分析 → サンプリング → 入力範囲(図4-2) → 標本の採取方法「ランダム」 → 標本数「10」 → 出力先 → 同一画面に分析結果出力

8

2.2 サンプリングの標準誤差

- n 個の単純サンプルを採取し、 σ をサンプルの標準偏差とする。

(1) 無限母集団のサンプルからの標準誤差

$$\text{標準誤差} = \frac{\sigma}{\sqrt{n}}$$

(2) N 個の母集団のサンプルからの標準誤差

$$\text{標準誤差} = \sqrt{1 - \frac{n}{N}} \frac{\sigma}{\sqrt{n}}$$

9

2.3 標準誤差の計算例

- F4-1.xlsxのシート2の図4.8のサンプリングデータを使用して標準誤差を計算する
- 準備(サンプルの平均と標準偏差)
 - 平均(average)、標準偏差(stdev)を計算
 - 無限母集団($n=30$)
 - 標準誤差 = 0.19943
 - ロット100個の有限母集団($N=100, n=30$)
 - 標準誤差 = 0.16686

10

2.4 母集団の平均の検定

- 母集団の平均(μ)が a である確率は標準誤差とnormdist関数で求められる。

確率 = normdist(μ , a , 標準誤差, TRUE)ただし、 $\mu < a$ とする、小さい方を先頭へ
確率が5%以上 → 母集団平均は a と言える

(1) 無限母集団(平均が50と言えるか)

確率=36% → 平均は50と言える

(2) ロット100個の有限母集団

確率=33% → 平均は50と言える

11

3. 2段サンプリング

- 製品の小箱がいくつか集まってロットを形成している時などに使用する。
- 単純サンプリングに比較して、開ける小箱の数を少なくする利点がある
- 1ロットに小箱が M 箱、小箱の中に N 個の製品があり、小箱を m 箱選び、各小箱から n 個の製品をサンプリングする

12

3.1 2段サンプリングの標準誤差

- 1ロットに小箱がM箱、小箱の中にN個の製品があり、小箱をm箱選び、各小箱からn個の製品をサンプリングする
- σ_m : 小箱同士の標準偏差
 σ_n : 小箱内の標準偏差
 「分散分析:一元配置」の分散分析表から求める
- ロットの標準誤差

$$\text{標準誤差} = \sqrt{\left(1 - \frac{m}{M}\right) \frac{\sigma_m^2}{m} + \left(1 - \frac{n}{N}\right) \frac{\sigma_n^2}{mn}}$$

13

4. 層別サンプリング

- 母集団が、異質な成分の混ざった状態で構成されている時、同じ成分同士で分類し、それぞれ区分した中からサンプリングする方法
- 例えば、対象者を性別、年代でグループ分けして、それぞれのグループからランダムに抽出した人にアンケートを実施する場合
- 層別サンプリングは、サンプリングの精度を高めたいときに利用される。単純サンプリングより手間はかかるが推定精度が高い

14

4.1 層別サンプリングの種類

- (1) 層別比例サンプリング
 それぞれのグループの個数に比例してサンプル数を決める方法
- (2) ネイマンサンプリング
 母集団の推定精度が最もよくなるように、それぞれのグループの個数×標準偏差の数に比例
- (3) デミングサンプリング
 母集団の推定精度を一定のレベルに保ちながら、調査費用も考慮して、抽出数を決める

15

5. 系統サンプリング

- 単純ランダムサンプリングより、より簡単に実施するための方法
- 例として、順次流れている製品や棚に並んでいる製品に対して、一定間隔ごとにサンプリングする方法
- 最初のスタートは、実施するたびにランダムに変更する必要がある
- 周期性があるものは、サンプリング周期を適時変更する必要がある

16

6. 演習問題

- F4-1A.xlsxのシート1に示す、「図4-1 A社の自動車の燃費データ」は無限母集団とみなされるA社の車の中から単純ランダムサンプリングで抽出したサンプリングデータである。
- このサンプリングデータから、無限母集団のサンプルとして求められる標準誤差を求めよ
- A社のこの車の平均燃費は12km/lと言えるか検定せよ。有意水準は5%とする。また、平均燃費がより大きい12.5km/lと言えるか。

17