

レポート提出票

科目名: 情報工学実験2

実験テーマ: 実験テーマ4 統計的推測と単回帰分析

実施日: 2020年 9月 21日

学籍番号: 4619055

氏名: 辰川力駆

共同実験者: _____

1 はじめに

本実験では、推定と検定の考え方と手順を理解することを目標とする。

2 目的

1. 推定と検定の考え方と手順

推定と検定の目的、考え方、手順を理解する。

2. t 分布

t 分布の定義とその性質を理解する。

3. 母分散が未知の場合の母平均の推定と検定

母分散が未知の場合の母平均の推定と検定の考え方と手順を理解する。さらに、推定と検定の関係を理解し、各方法の特徴を理解する。

3 実験方法

3.1 実験 1 推定と検定の考え方と手順

1. 平均 $\mu = 20.0$, 分散 $\sigma^2 = 0.4^2$ の正規分布 $N(20.0, 0.4^2)$ に従う乱数データを 10 個 \times 1000 セット 発生させる。

2. 10 個の乱数データの平均値 $\bar{X}_s = \frac{\sum_{i=1}^n X_{is}}{n}$ をそれぞれ (s ごとに) 計算する。

3. 1000 個の \bar{X}_s の平均値 $\bar{X} = \frac{\sum_{s=1}^S \bar{X}_s}{S}$ と次式の分散を求める。

$$\frac{\sum_{s=1}^S (\bar{X}_s - \bar{X})^2}{S - 1}$$

\bar{X}_s が理論的に平均 20.0, 分散 $\frac{0.4^2}{n}$ の正規分布に従うことを踏まえて、結果を考察する。

4. 次式に基づいて、 μ に対する 95 % 信頼区間をそれぞれ計算する。分散は σ^2 は既知 ($= 0.4^2$) とする。

$$\text{下限: } \bar{X}_s - z_{\alpha/2} \sqrt{\frac{\sigma^2}{n}}, \quad \text{上限: } \bar{X}_s + z_{\alpha/2} \sqrt{\frac{\sigma^2}{n}}$$

$z_{\alpha/2}$ は標準正規分布の上側 $\alpha/2$ % 点である。ここで $\alpha = 0.05$ とする。

5. 95 % 信頼区間に母平均 $\mu = 20.0$ が含まれる回数と割合を調べて、結果を考察する。

6. 両側検定における帰無仮説を $H_0 : \mu = \mu_0 = 20.0$, 対立仮説を $H_1 : \mu \neq \mu_0 = 20.0$ とする。分散は σ^2 は既知 ($= 0.4^2$) とする。10 個の乱数データから、帰無仮説 H_0 に対する次式の検定統計量 Z_s をそれぞれ計算する。

$$Z_s = \frac{\bar{X}_s - \mu_0}{\sqrt{\sigma^2/n}}$$

7. 検定の有意水準 α は 5 % とする。 $|Z_s| > z_{\alpha/2}$ のとき、帰無仮説 H_0 を棄却する。帰無仮説 H_0 が棄却された回数と割合を調べて、結果を考察する。
8. 手順 5 と手順 7 の結果を班のメンバーで共有 (合計) して、結果を考察する。
9. 平均 $\mu = 20.1 + a$, 分散 $\sigma^2 = 0.4^2$ の正規分布 $N(20.1 + a, 0.4^2)$ に従う乱数データを 10 個 \times 1000 セット 発生させて、手順 6,7 を再度行なう。ただし、 $a =$ 学籍番号の下一桁/10 とする。
10. 手順 9 の結果を班の他のメンバーの結果と比較して、その違いを考察する。

3.2 実験 2 t 分布

1. 平均 $\mu = 0$, 分散 $\sigma^2 = 1$ の正規分布 $N(0, 1)$ に従う乱数データを $c = b + 4$ 個 \times 1000 セット 発生させる。 b は学籍番号の下一桁とする。
2. 最初の $\phi = c - 1 = b + 3$ 個の乱数データ $X_{ij} (j = 1, \dots, \phi)$ から、次式により、自由度 ϕ のカイ二乗分布に従う乱数データを 1000 個作成する。

$$V_i(\phi) = X_{i1}^2 + \dots + X_{i\phi}^2$$

3. 最後の c 番目の乱数データを U とする。式 (1) から、自由度 ϕ の t 分布に従う乱数データを 1000 個作成する。

$$T = \frac{U}{\sqrt{V/\phi}} \quad (1)$$

4. 得られた乱数データの上側 2.5 % 点を計算し、理論値と比較する。

3.3 実験 3 母分散が未知の場合の母平均の推定と検定

1. データシートから、学籍番号の下一桁に対応したデータ (7 個) を抽出する。
2. 7 個のデータを基本統計量 (データ数、平均、標準偏差、最小値、最大値) を計算する。

3. 次式に基づいて、母平均 μ に対する 95 % の信頼区間を求める。

$$\text{下限: } \bar{X} - t_{\alpha/2, v} \sqrt{\frac{s^2}{n}}, \quad \text{上限: } \bar{X} + t_{\alpha/2, v} \sqrt{\frac{s^2}{n}}$$

ここで、

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

である。 $t_{\alpha/2, v}$ は自由度 $v = n - 1$ の t 分布の上側 $\alpha/2$ % 点である。ここで $\alpha = 0.05$ とする。

4. 両側検定における帰無仮説を $H_0: \mu = \mu_0 = 12.5$, 対立仮説を $H_0: \mu \neq \mu_0 = 12.5$ とする。
7 個のデータから、帰無仮説 H_0 に対する次式の検定統計量 T を計算する。

$$T = \frac{|\bar{X} - \mu_0|}{\sqrt{s^2/n}}$$

5. 両側検定における P 値 を計算する。
6. 帰無仮説 H_0 が棄却されるかどうか判定して、検定の結果を解釈する。
7. 信頼区間と検定の結果にどのような関係があるかを数式に基づいて考察する。

4 結果・考察

4.1 実験 1 推定と検定の考え方と手順

- Excel で乱数データを発生させた。
- 10 個ごとの和を 1000 個求めた。
- 1000 個の \bar{X}_s の平均値は $\bar{X} = 20.00019272$, 分散は 0.016151084 となった。平均 $\mu = 20.0$, 分散 $\sigma^2 = 0.4^2$ の正規分布で乱数を発生させたので、ほぼ一致するはずであるが、確かに近い値を取っている。さらに、多くの乱数データを用意すれば、さらに近い値になると考える。
- μ に対する 95 % 信頼区間を求めた。
- 95 % 信頼区間に母平均 $\mu = 20.0$ が含まれる回数は 943 回となった。よって、割合は 0.943 である。正規分布の 95 % 信頼区間で計算したから 0.95 に近い値になっていると考える。
- 検定統計量 Z_s を 1000 個に対し求めた。
- 帰無仮説 H_0 が棄却された回数は 57 回となった。したがって、割合は 0.057 である。これは、元々正規分布のデータを有意水準を 5 % で考えたから近づいていると考える。

8. 自分のデータの手順 5,7 で計算した回数の和を求めると 1000 回であった。また、他の人のデータの和を考えると同様に 1000 回であった。“Z 検定の場合は”帰無仮説が棄却されないことと 95 % 信頼区間に母平均が含まれることは同様と考えられる。
9. 学籍番号は 4619055 なので $a = 5/10 = 0.5$ として、 $N(20.6, 0.4^2)$ の正規分布を考えて手順 6,7 を再度行った。結果は、帰無仮説が棄却された回数は 57 回となった。よって割合は 0.057 である。
10. 手順 9 の結果を他のメンバーと比較した。比較した人は近持さんである。近持さんの結果は、帰無仮説が棄却された回数は 42 回となった。よって割合は 0.042 である。近持さんは、平均 20.9 で考えているのでもちろん平均値は違うものの、帰無仮説が棄却された回数はほぼ同じで、5 % に近かった。

4.2 実験 2 t 分布

1. b は学籍番号の下一桁なので $b = 5$ として、正規分布 $N(0, 1)$ に従う乱数データを $c = b + 4 = 9$ 個 \times 1000 セット 発生させた。
2. 自由度 $\phi = 8$ のカイ二乗分布に従う乱数データを 1000 個作成した。
3. 自由度 $\phi = 8$ の t 分布に従う乱数データを 1000 個作成した。
4. 得られた乱数データの上側 2.5 % 点は 2.30584946 であった。上側確率 2.5 % で自由度 8 の理論値は、2.306 なので比較すると、ほぼ等しくなった。

4.3 実験 3 母分散が未知の場合の母平均の推定と検定

1. 学籍番号の下一桁は 5 なので、データシートから対応したデータ (7 個) を抽出した。
2. 7 個のデータから基本統計量を求めたものを表 1 にまとめた。

表 1: データから求めた基本統計量

基本統計量	値
データ数	7
平均	12.2
標準偏差	0.429839394
最小値	11.7
最大値	12.7

3. 母平均 μ に対する 95 % の信頼区間は、次のようになった。

$$\text{下限} : \bar{X} - t_{\alpha/2, v} \sqrt{\frac{s^2}{n}} = 11.81675058$$

$$\text{上限} : \bar{X} + t_{\alpha/2, v} \sqrt{\frac{s^2}{n}} = 12.61182085$$

4. 7個のデータから、帰無仮説 H_0 に対する検定統計量 T を計算すると 1.758631145 であった。

5. 両側検定における P 値 は、0.064569771 となった。

6. P 値 は、 $0.064569771 > 0.025$ なので、帰無仮説 H_0 は棄却できない。よって、作業時間の平均値が変化しているとは言えない。

7. 帰無仮説で使った $H_0 : \mu = \mu_0 = 12.5$ の値 12.5 が信頼区間

$$\text{下限} : \bar{X} - t_{\alpha/2, v} \sqrt{\frac{s^2}{n}}, \quad \text{上限} : \bar{X} + t_{\alpha/2, v} \sqrt{\frac{s^2}{n}}$$

に入っているから棄却されていないと考える。

5 まとめ

- 推定と検定の考え方を学んだ
 - － 点推定
 - － 区間推定
 - － 仮説検定
- t 分布について学んだ
 - － t 分布の定義と性質
- 母分散が未知の場合の母平均の推定と検定について学んだ
 - － 標本分散を用いる
 - － 検定統計量が t 分布に従うことを利用した推定と検定

6 感想

Excel で T.DIST.RT などの関数を用いて t 分布の上側確率などを求めたが、他授業 (確率統計) で使った関数があり、検定の仕方も習っていたので、スムーズにすることができた。

参考文献

- [1] 東京理科大学工学部情報工学科 情報工学実験 2 2020 年度東京理科大学工学部情報工学科
出版