データ解析実習
（ ２回目 ）

2011.10.17 月曜、４・５限

担当教員：高木英至、田端章明

■本日の授業

- 基礎概念導入
- 前半(４限)：基本的概念
  - 統計的推測：推定と検定
  - 仮説検定
- 後半(５限)
  - 推定・検定の実際
  - 相関係数など

2

■１．統計的推測

- （１）標本抽出（sampling）
  - 母集団（population）：情報を得るべき対象の全体
  - 標本（sample）：研究のために選択された母集団の一部
- 代表性のある標本（representative sample）＝母集団を正確に反映する標本
  - 無作為性（randomness）＝代表性のある標本を得る唯一の方法（無作為抽出、random sampling）
  - → 確率論の応用
  - → 母集団に対する統計的推測が可能になる。

（２）統計的推測(statistical inference)

- 1.分布型の推測　ー　適合度検定
- 2.母数の推測
  - a.推定
    - 点推定(point estimation)
    - 区間推定(interval estimation)
  - b.検定

（３）　信頼区間と標本誤差

- ［例］比率（e.g.,内閣支持率）の信頼区間

P_0 : 標本の比率

N : 標本サイズ（サンプル数）

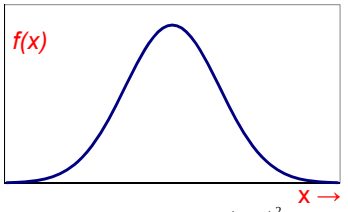
$Z_{\alpha/2}$: 正規偏差

ε : 比率の標本誤差

$$\varepsilon = Z_{\alpha/2}\sqrt{\frac{P_0(1-P_0)}{N}}$$

推定の考え方ーまず、

- 正規分布(normal distribution)　テキストp.103



- 連続分布
- ←→離散分布
- 対称分布
- 単峰形

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ: 期待値　　σ: 標準偏差

6

中心極限定理(テキスト　p.110)

- 期待値μ、分散σ^2の母集団からＮ個の観測値を標本として抽出したとき、Ｎが大きくなるほど、標本平均値（ $\overline{X}$ ）の分布は次の正規分布に近づく

$$N\left(\mu, \frac{\sigma^2}{N}\right)$$

7

- 確率95% で　z は [-1.96, 1.96]

$$p \sim N(\mu_p, \sigma^2/N) = N(\mu_p, {\sigma_p}^2)$$

- とすると、zに対応するpは　$\left[\mu_p - z\sqrt{\frac{\sigma^2}{N}}, \mu_p + z\sqrt{\frac{\sigma^2}{N}}\right]$
- σ_pには　$\sqrt{\frac{p_o(1-p_o)}{N}}$　をあてる。(p.132)
- μ_pにはp_0を推定値としてあてる。
- p_0=0.5, N=1600 とすれば、
- 確率95% でp は [47.55%, 52.45%]
  - ε=±2.45%

8

より一般的に

- 有意水準αのとき、（例：α=0.05）
- 平均値x'の信頼度(1-α)×100%（例：95%）の信頼区間は、正規分布を用いて、

$$\left[\bar{x} - Z_{\alpha/2}\sigma_{\bar{x}}, \bar{x} + Z_{\alpha/2}\sigma_{\bar{x}}\right]$$

- ただし上式は常には使えない
- $\sigma_{\bar{x}}$ が分らないことが多い
- → t 分布を使う

■２．仮説検定

- 1.仮説を立てる。
  - 帰無仮説（H_0,null hypothesis）：検定されるべき仮説、検定仮説
  - 対立仮説（H_1,alternative hypothesis）：帰無仮説の対立事象を主張する仮説
- 2.測定
  - 測定結果がその仮説の下では希にしか起こらないものであるとき
    - → 仮説を否定する（帰無仮説を棄却し対立仮説を採用する）
  - 測定結果がその仮説の下でもある程度の（小さくない）確率で起こり得るとき
    - → 仮説（帰無仮説）は否定できない、と判断する。
  - 帰無仮説を否定できるか否かは確率問題

		棄却	棄却しない
母集団で帰無仮説は	真	第一種の過誤（α過誤）	正しい判断
	偽	正しい判断	第二種の過誤（β過誤）

有意水準（significance level）：第一種の過誤が生じる確率（危険率）
＝ 帰無仮説を棄却する（対立仮説を採用する）ことが誤りである確率
慣例的に、0.05（5%）以下の有意水準を設定する。（5%, 2.5%, 1%, 0.5%, ･･･）

学年と性別は関連する。
男は1年が多く、女は2年以上が多い傾向がある。

性別 と 学年 のｸﾛｽ表

			学年 1年	学年 2年以上	合計
性別	男	度数	122	16	138
		期待度数	115.2	22.8	138.0
	女	度数	100	28	128
		期待度数	106.8	21.2	128.0
合計		度数	222	44	266
		期待度数	222.0	44.0	266.0

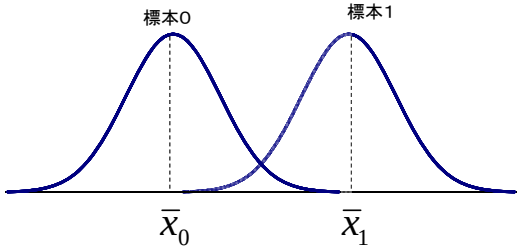
ｶｲ2乗検定

	値	自由度	漸近有意確率（両側）	正確有意確率（両側）	正確有意確率（片側）
Pearson のｶｲ2乗	5.084[b]	1	.024		
連続修正[a]	4.367	1	.037		
尤度比	5.122	1	.024		
Fisher の直接法				.031	.018
線型と線型による連関	5.065	1	.024		
有効なｹｰｽの数	266				

a. 2x2 表に対してのみ計算
b. 0 セル (.0%) は期待度数が 5 未満です。最小期待度数は 21.17 です。

5%水準で有意

集団間の平均値の差の検定



- H_0：母集団では平均値に差がない
- 考え方１：標本の平均値の分布（正規分布、ｔ分布）を利用する考え方
  - ｔ　検定

13

4限は
おしまい

14