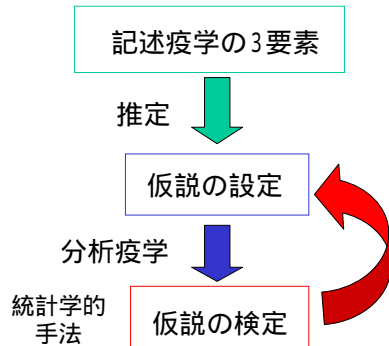


Lesson 11. 記述疫学

§ C. 仮説の設定

記述疫学の3要素

- 人(だれが): 疾患がどのような特性を持った人に起こったか
 - 生物学的要因(性、年齢、など)
 - 社会的要因(職業、生活水準、など)
- 時(いつ): 疾患がいつ起こったか
 - 時間経過、変動、要因との時間的關係など
- 場所(どこで): 疾患がどの場所に多いか
 - 地域差、環境要因(気候、風土、文化、など)



仮説

- 求めたい事実が「AならばBという結果になる」ならば、「Aを持っているC群(曝露群、介入群)と持っていないD群(非曝露群、対照群)ではBとなる結果に差がある」というのと同じ
- 従って、知りたいのは「C群のBとD群のBに差がある」、つまり「C = D」
- この証明は統計学的(確率論的)に困難

仮説(続き)

- 「C = D」を証明するかわりに「C ≠ D」を否定する
- 「C = D」:**帰無仮説 Null Hypothesis: H₀**
- 「C ≠ D」:**対立仮説 Alternative Hypothesis :H₁**
- 検定の目的は帰無仮説を否定、つまり**棄却 reject**すること

過誤

- 仮説検定は標本集団で行われる。(対象者が無数の母集団では不可能)
- 母集団での真実とは異なる可能性がある。
- 母集団の結果と異なる可能性、つまり真実が誤られる可能性を**過誤 Error**という。

第1種の過誤 Type Error

- 本当は(母集団で)正しい(真実である)帰無仮説を棄却してしまう(従って、誤って対立仮説を採用する) = **第1種の過誤**
- この過誤が生じる確率 = **危険率**
- は確率 p で表される。
- 20回抽出した標本のうち1回に過誤が生じる確率は $p=0.05$ 、これ以下なら $p<0.05$
- は**有意水準 Significance level**

第2種の過誤 Type Error

- 本当は(母集団で)誤っている(真実でない)帰無仮説を棄却できずに採用してしまう(従って、誤って対立仮説を否定する) = **第2種の過誤**
- この過誤が生じる確率 = **危険率**
- 判断が正しくてこの過誤が生じない確率 = $1 -$ **検出力 Power**
- 検出力は分析の能力(通常は80%以上に設定)

第1種と第2種の過誤の関係

		母集団の真の結果	
		差がある (H_0 は誤り)	差がない (H_0 は正しい)
仮説 検定 の 結果	差がある (H_0 を棄却)	正しい ($1 -$ 、検出力)	第1種の過誤 (、有意水準)
	差があるとはいえない (H_0 を採用)	第2種の過誤 ()	正しい

H_0 は帰無仮説

区間推定

- 仮説の検定の最終目標は、標本集団の結果より母集団の結果を推測すること
- 点推定 Point Estimation
 - ひとつの値を推定
 - はずれる危険が大きい
- 区間推定 Interval Estimation
 - ある幅(区間)を持って推定
 - 確率的にその幅に母集団の値がある可能性

区間推定(続き)

- 標本集団の代表値(平均値など)は正規分布(分布と標準誤差を参照)
- 95%の確率で母集団の値がその区間にある = 5%の確率で誤り(p=0.05)
- つまり、有意水準は母集団の区間推定が誤りである確率
- 区間推定で求める値の幅
= 信頼区間 Confidence Interval: CI

信頼区間

- 平均値の信頼区間

$$m \pm 1.96 \times SE$$

mは標本平均、SEは標準誤差

- 比率の信頼区間

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{N}}$$

pは標本の比率、Nは標本数
