

Public Health

18. 보건통계

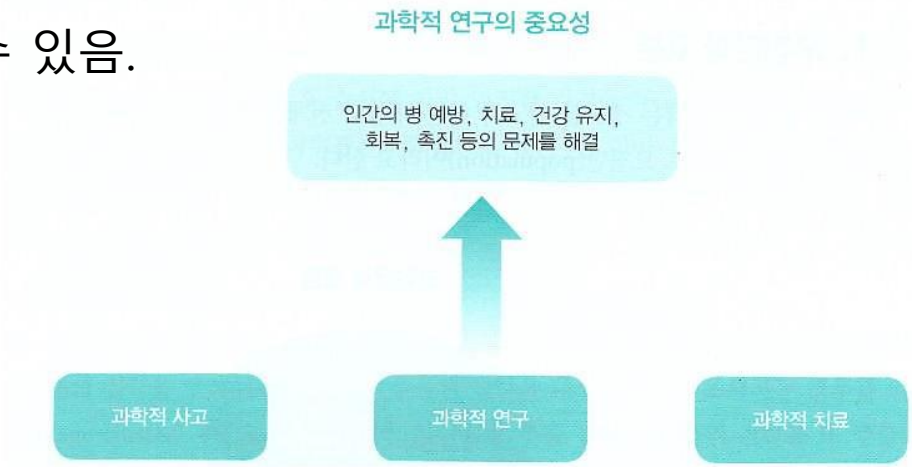
1. 보건통계학

■ 보건통계의 개념

• 보건통계 정의

- 통계학이란 '관측자료의 수집, 정리, 요약, 제시, 분석하는 방법과 이론을 다루며, 자료분석을 통하여 결론을 유도해 내거나 판단의 근거를 제공하는 과학' 이라고 표현할 수 있음

=> 보건통계란 출생, 사망, 질병, 인구변동 등 인구의 특성을 연구하는 일과 보건에 관한 여러가지 현상 및 대상물을 다량 관측 또는 계측하여 얻은 숫자를 집계하고, 정리, 분석하여 건강과 관련된 현상의 변동을 확률론적으로 추론하는 학문이라고 할 수 있음.



■ 보건통계의 목적

- 인구, 출생, 사망, 사산, 혼인, 질병 등의 인구변동에 대한 인구학적 특성을 연구하고 생명, 건강 등 보건에 관한 여러가지 현상과 대상물들을 측정하고 이를 분석하여 그 특성을 알아보는데 있음

■ 보건통계의 이용

- 보건통계는 어느 지역사회 또는 한 국가의 보건수준을 나타내거나 질병 관리와 보건향상을 위한 연구와 조사의 분석에 기초자료로 이용되고, 보건 사업의 계획과 사업진행 및 결과평가에 활용.

① 보건입법의 당위성 제공과 보건사업에 대한 공공지원 촉구 가능
② 보건사업의 우선순위를 결정하여 보건사업 수행상 지휘, 관제에 도움을 주고 보건사업의 기술발전에 도움을 주며 또한 절차, 규정, 분류, 기술의 발전에 도움을 줌
③ 보건사업의 행정활동에 지침이 될 수 있음
④ 보건사업의 성패를 결정하는 자료가 되고, 보건사업의 기초자료가 됨
⑤ 지역사회나 국가의 보건수준 및 보건상태의 평가에 이용 됨
⑥ 보건사업의 필요성을 결정해 주고, 사업의 평가에 이용

1) 통계자료의 수집과 표본추출

■ 자료수집

(1) 모집단(population)과 표본(sample)

- 자료수집 첫단계는 정보를 얻고자 하는 대상의 전체 집합을 규정하는 것.
  - 정보를 얻고자 하는 그 대상의 전체 집합을 모집단(population)이라 함.

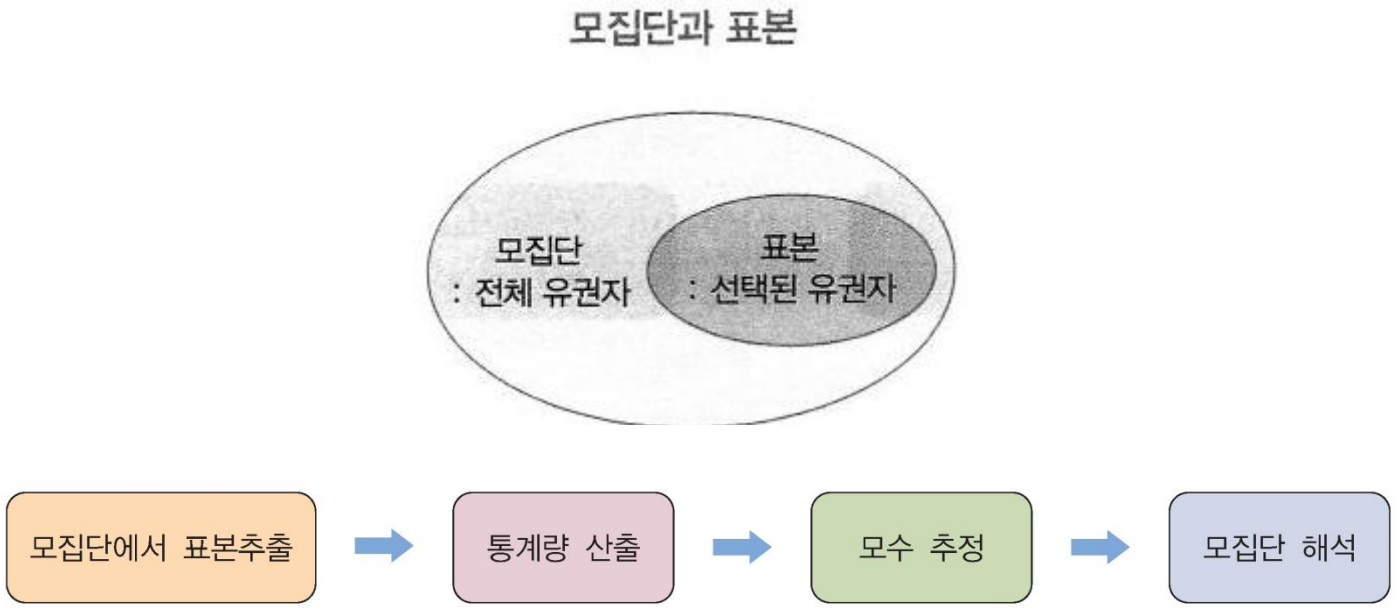


그림 18-1 : 표본으로부터 모집단을 해석하는 과정

■ 표본조사를 하는 이유 및 상황

- 전수조사를 실시하는 것이 현실적으로 어려울 때(시간, 비용, 전문인력 등)
- 무한모집단을 가정할 때(미래 상황-작물의 매년 수확량 예측, 내원 환자 수의 추정).
- 시급한 조치가 요구될 때(질병의 집단유행, 시장점유율의 급변).
- 대상이 파괴되어야만 측정이 가능할 경우(전기 퓨즈 내구시간 예측, 시약 검정).
- 전수조사 시 오히려 비표본추출 오차가 커져 정확성이 감소될 우려가 있을 때 (입력 오류)
- 표본조사만으로도 적당한 오차한계 내에서 모수를 추정할 수 있기 때문

(2) 표본추출의 방법.

- 모든 개체가 표본으로 선택될 기회가 동일하도록 표본선택 과정에 반드시 '무작위성(randomness)' 이 있어야 함.

• 확률표본추출법(probability sampling) :
모집단의 모든 대상들이 표본으로 선택될 확률이 동일한 방법으로 무작위로 표본을 추출하는 방식으로 객관성 유지.

① 단순 무작위 표본추출법 : 경품권 추첨방식 등 해당, 때로는 난수표 이용.

② 계통적 무작위 표본추출법 : 모집단에 일련번호를 부여한 후 순서에 따라 일정한 간격을 두고 표본을 하나씩 추출하는 경우(1번, 11번, 21번, 31번... 과 같이 10번째마다 표본을 뽑게 됨).

③ 층화 무작위 표본추출법 : 먼저 모집단을 성별, 연령 등을 기준으로 하부집단으로 구분해 두고 그 하부집단에서 고르게 표본을 추출하는 경우.

④ 집락추출법 : 모집단을 여러 개 집락으로 구분해 놓은 다음 그 집락들 중에서 몇 개의 집락만을 대상으로 전수조사하는 경우.

(3) 표본조사의 오차

- 오차(error) : 표본통계량과 모수와의 차이.
- 확률표본을 사용한다 할지라도 표본에서 얻은 통계량으로부터 모수를 추정하는 과정에는 항상 오차가 개입 될 수 있음 -> 따라서 표본으로부터 얻은 정보가 모집단을 항상 정확하게 반영하지 못함.

① 표본추출 오차

- 무작위 오차(random error) : 우연에 의해 발생한 오차

ex) 적색과 백색 구슬이 있는 큰 주머니에서 25개의 구슬을 꺼내는 경우 적색 구슬일 확률이 1/25 이어야 하지만 그렇지 않고 몇 개씩 차이가 나는 것과 같은 경우

- 계통적 오차(systematic error) : 기본 설정이 잘 못되어 '한쪽 방향으로만 벗어나는 오차(bias)'

ex) 1936년 미국 대통령 선거 전 여론조사에서의 표본 추출 오류.

② 비표본추출 오차

- 응답오차(response error), 자료처리오차(data processing error), 자료 획득방법 차이에 의한 오차.

2) 통계자료의 정의

- 통계자료란 변수의 값들을 측정한 순서대로 나열한 값들로 방법론상 측정 기준을 미리 정의하거나 설정해 두고 측정된 값들을 의미

■ 변수의 측정척도

- 분석과정에 이용되는 변수들은 척도의 종류에 따라 그 변수가 갖는 정보의 수준에 차이가 나고 또한 그척도에 따라 통계분석 기법이 결정되기 때문에 사전에 어떤 척도로 측정할 것인가에 충분히 고려하는 것이 매우 중요
- 변수들을 측정하여 그 결과를 표시하는 경우, 다음 네 가지 척도(scale) 중의 하나를 사용

- 범주형(질적) 척도 자료
  ① 명목척도(nominal scale) : 가장 간단한 형태의 분류로 어떤 범주에 속해 있는지만 표현[예: 성별(남, 여) , 혈액형(A, B, 0, AB), 인종(백인, 흑인, 기타]
  ② 순위척도(ordinal scale) : 사물의 분류와 순위를 정하기 위한 척도로서 각 범주 간의 대소관계를 구분하지만 그 간격의 차가 어느 정도인지 뚜렷하지 않은 것이 특징이다[예: 경제상태(상,중,하), 군 졸병들의 계급(일등병, 상병, 병장)].
- 연속형(양적) 척도 자료
  ③ 간격척도(interval scale) : 어떤 표본단위와 다른 표본단위와의 '얼마 만한 차이가 있는지' 그 간격이 표현된 것으로 '영 '의 개념(몇 배를 나타낼 수 있는)이 포함되지 않는 것이 특징[예:온천장의 온천수 온도에서 23℃ 물과 46℃ 물의 온도는 2배지만 실제 뜨거운 정도는 두 배 차이가 아니다].
  ④ 비율척도(ratio scale) : 대소의 관계뿐만 아니라 그 대소의 차이, 그리고 '몇 배의 차이'까지 표현 가능[예: 100 cm 아들 키와 180 cm 아빠의 키 (아빠는 아들보다 1.8배 크다고 표현]

4) 통계분석의 활용

① 대표값(중심경향성)

- 측정값을 대표할 수 있는 객관적인 하나의 값.
- 평균값(mean), 중위수(median), 최빈수(mode) 의 대표값을 사용

- 평균값(mean, M) : 보통은 산술평균이 많이 이용되는데, 자료 실측값에 가중치를 곱한 후 이들을 모두 더한 값을 가중치의 합으로 나눈 것
- 중위수(median) : 중위수는 중간이란 뜻으로 측정값을 크기순으로 나열하고, 중간에 위치하는 값을 말함(예: 2, 5, 6, 8, 9에서 6이 중위수가 됨)
- 최빈값(mode) :- 최빈수는 측정값 중에서 가장 많은 빈도(자주 나오는 수)의 값을 의미(주로 질적 변수에 많이 쓰임)
  - 1년동안 병원 방문한 환자의 질병, 나이, 성별등에 사용

가) 산술평균(arithmetic mean).

- 한 집단에서 모든 측정값의 합을 구하고 수집된 자료의 전체 개수로 나눈 측정값.

$$\bar{x} = \frac{X_1 + X_2 + X_3 + \cdots\cdots + X_n}{n} = \frac{\Sigma X_i}{n}$$

나) 기하평균(geometric mean, g)

- 주어진 n개의 양수의 곱의 n 제곱근의 값.
- 대칭적 분포가 아니고, 중앙치가 한쪽으로 몰릴 경우 기하평균 이용시 정규분포가 되는데, 이를 대수정규분포라 함.

$$g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots\cdots \cdot x_n}$$

② 산포도(measure of dispersion) - "퍼짐(흩어짐)의 지표"

- 대표값만으로는 자료들의 집합적 양상을 알 수 없기 때문에 그 집단의 특성을 알기 위해 측정값이 어느 정도 흩어져 있는가를 보게 되는데 이를 산포정도(대표값으로부터 얼마나 떨어져 있는지의 정도)라 함.
- 범위(range), 편차와 분산(deviation and variance), 표준편차(standard deviation, SD)

• 범위(range) : 가장 초보적 산포도로서 자료의 최대값에서 최소값을 뺀 것
  - 범위를 산포의 측도로 사용할 때의 장점은 계산이 매우 쉽다는 것.

• 분산과 편차(variance & deviation)
  - 편차와 분산은 측정치들이 평균으로부터 얼마나 떨어져 있는가를 보여주는 것으로 퍼짐의 정도를 표현하는데 있어 수학적 조작이 용이하고, 자료의 분포가 대칭일 때 훌륭한 지표가 될 수 있음

- 표준편차(standard deviation, SD)

  - 자료가 평균을 중심으로 얼마나 퍼져 있는지를 나타내는 대표적인 수치
  - 평균과의 차이를 제곱한 값의 평균을 계산 후, 다시 루트(제곱근)를 취한 값.
  - 표준편차가 클수록 자료 값들이 평균값에서 널리 퍼져 있음을 의미.

  => 표준편차는 산포성을 가장 많이 나타내고, 가장 신뢰할 수 있는 통계값.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

s: 표본의 표준편차, n: 표본 수, x_i: 개체의 값, $\bar{x}$: 산술평균

③ 변이계수(coefficient of variation, CV)

- 표준편차를 산술평균으로 나눈 값으로 상대적인 분산 정도를 알아보기 위해 사용 => 검사실의 검사기기 정밀도 평가 및 정도관리지표로 이용.

$$\mathrm{CV}(\%) = \frac{\mathrm{SD}}{\bar{x}} \times 100$$

※ 표준편차(SD)
예)

	수민	승우	시우	미선	숙경	윤겸	진명	미진	영신	마리
키 (cm)	177	167	160	162	174	180	176	158	172	184

X의 평균 $\mu = E(X)$는

$\mu = \frac{1}{10}(177+167+160+162+174+180+176+158+172+184) = 171$

$|X - \mu|$ 와 $(X - \mu)^2$를 구하면

	수민	승우	시우	미선	숙경	윤겸	진명	미진	영신	마리	합계
$\|X - \mu\|$	6	4	11	9	3	9	5	13	1	13	74
$(X - \mu)^2$	36	16	121	81	9	81	25	169	1	169	708

따라서, **분산**은

$V(X) = E(((X - \mu)^2) = \frac{708}{10} =$ **70.8**

표준편차는

$\Sigma x = \sqrt{V(X)} = \sqrt{70.8} \approx$ **8.4**

6) 가설검정

- 모집단의 산술평균값이 어떤 값과 같은가 또는 같지 않은가를 판정하는 형식으로 이러한 검정에 이용되는 가설은 귀무가설과 대립가설이 있음.

(1) 귀무가설과 대립가설

① 귀무가설(H_0) : 연구과정에서 실제 검정의 대상이 되는 것.

② 대립가설(H_a) : 귀무가설이 받아들여질 수 없을 때 대신 받아들이는 것.

(2) 유의수준(α)

- 어느 수준에서 귀무가설을 기각하고, 대립가설을 채택할 것인가의 수준

예) 0.01, 0.05, 0.001 => 1%, 5%, 0.1%를 의미하는 것으로 수준은 본인이 임의로 정하는 것임

(3) 가설검정의 순서

① 귀무가설과 대립가설의 설정
② 유의수준 결정(α = 0.05)
③ 유의수준과 유의확률(p-value) 비교
④ 결과해석

3. 보건지표

1) 건강 관련 지표.

(1) 발생률(incidence rate)

- 일정한 기간에 한 인구집단 내에서 어떤 질병 또는 사건이 새롭게 일어난 횟수가 얼마나 되는가를 나타낸 것으로서 이 질병에 걸릴 확률 또는 위험도를 나타냄

$$\text{발생률} = \frac{\text{일정기간 동안 새롭게 발생된 환자 건수(occurrence of new cases)}}{\text{일정기간 동안 해당 질병에 노출된 인구(observed person-time)}} \text{ X 1,000}$$

=> 대상 인구는 질병이 없는 사람이거나 해당 질병에 걸릴 위험성이 있는 사람만이 포함되어야 함(지역 전체인구를 대상).

※ 발병률(attack rate) : 어떤 집단이 한정된 기간에 한해서만 어떤 질병에 걸릴 위험에 놓여 있을 때 기간 중 주어진 집단 내에 새로 발병한 총수의 비율(분모가 지역전체 인구가 아니라 감염에 폭로될 수 있는 제한된 인구집단만 해당)

$$\text{발병률} = \frac{\text{유행기간 중 새로운 발병자 수}}{\text{위험에 폭로된 인구 수}} \times 1{,}000$$

(2) 유병률(prevalence rate)

- 어떤 일정 시점이나 기간에 있어서 전체 인구집단 중에서 특정 질환을 보유하는 사람의 비율(proportion)을 나타낸 것.
- 조사방법에 따라 시점유병률(point prevalence rate)과 기간유병률(period prevalence rate)로 나뉨.

$$\text{시점(기간)유병률} = \frac{\text{일정시점(기간)에서의 특정질환에 이환된 환자 수}}{\text{일정시점(기간)에서의 특정집단의 인구 수}} \times 1{,}000$$

※ • 만성 감염병 : 유병률이 높음 • 급성감염병 : 발생률이 높음

• 유병률 = 발생률 : 감염병 유행기간이 짧을 때

(3) 치명률(case fatality rate)

- 질병의 심각한 정도를 나타내는 수치로서 특정 질병에 이환된 사람 중 그 질병으로 인해 사망한 사람의 비율로 나타낸 것

$$\text{치명률} = \frac{\text{일정기간 동안 발생된 해당 질병에 의한 사망자 수}}{\text{특정질병이 발생된 환자 수}} \times 1{,}000$$

2) 사망에 관한 지표.

(1) 조사망, 보통사망률(crude death rate, CDR)

- 한 인구집단의 사망 수준을 나타내는 기본적인 지표로서 인구 1,000명당 1년간의 발생 사망자 수로 표시하는 비율
- 그 나라의 건강수준, 보건의료수준 이외에 연령의 구성정도에 따라 큰 영향을 받음

$$\text{보통사망률} = \frac{\text{해당 연도의 해당 집단의 사망자 수}}{\text{특정 연도의 특정집단 중앙인구 수}} \times 1{,}000$$

(2) 영아사망률(infant mortality rate)

- 출생아 1,000명당 1년간 1세 미만(O 세)의 사망아 수
- 영아사망률은 국가사회나 지역사회의 보건수준을 나타내는 대표적 지표.

$$\text{영아사망률} = \frac{\text{해당 연도의 1세 미만의 사망아 수}}{\text{특정 연도의 정상 출생아 수}} \times 1{,}000$$

(3) 신생아사망률(neonatal mortality rate)

- 신생아사망은 신생아 고유질환이나 분만 시 사고, 조산아 등이 원인
- 영아사망의 원인에 따라 보건수준의 척도로서 중요한 의미를 지님.

$$\text{신생아사망률} = \frac{\text{해당 연도의 생후 28일 미만의 사망아 수}}{\text{특정 연도의 정상 출생아 수}} \times 1{,}000$$

(4) 모성사망률(maternal mortality rate, MMR)

- 임신 또는 분만 후 42일 이내에 발생하는 산모 사망.

$$\text{모성사망률} = \frac{\text{해당 연도의 임신, 분만, 산욕 합병증에 의한 모성사망 수}}{\text{특정 연도의 정상 출생아 수}} \times 100{,}000$$

※ 출생아수가 분모에 해당하는 보건지표는? 영아, 신생아, 모성 사망률

(5) 주산기사망률(perinatal mortality rate, PMR)

- 출생아 수에 대한 임신 28주 이후에서 출생 후 1주 이내에 사망아 수
- 이 기간의 사망은 주로 임신중독, 출생 시 손상, 난산, 조산아 무산소증 및 저산소증 등 그 사망원인이 서로 공통성이 인정되므로 주산기사망이라 함

$$\text{주산기사망률} = \frac{\text{해당 연도의 28주 이후부터 생후 1주 이내의 사망아 수}}{\text{특정 연도의 정상 출생아 수} + \text{사산 수}} \times 1{,}000$$

(6) 비례사망률(proportional mortality rate)

$$\text{비례사망률} = \frac{\text{해당 연도의 특정 원인에 의한 사망 수}}{\text{특정 연도의 사망 수}} \times 100$$

(7) 비례사망지수(proportional mortality index)

- 총 사망자 중 50세 이상 사망자 수의 비
- 비례사망지수가 크면, 건강수준이 높고 장수 인구가 많다는 의미.

$$\text{비례사망지수} = \frac{\text{해당 연도의 50세 이상 사망 수}}{\text{특정 연도의 사망 수}} \times 100$$

※ 유아사망률

- 1~4세의 사망은 보건수준이 낮은 나라에서는 주로 감염증, 영양실조 등이 원인이고, 선진국에서는 사고가 주원인

$$\text{유아사망률} = \frac{\text{연간 1~4세 사망자 수}}{\text{1~4세 인구}} \times 1{,}000$$

※ 출산통계

- 모자보건, 가족계획과 관련하여 건강평가의 수단으로 사용되며, 경제성장률을 예측함에 있어서 국가경쟁력의 중요한 지표로 사용되는 보건지표

• 조 출생률(보통출생률, Crude Live-Birth Rate, CBR)

- 해당 연도의 총 출생아수(사산 미포함)를 특정 기간의 중앙인구로 나눈 값.
- 보통출생률은 한 나라의 건강수준, 보건의료 수준, 연령별 결혼률, 인구의 연령 구성에 따라 크게 영향을 받음
- 가족계획사업을 실시하고, 그 정도를 효과적으로 판정할 때 가장 많이 사용되는 보건지표

$$\text{보통출생률(CBR)} = \frac{\text{해당 연도의 정상출생아 수}}{\text{특정 연도의 중앙인구}} \times 1{,}000$$

3) 인구 재생산 통계

(1) 총재생산율 (Gross Reproduction Rate)

- 한 여성이 가임 기간(15~49세)동안에 출산한 여아(딸) 수의 평균값
- 즉, 한 여성이 일생 동안 몇 명의 여아를 낳는지를 나타내는 것
- 가임 여성의 사망률 무시.

(2) 순재생산율 (Net Reproduction Rate)

- 가임 여성의 사망률을 고려한 재생산율.

=> 생산율 1.0 : 인구 증감 없음, 1.0 이하 : 인구 감소, 1.0 이상 : 인구 증가

※ 병원통계(병원 행정통계, 병원 이용통계, 병원 운영통계 등을 총칭)

1. 입원율(Admission rate)

- 대상인구 중 1,000명당 입원 환자 수의 비율

$$\text{입원율} = \frac{\text{대상인구 중 입원환자 수}}{\text{대상인구}} \times 1{,}000$$

2. 병상점유율(Bed occupancy ratio)

- 단위 인구가 하루에 점유하고 있는 병상의 비로서, 보통 1,000명당 1일간의 재원일수

$$\text{병상점유율} = \frac{\text{1일평균 병상점유 수}}{\text{인구}} \times 1{,}000$$

3. 평균재원일수(Average length of stay)

- 입원환자당 평균재원기간을 나타내는 것으로 기간 중 퇴원환자들의 총재원일수를 그 퇴원환자수로 나누어 계산

$$\text{평균재원일수} = \frac{\text{기간 중 퇴원환자의 총재원일수}}{\text{일정 기간 중 총퇴원환자}}$$

4. 병상이용률(Bed occupancy rate)

- 가용병상 수에 대한 병상사용 수의 비율로서, 병상이용률이라고 함

$$\text{병상이용률} = \frac{\text{1일 평균 재원환자 수}}{\text{병상 수}} \times \mathbf{100}$$

5. 병상회전율(Turn over ratio)

- 일정기간내에 한 개의 병상을 사용한 평균 환자수를 나타내고, 병상이용의 정도를 측정하는 자료로 사용
- 일반적으로 평균재원일수가 긴 병원의 병상회전율은 낮고, 평균재원일수가 짧은 병원의 병상회전율은 높음

$$\text{병상회전율} = \frac{\text{해당기간 중 총퇴원환자 수(사망포함)}}{\text{기간 중 사용가능한 평균 병상 수}} \times \mathbf{100}$$