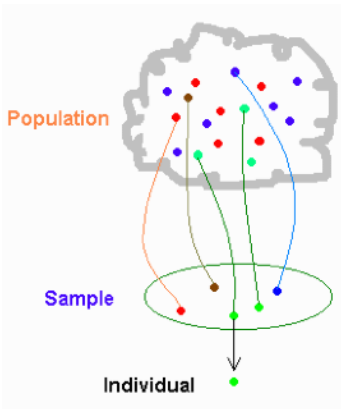


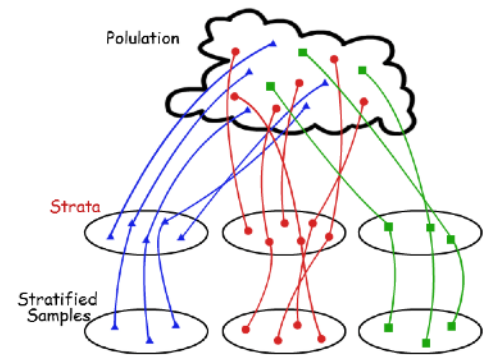
모집단 확률분포함수(Probability Density Function)와 모수 parameter



모집단의 개체가 표본으로 추출될 가능성이 모두 동일(1/N)하도록 하여 추출된 표본

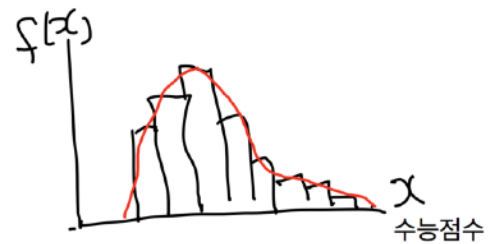
<- Simple Random Sampling 단순임의 추출

Stratified Sampling 층화 추출 ->



* H대 신입생 3,000명에 대한 수능점수를 알고자 한다(모집단 크기 $N=3,000$). 이를 위하여 신입생 200명을 (단과대학, 성별) 층화하여 층화표본 추출 하였다고 하자(표본크기).

* 모집단 관심 특성 = 확률변수 를 X (수능점수)라 하자. 모집단 1번째 개체(학생) 수능성적을 X_1 , 2번째 학생 수능성적 X_2 , ... 그리고 마지막 $N(3,000)$ 번째 학생의 수능성적을 X_{3000} 라 하자.



* 실제 우리는 3,000명 전체의 개별 수능성적에 관심이 있는가? 그렇다면 우리는 3,000명 수능점수를 모두 조사하여 확률분포함수($f(x)$)를 만들어야 한다. -> 막대빈도표(히스토그램)을 만들어 구간의 중앙을 연결

* 확률분포함수를 알면 모집단 정보(한남대 학생의 수능성적)에 대한 모든 것을 알 수 있다. 그러나 모집단 전수를 조사하지 않는다면 $f(x)$ 를 얻는 것은 불가능하다.

* 그리고 모집단 개체의 개별 값들에 대한 정보($f(x)$)보다는 요약된 하나의 값, 이를 모수 parameter (θ)라 한다. (예) 수능점수 평균, 수능점수 최고, 최저 등

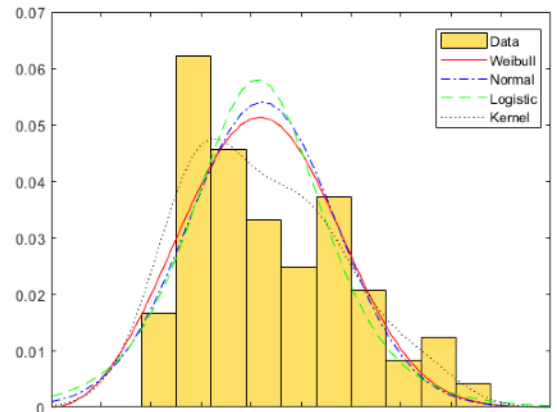
* 모집단 확률변수 $X \sim f(x; \theta)$ 로 표시한다. 모집단 전수를 조사하기 전에는 확률분포함수의 형태 $f(x)$, 모수 θ 둘 다 알 수 없고, 표본조사를 통하여 모집단의 $f(x)$, θ 를 추정할 수 있다.

* $f(x)$ 에 대한 추정은 분포의 적합성 검정으로, 귀무가설 하에 설정된 $f(x)$ 가 맞는지, 아닌지를 검정할 수 있을 뿐 실제 분포함수는 알 수 없다.

* 모수 θ 에 대한 추정과 검정은 모집단의 확률분포함수와 표본의 확률분포함수가 동일하다는 가정 하에 이루어진다.

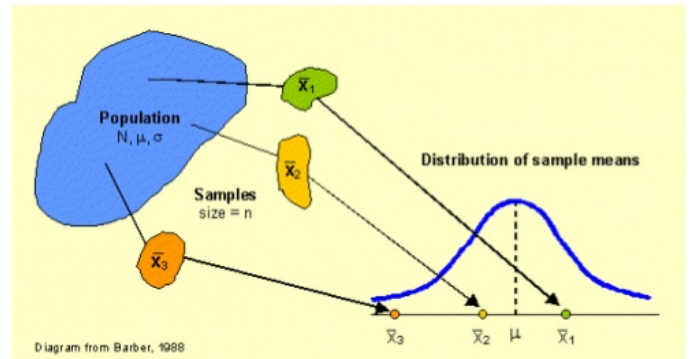
확률표본 random sample, 표본확률분포함수 sample PDF

- * 모집단 모수 θ (모집단 수능점수 평균, 최고점수, 최저점수)를 추정하거나 관련 가설을 검정 하려면 모집단의 확률분포함수와 표본의 확률분포함수는 동일하다는 가정이 필요하다.
- * 표본을 추출할 때 모집단의 개체가 표본으로 선택될 가능성이 동일하도록 하는 경우 이를 확률 표본이라 하는데, 확률표본(데이터)의 확률분포함수는 모집단의 확률분포함수와 동일하다.
- * 표본크기 $n=200$ 확률표본 데이터(수능점수) $(x_1, x_2, \dots, x_{200})$ 의 확률분포함수는 모집단(한 남대신입생 3,000명)의 확률분포함수 $f(x; \theta)$ 와 동일하다. $x_i \sim f(x; \theta)$
- * 오른쪽 히스토그램은 예제 표본 확률분포함수의 예이다. 확률표본이면 히스토그램의 분포와 모집단의 확률분포는 동일하다.



통계량 statistic, 샘플링확률분포함수 sampling PDF

- * 확률표본으로부터 계산되는 요약 값을 모두 통계량이라 한다. 통계량은 관측 데이터의 함수이다. $T = f(x_1, x_2, \dots, x_n)$ (예) 표본 평균($\theta = \mu$), 표본 최대값, 표본 최소값
- * 통계량을 추정에 사용되면 추정량 $\hat{\theta} = T(x_1, x_2, \dots, x_n) = \bar{x}$, 가설 검정에 사용되면 검정통계량 $TS = \frac{\bar{x}}{s/\sqrt{n}}$ 이라 한다.



- * 통계량을 활용하여 가설검정하거나 신뢰구간 추정을 하려면 통계량의 확률분포함수를 알아야 한다. 통계량의 확률분포함수를 샘플링 확률분포함수라 한다.
- * 모집단의 확률분포=표본확률분포이나 샘플링 확률분포함수는 상이하다. 모집단의 확률분포함수는 몰라도 샘플링 확률분포함수는 알 수 있어, 이를 이용하여 추정과 검정을 한다. (예) 모집단의 확률분포함수와 무관하게 표본평균의 확률분포함수는 중심극한정리에 의해 정규분포에 근사한다.