

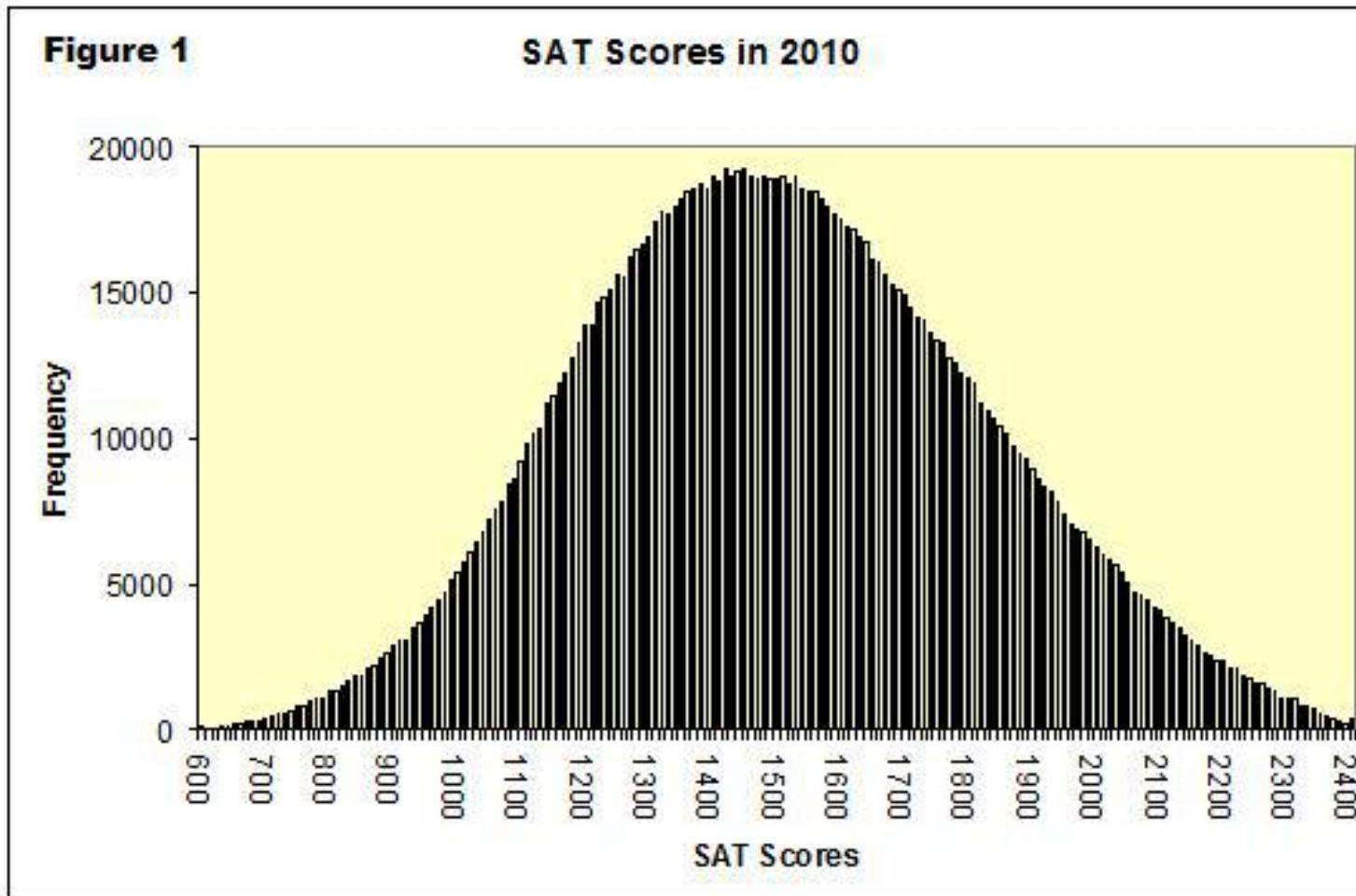
과제 5

1. 과제 2의 uniform random number generator 사용
2. 앞에서 배운 Box-Muller 변환을 이용하여 Gaussian 분포를 가지는 “효율적인” 난수 발생기(함수)를 작성
3. $N(10, 2^2)$ 분포를 가지는 10000 개의 난수 발생
4. 결과를 히스토그램으로 표현
5. `np.random.normal`을 사용한 결과와 비교. (고찰: 두 분포가 동일함을 어떻게 보일 수 있을까?)
6. Box-Muller 변환에는 어째서 로그와 삼각함수가 필요할까?

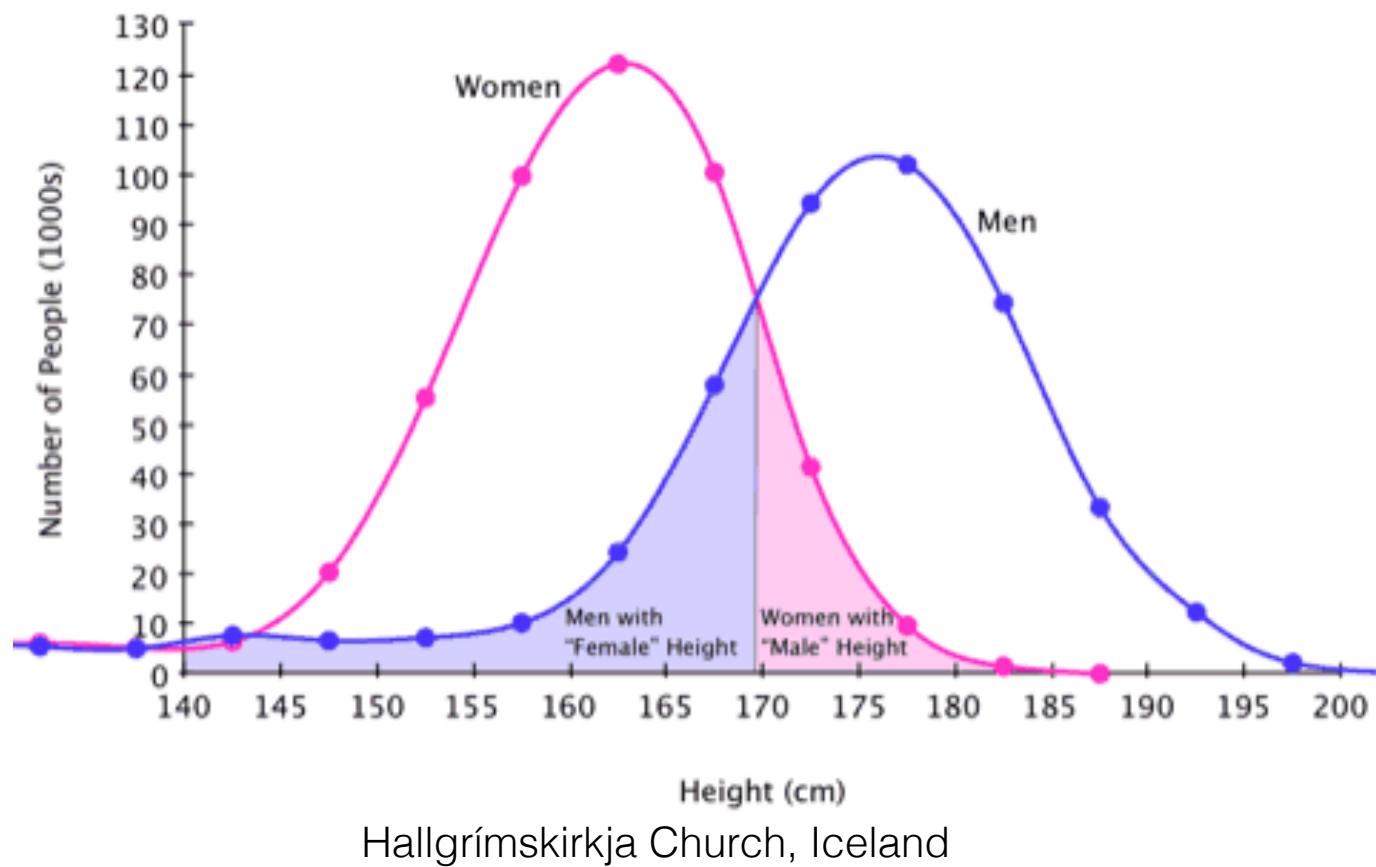
Central Limit Theorem

중심 극한 정리

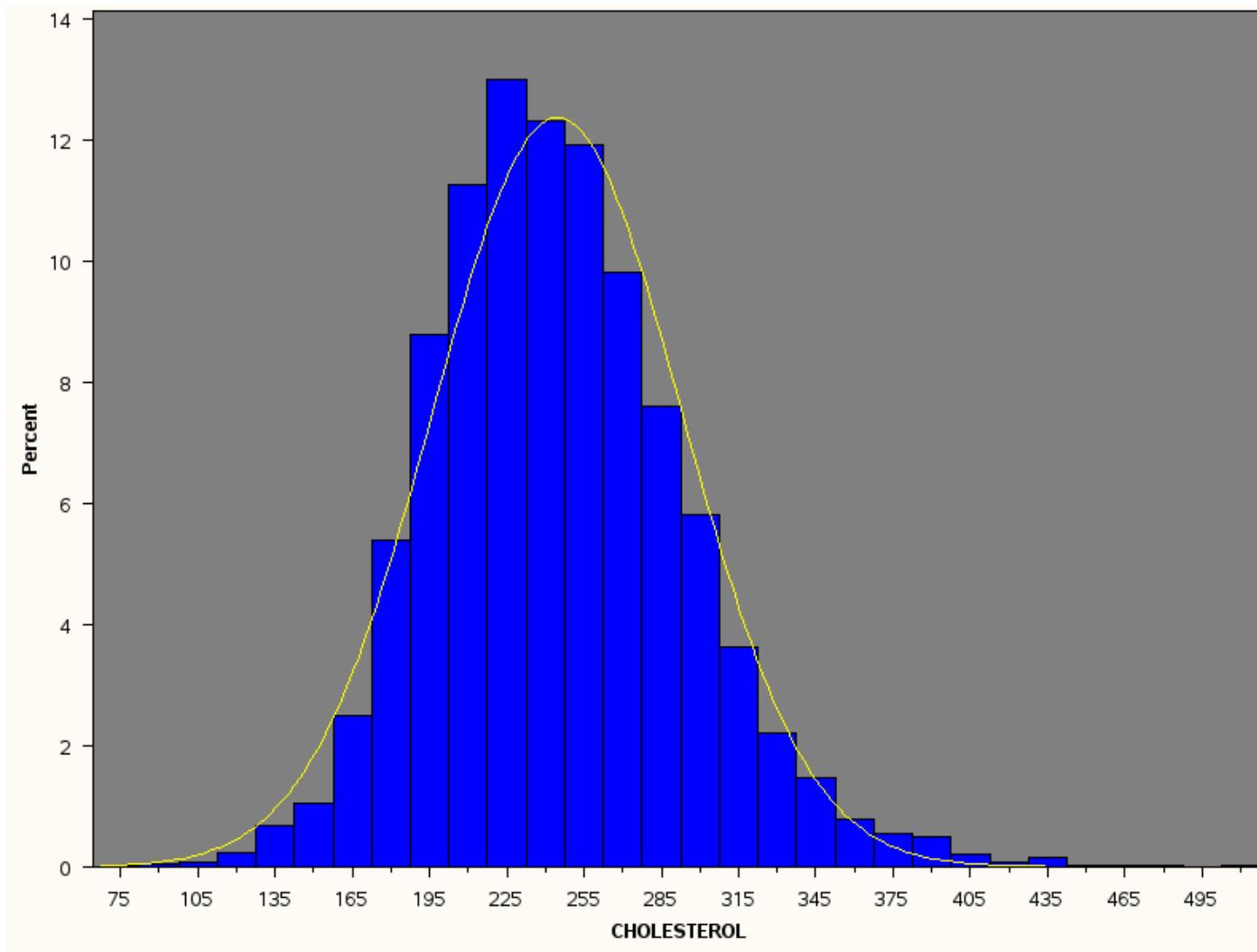
왜 자연계의 많은 현상은 정규 분포를 따르는가?



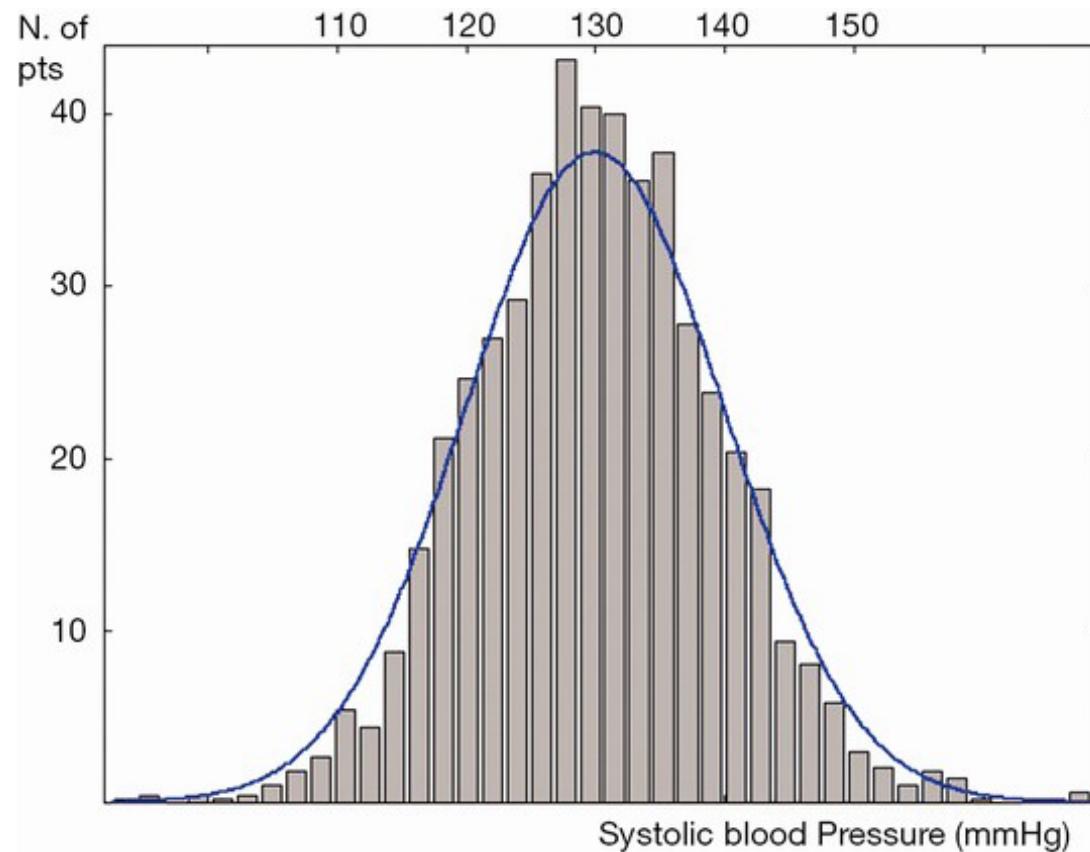
왜 자연계의 많은 현상은 정규 분포를 따르는가?



왜 자연계의 많은 현상은 정규 분포를 따르는가?



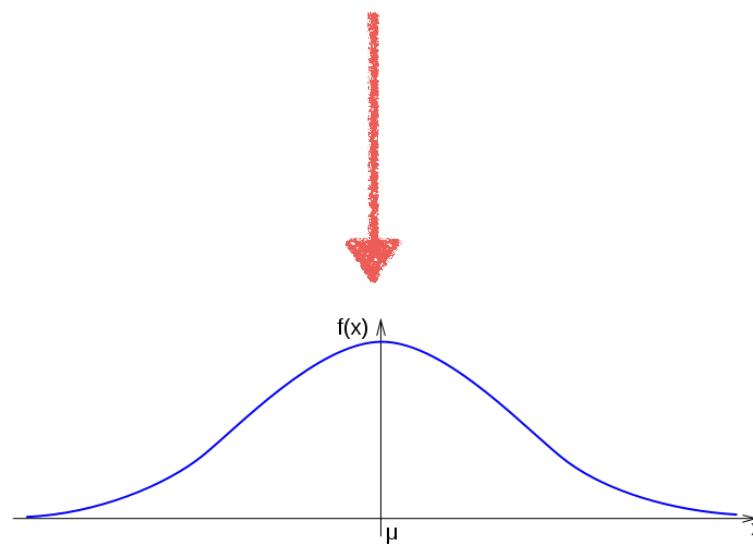
왜 자연계의 많은 현상은 정규 분포를 따르는가?



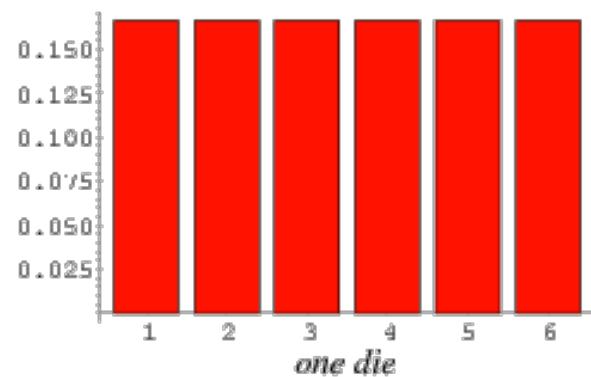
Central Limit Theorem

The central limit theorem states that the distribution of the sum of a large number of independent, identically distributed variables will approach the normal distribution if the number goes to infinity.

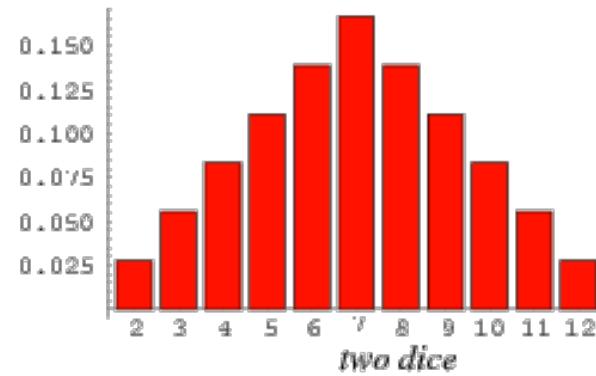
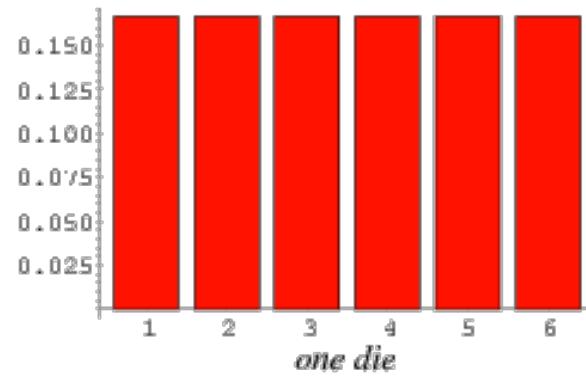
$$x_1 + x_2 + x_3 + x_4 + x_5 \dots$$



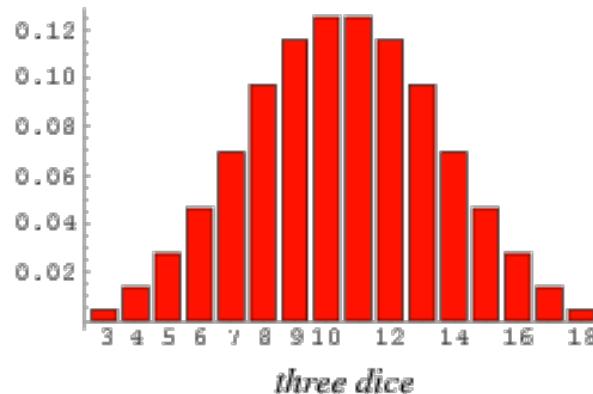
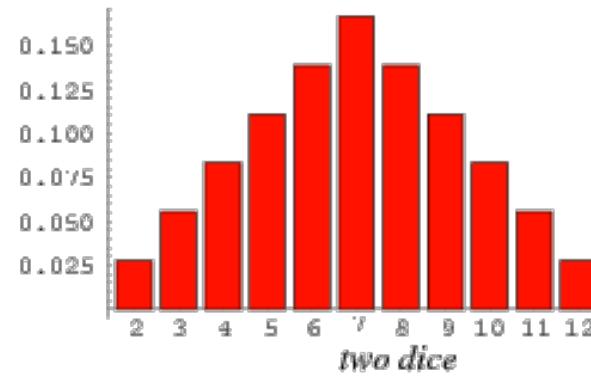
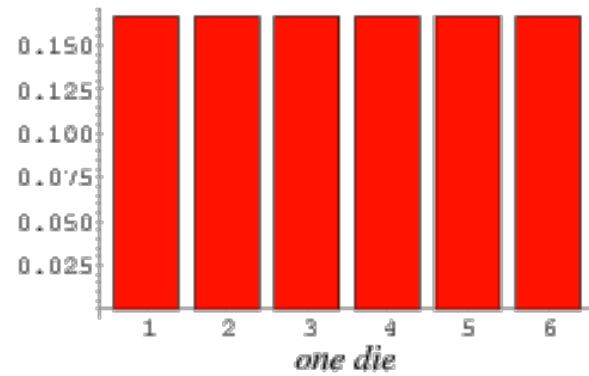
Sum of Many Dice



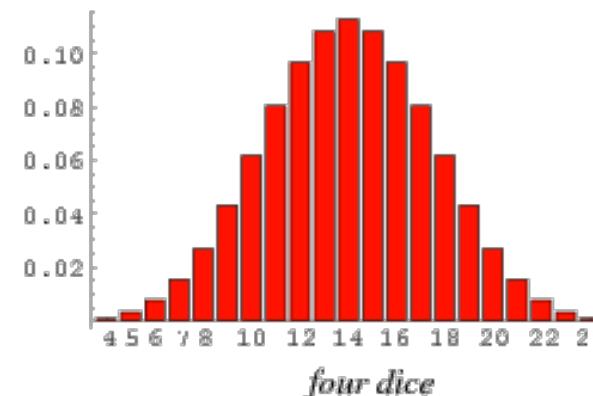
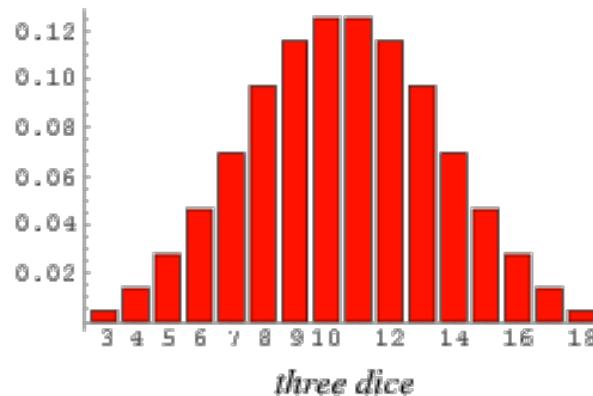
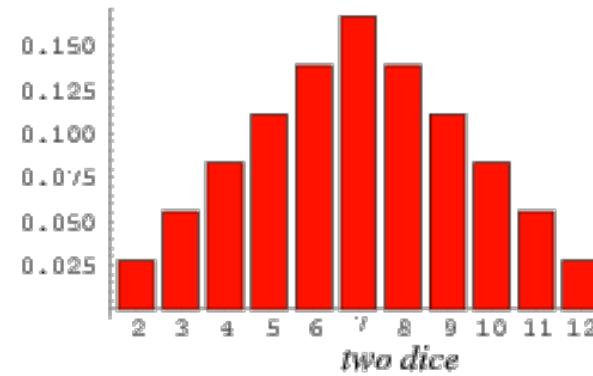
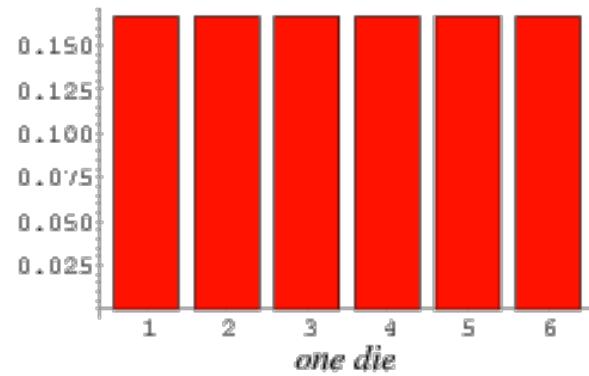
Sum of Many Dice



Sum of Many Dice

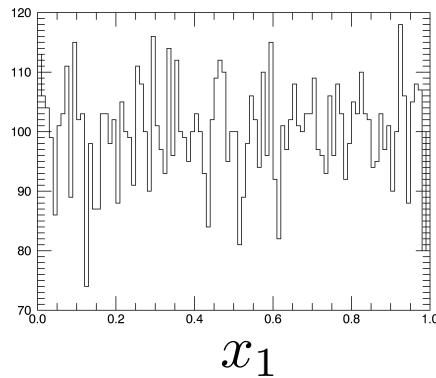


Sum of Many Dice

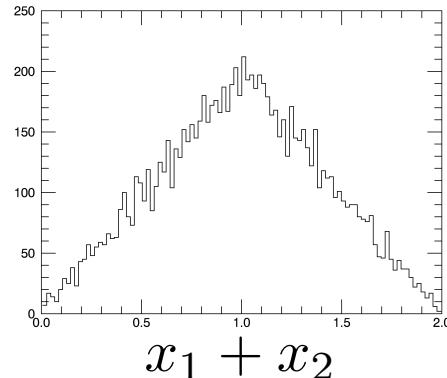


Central Limit Theorem

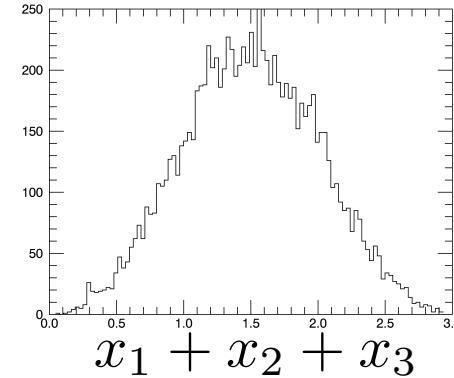
one



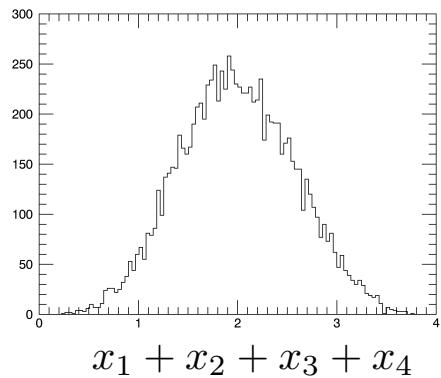
two



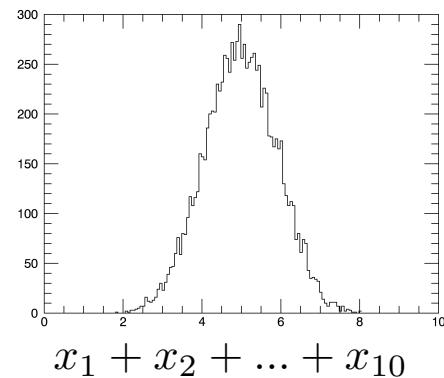
three



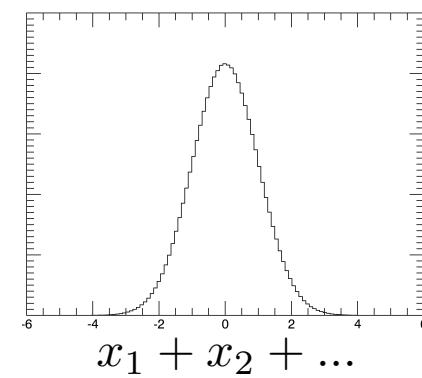
four



ten

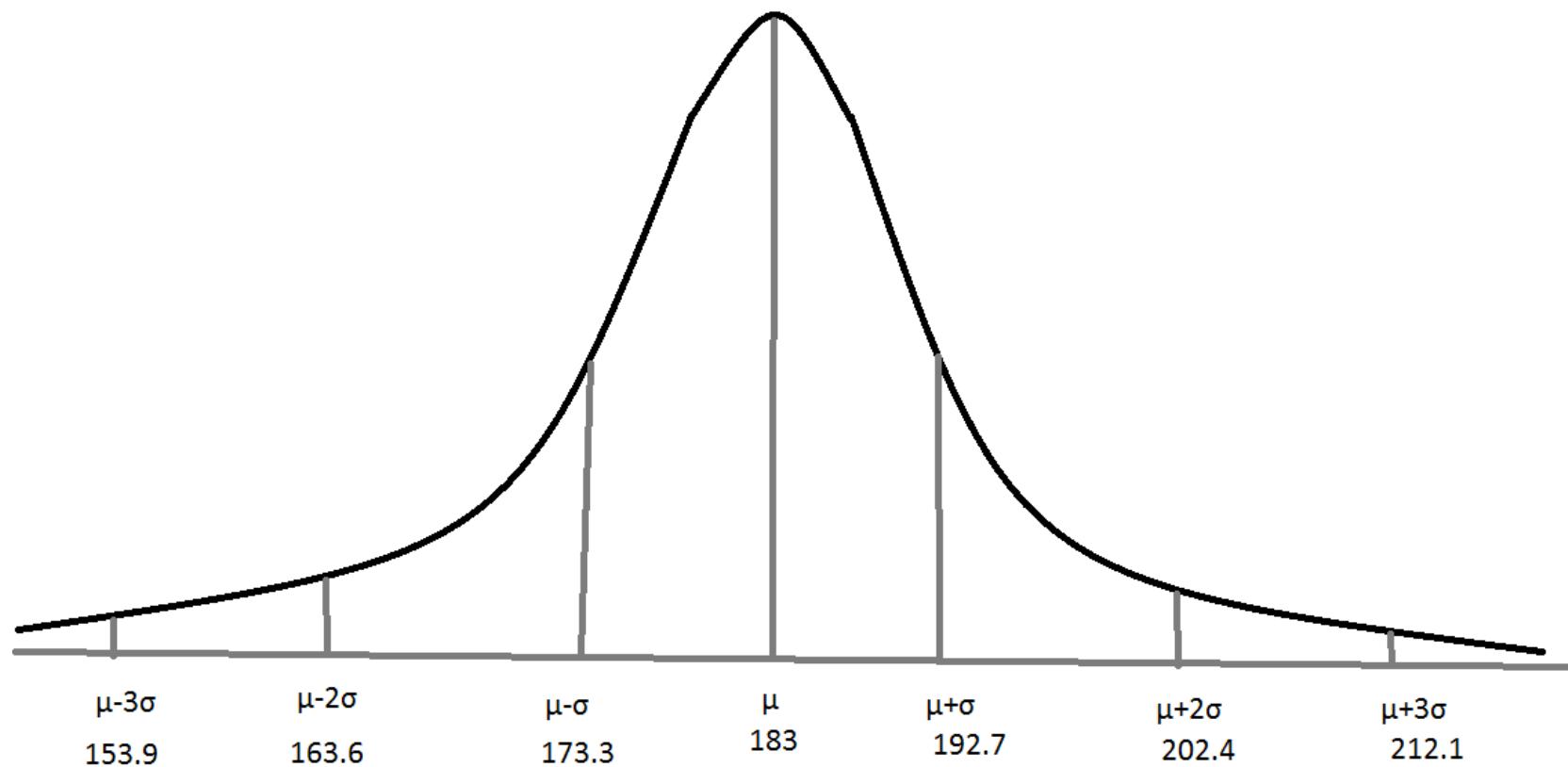


Many

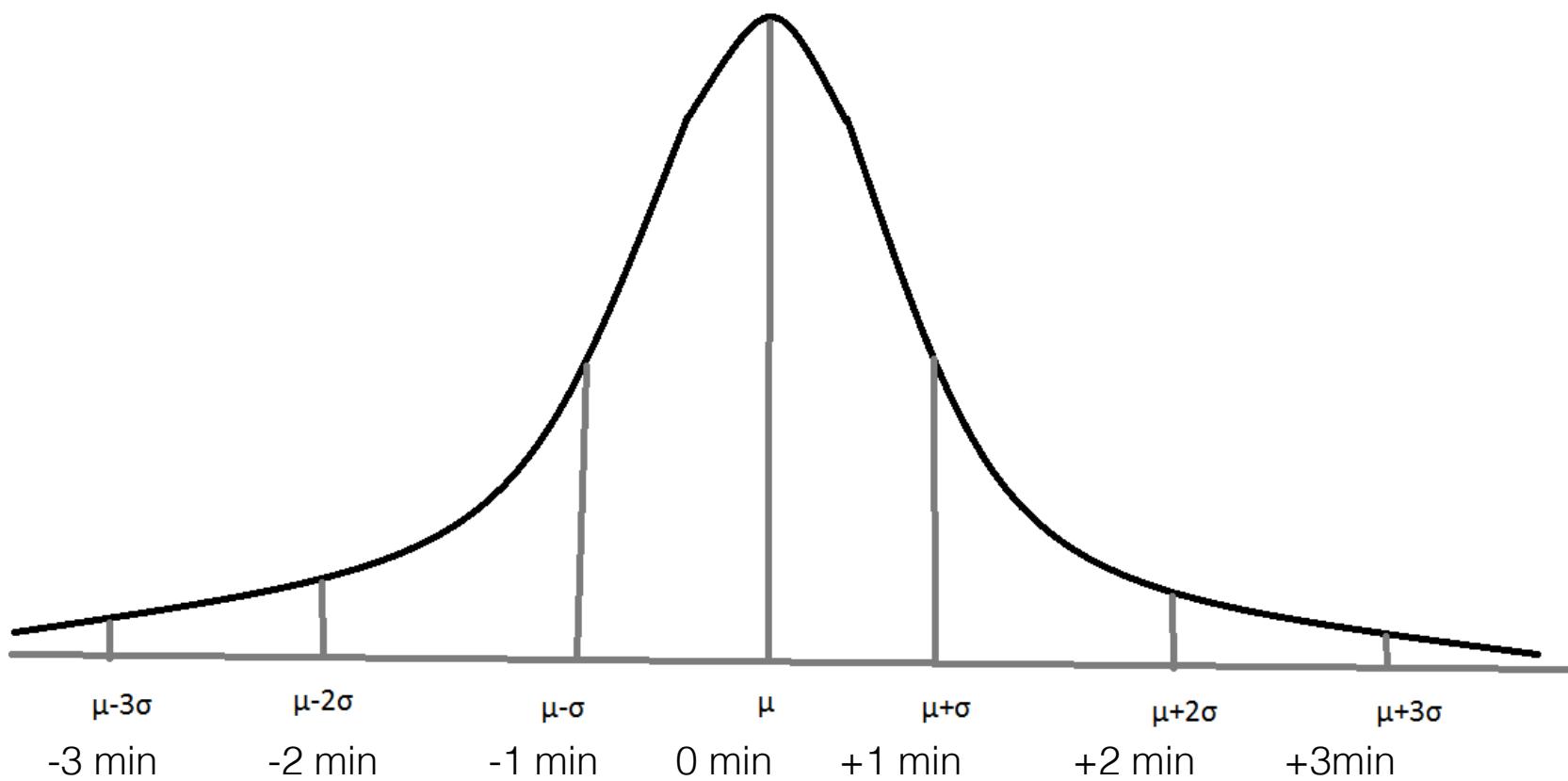


아이슬랜드 20 세 성인 남자의 키 =

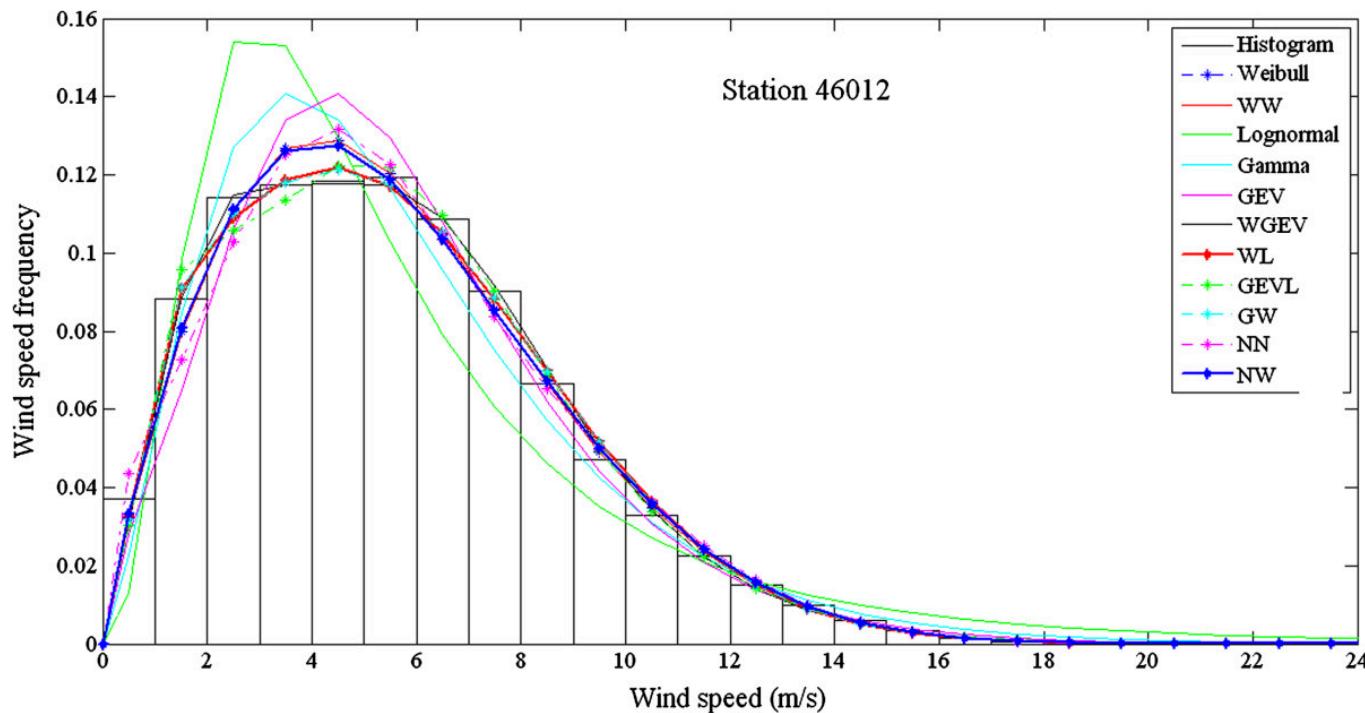
1st year growth + 2nd year growth + ... + 20th year growth



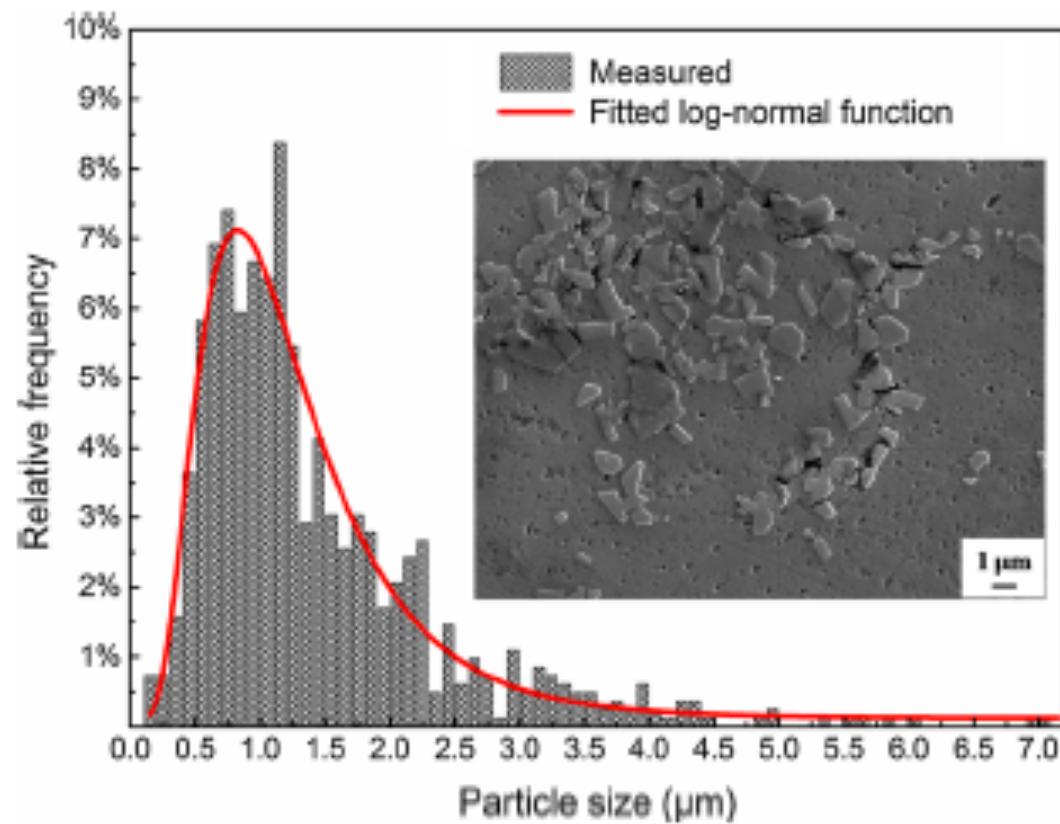
통근시간=기상시간 + 식사시간 + 지하철 도착시간 + 지하철 이동시간 + 신호등 대기시간 + 과학관 이동시간 + 엘리베이터 대기시간



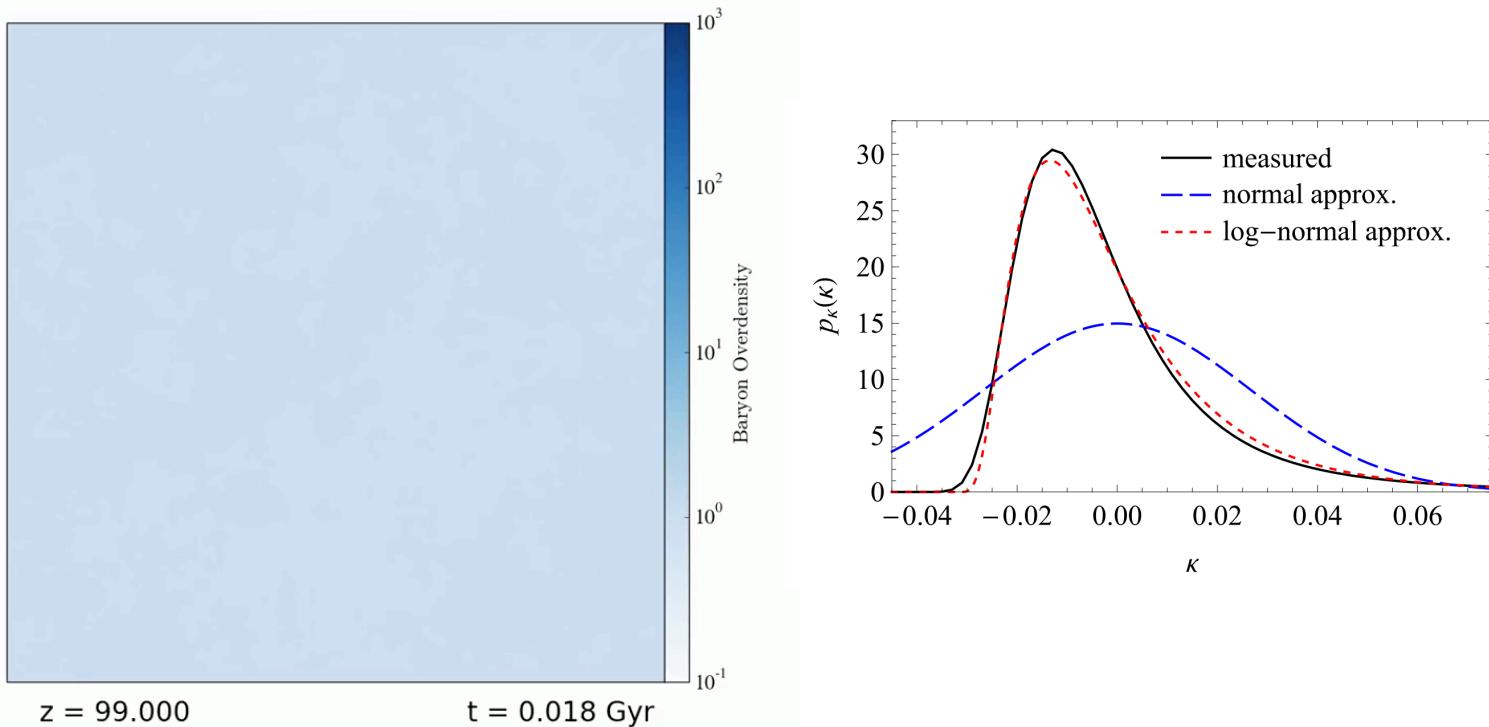
Log-normal Distribution



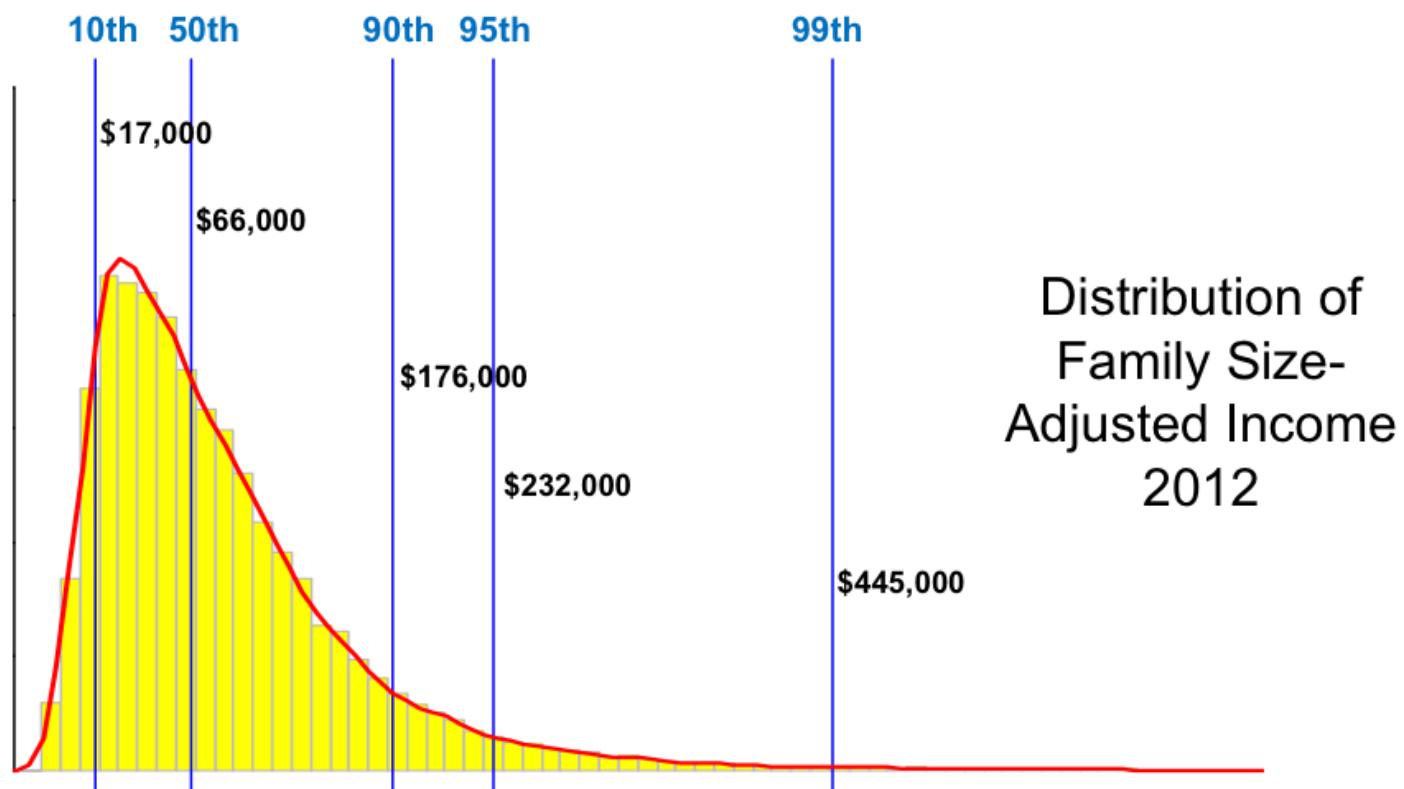
Log-normal Distribution



Log-normal Distribution



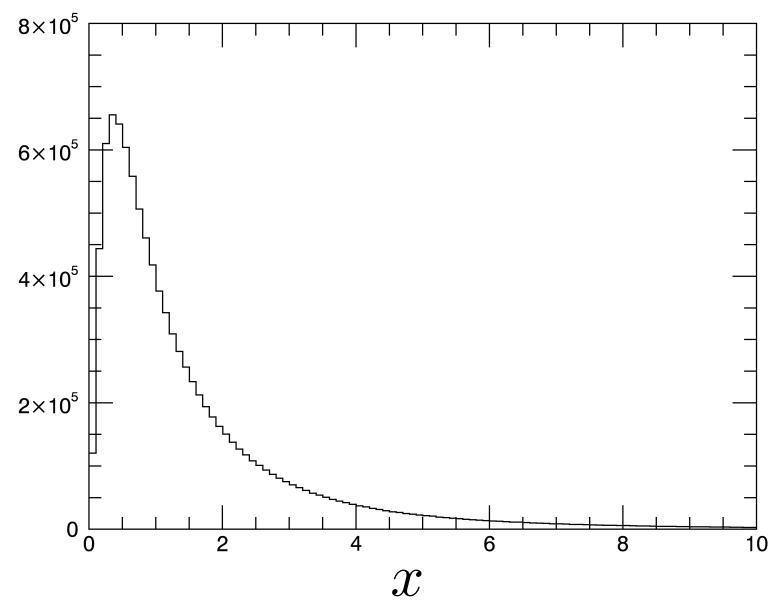
Log-normal Distribution



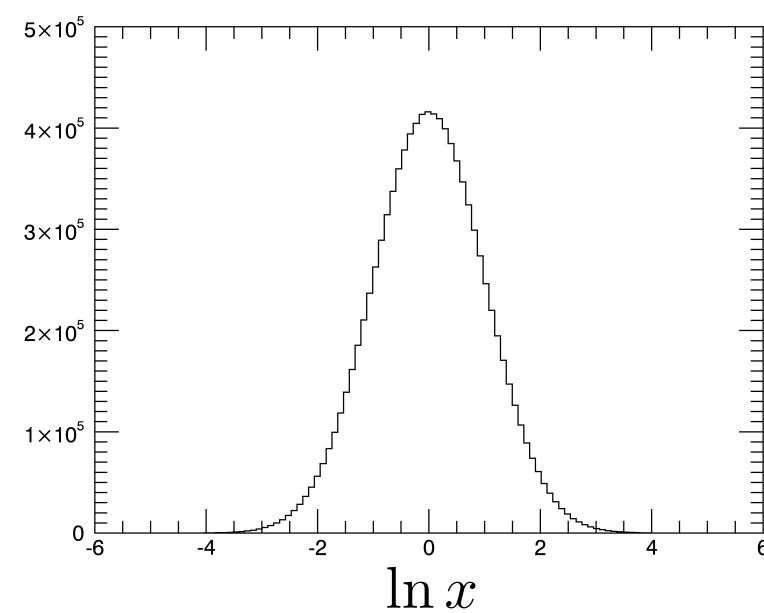
Log-normal Distribution

a log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed.

log-normal

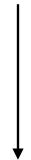


normal



왜 log-normal 분포가 자주 일어나는가?

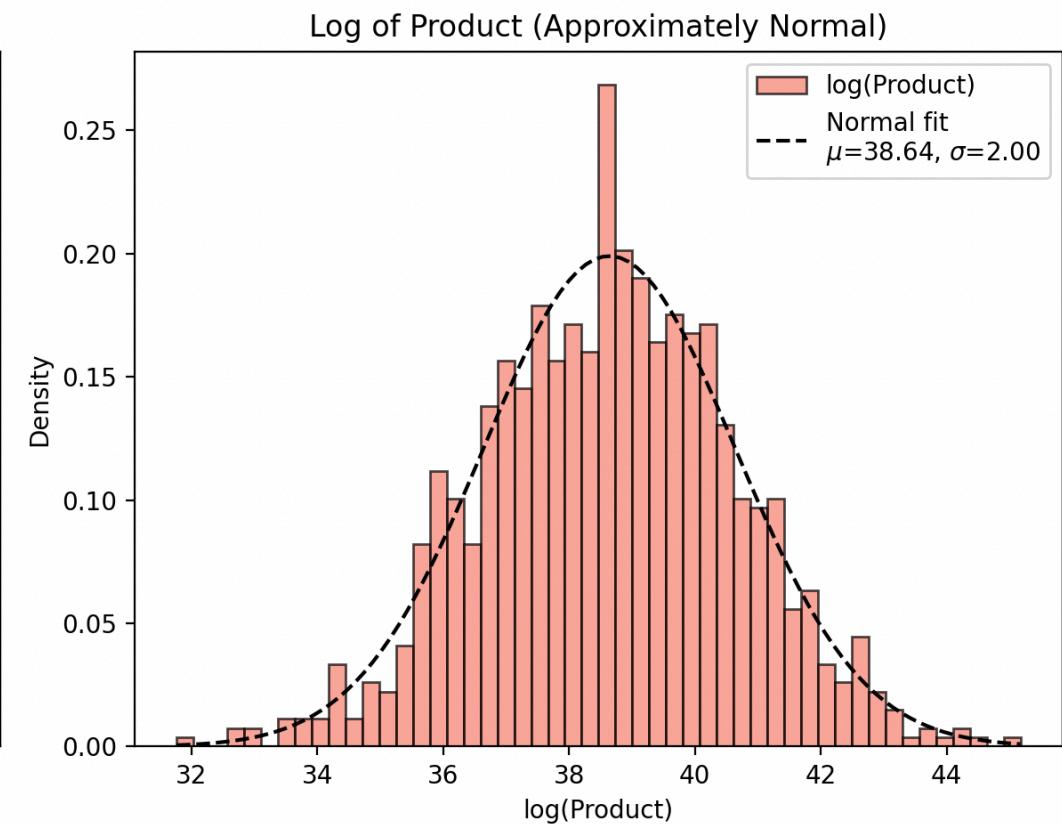
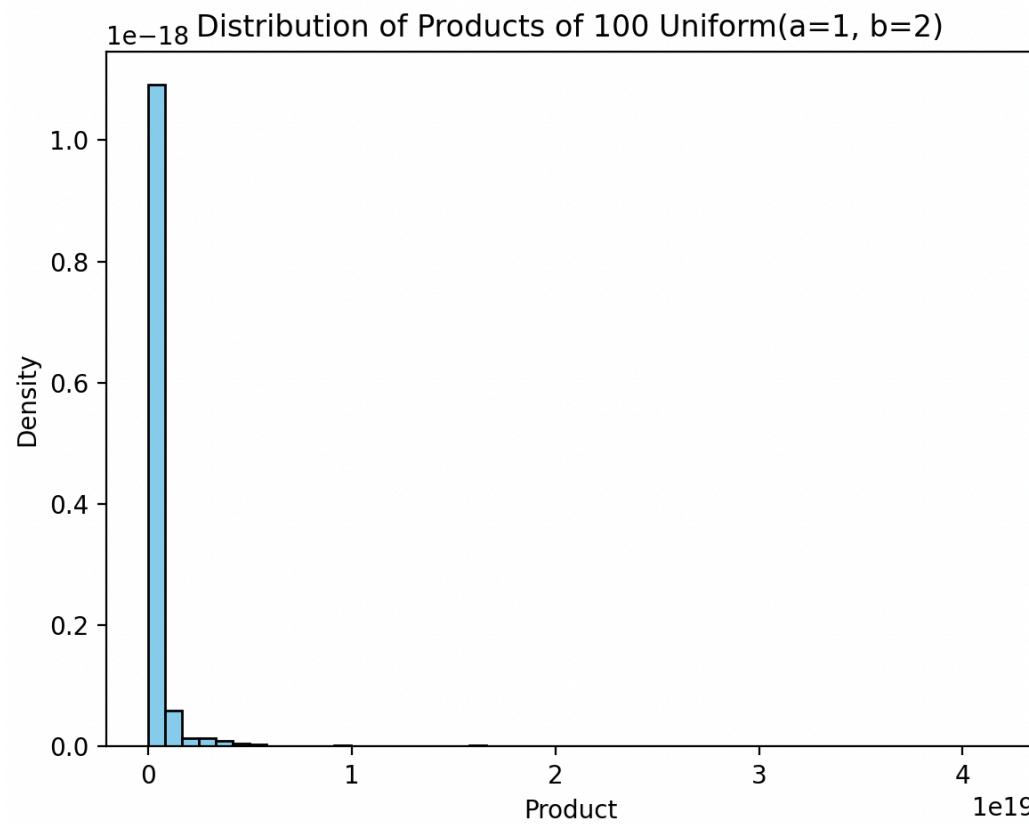
$$x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6 \cdot x_7 \dots$$

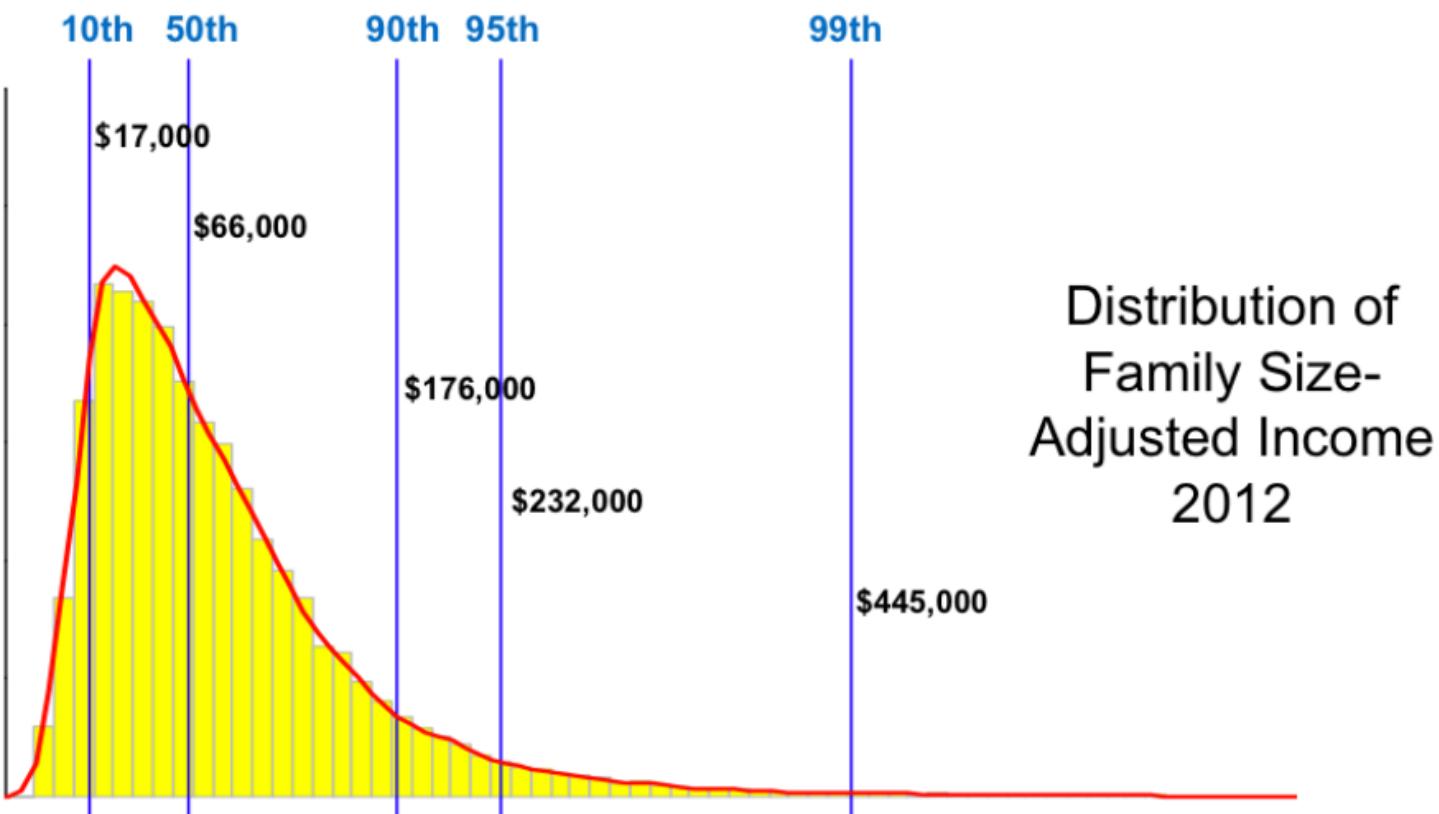


$$\log(x_1) + \log(x_2) + \log(x_3) + \log(x_4) + \log(x_5) + \log(x_6) + \log(x_7) \dots$$

central limit theorem 의 다른 발현

1부터 2 사이의 uniform 분포를 가지는 난수 100개를 곱해서 1000개의 난수를 생성



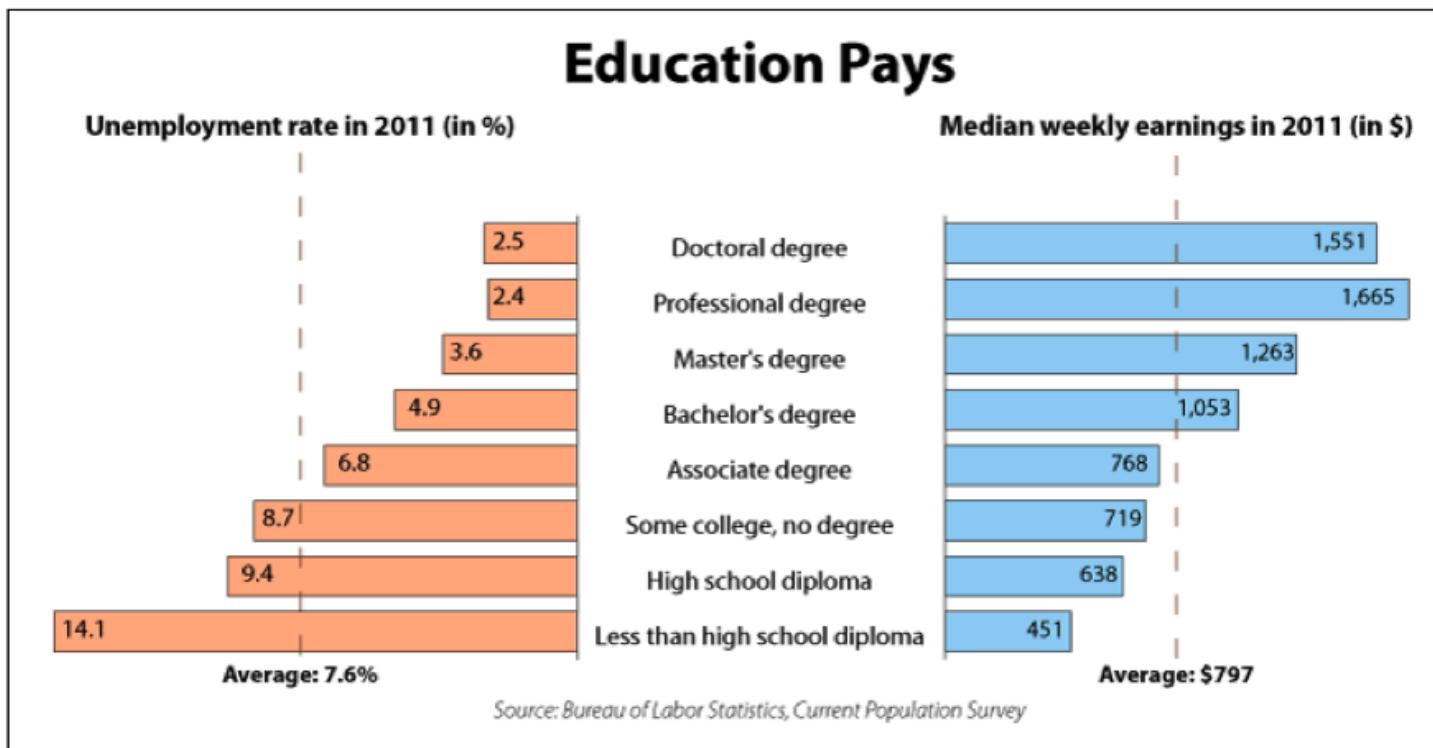


The income level may depend on many factors.

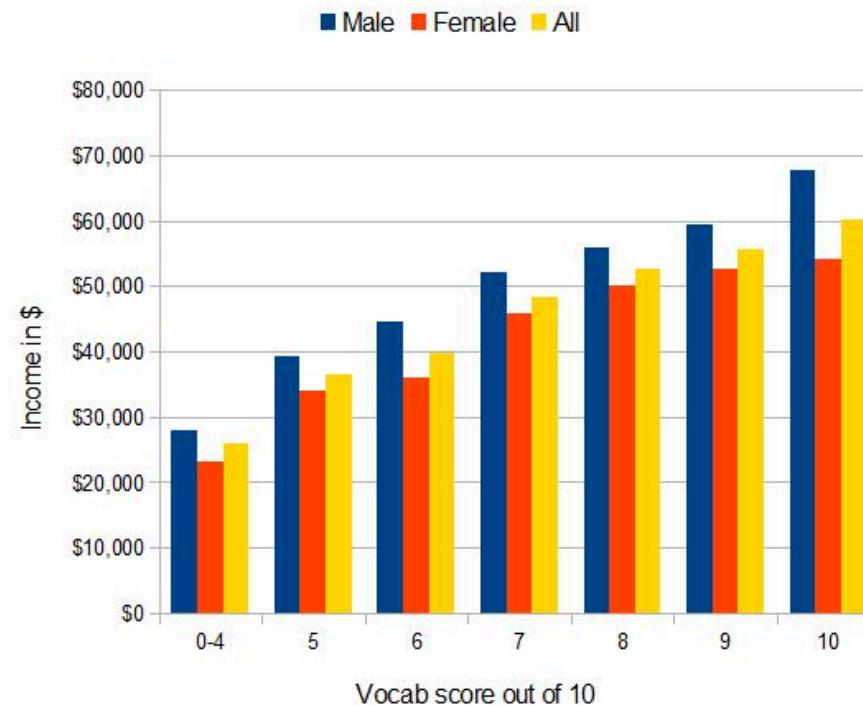
income=education x IQ x family wealth x health x self-control x luck x ...

Education pays...

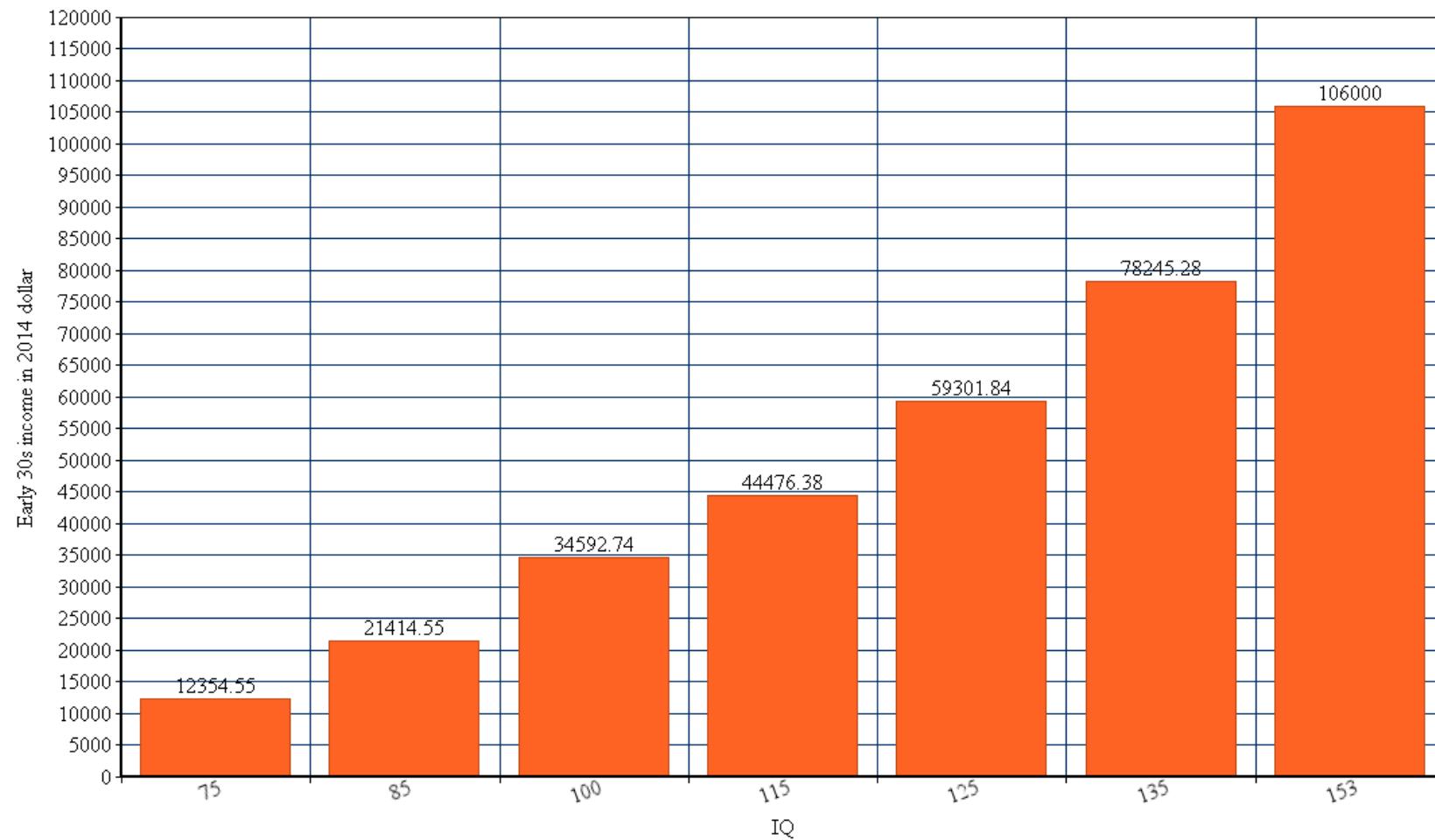
Education pays in higher earnings and lower unemployment rates

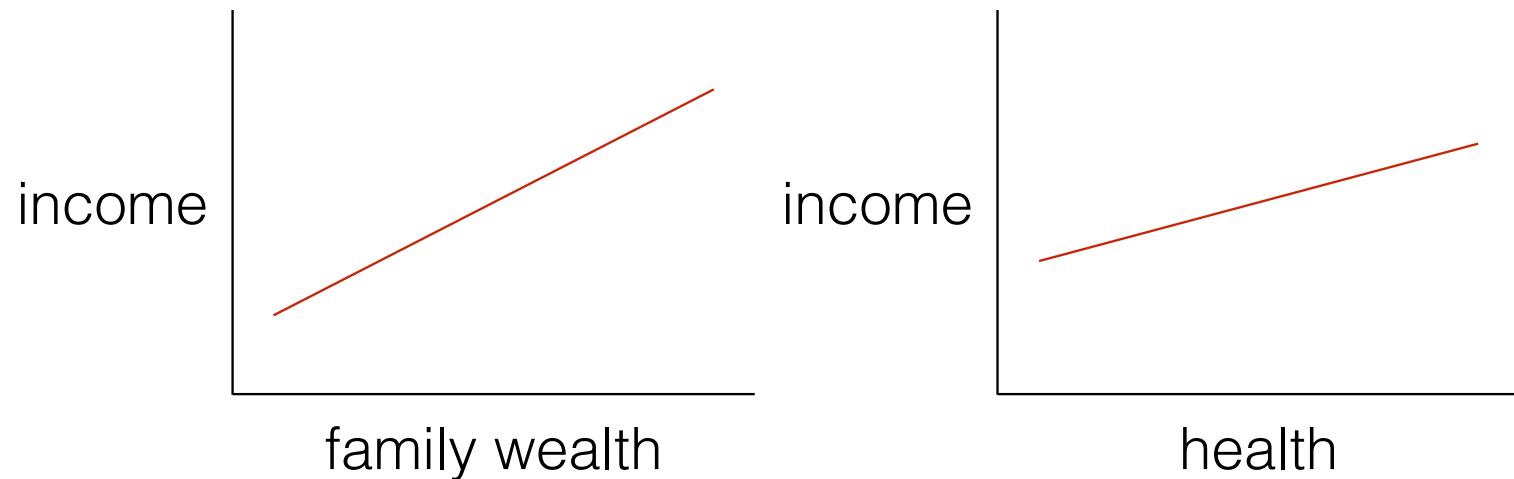


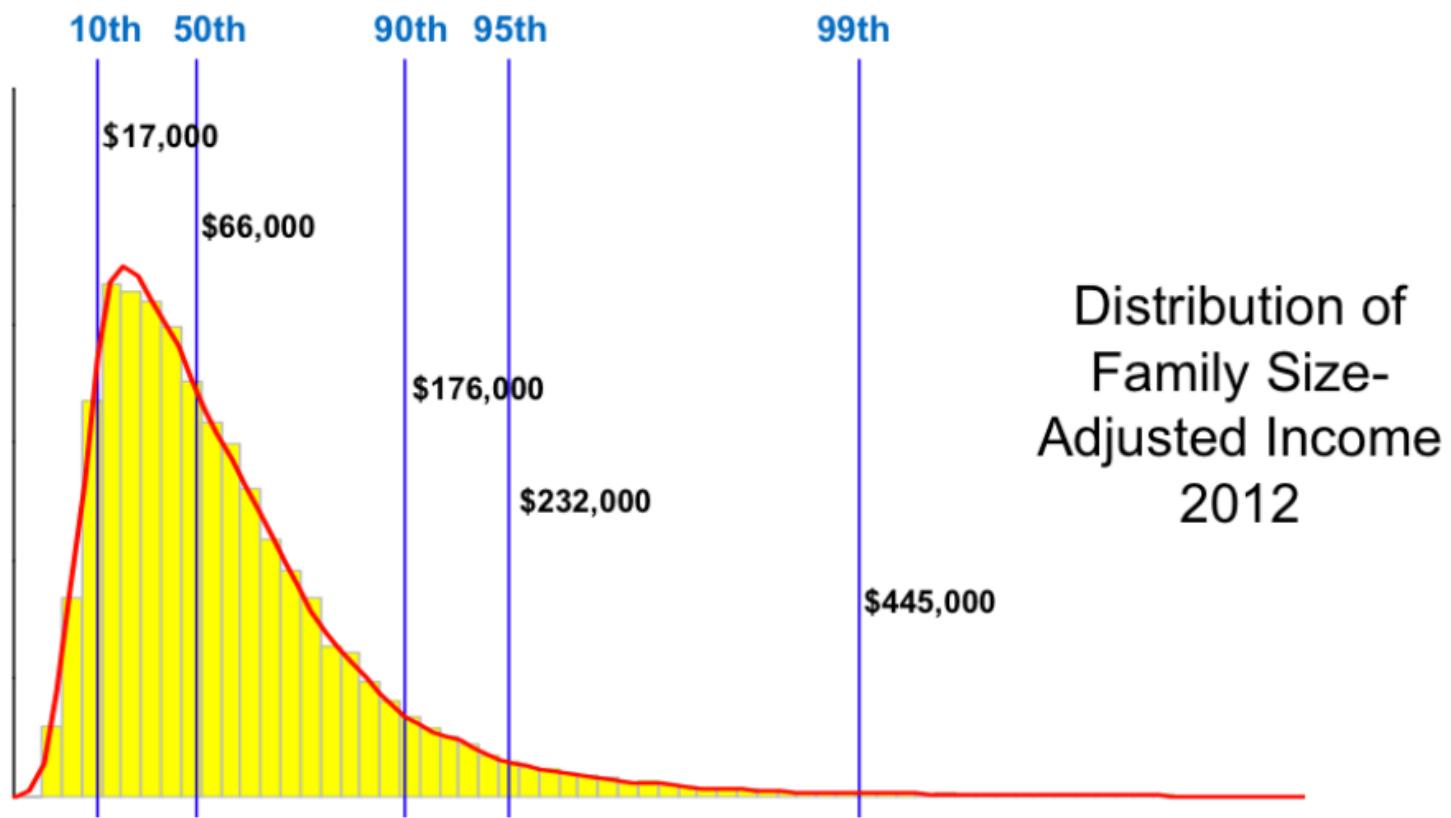
Household income for non-Hispanic whites by vocabulary score



Median income by IQ level





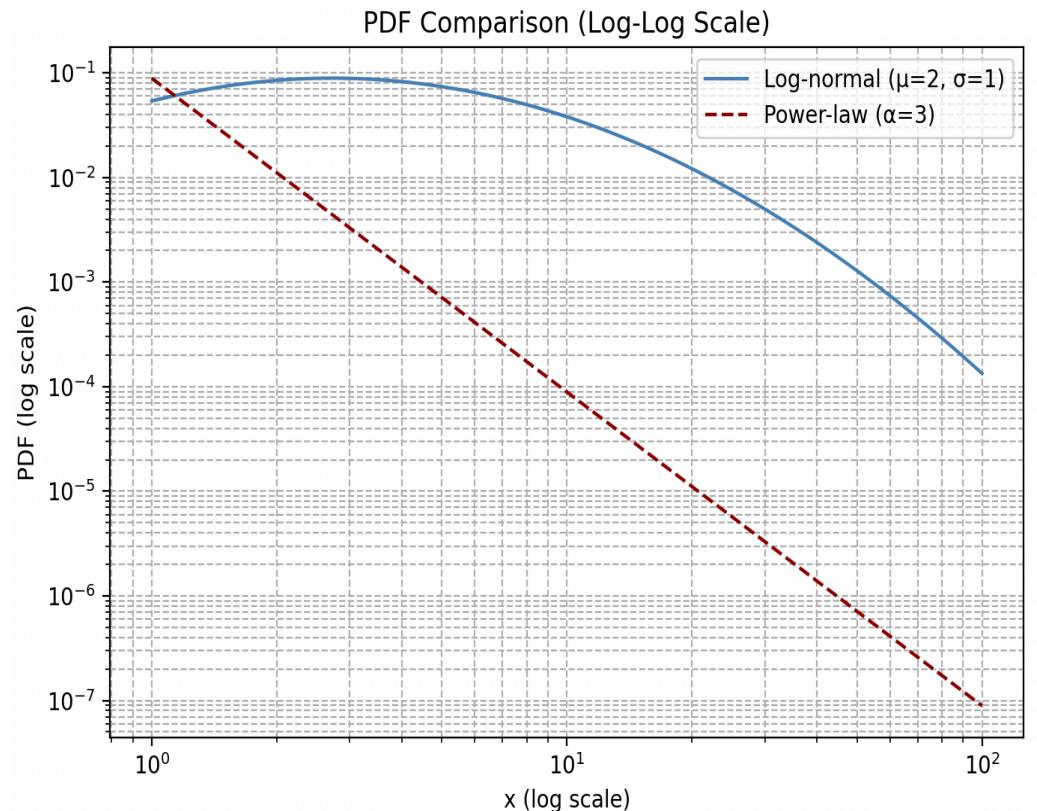


The log-normal distribution in income may be a natural outcome when the society is based on capitalism.

Some forms of external controls are needed to narrow the distribution.

Log-Normal vs Power Law

- **Heavy tails:** Both distributions can show **right-skewed, long-tailed behavior.**
- **Straight lines on log-log plots**
 - **A power-law is a straight line on a log-log plot over its *whole tail***
 - **A log-normal can appear *approximately linear* on a log-log plot over a limited range — this leads to confusion.**
- **Appear in empirical data:** Both are candidates for modeling phenomena like **wealth, city sizes, internet traffic, etc.**

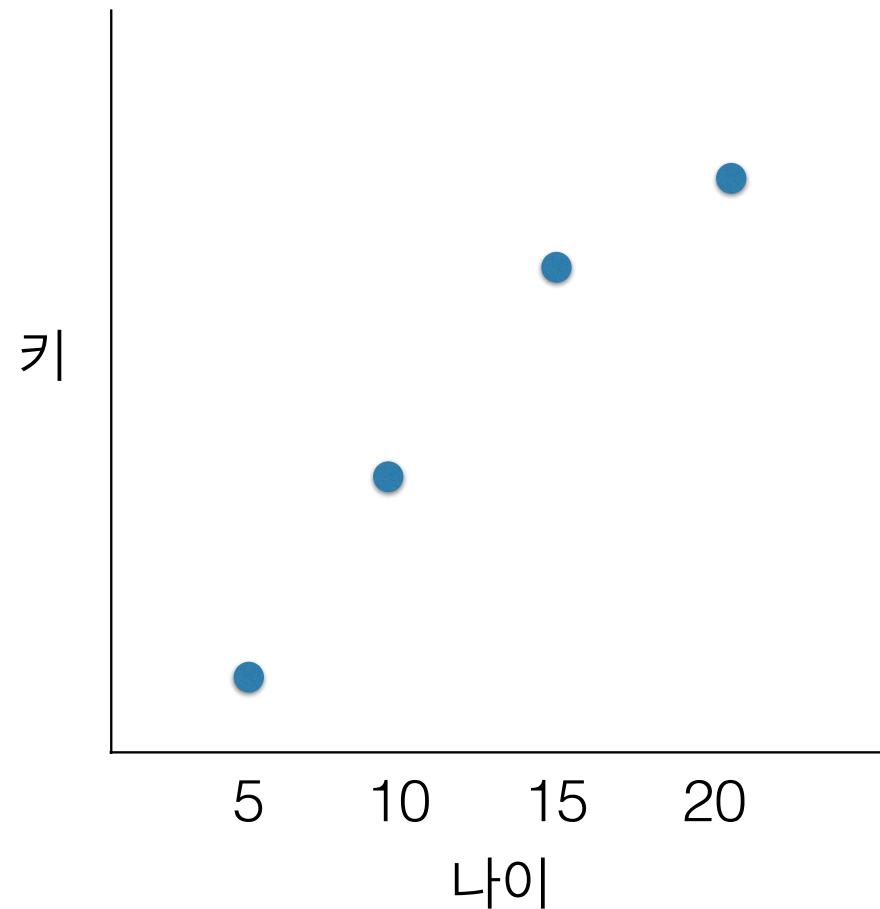


과제 6

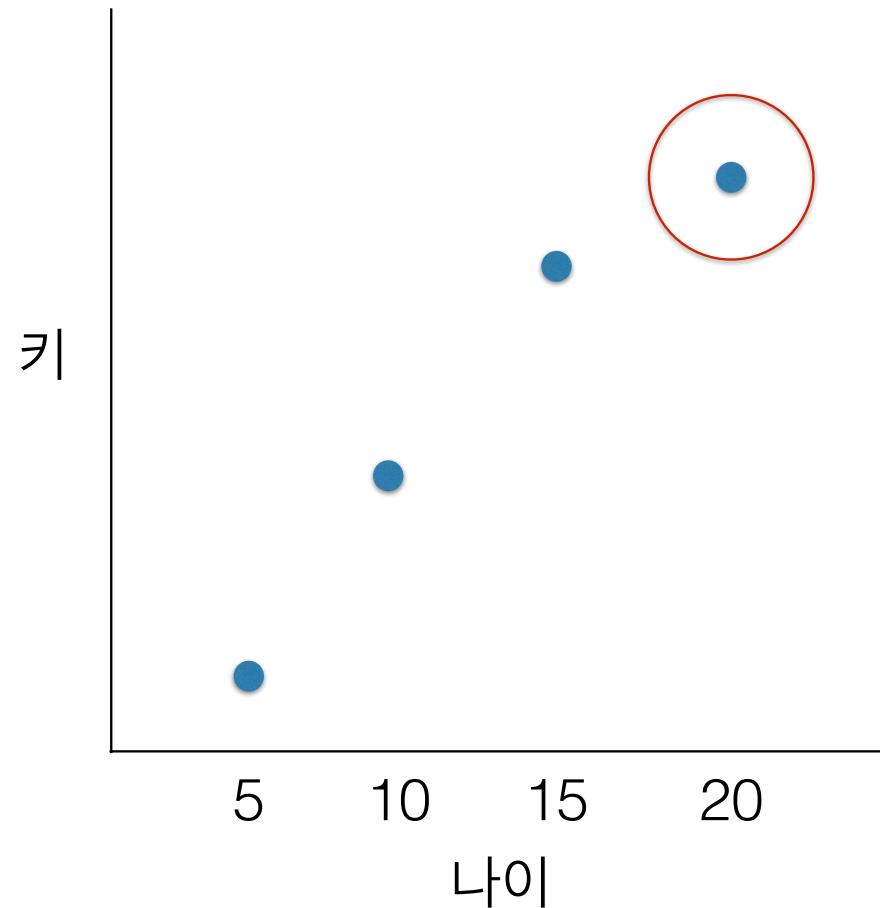
- uniform random number generator 사용 ($0 < x < 1$).
- random variable 이 1개일때, 2개일때, 3개일때, 10개일때 경우에 대하여 아래의 작업을 수행
- random number의 갯수는 각 variable당 10,000개.
- random variable을 모두 합했을때 나오는 분포를 히스토그램으로 표현 (normal distribution).
- random variable을 모두 곱했을때 나오는 분포를 히스토그램으로 표현 (log-normal distribution).
- 자연 또는 우리 일상 생활가운데 나타나는 분포 중 CLT때문에 나타나는 현상을 찾아 설명 (normal 또는 log-normal).

Interpolation

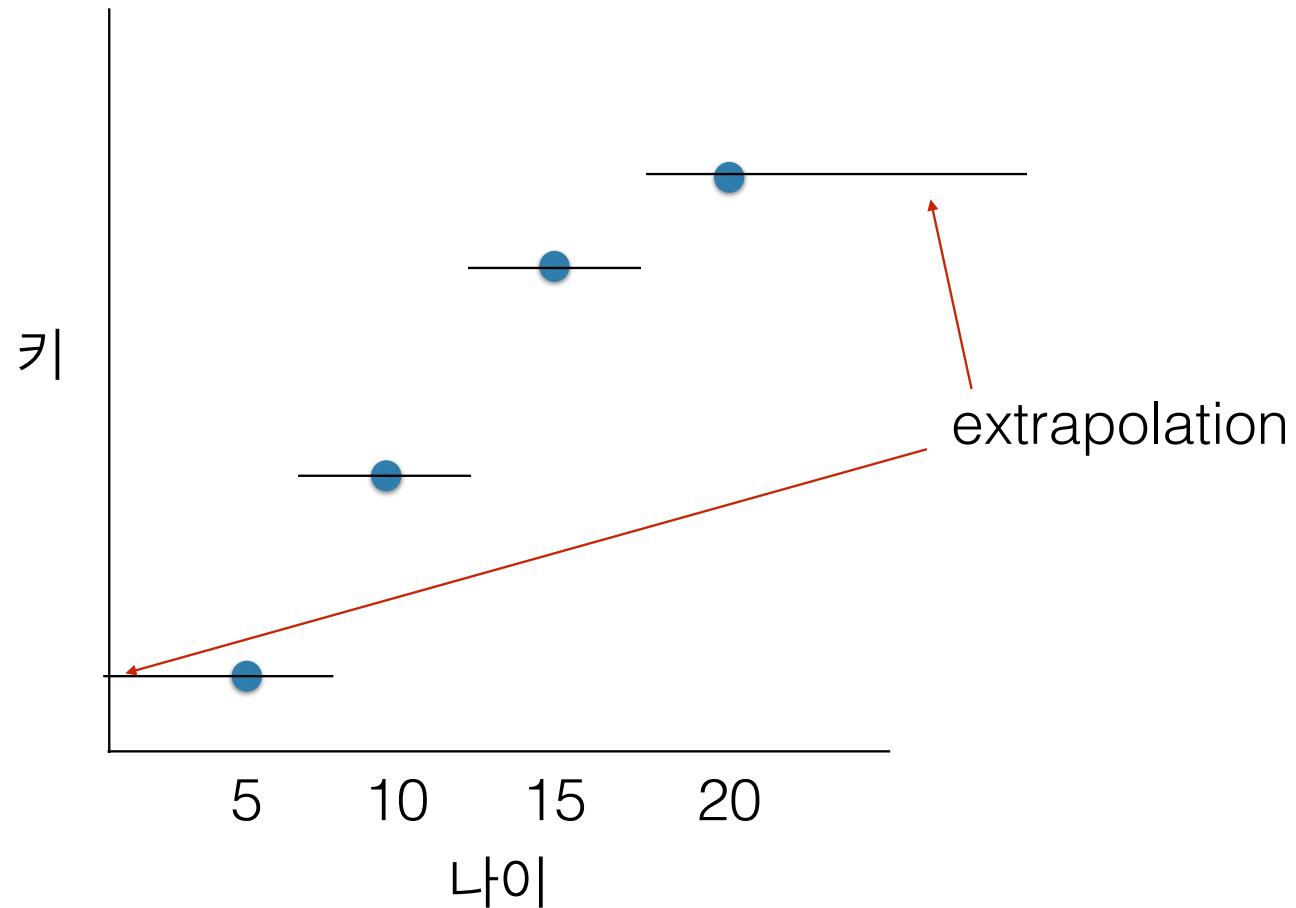
18 살때의 키는 얼마인가?

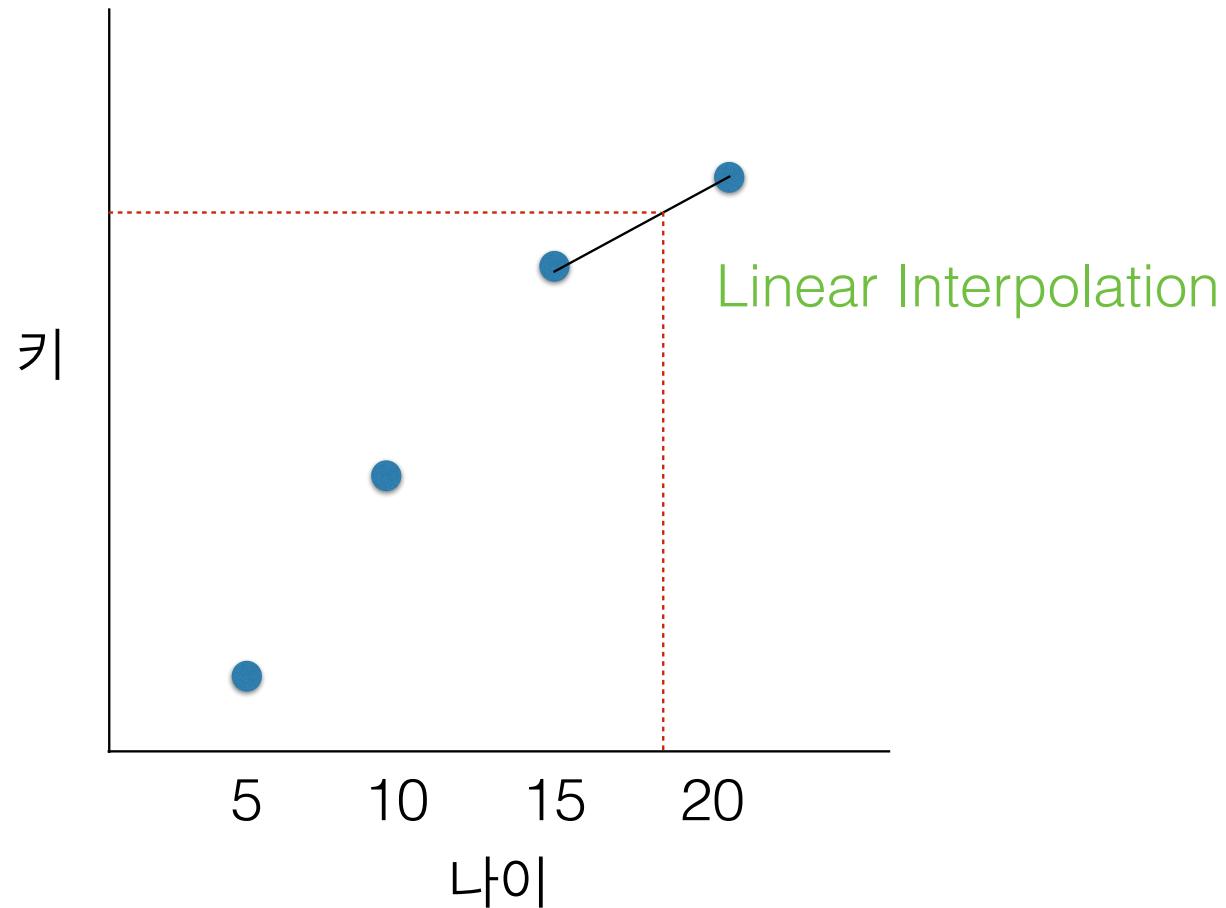


Nearest Neighbor Interpolation

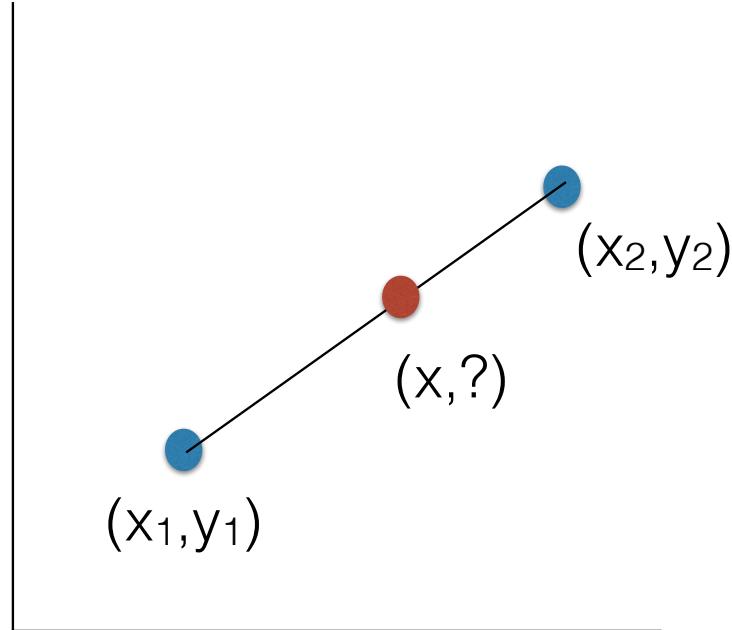


Nearest Neighbor Interpolation for the entire range





Linear Interpolation

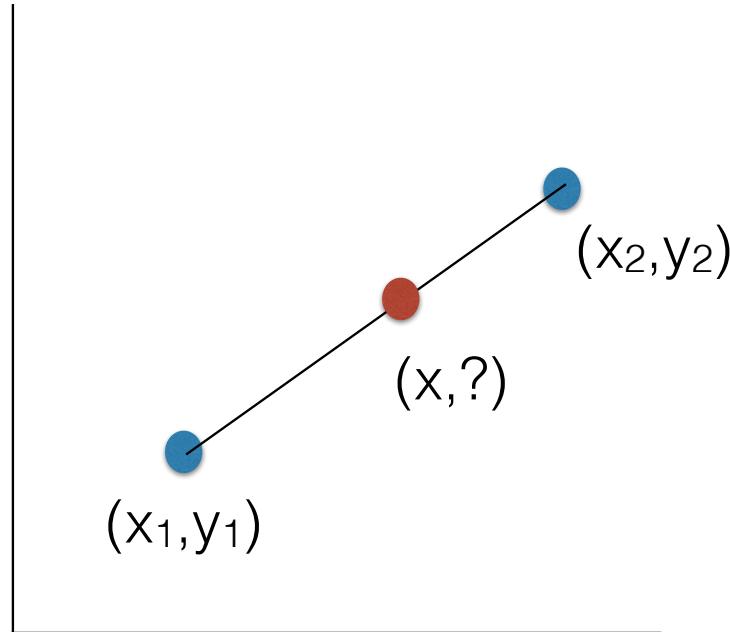


$$\text{기울기: } \frac{y_2 - y_1}{x_2 - x_1}$$

$$\text{그래프 방정식: } y - y_1 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1)$$

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} (x - x_1)$$

Linear Interpolation 의 기하학적 의미



$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} (x - x_1) \longrightarrow y = \frac{x_2 - x}{x_2 - x_1} y_1 + \frac{x - x_1}{x_2 - x_1} y_2$$

$$y = w_1 y_1 + w_2 y_2$$

Linear interpolation gives a weighted average.