통계입문 1부

2021.06.17.

1. 주요 확률분포들

<Bernoulli distribution>

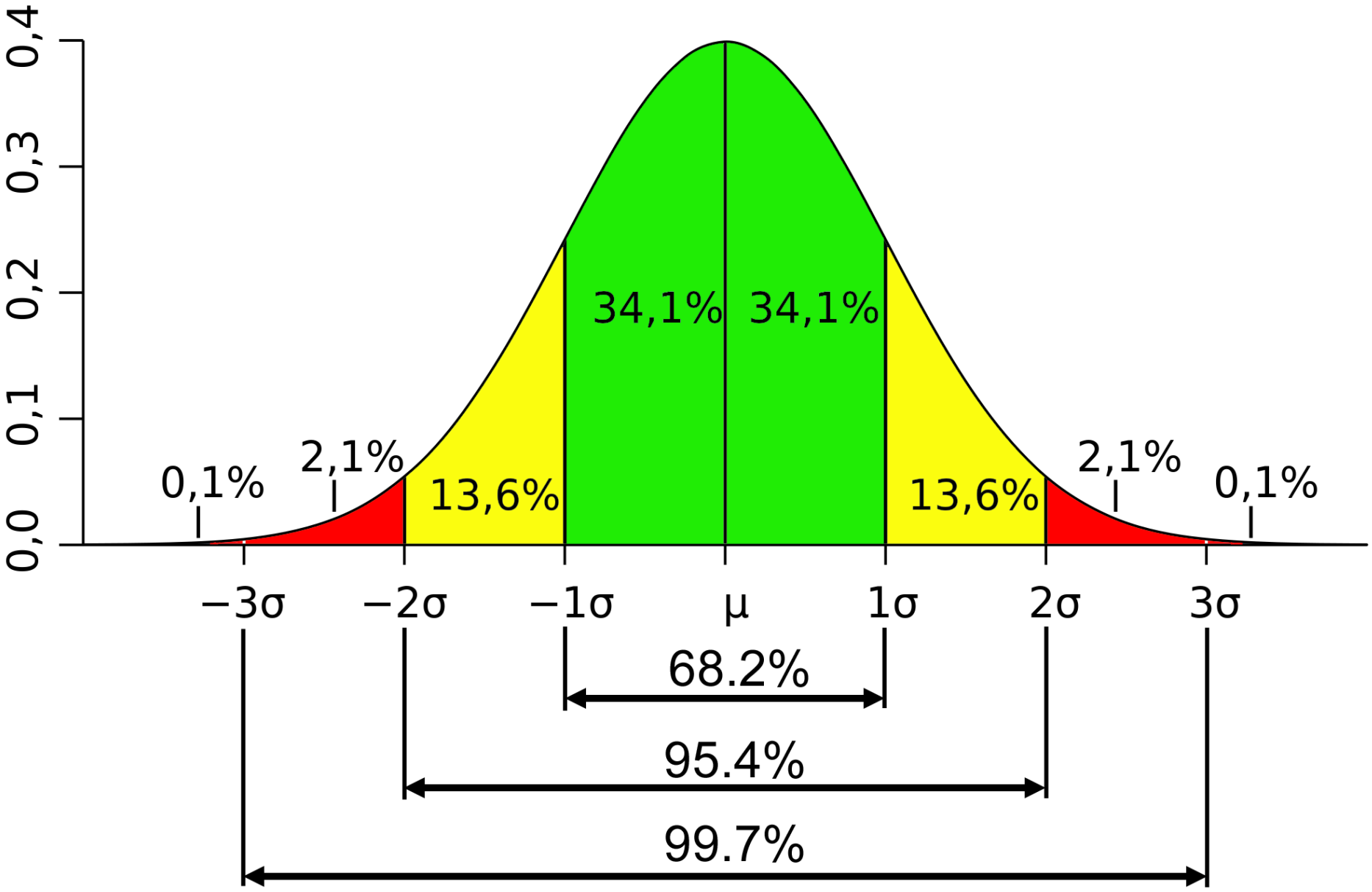
$$f(x) = \begin{cases} p\ (x=1) \\ \\ 1-p\ (x=0) \end{cases}$$

<Binomial distribution>

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

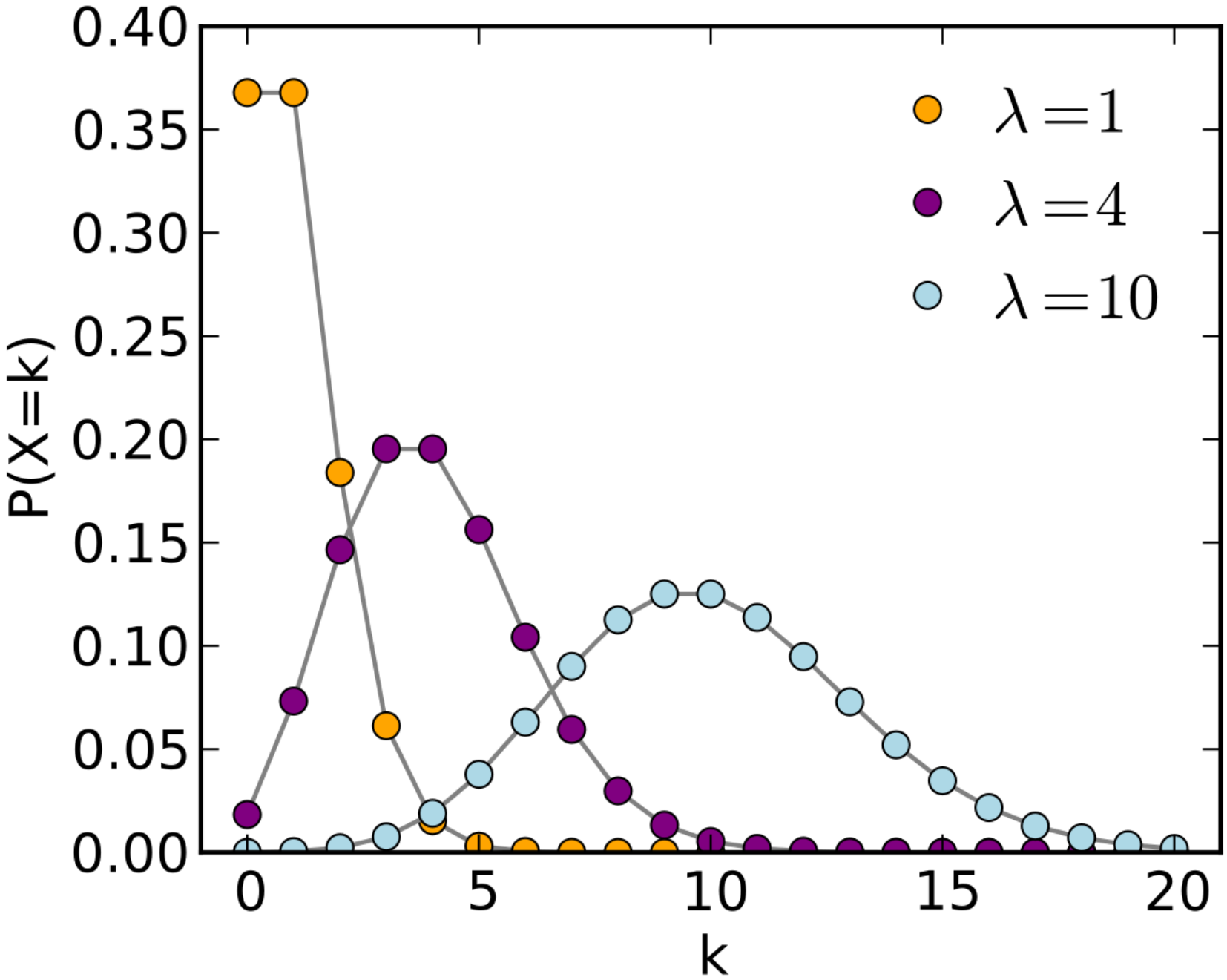
<Gaussian distribution>

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

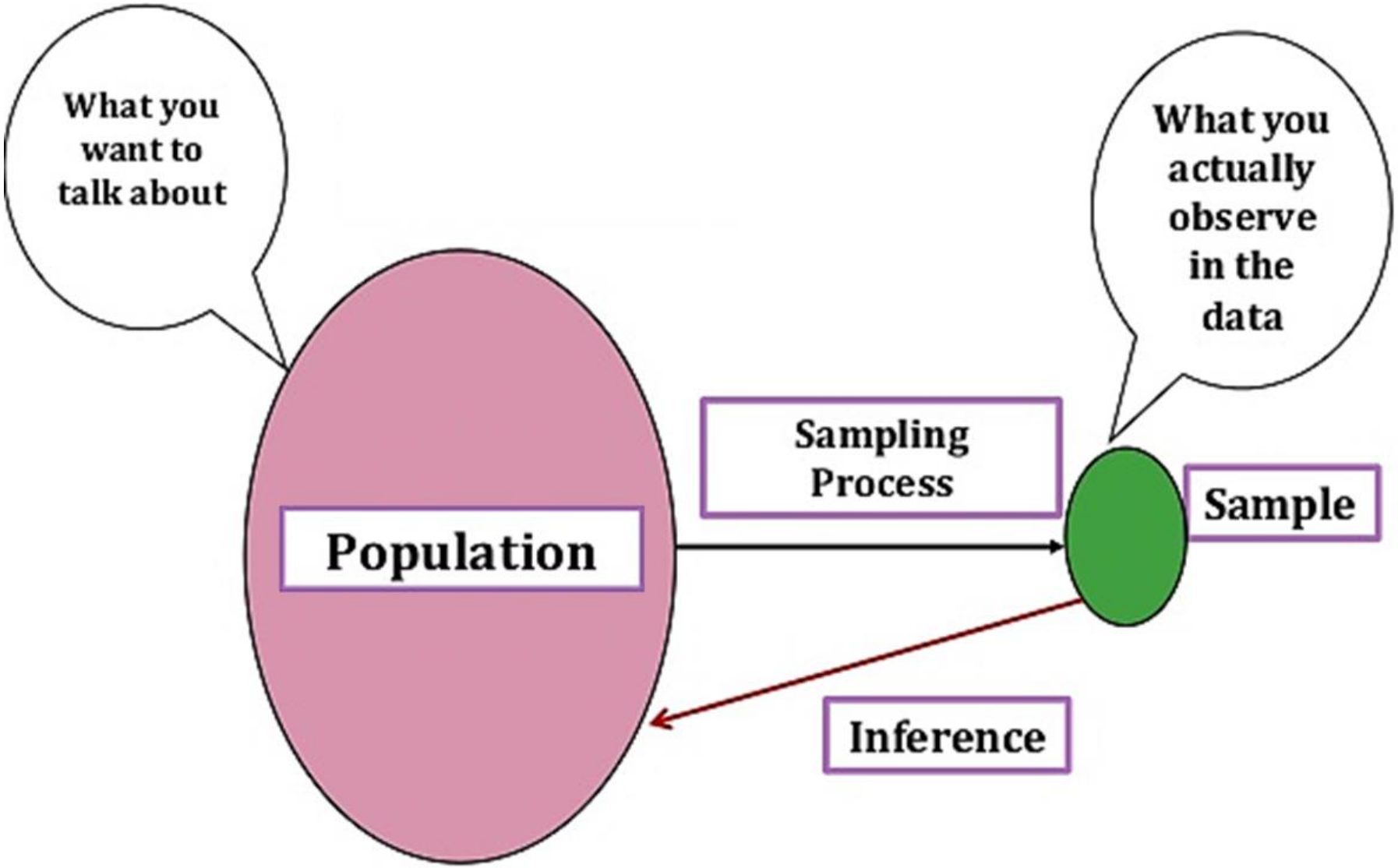

0,4
0,3
0,2
0,1
0,0
34,1%
34,1%
13,6%
13,6%
2,1%
2,1%
0,1%
0,1%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ
68.2%
95.4%
99.7%

<Poisson distribution>

$$f(n;\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(X=k)
0
5
10
15
20
k
$\lambda = 1$
$\lambda = 4$
$\lambda = 10$

2. population(모집단), sample(표본), sampling(표집)



우리가 분석할 때 손에 들고 있는 데이터는 **sample**.

근데 우리가 궁금한 건 지금 당장 들고 있는 sample이 아니라
이 sample이 원래 속해 있는 **population**의 정보다!
(즉 이 과정에서 내가 손에 든 sample은 **무한한 sampling의 산물 중 하나**일 뿐이라는 걸 감안해야한다!)
→ 이렇게 sample로 population에 대해 알아내는 과정이
"통계적 추론(Statistical Inference)"

물론 제대로 된 statistical inference를 하려면, 애초에 우리가 **random sampling**을 실시해서 얻은 데이터를 갖고 있어야 한다.
(구글독스나 출처를 알 수 없는 요상한 데이터 같은 건 좀…)

4. 평균

사회학자 A의 연봉이 7000만원
통계학자 B의 연봉이 800만원
물리학자 C의 연봉이 1200만원

이들 세 사람의 평균 연봉은?
→ (7000+800+1200)/3명=3000만원

근데 정확히 **'평균'**의 개념을 사용하는 이유가 뭘까?

평균이란

(1) 제곱법에 기초하여

(2) 측정값에 포함되어 있는 차이를

(3) 가장 작게 만듬으로서

특정 데이터의 정보를 가장 잘 보여주는 값!

분산이란 데이터가 퍼져있는 정도!

$(x-\bar{x})$ ← 편차 (deviation)

$(x-\bar{x})^2$ ← 편차 제곱

$\sum(x-\bar{x})^2$ ← 편차 제곱합

$\frac{\sum(x-\bar{x})^2}{n}$ ← 편차 제곱의 평균(분산) (variance)

모평균: $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = m$

모분산: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$

표본평균: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

표본분산: $s^{*2} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$ $s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$

5-1. 대수의 법칙 (Law of Large Numbers)

서로 독립인 확률 변수 $X_1, X_2, X_3, \cdots X_n \cdots$ 들이

평균이 μ인 동일한 확률 분포를 따를 때, (ϵ>0)

$$\lim_{n \to \infty} \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1$$

데이터 수가 많은 sample일수록 그 표본평균(sample mean)값이

모집단 평균(population mean)의 값에 가까워진다!

(위에 식 몰라도 되니 요 정의만 제대로 이해하세요!)

5-2. 중심극한정리 (Central Limit Theorem)

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)-\mu\right)\xrightarrow{d}N\left(0,\sigma^2\right)$$

동일한 확률분포를 가진 독립 확률 변수 k개의 평균의 분포는 k가 적당히 크다면 정규분포에 가까워진다. 그냥 이렇게만 보면 무슨 말인지 잘 모르겠으니까 R 코드와 함께 알아보자! 아 그리고 위에 저 정의는 지금 달달 외우거나 심각하게 보지 말자. 맨 뒤에 붙어있는 '수업 보충 노트'에 좀은 의미의(?) CLT 정의를 좀 더 쉽게 이해할 수 있게 다시 써뒀다.

부록 : 표본분산 분모에서 1을 빼도 어떻게 모분산과 잘 맞아 떨어지는지의 증명

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

$$E[s^2] = \frac{1}{n-1}E[\sum_{k=1}^{n}(X_k - \overline{X})^2]$$

다음 페이지(P.23)에서는 딱 이 부분만 전개하는걸 보여준다

$$\sum_{k=1}^{n}(X_k-\overline{X})^2=\sum_{k=1}^{n}((X_k-m)+(m-\overline{X}))^2$$

$$=\sum_{k=1}^{n}((X_k-m)^2+2(X_k-m)(m-\overline{X})+(m-\overline{X})^2)$$

$$=\sum_{k=1}^{n}(X_k-m)^2+2(\overline{X}-m)n(m-\overline{X})+n(m-\overline{X})^2$$

$$=\sum_{k=1}^{n}(X_k-m)^2-2(\overline{X}-m)n(\overline{X}-m)+n(\overline{X}-m)^2$$

$$=\sum_{k=1}^{n}(X_k-m)^2-n(\overline{X}-m)^2$$

이 파트가 전개되는게 좀 아리까리할 수 있을 것 같아서 P.36에 과하게 친절한 풀이를 써보겠습니다.

물론 그걸 안 보고 혼자 먼저 위 수식을 대강 눈으로 이해해보세요! 그 다음에 P.43을 봐도 안 늦습니다.

$$\therefore E[s^2] = \frac{1}{n-1} E[\sum_{k=1}^{n} (X_k - \overline{X})^2] = \frac{1}{n-1} E[\sum_{k=1}^{n} (X_k - m)^2 - n(\overline{X} - m)^2]$$

$$E[(X_k - m)^2] = \sigma^2$$

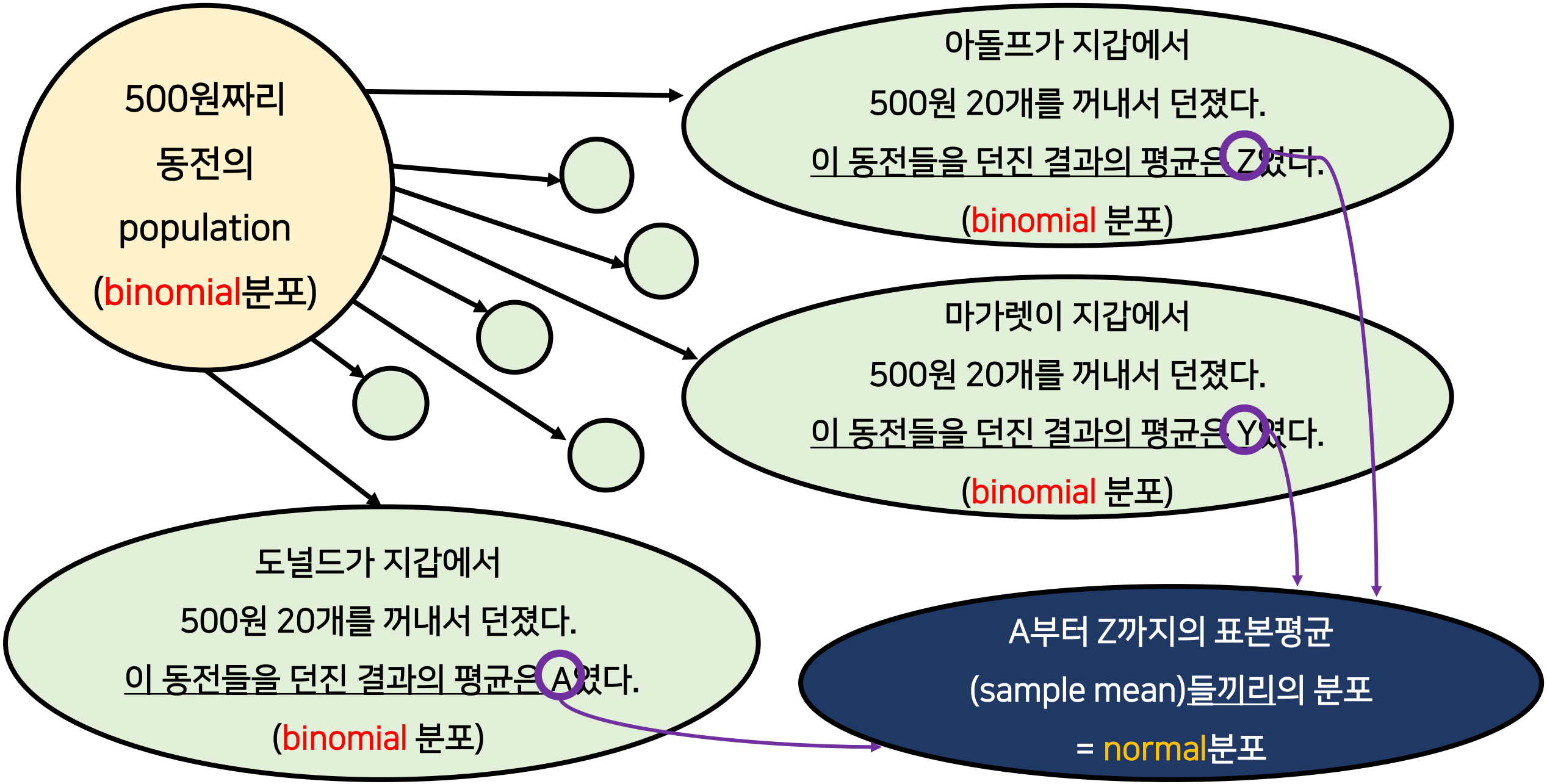
$$E[(\overline{X} - m)^2] = V(\overline{X}) = \sigma^2 / n$$

$$= \frac{1}{n-1} \left[\sum_{k=1}^{n} E[(X_k - m)^2] - n(E[(\overline{X} - m)^2]) \right]$$

$$\frac{1}{n-1} \left(n\sigma^2 - n\frac{\sigma^2}{n} \right) = \sigma^2$$

수업 보충 노트

<CLT 한 짤 요약 : population과 sample, sampling을 구분해서 표시해봤다>

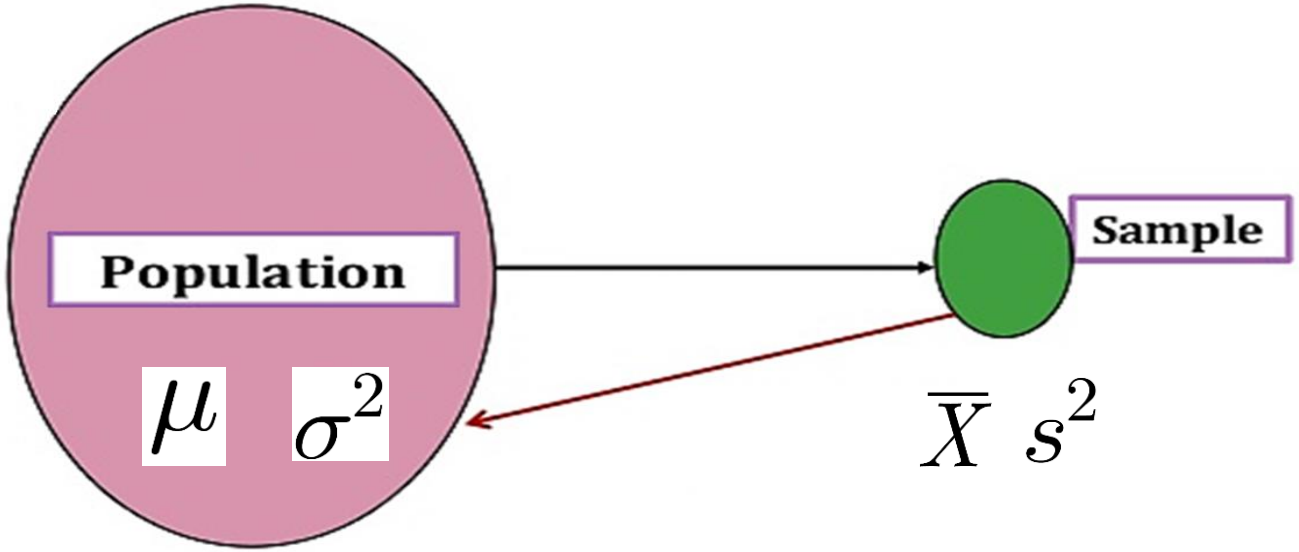


2. LLN 보충설명

대수의 법칙은 오히려 신경쓸게 더 적어요. CLT에서는 population, sample, sampling 세 개를 다 신경써야 했었죠? LLN은 걍 population과 sample의 관계만 보면 됩니다. 동전을 지갑에서 5개 꺼내서 던져볼 때 보다는 10개 꺼내서 던져볼 때, 그리고 그 보다는 한 10000개 쯤 꺼내서 던져 볼 때, 그 던진 결과의 평균값이 모집단의 평균값에 가까워진다는 겁니다. 이거는 코드 직접 보시면 이해가 잘 될 거에요. (그래도 헷갈리면 http://jaekwangkim.com/articles/2016-09/Law-of-Large-Numbers 이 글을 읽어보세요. 수식이 좀 깨지긴 했지만 이해하는 데는 큰 문제 없을 겁니다.)

3. 모수와 추정량

모평균(기호는 mu라고 읽습니다)과 모분산(기호는 sigma^2로 읽습니다)은 모집단(population)의 정보를 알려주는 값들, 즉 **"모수(parameter)"**입니다. 이건 확실하게 고정된, 정해진 값이 있습니다. (동전의 앞면 나올 확률은 ½로 정해져 있죠.) 그치만 표본평균과 표본분산은 sampling 결과에 따라서 쪼끔 쪼끔씩 달라집니다. (동전을 n번 던졌을 때 무조건 n/2번 짝수가 나오지는 않습니다) 그래서 이런 **"추정량(estimator)"**은 고정되지 않은 값입니다. 물론 우리는 구체적인 개별 sample들을 통해 고정된 **"추정치(estimates)"**를 계산할 수 있습니다. 통계학적으로는 조금 부정확한 설명일 수 있지만 직관적인 이해를 위해 보충 설명하자면, 고정되어 있는 "모수"값을 알기 위해 우리는 고정되지 않은 "추정량"이라는 추상적 개념(ex. 표본평균이라는 추상적인 개념 그 자체)을 통해 통계적 추론을 해야하고, 이를 위해 실험이나 서베이 등을 통해 상수로 표현되는 구체적인 추정량의 값, 즉 추정값(ex. 정확한 표본평균 값)을 계산합니다.



※ 잠깐 부록! : 여기는 앞 내용을 완전히 이해 못하셨다면 굳이 안 읽으셔도 됩니다.

방금 막 앞 페이지에서 고정된 값인 estimates(추정치)와 달리 parameter(모수)는 확실히 고정되어 있다고 이야기했는데요. 이런 식의 설명에 반대하는 통계학의 학파가 있습니다. 여기서 잠깐 통계학의 양대 학파에 대한 설명이 필요한데요, 전통적으로 우리가 배우는 (그리고 앞 슬라이드에서도 그에 맞게 설명한) 관점은 "빈도주의(frequentist)"의 입장입니다. 우리가 대학 다니면서 배우는 99%의 통계학 수업은 다 빈도주의 관점에 기반합니다. 반대로 "베이지언(Bayesian)"들은 모수 역시 고정된 값이 아니라 확률분포를 지니는 움직이는 개념, 즉 확률변수(random variable)라고 주장합니다. 경영학과나 통계학과에서 베이지언을 다루는 일부 수업들이 개설되고는 하니 이쪽에서 찾아서 배우실 수는 있습니다. 여튼 제 강의의 주요 대상인 양방/계량사회과학에서는 아직까지 비주류인 게 사실인데요, 베이지언은 머신러닝/딥러닝 쪽으로 조금만 넘어가면 상당한 비중을 차지하는 관점입니다. 그 차이가 전통적인 계량사회과학/수리통계학과 얼마나 다르게 나타나는지 다음 페이지에서 알아봅시다.

4. P.17~19 증명을 볼 때 궁금해 할 수 있는 것들 정리…

아마 다음 두 질문들을 가장 궁금해하실 것 같습니다.
첫 번째, 아니 그래서 표본분산(s^2)이 모분산(σ^2)이랑 잘 '맞아떨어진다' 는게 왜 $E(s^2)= \sigma^2$ 로 설명이 돼요?
두 번째, 저 식들 전개할 때 X_k랑 $\bar{X}$ 를 정확히 어떻게 취급해야돼요? 아직 잘 이해가 안 됩니다 ㅠㅠ

자. 첫 번째 질문에 대한 답변. 우리가 어떤 **추정량(estimator)**을 평가할 때 중요한 요소는 크게 (1)불편성(unbiasedness)과 (2)일치성(consistency)입니다. 일치성이 뭔지는 지금 알 필요는 전혀 없고 지금은 **불편성**에 대해서만 살펴봅시다. 불편성이라고 하니까 이게 번역투여서 잘 안 와닿습니다. 영어로 하면 unbiasedness, 즉 bias(편향)이 없다는거죠. 그리고 bias(편향)이 없다는건 수학적으로 "estimator 자체가 모수랑 일치할 필요는 없지만 estimator의 '평균'은 모수와 일치해야 한다"를 의미합니다.

E(추정량) = 모수

(영어로 하면…)

E(estimator) = parameter

이제 두 번째 질문에 대한 답변. 29~31페이지에서 증명하는 건 우리가 들고 있는 하나의 sample의 표본 분산 값에 대한 증명입니다. 전체 n개의 데이터가 있는거고요. X_k 는 X라는 데이터의 k번째 값입니다. 그리고 $\bar{X}$(X bar라고 읽습니다)는 우리가 들고 있는 샘플의 n개 데이터들의 (표본)평균입니다. 즉 상수값으로 딱 계산이 떨어지죠. 그래서 시그마(수열의 합)에서 넣고 빼고 할 때 X bar는 상수 취급하면 되고 X_k 는 변하는 값이니까 상수 취급하면 안 됩니다. 딱 이것만 유의하고 보면 됩니다. 그러면 30페이지에 전개된 수식은 다음 페이지에서와 같이 좀 더 구체적으로 풀어서 써볼 수 있습니다.

$$\sum_{k=1}^{n}(X_k-\overline{X})^2=\sum_{k=1}^{n}((X_k-m)+(m-\overline{X}))^2$$

$$=\sum_{k=1}^{n}((X_k-m)^2+2(X_k-m)(m-\overline{X})+(m-\overline{X})^2)$$

$$=\sum_{k=1}^{n}(X_k-m)^2+\sum_{k=1}^{n}(2(X_k-m)(m-\overline{X}))+\sum_{k=1}^{n}(m-\overline{X})^2$$

Let $(m-\overline{X})=c$, (m도 상수고 $\overline{X}$도 상수니까 둘이 사칙연산해도 상수) then,

$$=\sum_{k=1}^{n}(X_k-m)^2+\sum_{k=1}^{n}2c(X_k-m)+\sum_{k=1}^{n}c^2$$

$$=\sum_{k=1}^{n}(X_k-m)^2+2c\sum_{k=1}^{n}X_k-2ncm+nc^2$$

Because $\sum_{k=1}^{n}X_k=n\overline{X}$ ($\because$ 표본평균 = $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$),

$$=\sum_{k=1}^{n}(X_k-m)^2+2cn\overline{X}-2cnm+nc^2=\sum_{k=1}^{n}(X_k-m)^2+2cn(\overline{X}-m)+nc^2$$

$$=\sum_{k=1}^{n}(X_k-m)^2-2c^2n+nc^2=\sum_{k=1}^{n}(X_k-m)^2-nc^2=\sum_{k=1}^{n}(X_k-m)^2-n(m-\overline{X})^2$$

$$=\sum_{k=1}^{n}(X_k-m)^2-n(\overline{X}-m)^2$$

5. 편미분 (다음 수업에 오기 전에 알아야 할 것)

편미분이라고 이름이 뭔가 무시무시해 보이지만 별거 아닙니다. 고등학교 때는 식에 변수가 한 개 밖에 없었지만 한 식에 변수가 여러 개 섞여 있으면 어떻게 미분해야 할까요? 편미분은 그런 상황을 다룹니다. 아래 식의 풀이를 보면 직관적으로 이해가 될겁니다. 풀이의 첫번째 줄은 x1에 대한 편미분을, 두번째 줄은 x2에 대한 편미분을 나타낸 겁니다. 윗줄은 x1을 편미분이니까 관심 대상이 아닌 x2를 그냥 상수 취급해버렸고 두번째줄은 반대로 x1을 상수 취급해버렸습니다. 호옥시 오른쪽 풀이를 보고도 바로 이해가 안 되면 그냥 구글링해보시면 좀 더 친절하고 자세한 설명이 나올겁니다. (근데 보면 아시겠지만 복잡한거 전혀 아닙니다)

예제) 다음을 편미분 하여라.

$$y = f(x_1, x_2) = 3x_1^4 + 2x_1^2x_2^2 + 7x_2^4$$

풀이)

$$\frac{\partial y}{\partial x_1} = 12x_1^3 + 4x_1x_2^2$$

$$\frac{\partial y}{\partial x_2} = 4x_1^2x_2 + 28x_2^3$$

통계입문 2부

2021.06.17.

<회귀분석의 기본 개념 I>

(우선은 쉬운 버전으로 알아보자..)

신입생들의 **첫 학기 GPA**에는 어떤 요인들이 영향을 미칠까?

인간의 키는 어떠한 생물학적, 환경적 요인들에 의해 결정될까?

특정한 개인의 **정치적 성향**은 무엇에 의해 결정될까?

GPA = 월별 마신 술의 양 + 공부 시간 + 대입 만족도

키 = 영양 상태 + 운동량 + 부모의 키 + 재력

정치적 보수성 = 연령 + 소득 수준 + 학력 수준

GPA = -3×월별 마신 술의 양 + 5×공부 시간 + 2×대입 만족도

키 = 4×영양 상태 + 3×운동량 + 1.2×부모의 키 + 5×재력

정치적 보수성 = 5×연령 + (-7)×소득 수준 + (-11)× 학력 수준

GPA = -3×500cc 한 잔+ 5×열람실체류시간 + 2×10점만점설문

키 = 4×일평균칼로리+ 3×조깅시간 + 1.2×부모키 + 5×소득10분위

정치적 보수성 = -5×출생년도 + (-7)×소득10분위+ (-11)×교육년수

회귀모형

(regression model)

종속변수 = 독립변수1 +독립변수2 + … +독립변수n

$$Y_i = \mu + \beta_1 X_i + \beta_2 X_i + ... + \beta_k X_i$$

↑

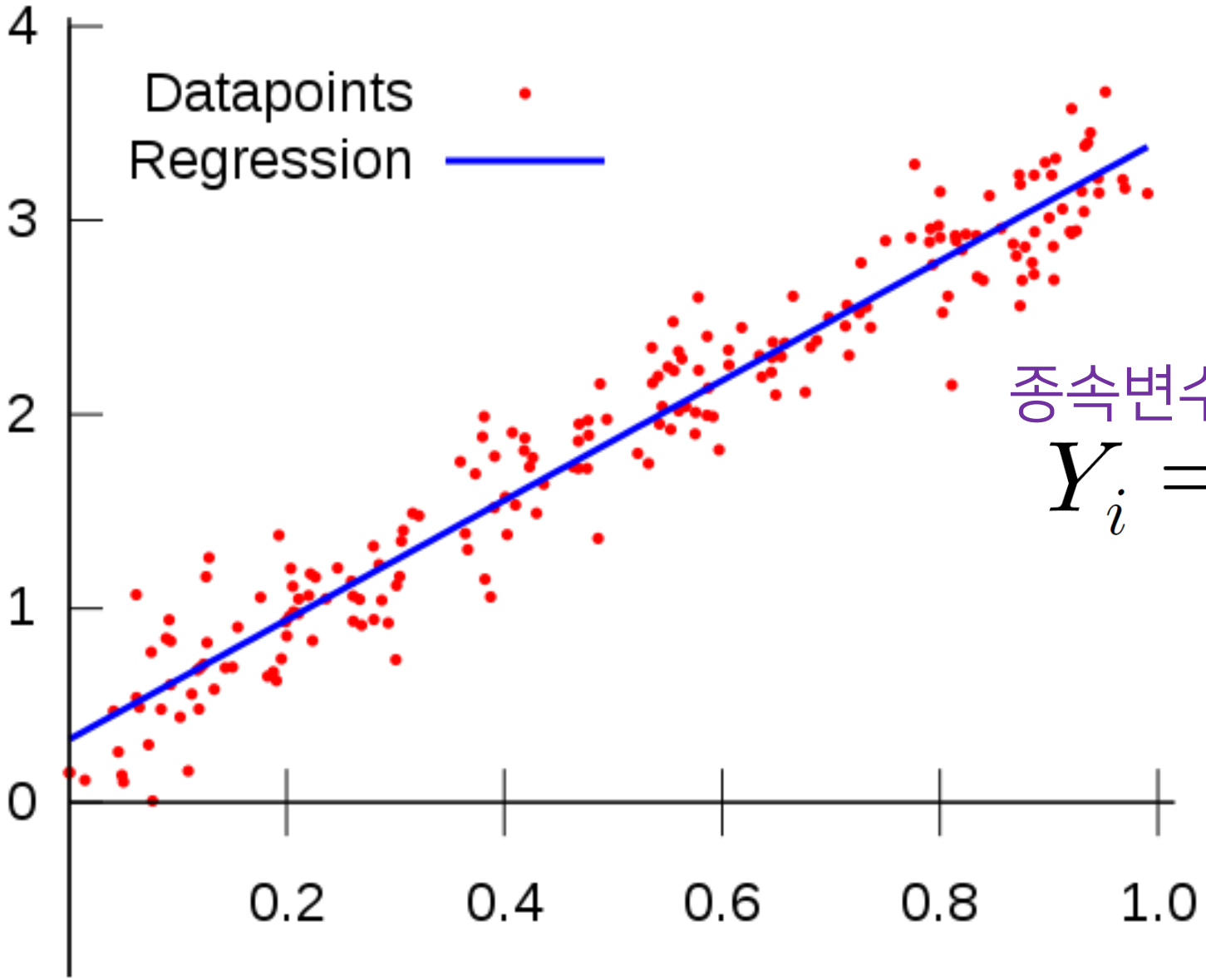
$$Y_i = \widehat{Y}_i + e_i$$

$$Y_i = \hat{\mu} + \widehat{\beta}_1 X_i + \widehat{\beta}_2 X_i + ... + \widehat{\beta}_k X_i + e_i$$

$$Y_i = a + b_1 X_i + b_2 X_i + ... + b_k X_i + e_i$$

종속변수 : 삶의 만족도

독립변수 : 연령, 성별, 교육수준, **소득수준** 등등…



종속변수 기울기 잔차

$$Y_i = a + b_1 X_i + e_i$$

y 절편 독립변수

<평균 개념의 복습>

(with least squares approach)

평균이란

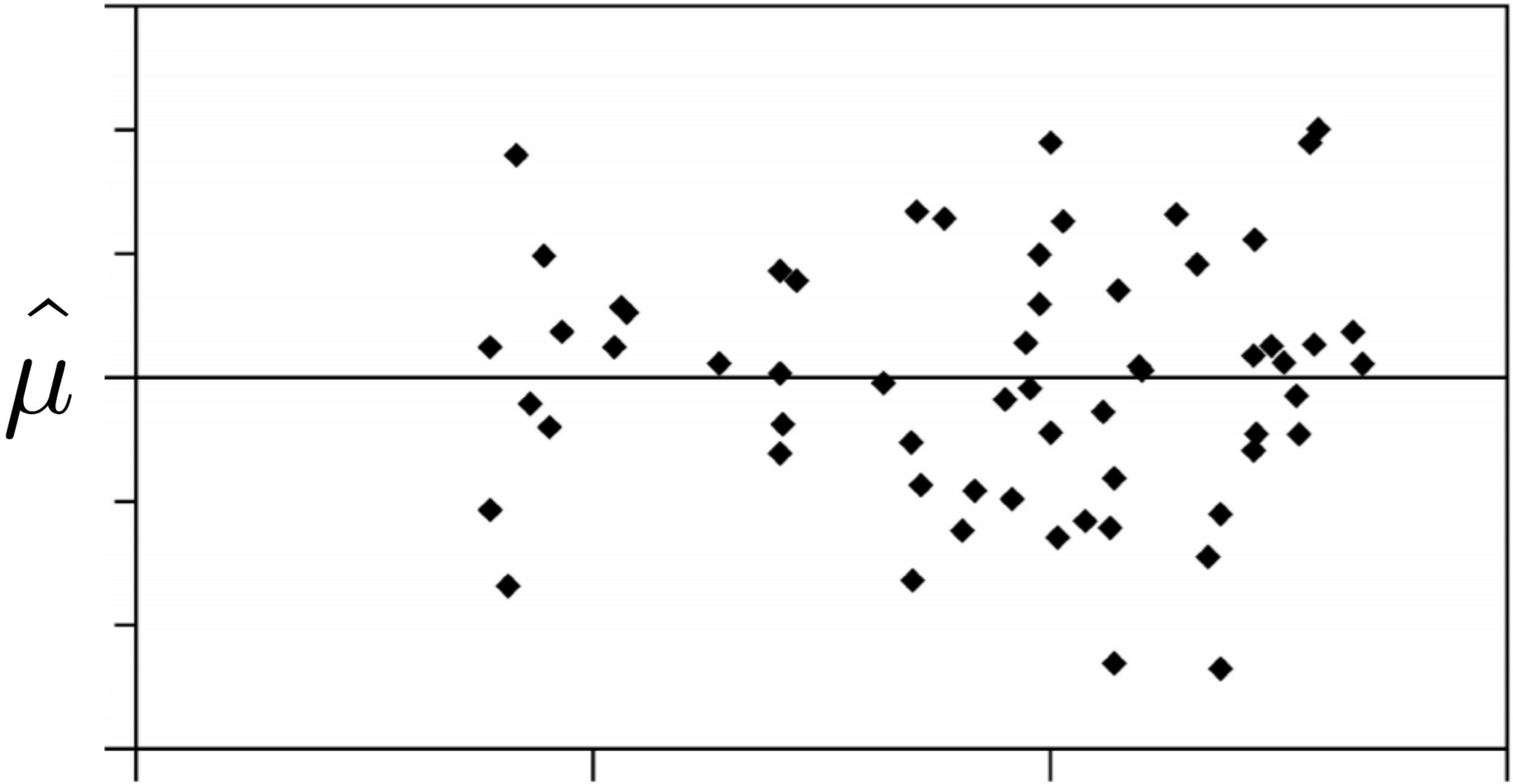
(1) 제곱법에 기초하여

(2) 측정값에 포함되어 있는 차이를

(3) 가장 작게 만듬으로서

특정 데이터의 정보를 가장 잘 보여주는 값!

분산이란 데이터가 퍼져있는 정도!


$\hat{\mu}$

$$Y_i = \hat{\mu} + \hat{e}_i$$

$$\hat{e}_i = Y_i - \hat{\mu}$$

$\hat{\mu}$에 대해서... $\min \sum_{i=1}^{n} \hat{e}_i^2 = \min \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$

$$\frac{\partial \sum_{i=1}^{n} \hat{e}_i^2}{\partial \hat{\mu}} = \sum_{i=1}^{n} 2(Y_i - \hat{\mu})(-1) = 0$$

$$\sum_{i=1}^{n} (Y_i - \hat{\mu}) = \sum_{i=1}^{n} Y_i - n\hat{\mu} = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

<회귀분석의 기본 개념 II>

(이번엔 analytic하게 알아보자..)

$$Y_i = a + bX_i + e_i$$

a와 b에 대해서 $\min \sum_{i=1}^{n} \widehat{e_i^2} = \min \sum_{i=1}^{n} (Y_i - a - bX_i)^2$

$$\frac{\partial \sum_{i=1}^{n} \widehat{e_i^2}}{\partial a} = \sum_{i=1}^{n} 2(Y_i - a - bX_i)(-1) = 0$$

$$\sum_{i=1}^{n} (Y_i - a - bX_i) = 0$$

$$\sum_{i=1}^{n} Y_i - na - b\sum_{i=1}^{n} X_i = 0$$

$$a = \frac{1}{n}\sum_{i=1}^{n} Y_i - b\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{Y} - b\overline{X}$$

$$\frac{\partial \sum_{i=1}^{n} \widehat{e}_i^2}{\partial b} = 2\sum_{i=1}^{n}(Y_i - a - bX_i)(-X_i) = 0$$

$$\sum_{i=1}^{n}(Y_iX_i - aX_i - bX_i^2) = \sum_{i=1}^{n}\left\{Y_iX_i - (\overline{Y} - b\overline{X})X_i - bX_i^2)\right\} = 0$$

$$\sum_{i=1}^{n}\left\{Y_iX_i - \overline{Y}X_i + b\overline{X}X_i - bX_i^2)\right\} = 0$$

$$\sum_{i=1}^{n}\left\{(Y_i - \overline{Y})X_i + b(\overline{X}X_i - X_i^2)\right\} = 0$$

$$b\sum_{i=1}^{n}(X_i^2 - \overline{X}X_i) = \sum_{i=1}^{n}\left\{X_i(Y_i - \overline{Y})\right\}$$

$$\text{LHS}: b\sum_{i=1}^{n}(X_i^2-\overline{X}X_i)=b\left\{\sum_{i=1}^{n}(X_i^2-\overline{X}^2)\right\}=b\left\{\sum_{i=1}^{n}X_i^2-n\overline{X^2}\right\}$$

$$=b\left\{\sum_{i=1}^{n}(X_i^2-2\overline{X}X_i+\overline{X}^2)\right\}=b\left\{\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}$$

$$\text{RHS}: \sum_{i=1}^{n}X_iY_i-n\overline{X}\,\overline{Y}=\sum_{i=1}^{n}(X_iY_i-\overline{Y}X_i-\overline{X}Y_i+\overline{X}\,\overline{Y})$$

$$=\sum_{i=1}^{n}\left\{(X_i-\overline{X})(Y_i-\overline{Y})\right\}$$

쭈욱 정리하면..

$$b = \frac{\sum_{i=1}^{n} \left\{ (X_i - \overline{X})(Y_i - \overline{Y}) \right\}}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

<회귀분석의 기본 개념 III>

(이번엔 matrix algebra로 알아보자..)

A^T	transposed matrix of A
A^{-1}	inverse matrix of A
$AA^{-1} = I$	product of A and its inverse is identity matrix
$A_{n \times n}$	square matrix
$I_{n \times n}$	identity matrix
$A^T A = A$	idempotent matrix
$A \times A \times \ldots \times A = A$	
$(A+B)^T = A^T + B^T$	
$(AB)^{-1} = B^{-1}A^{-1}$	잠깐! A와 B의 inverse가 존재하지 않는다면 무효!
$(AB)^T = B^T A^T$	

$$y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \epsilon_{n\times 1}$$

$$SSE(b) = \sum_{i=1}^{n} e_i^2 = e^T e = (y - Xb)^T (y - Xb) = (y^T - b^T X^T)(y - Xb)$$

$$= y^T y - y^T Xb - b^T X^T y + b^T X^T Xb = y^T y - 2y^T Xb + b^T X^T Xb$$

$$\min SSE(b)\ w.r.t.\ b = \frac{\partial SSE(b)}{\partial b} = -2X^T y + 2X^T Xb = 0$$

$$2X^TXb = 2X^Ty$$

$$X^TXb = X^Ty$$

$$b = (X^TX)^{-1}X^Ty$$

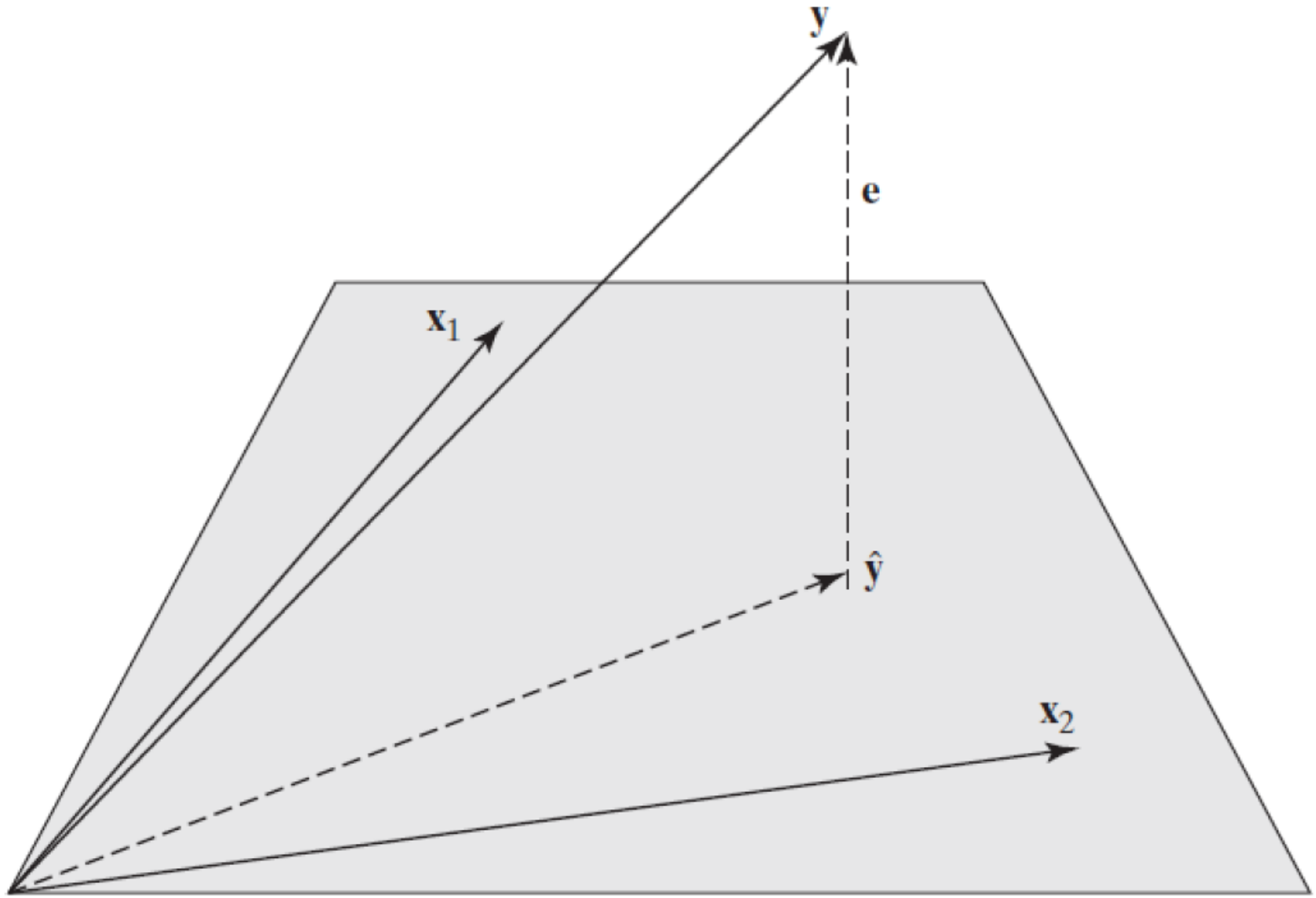
$$\hat{y} = Xb = X(X^TX)^{-1}X^Ty = Py$$

$$e = y - Xb = y - My = (I - P)y = My$$

P = projection matrix

M = residual maker matrix

$$y = \hat{y} + e = Py + My$$



< Projection of y into the column space of X >

p-value VS coefficient size

Logistic regression

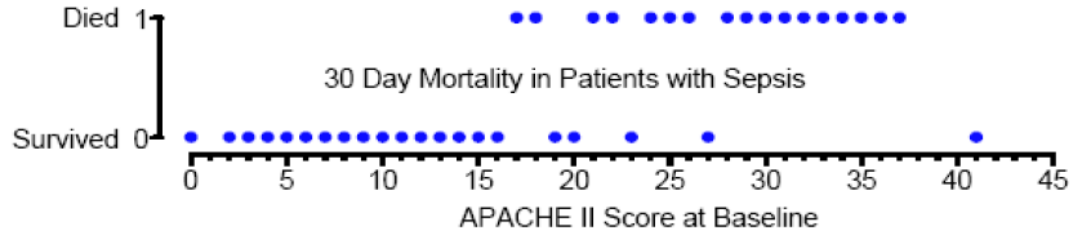
Definition

- Logistic regression is a technique for relating a binary variable to explanatory variables. The explanatory variables may be categorical, continuous, or both.
- We will look at the logistic regression model with one predictor variable:

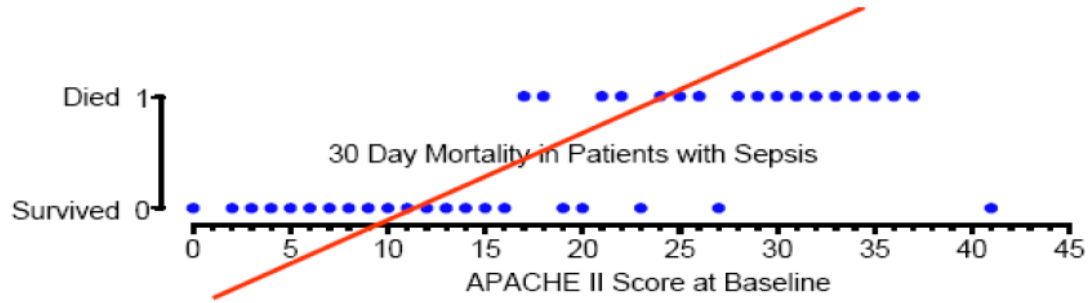
  Y: binary response variable (Y=1 success, Y=0 fail)

  X: quantitative explanatory variable
- We want to model $\pi(x) = P(Y = 1 \mid X = x)$. This is the probability of a success when $X = x$
- The logistic regression model has a linear form for logarithm of the odds, or logit function,

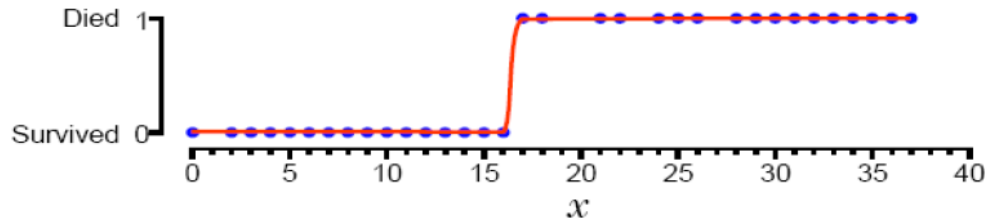
$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x$$



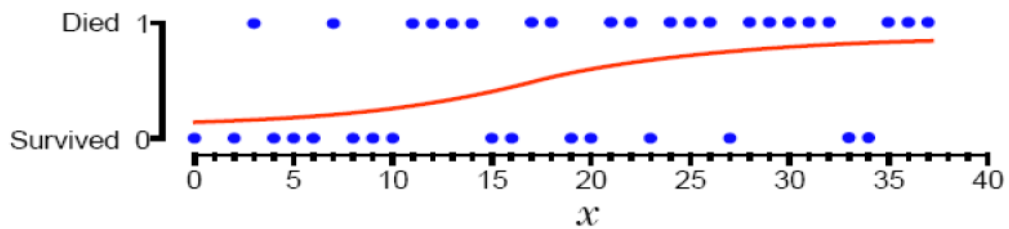
Note that linear regression would not work well here since it could produce probabilities less than zero or bigger than one.



Data have sharp survival cut off points between patients who live or die should have a large value of β.

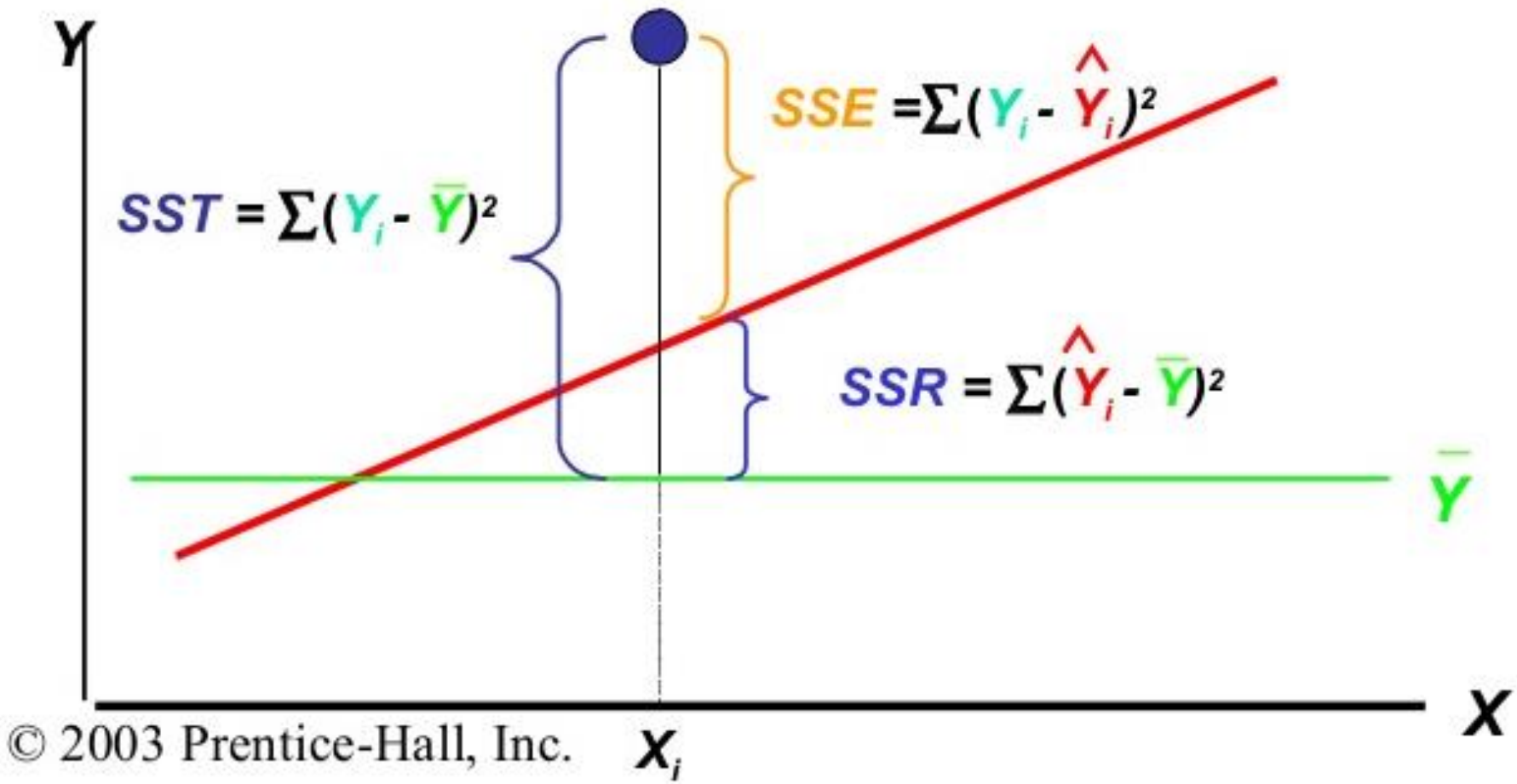


Data with lengthy transition from survival to death should have a low value of β.



Model Selection

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$



$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$