Probability Distributions

2016. 06. 22.
정태승

Contents

Probability distribution

확률분포

위키백과, 우리 모두의 백과사전.

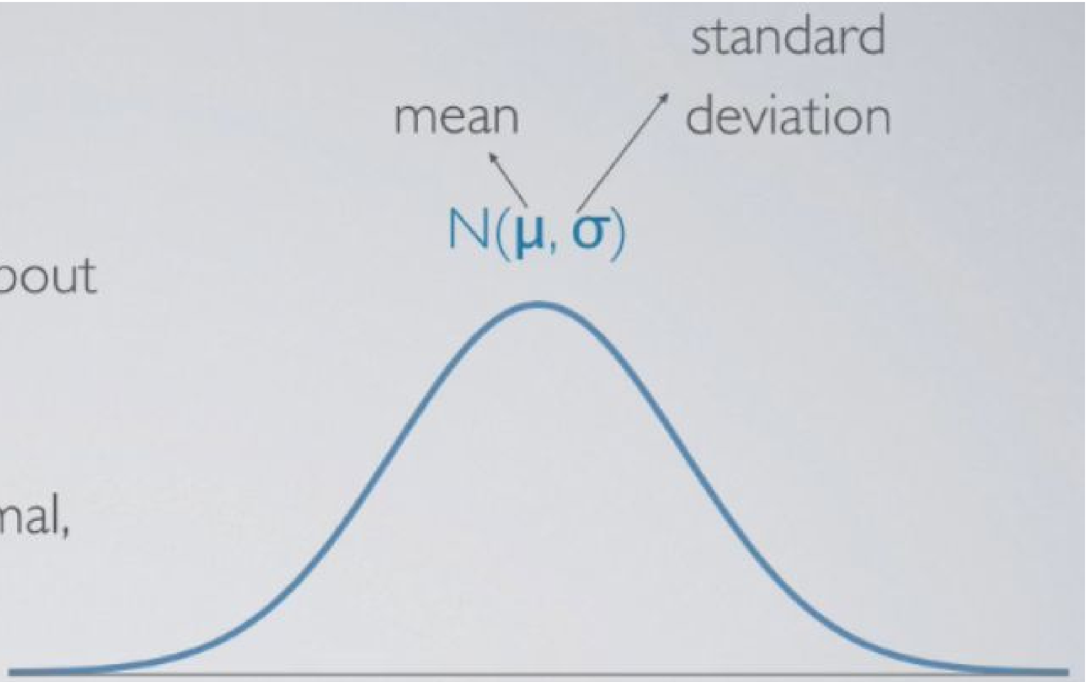
확률분포(確率分布)는 확률변수가 특정한 값을 가질 확률을 나타내는 함수를 의미한다. 예를 들어, 주사위를 던졌을 때 나오는 눈에 대한 확률변수가 있을 때, 그 변수의 확률분포는 이산균등분포가 된다.

확률분포는 확률변수가 어떤 종류의 값을 가지는가에 따라서 크게 이산 확률분포와 연속 확률분포 중 하나에 속하며, 둘 중 어디에도 속하지 않는 경우도 존재한다.

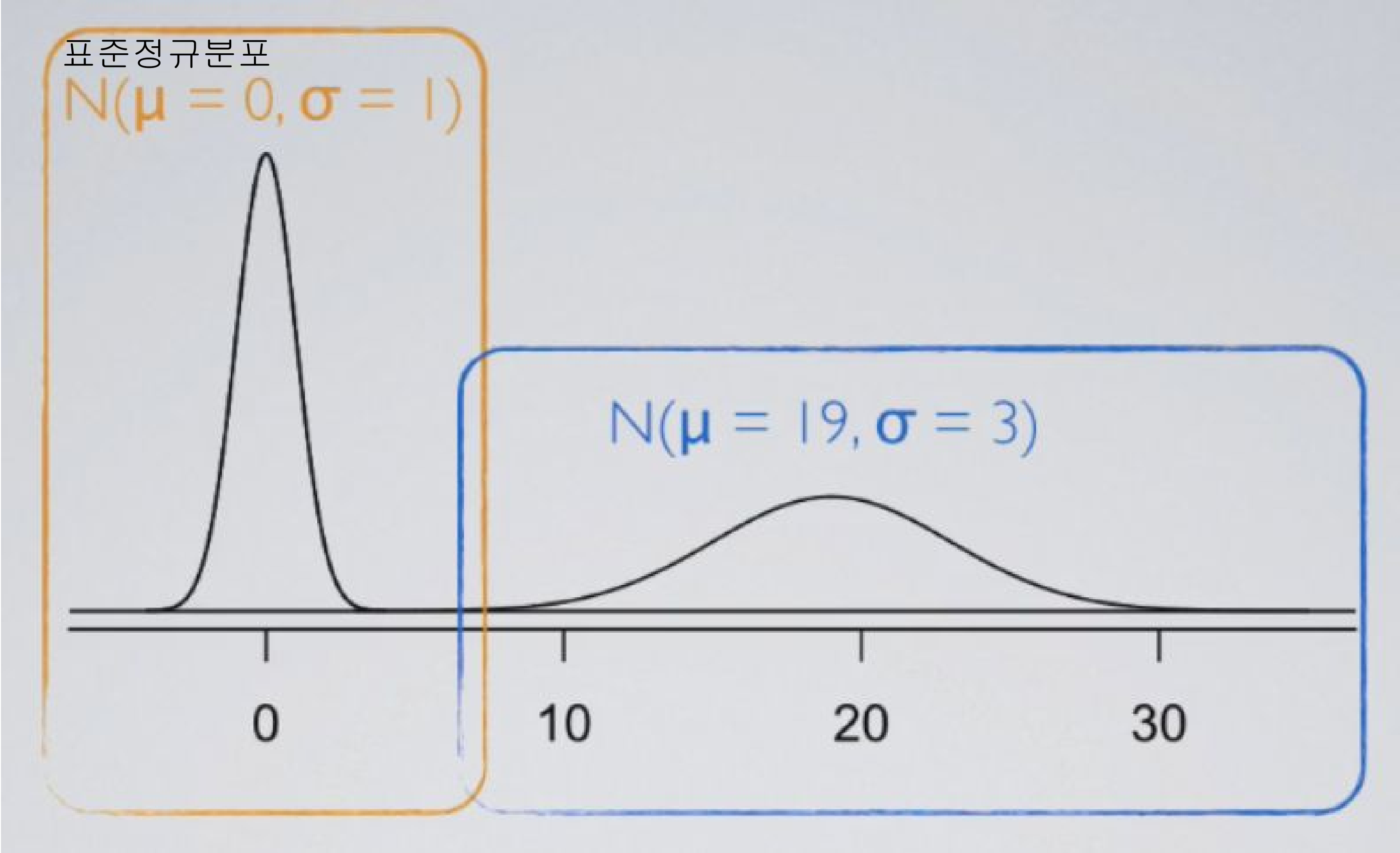
The Normal Distribution

The Normal Distribution

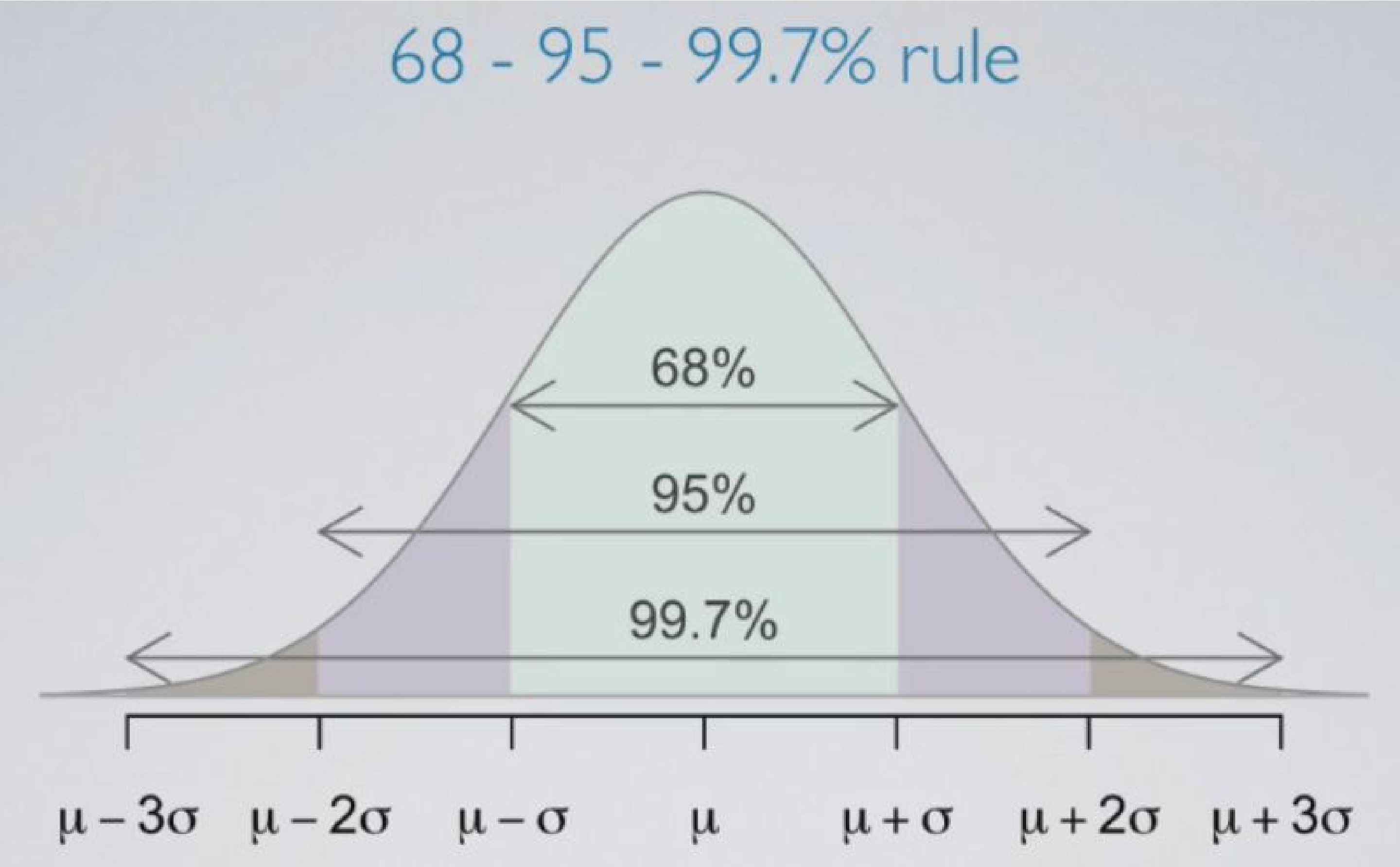
- unimodal and symmetric
  - bell curve
- follows very strict guidelines about how variably the data are distributed around the mean
- many variables are nearly normal, but none are exactly normal



The Normal Distribution



The Normal Distribution



The Normal Distribution

- standardized (Z) score of an observation is the number of standard deviations it falls above or below the mean
- Z score of mean = 0
- unusual observation: $|Z| > 2$
- defined for distributions of any shape

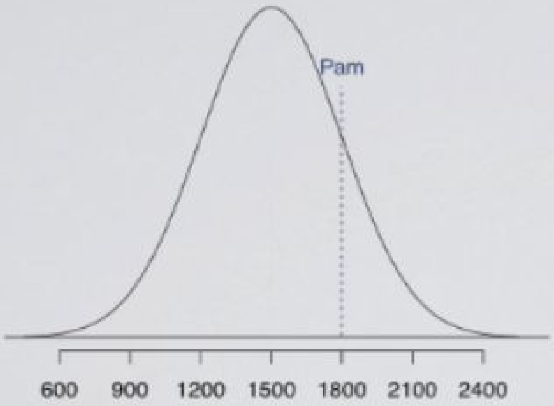
$$Z = \frac{\text{observation - mean}}{\text{SD}}$$

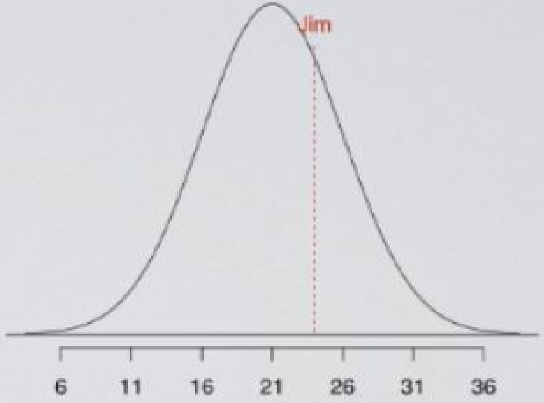
The Normal Distribution

A college admissions officer wants to determine which of the two applicants scored better on their standardized test with respect to the other test takers: Pam, who earned an 1800 on her SAT, or Jim, who scored a 24 on his ACT?

SAT scores ~ N(mean = 1500, SD = 300)

ACT scores ~ N(mean = 21, SD = 5)





The Normal Distribution

A college admissions officer wants to determine which of the two applicants scored better on their standardized test with respect to the other test takers: Pam, who earned an 1800 on her SAT, or Jim, who scored a 24 on his ACT?

SAT scores ~ N(mean = 1500, SD = 300)

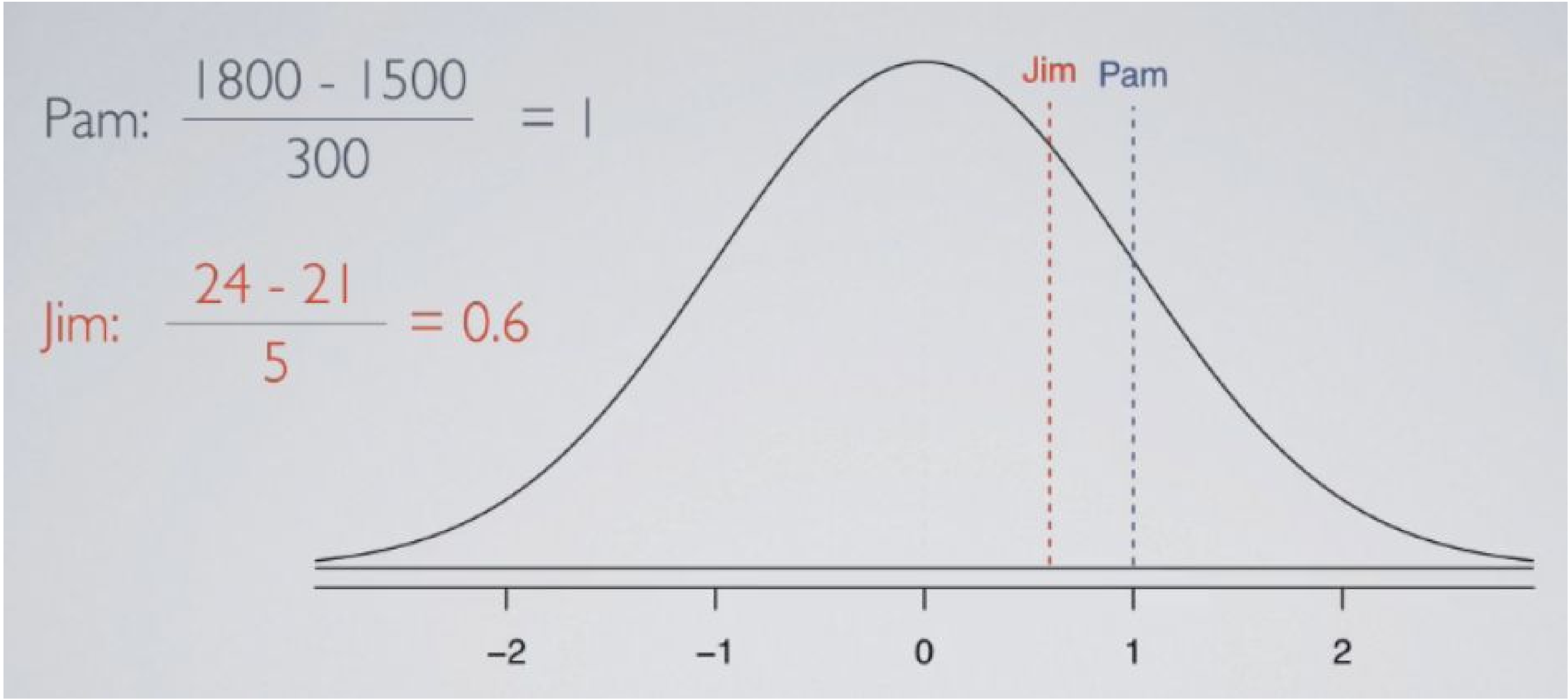
ACT scores ~ N(mean = 21, SD = 5)

$$Z = \frac{\text{observation - mean}}{\text{SD}}$$

Pam: $\frac{1800 - 1500}{300} = 1$

Jim: $\frac{24 - 21}{5} = 0.6$

The Normal Distribution

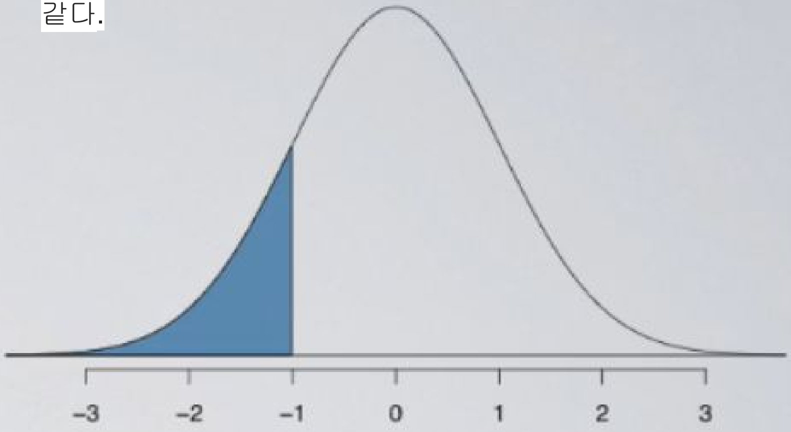


The Normal Distribution

percentiles

- when the distribution is normal, Z scores can be used to calculate percentiles
- percentile is the percentage of observations that fall below a given data point
- graphically, percentile is the area below the probability distribution curve to the left of that observation.

백분위수(Percentile. **百分位數**)는 크기가 있는 값들로 이뤄진 자료를 순서대로 나열했을 때 백분율로 나타낸 특정 위치의 값을 이르는 용어이다. 일반적으로 크기가 작은 것부터 나열하여 가장 작은 것을 0, 가장 큰 것을 100으로 한다. 100개의 값을 가진 어떤 자료의 20 백분위수는 그 자료의 값들 중 20번째로 작은 값을 뜻한다. 50 백분위수는 중앙값과 같다.



The Normal Distribution

computing percentiles

Second decimal place of Z					
0.04	0.03	0.02	0.01	0.00	Z
0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0618	0.0630	0.0643	0.0655	0.0668	−1.5
0.0749	0.0764	0.0778	0.0793	0.0808	−1.4
0.0901	0.0918	0.0934	0.0951	0.0968	−1.3
0.1075	0.1093	0.1112	0.1131	0.1151	−1.2
0.1271	0.1292	0.1314	0.1335	0.1357	−1.1
0.1492	0.1515	0.1539	0.1562	0.1587	−1.0

The Normal Distribution

SAT scores are distributed normally with mean 1500 and SD 300. Pam earned an 1800 on her SAT. What is Pam's percentile score?

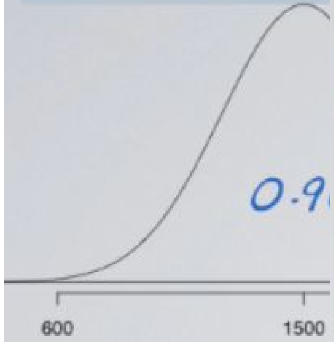
$$Z = \frac{1800 - 1500}{300} = 1$$

$$P(Z < 1) = 0.8413$$

	Second decimal place of Z			
Z	0.00	0.01	0.02	
0.0	0.5000	0.5040	0.5080	0.
0.1	0.5398	0.5438	0.5478	0.
0.2	0.5793	0.5832	0.5871	0.
0.8	0.7881	0.7910	0.7939	0.
0.9	0.8159	0.8186	0.8212	0.
1.0	0.8413	0.8438	0.8461	0.
1.1	0.8643	0.8665	0.8686	0.

The Normal Distribution

A friend of yours tells you that she
scored in the t
What is the lo
she could have

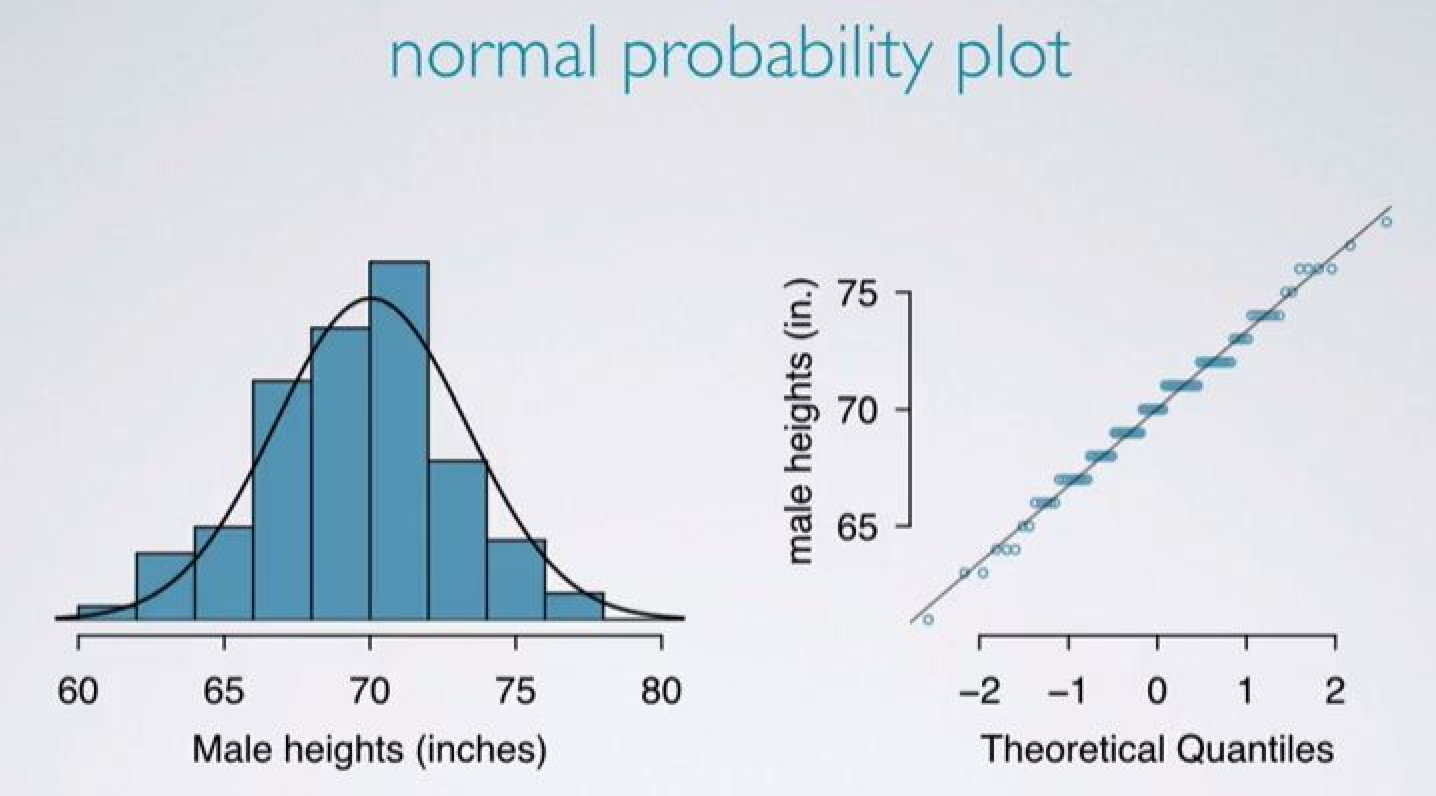


Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
	Second decimal place of Z									
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0				.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
				.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
				.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
				.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
				.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

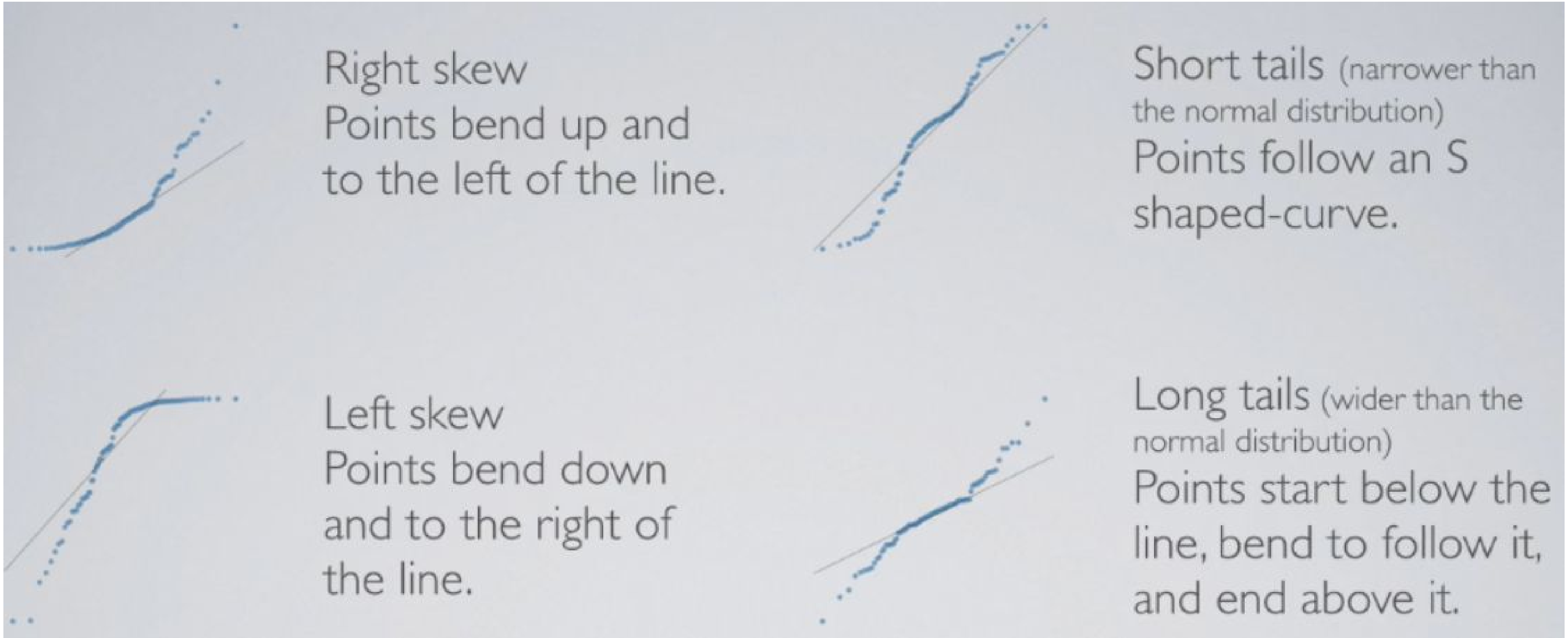
$$Z = 1.28 = \frac{X - 1500}{300}$$

$$X = (1.28 \times 300) + 1500 = 1884$$

The Normal Distribution

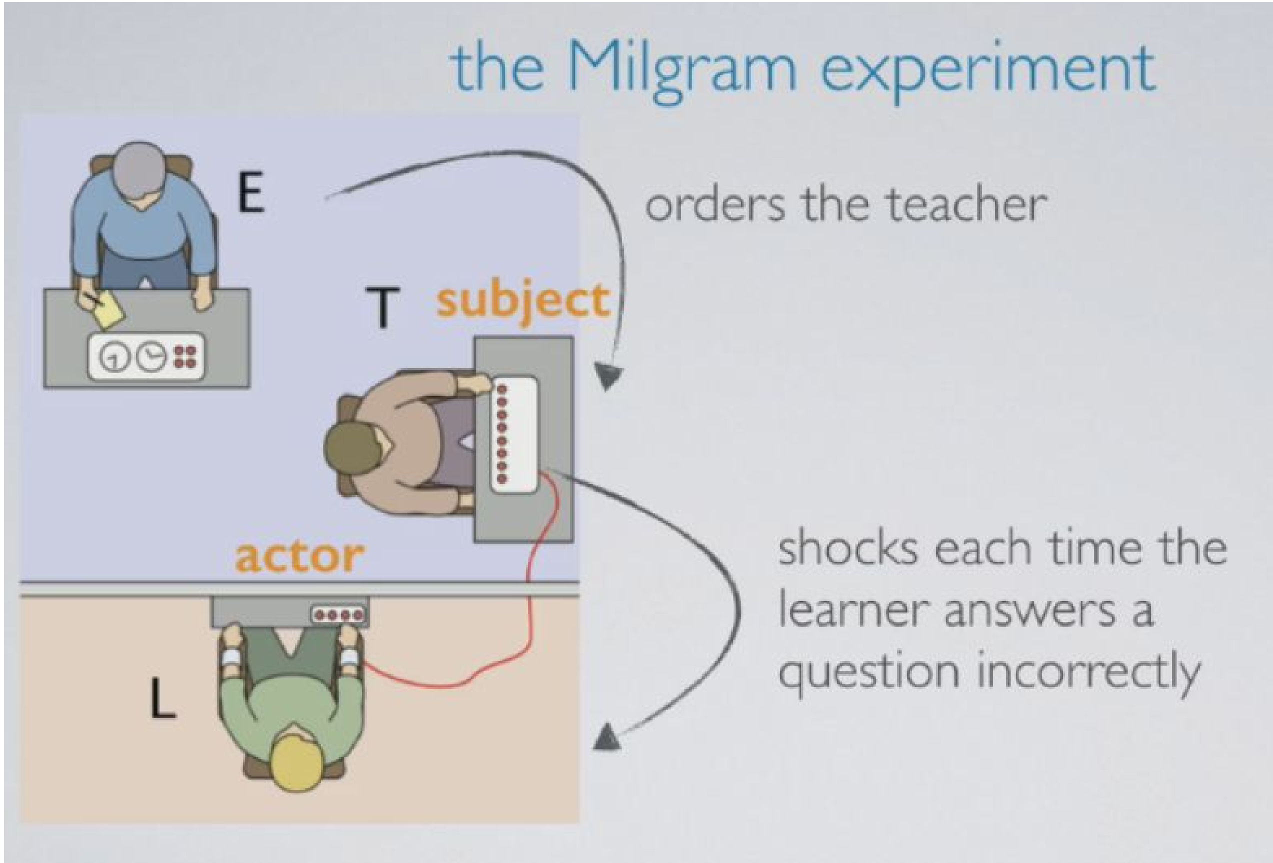


The Normal Distribution



Binomial Distribution

Binomial Disribution



Binomial Disribution

Bernouilli random variables

- each person in Milgram's experiment can be thought of as a trial
- a person is labeled a success if she refuses to administer a severe shock, and failure if she administers such shock
- since only 35% of people refused to administer a shock, probability of success is $p = 0.35$.
- when an individual trial has only two possible outcomes, it is called a Bernoulli random variable

Binomial Disribution

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Scenario 1: $\underset{\text{(A) refuse}}{\underline{0.35}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) refuse}}{\underline{0.35}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 3: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) refuse}}{\underline{0.35}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 4: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) refuse}}{\underline{0.35}} = 0.0961$

$$\mathbf{4 \times 0.0961 = 0.3844}$$

Binomial Disribution

binomial distribution

the binomial distribution describes the probability of having exactly k successes in n independent Bernouilli trials with probability of success p

# of scenarios x P(single scenario)

choose function

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

"n choose k"

$$p^k(1-p)^{(n-k)}$$

probability of success to the power of number of successes

probability of failure to the power of number of failures

Binomial Disribution

Binomial distribution:

If p represents probability of success, $(1-p)$ represents probability of failure, n represents number of independent trials, and k represents number of successes

$$P(\text{k successes in n trials}) = \binom{n}{k} p^k \, (1-p)^{(n-k)}$$

where $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Binomial Disribution

binomial conditions

1. the trials must be independent
2. the number of trials, n, must be fixed
3. each trial outcome must be classified as a success or a failure
4. the probability of success, p, must be the same for each trial

Binomial Disribution

According to a 2013 Gallup poll, worldwide only 13% of employees are engaged at work (psychologically committed to their jobs and likely to be making positive contributions to their organizations). Among a random sample of 10 employees, what is the probability that 8 of them are engaged at work?

$n = 10$

$p = 0.13$

$1 - p = 0.87$

$k = 8$

$$P(K = 8) = \binom{10}{8} 0.13^8 \times 0.87^2$$

$$= \frac{10!}{8! \times 2!} \times 0.13^8 \times 0.87^2$$

$$= \frac{10 \times 9 \times \cancel{8!}}{\cancel{8!} \times 2 \times 1} \times 0.13^8 \times 0.87^2$$

$$= 45 \times 0.13^8 \times 0.87^2$$

$$= 0.0000278$$

Binomial Disribution

Among a random sample of 100 employees, how many would you expect to be engaged at work? Remember: $p = 0.13$.

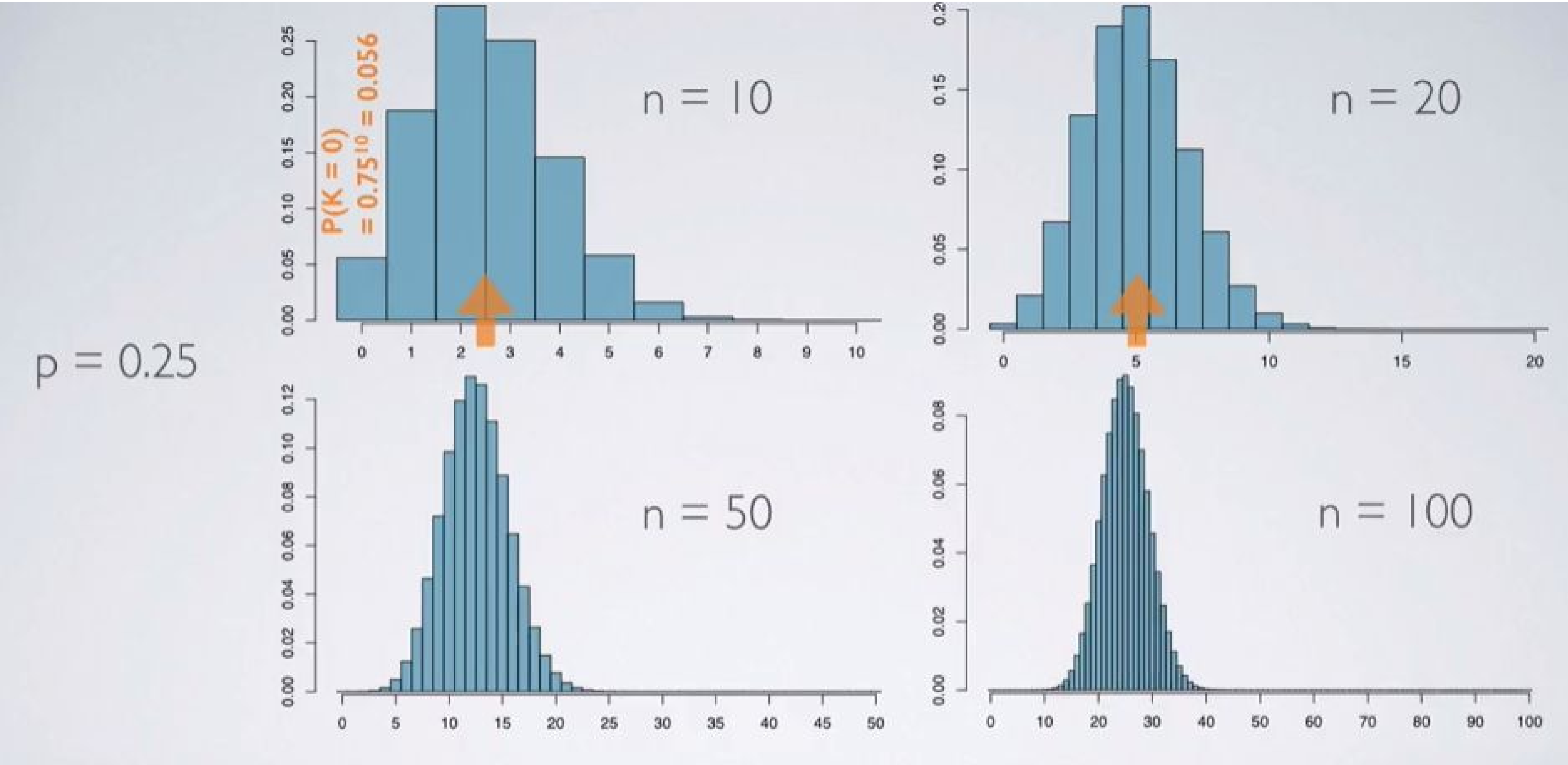
$\mu = 100 \times 0.13 = 13$

Expected value (mean) of binomial distribution: $\mu = np$

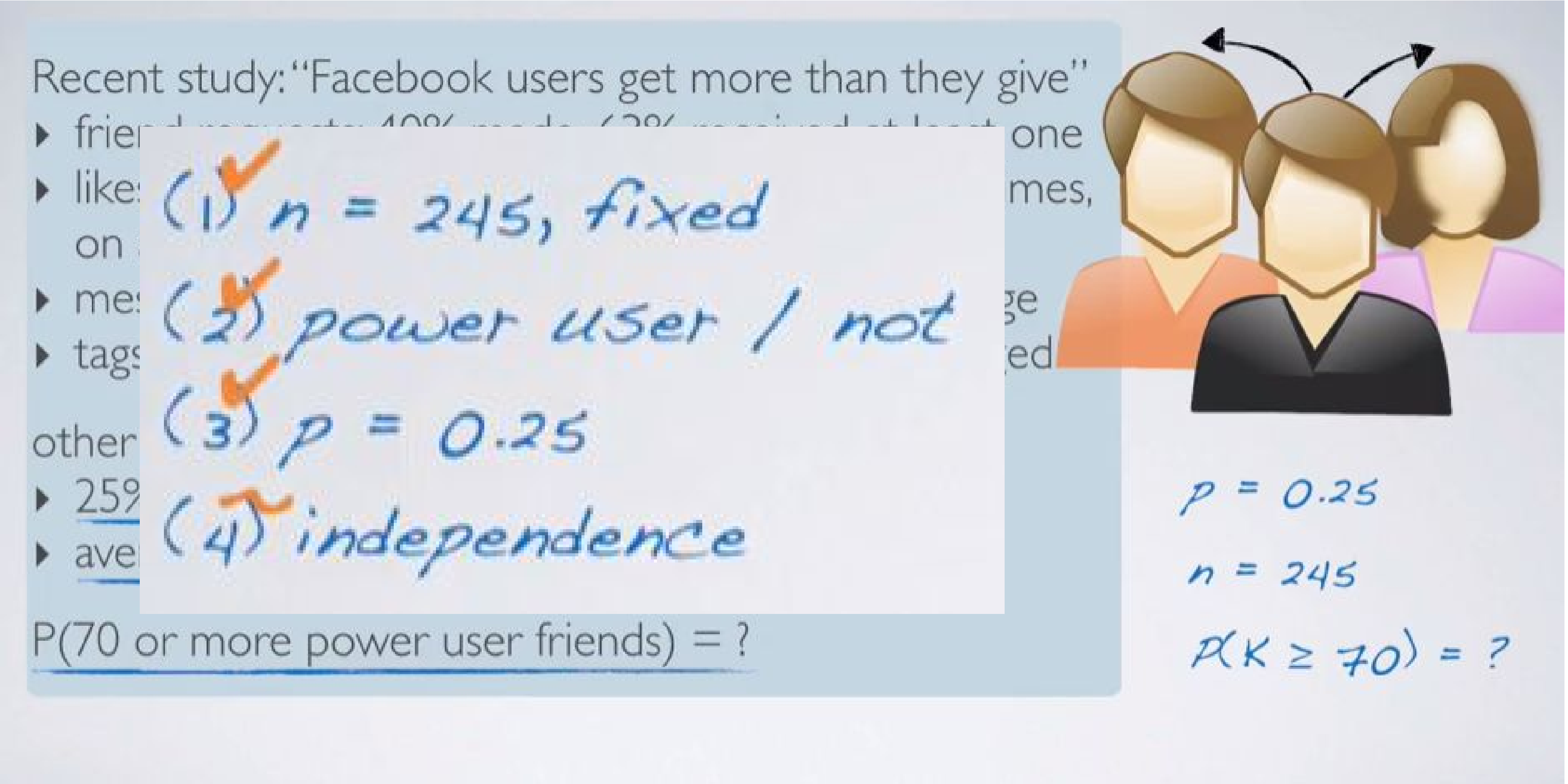
Standard deviation of binomial distribution: $\sigma = \sqrt{np(1-p)}$

$\sigma = \sqrt{100 \times 0.13 \times 0.87} = 3.36$

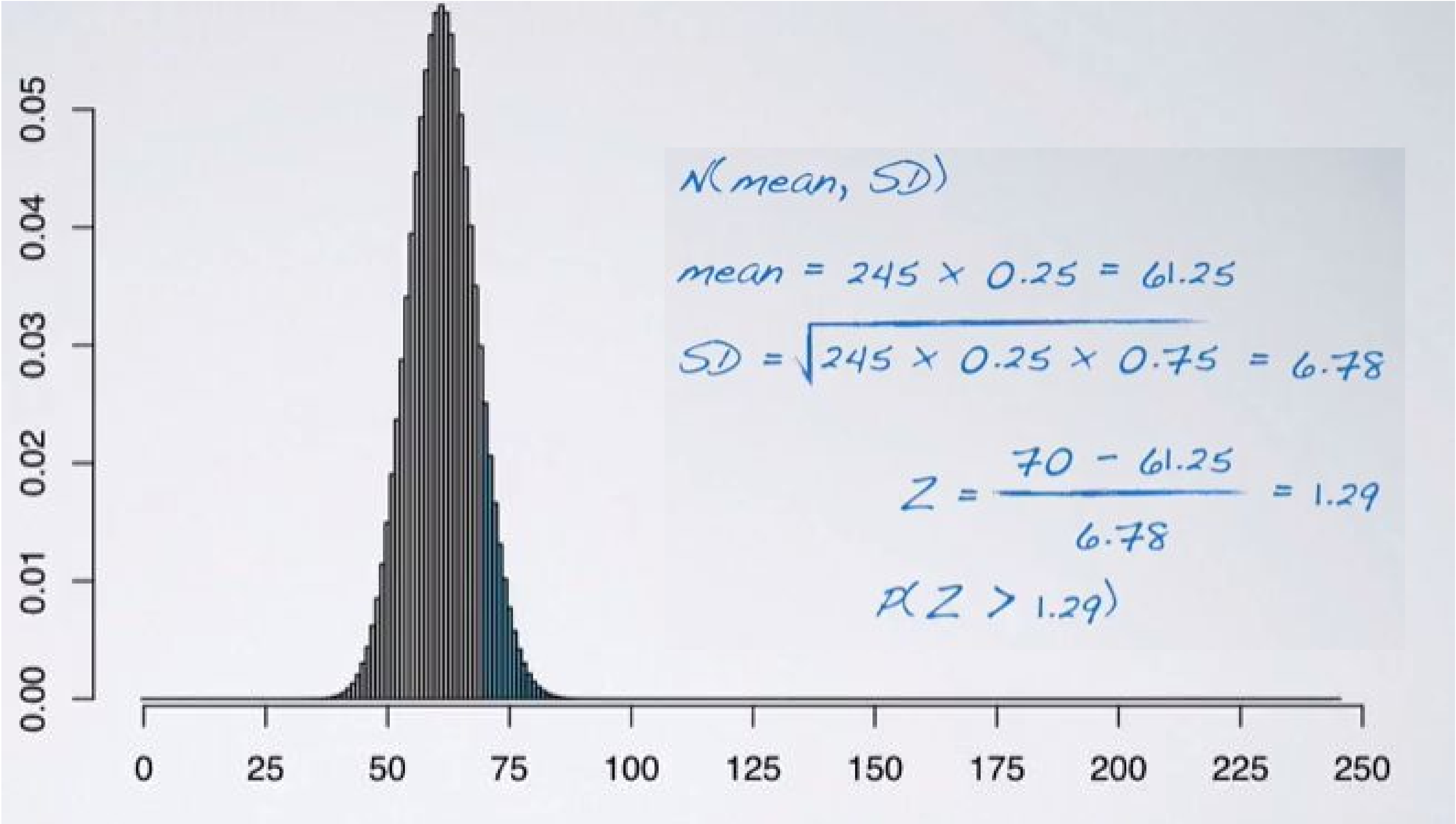
Binomial Disribution



Binomial Disribution



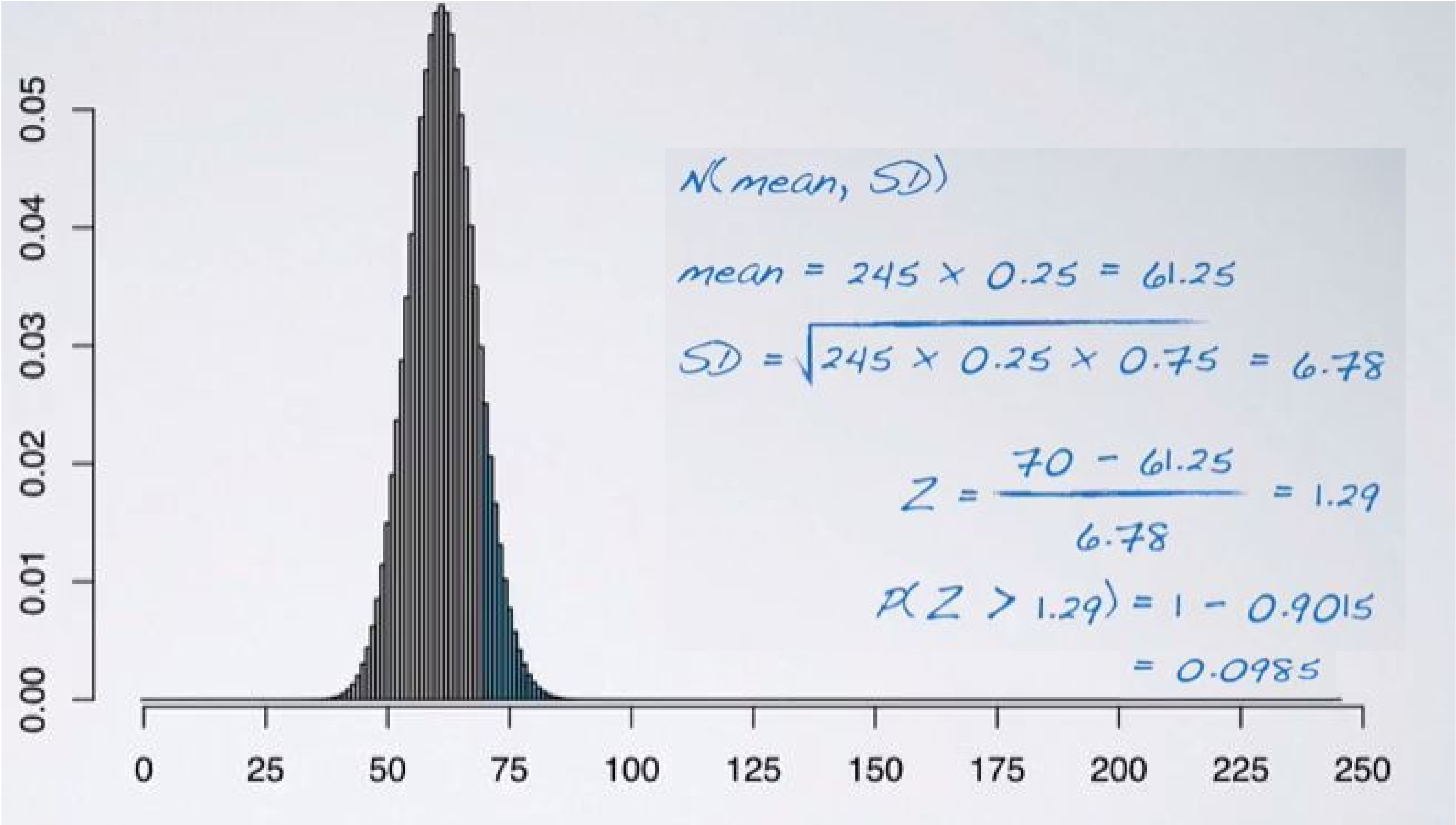
Binomial Disribution



Binomial Disribution

Z	Second decimal place of Z									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

Binomial Disribution



Binomial Disribution

Success-failure rule: A binomial distribution with at least 10 expected successes and 10 expected failures closely follows a normal distribution.

$$np \geq 10$$
$$n(1-p) \geq 10$$

Normal approximation to the binomial: If the success-failure condition holds,

$$\text{Binomial}(n,p) \sim \text{Normal}(\mu,\sigma)$$

where $\mu = np$ and $\sigma = \sqrt{np(1-p)}$

Thank you