



정규분포 및 표준정규 분포

대표적인 연속형 확률분포

▶ 연속형 확률분포

▶ 정규분포(normal distribution)

- ▶ 대표본 모집단의 합, 평균 추론 시 활용

▶ t-분포

- ▶ 소표본 모집단 평균 추론 시 활용
- ▶ 선형모형 회귀계수 추론(종속변수 정규분포 가정)시 활용

▶ 카이제곱(χ^2)분포

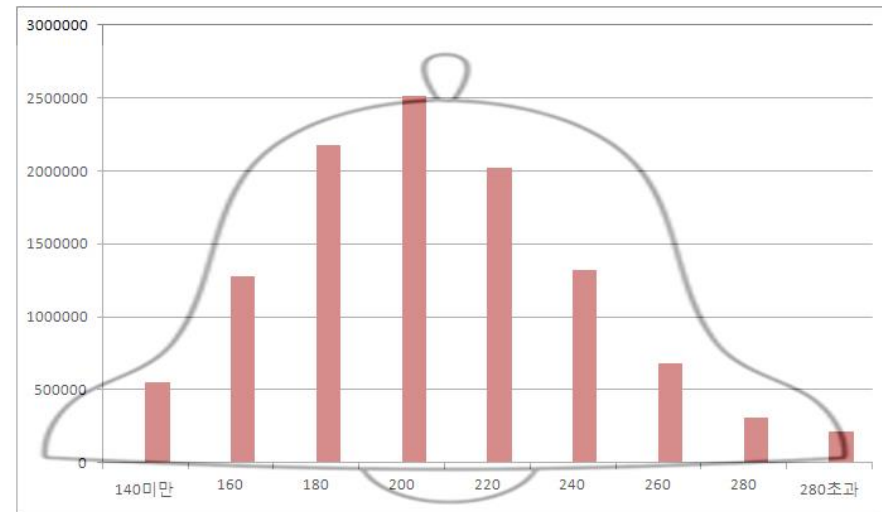
- ▶ 모집단 분산 추론 시(데이터 정규분포 가정) 활용
- ▶ 카이제곱 검정 시 활용

▶ F-분포

- ▶ 두 모집단 분산 차이 비교 시 활용(데이터 정규분포 가정)
- ▶ 분산분석 시 활용
- ▶ (설명하는 변동/설명하지 못하는 변동)이 F분포를 따름

정규분포

- ▶ 가장 대표적인 분포
- ▶ 가우지안(Gaussian) 분포라고도 불리움
- ▶ 연속형 확률변수에 대한 분포
- ▶ 평균(μ)과 표준편차(σ)로 모양이 결정
 - ▶ 좌우대칭, 종모양
 - ▶ 중앙 한 점이 뽕족
 - ▶ 평균 = 중앙값 = 최빈값
- ▶ 자연계, 사회현상에 많이 나타남

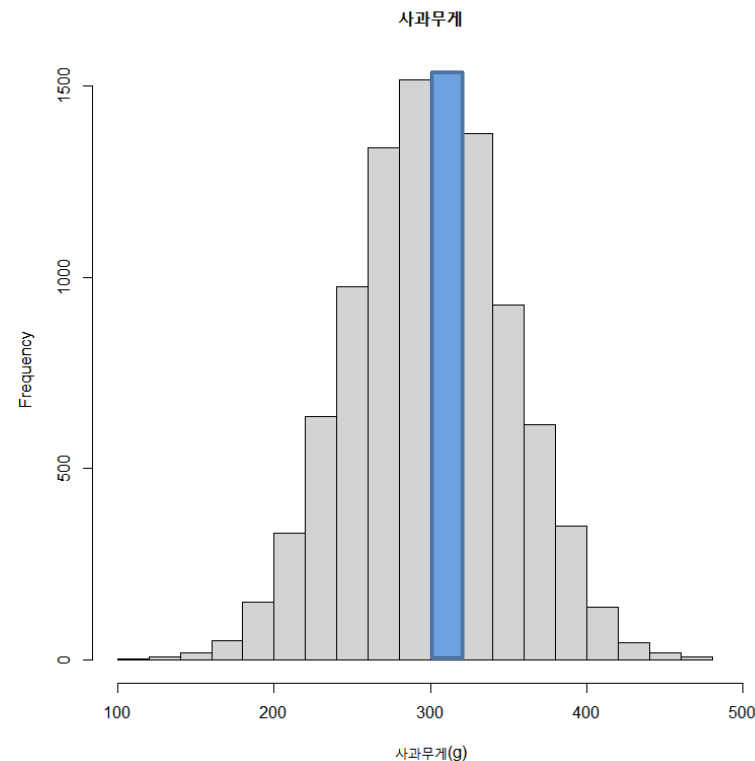


총콜레스테롤 분포 (출처 : 국민건강보험공단, 2011년)

확률밀도함수

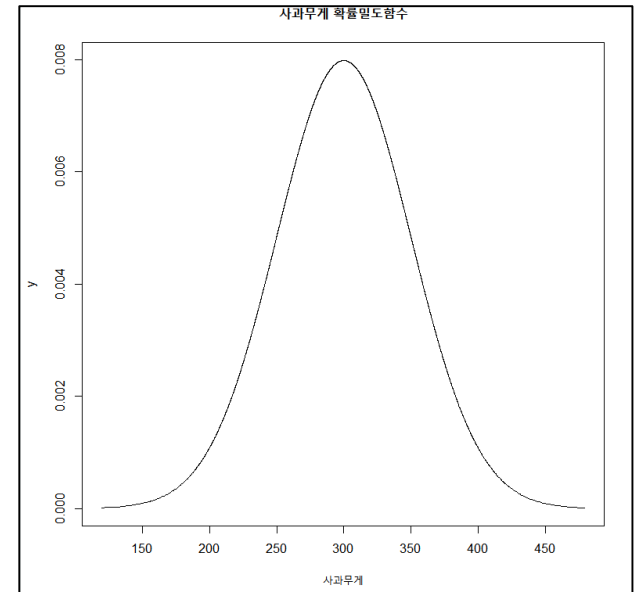
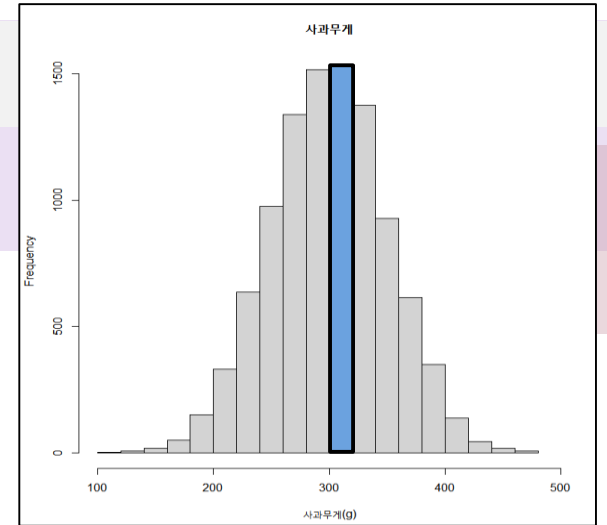
▶ 정규분포

- ▶ 평균으로부터 멀어질수록 출현 빈도수가 급격히 작아지는 형태
- ▶ A과수원에서 수확한 사과무게 히스토그램
 - ▶ 과수원에서 수확한 사과의 무게 평균은 300g
 - ▶ 사과의 무게를 20g 간격으로 나눠 빈도수를 표현한 히스토그램
 - ▶ 사과는 총 100000개이고 무게는 100g~500g 사이
- ▶ A농장에서 수확한 사과들은 큰 바구니에 모두 넣고 임의로 하나 꺼낸다고 할 때
 - ▶ 사과의 무게가 0~600g 사이일 확률
 - ▶ 사과의 무게가 300~320g일 확률
 - ▶ 히스토그램상 300~320은 1500개 이므로
 - ▶ 15% $[1500/10000]$ 가 됨
 - ▶ 히스토그램 총 면적을 100%라고 할 때 총 면적대비 해당구간 봉들의 면적 비율이 전체구간에서 해당구간이 선택될 확률
- ▶ 특정구간이 선택될 확률 = $\text{해당 봉 면적} / \text{전체봉면적} = \text{해당봉면적} / 1 = \text{해당봉면적}$



확률밀도함수

- ▶ 무게가 203.81g ~ 302.9g 사이에 해당하는 사과가 선택 될 확률
 - ▶ 봉 사이에 구간이 존재해 봉의 면적만으로는 구할 수 없음
 - ▶ 사과의 무게는 300g, 300.00198g, 300.9938374648482g 처럼 구간 안의 모든 값을 취할 수 있음
 - ▶ 이런 변수의 경우 확률적 분포는 구간으로 끊는것 보다 연속된 실선으로 표현하는 것이 더 자연스러움
- ▶ 즉, 사과무게에 대한 히스토그램의 총 면적을 1로하고 무게 분포를 실선으로 표현한 그래프로 연속변수에 대한 확률분포를 표현함
 - ▶ 특정 구간내에 포함될 확률을 확률밀도라고 하며
 - ▶ 이를 함수 형태로 표현한 것을 확률 밀도 함수라고 한다.
- ▶ 무게가 203.81g ~ 302.9g 사이에 해당하는 사과가 선택 될 확률을 구하려면 왼쪽 그래프에서 203.18~302.9g에 해당하는 부분의 면적을 구하면 됨



정규분포

▶ 정규분포의 확률밀도함수 $f(x)$

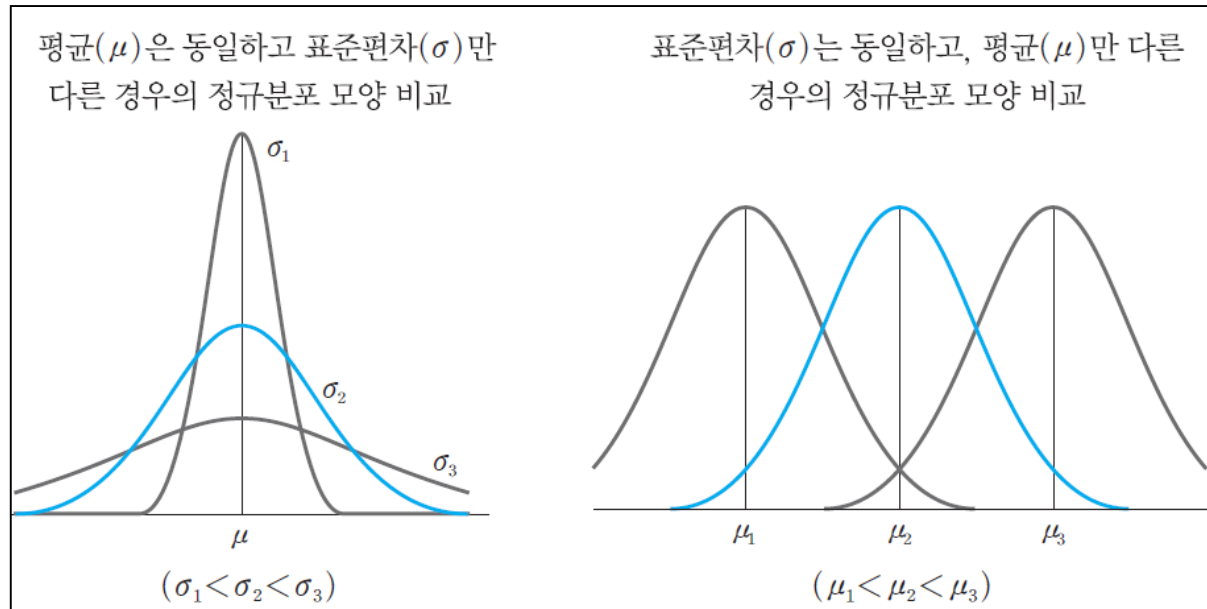
- ▶ 연속확률변수
- ▶ 넓이가 1이고 평균을 중심으로 좌우 대칭인 종모양의 특성을 활용해
- ▶ 평균과 표준편차만으로 전체 집단의 확률 밀도 함수를 정의 할 수 있음

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < +\infty$$

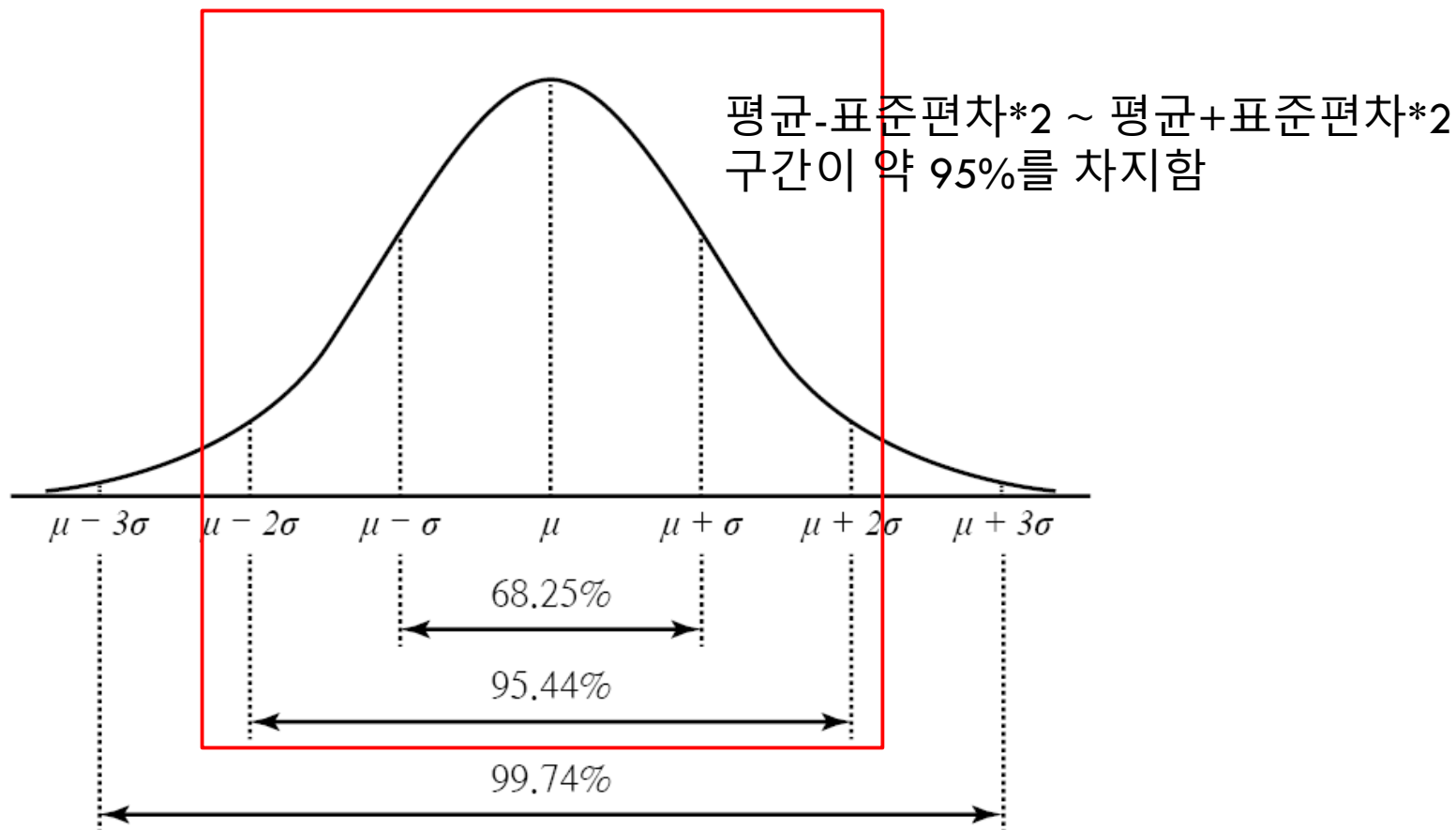
μ : 평균

σ : 표준편차



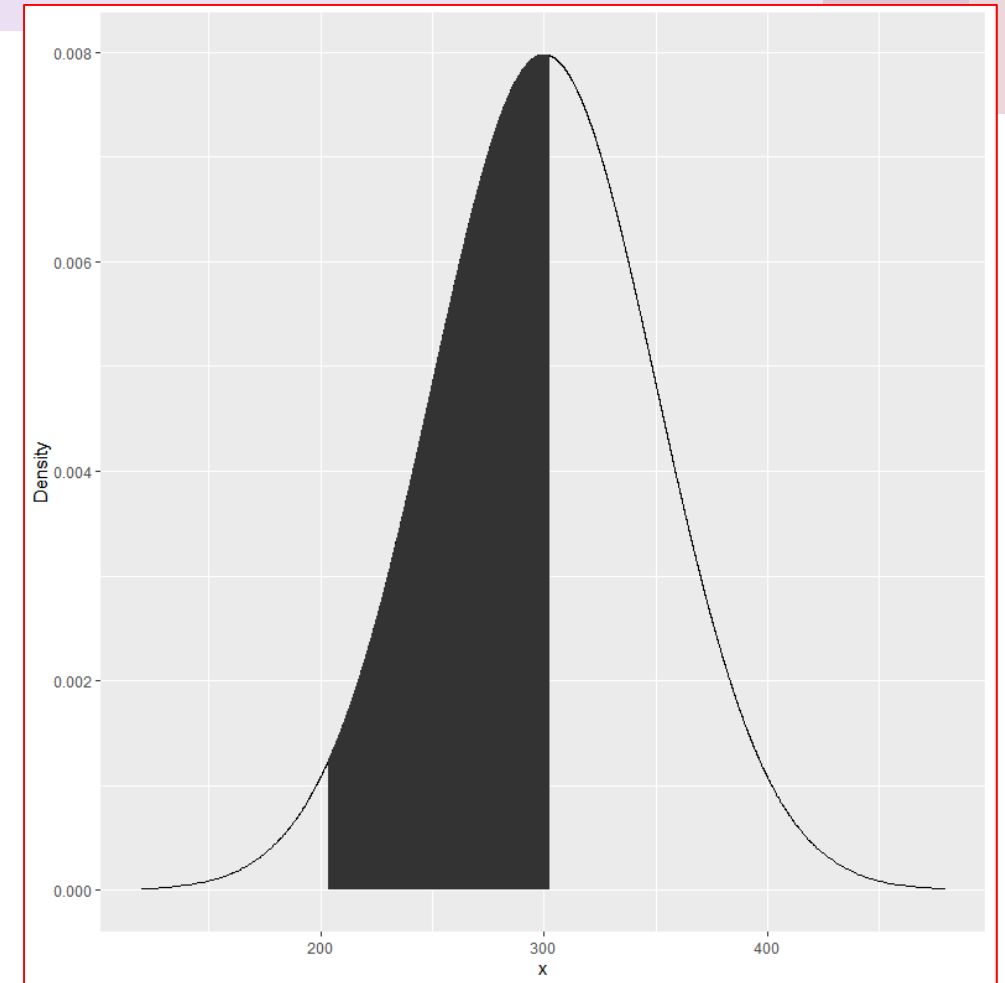
정규분포

- ▶ 검증된 정규분포의 다양한 특성들을 별도의 증명 없이 활용 할 수 있음
- ▶ 평균 주위로 $\pm \sigma$, $\pm 2\sigma$, $\pm 3\sigma$ 안에 약 68%, 95%, 99%의 데이터가 있음
- ▶ 아래 그림은 정규분포와 표준편차의 관계를 나타냄



확률 밀도 함수

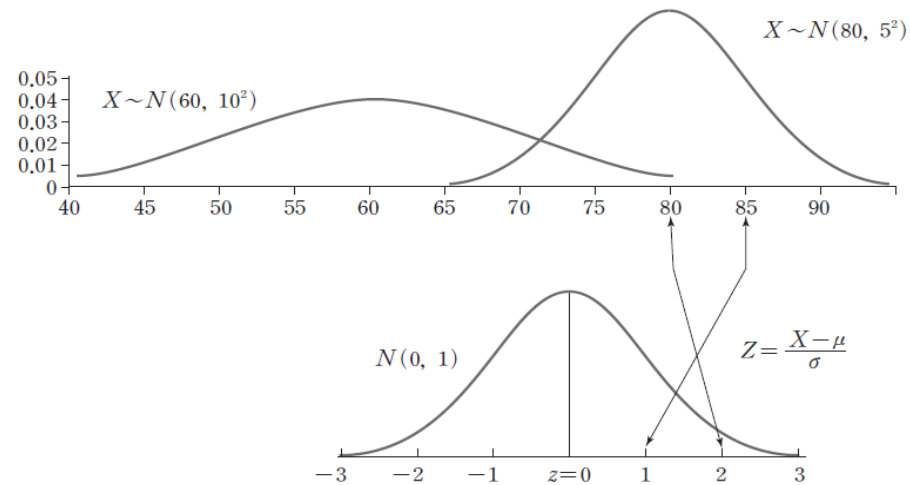
- ▶ A농장에서 무게가 203.81g ~ 302.9g 사이에 해당하는 사과가 선택될 확률
- ▶ 검정영역의 면적
- ▶ 정규분포로 가정한다면 이미 계산된 표준 정규 분포표를 활용해 쉽게 산출 가능



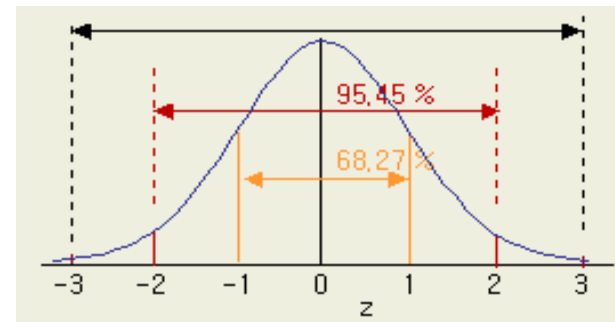
표준정규분포

- ▶ 평균을 0, 표준편차를 1로 변환하면 표준정규분포(Z분포)
- ▶ $z = (\text{확인할 값} - \text{평균}) / \text{표준편차}$: 표준점수
- ▶ 평균으로부터 표준편차의 몇배만큼 떨어졌는지에 대한 비율로 변환한 점수
- ▶ 서로 다른 정규분포들을 표준 정규분포라는 동일한 기준에서 비교 가능

$$Z = \frac{X - \mu}{\sigma}$$



- ▶ -1.96~1.96 안에는 95% 데이터가 있음



표준점수 Ex

- ▶ 철수는 영어와 중국어 시험을 응시해 800점과 85점을 받았다. 두 어학시험의 응시자들의 점수는 정규분포를 따르고 평균과 표준편차가 다음과 같을 때 철수는 영어와 중국어 중 어느 과목의 등수가 높을까?

과목	만점기준	평균	표준편차	취득점수
영어	1000	600	100	800
중국어	100	50	5	65

- ▶ 영어를 100점 만점으로 환산 한 후 비교

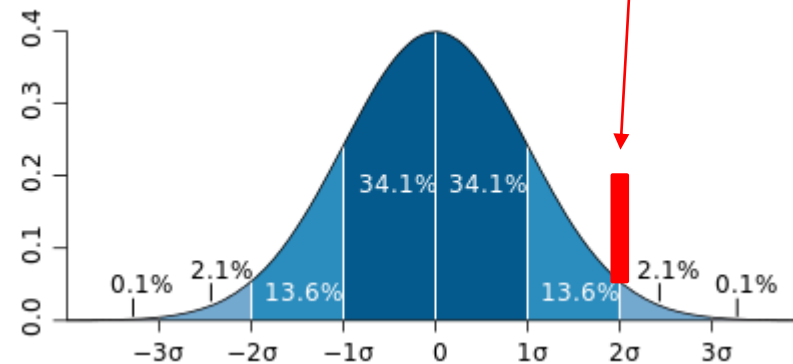
과목	취득점수	취득점수 100점 만점 환산	과목 평균점수	평균보다 몇 점 높나? (100점 만점 환산)
영어	800	80	600	20
중국어	65	65	50	25

- ▶ 영어 80점 > 중국어 60점
- ▶ 영어는 평균보다 20점 높음 > 중국어는 평균보다 15점 높음

표준점수 EX2

- ▶ 영어의 등수도 중국어보다 높을까?
- ▶ 표준편차까지 감안해 확인
 - ▶ 정규분포를 이루는 집단의 97.7%는 평균 + 표준편차*2 값보다 작은 값을 갖는다
- ▶ 각 점수에 상위 2.3%에 해당하는 점수를 확인

과목	취득점수	표준편차	상위2.3%(100~97.7%)
영어	800	100	$800(=600+100*2)$
중국어	65	5	$60(=50+5*2)$



표준점수 활용

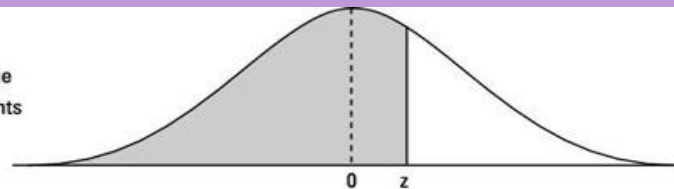
구간확률 확인

과목	취득점수	표준점수	표준점수 수식
영어	800	2	$(800-600)/100$
중국어	65	3	$(65-50)/5$

표준점수가 중국어가 더 높으므로 중국어 등수가 더 높다

표준점수가 1.65인 경우
표준 정규분포표의 값은 0.9505 이며
z값이 1.65보다 작은 구간은 전체 대비 95.05%를
차지한다는 것을 알 수 있음

Number in the
table represents
 $P(Z \leq z)$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999