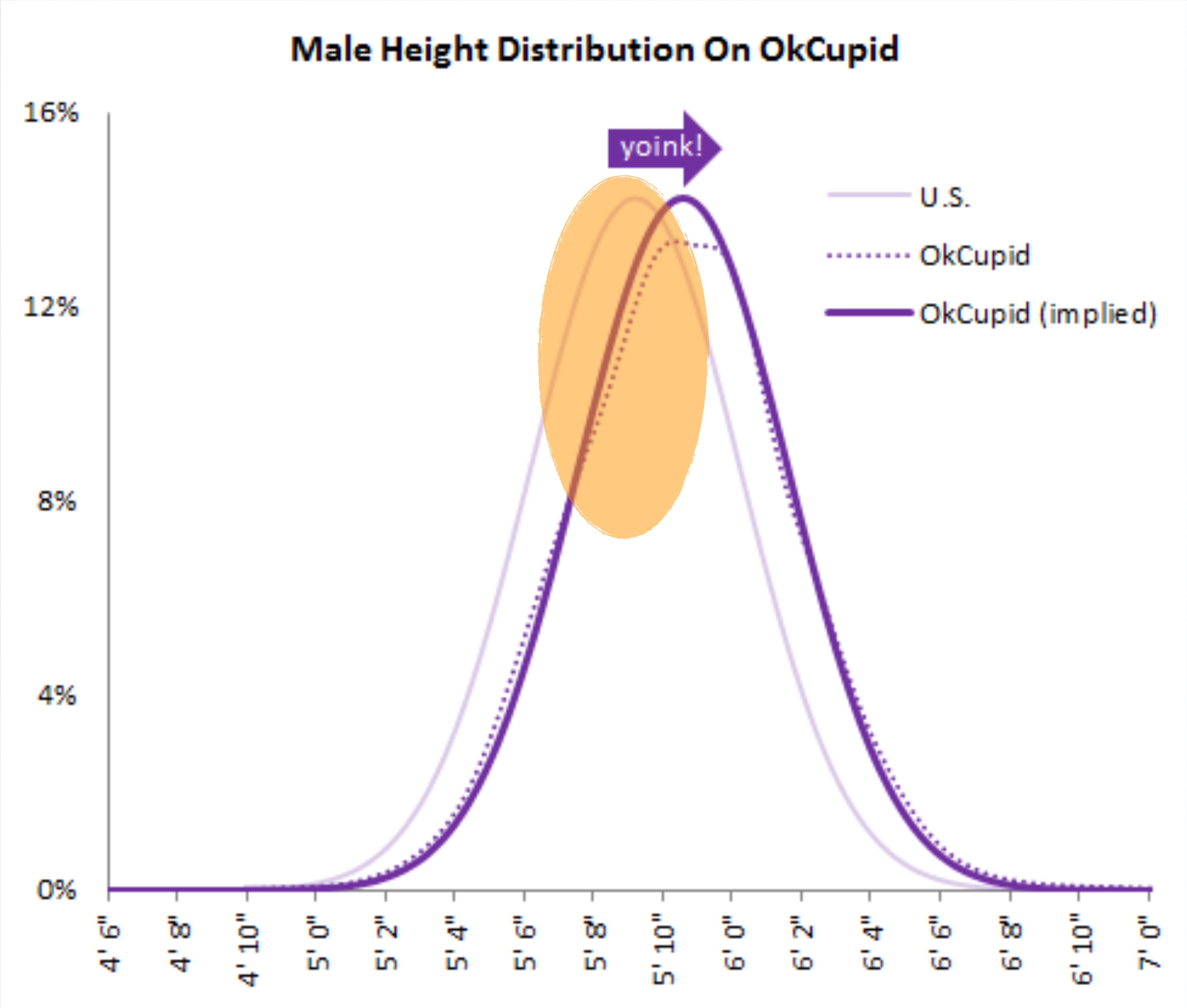
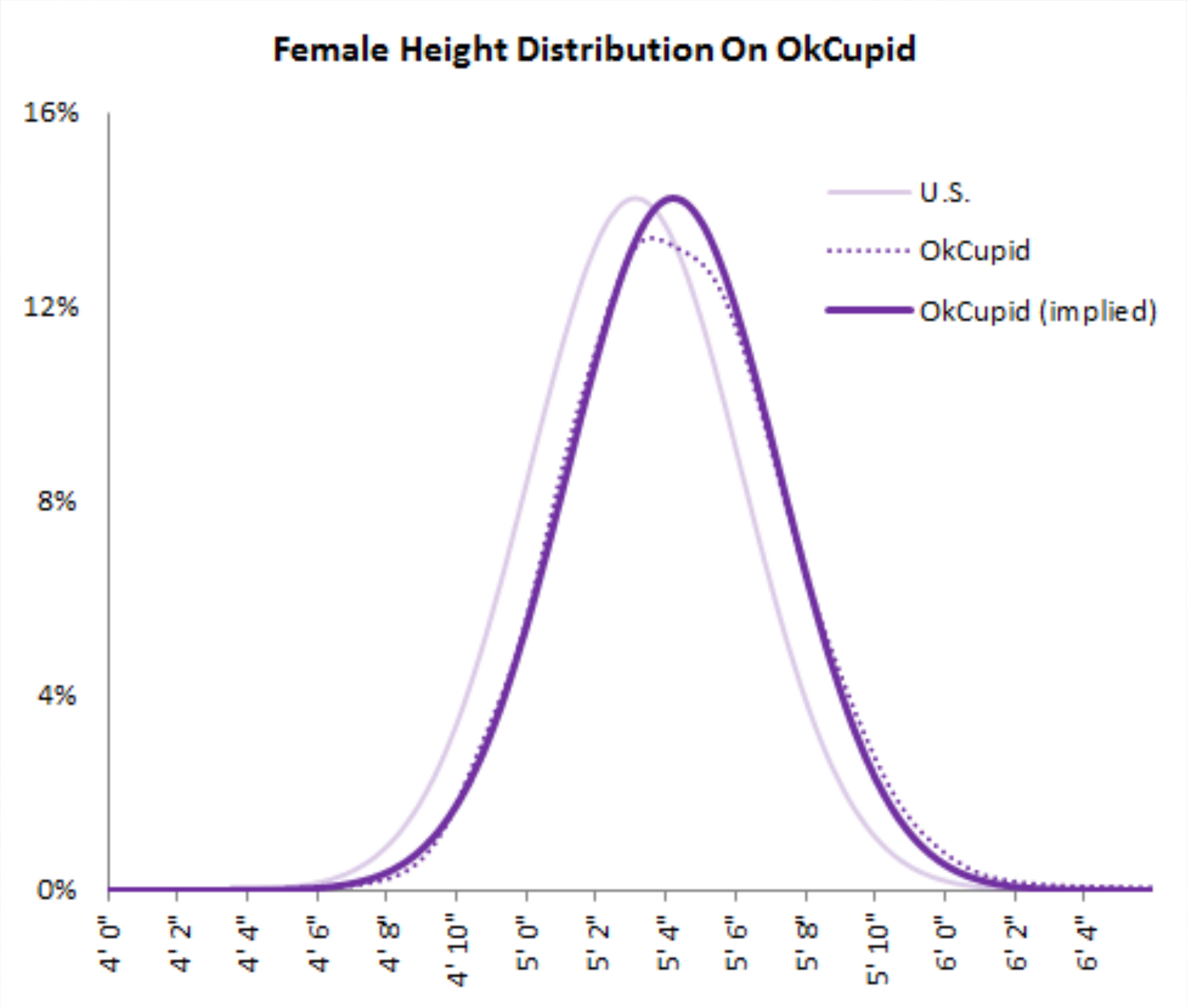
normal distribution

- normal distribution
- 68-95-99.7% rule
- standardized scores
- probabilities and percentiles

Duke University

Dr. Mine Çetinkaya-Rundel
Duke University

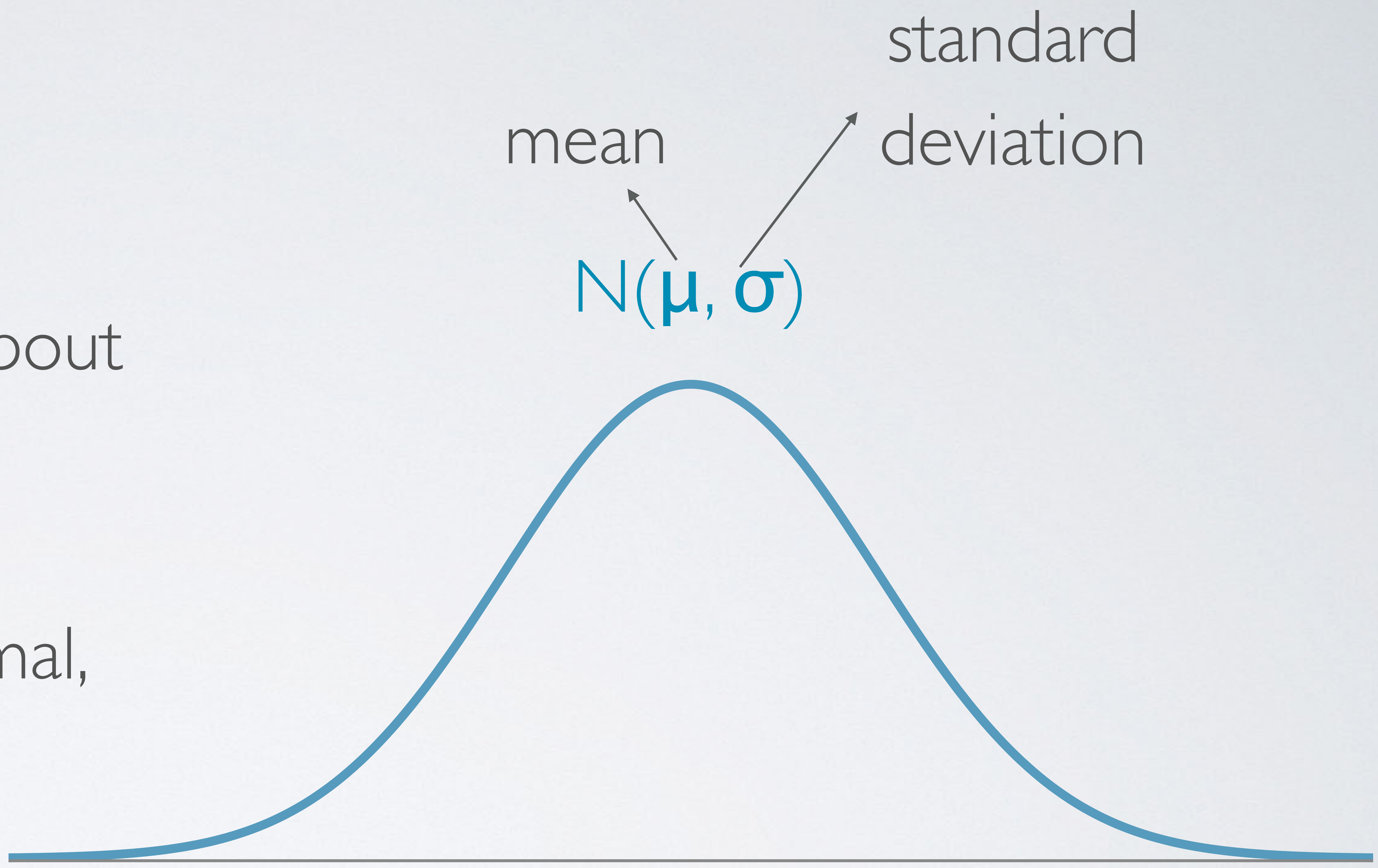


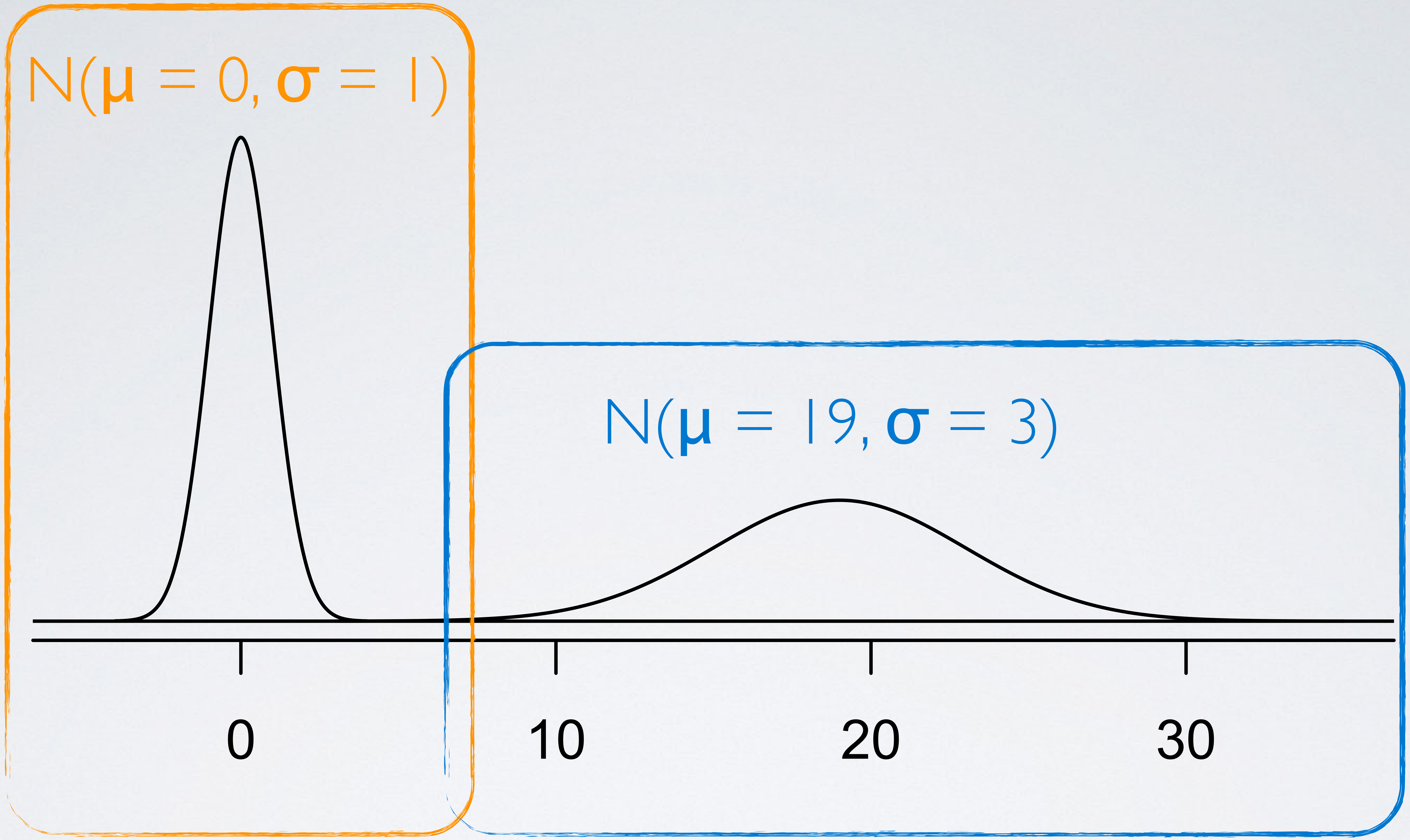


Source: http:// blog.okcupid.com/ index.php/ the- biggest- lies- in- online- dating/

normal distribution

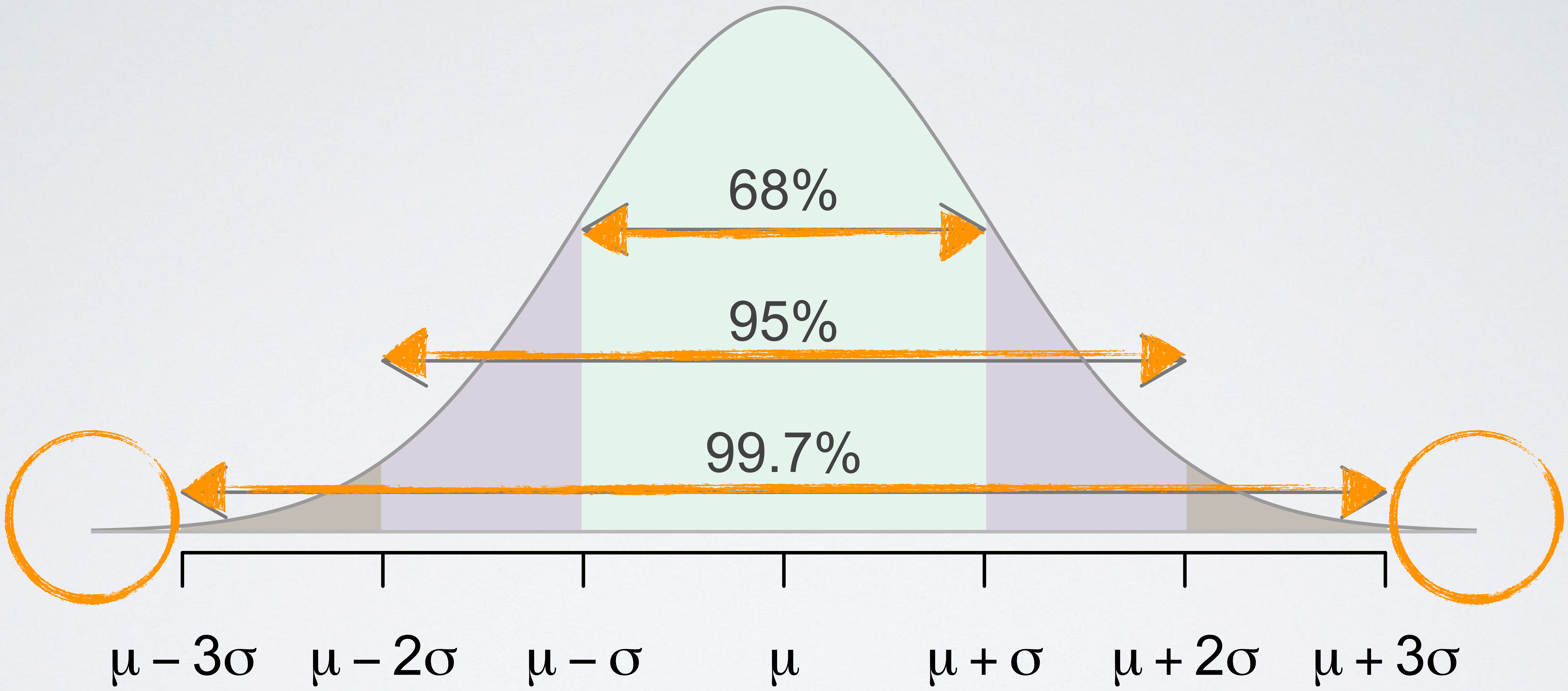
- unimodal and symmetric
  - bell curve
- follows very strict guidelines about how variably the data are distributed around the mean
- many variables are nearly normal, but none are exactly normal





N(μ = 0, σ = 1)
N(μ = 19, σ = 3)
0
10
20
30

68 - 95 - 99.7% rule



practice

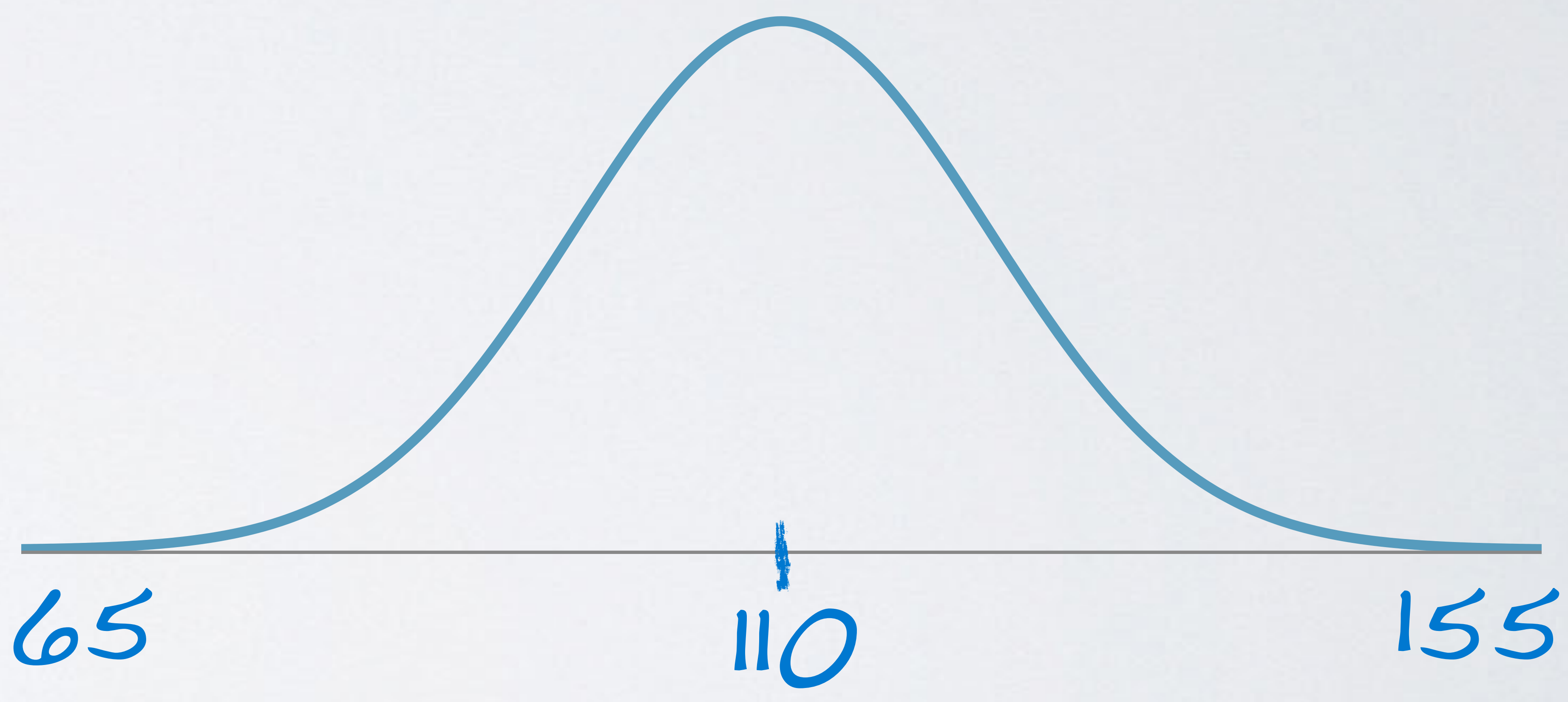
A doctor collects a large set of heart rate measurements that approximately follow a normal distribution. He only reports 3 statistics, the mean = 110 beats per minute, the minimum = 65 beats per minute, and the maximum = 155 beats per minute. Which of the following is most likely to be the standard deviation of the distribution?

(a) 5 $\longrightarrow$ $110 \pm (3 \times 5) = (95, 125)$

(b) 15 $\longrightarrow$ $110 \pm (3 \times 15) = (65, 155)$

(c) 35 $\longrightarrow$ $110 \pm (3 \times 35) = (5, 215)$

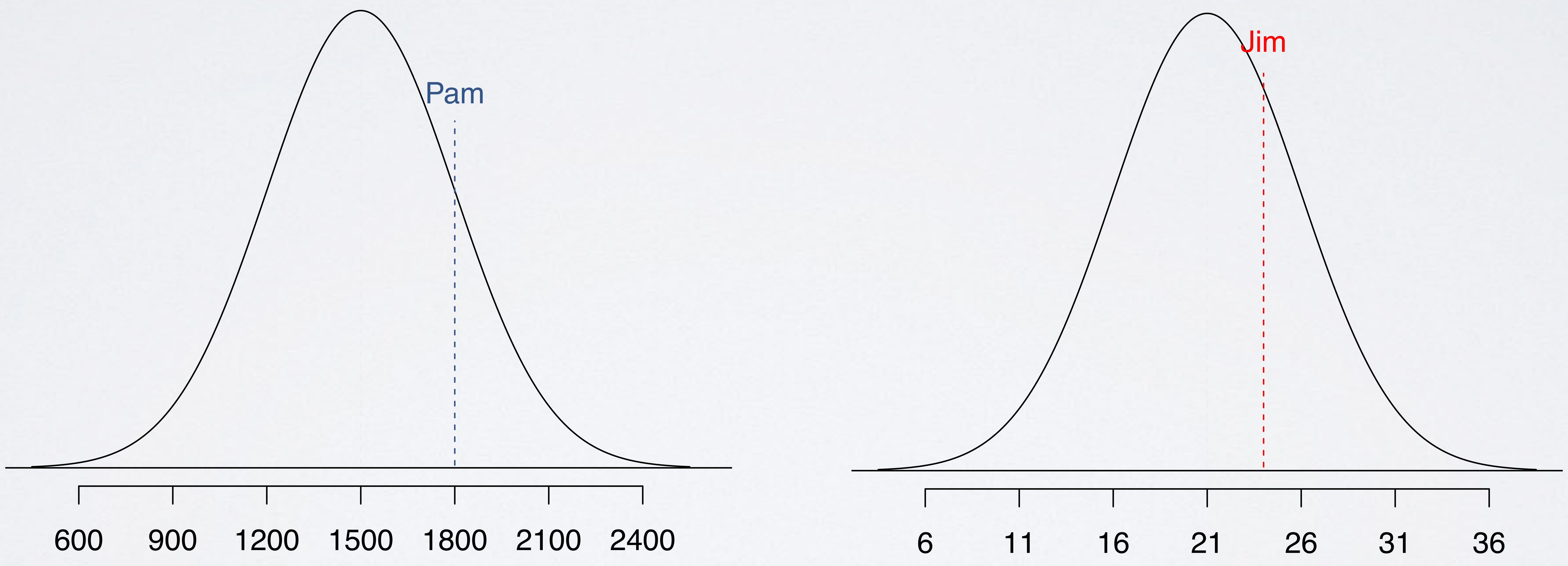
(d) 90 $\longrightarrow$ $110 \pm (3 \times 90) = (-160, 380)$



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?

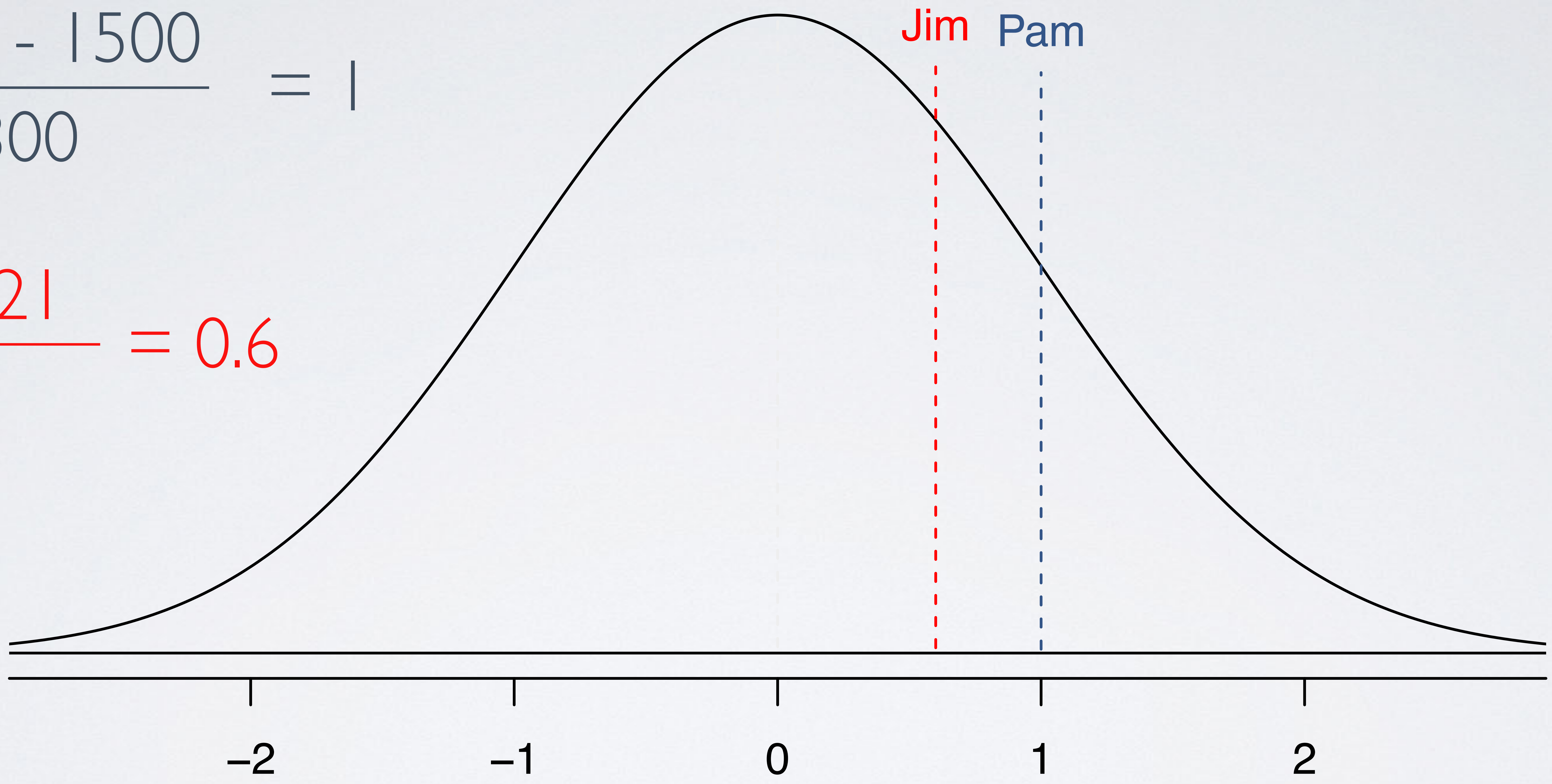
SAT scores ~ N(mean = 1500, SD = 300)

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



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

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



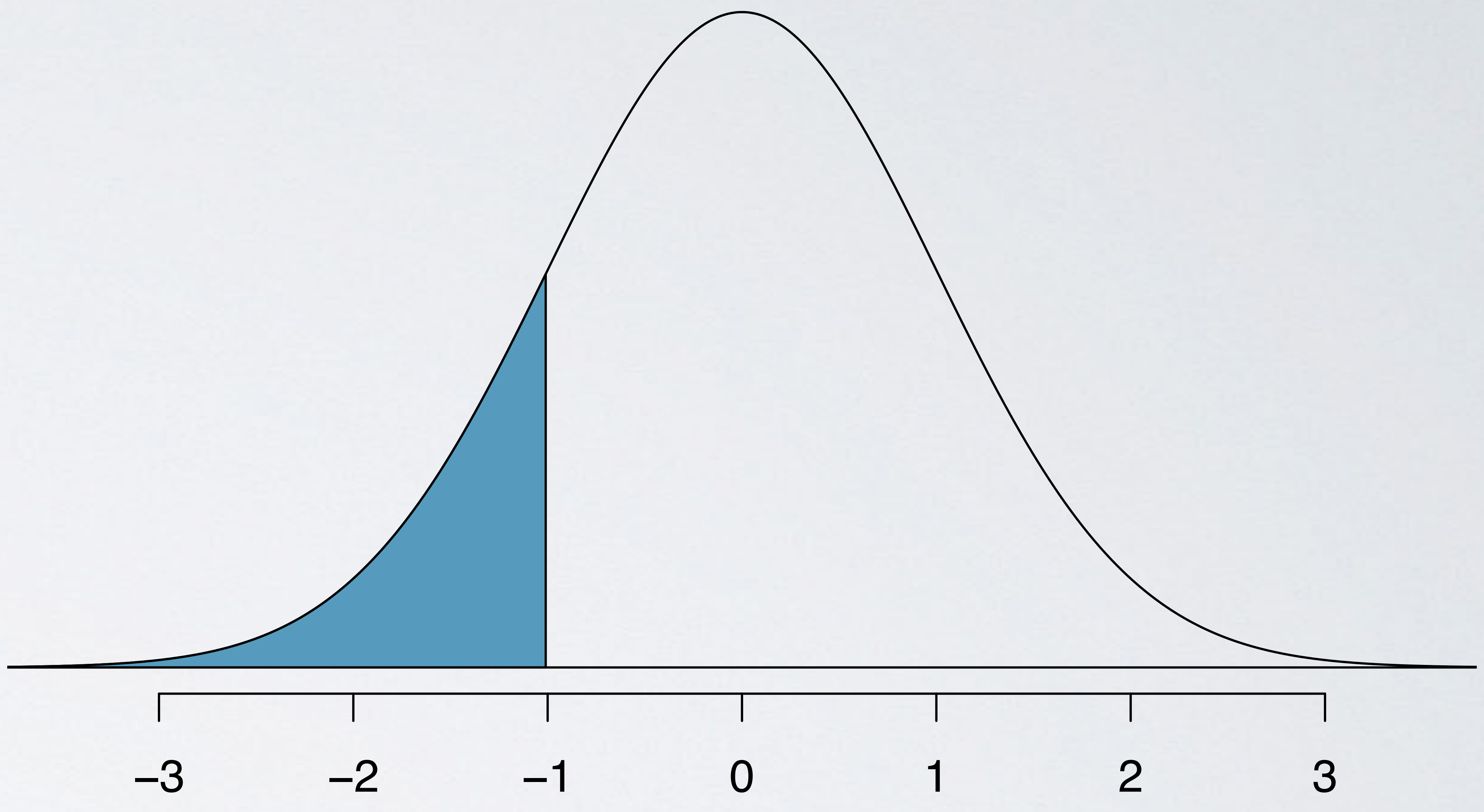
standardizing with Z scores

- 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

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

percentiles

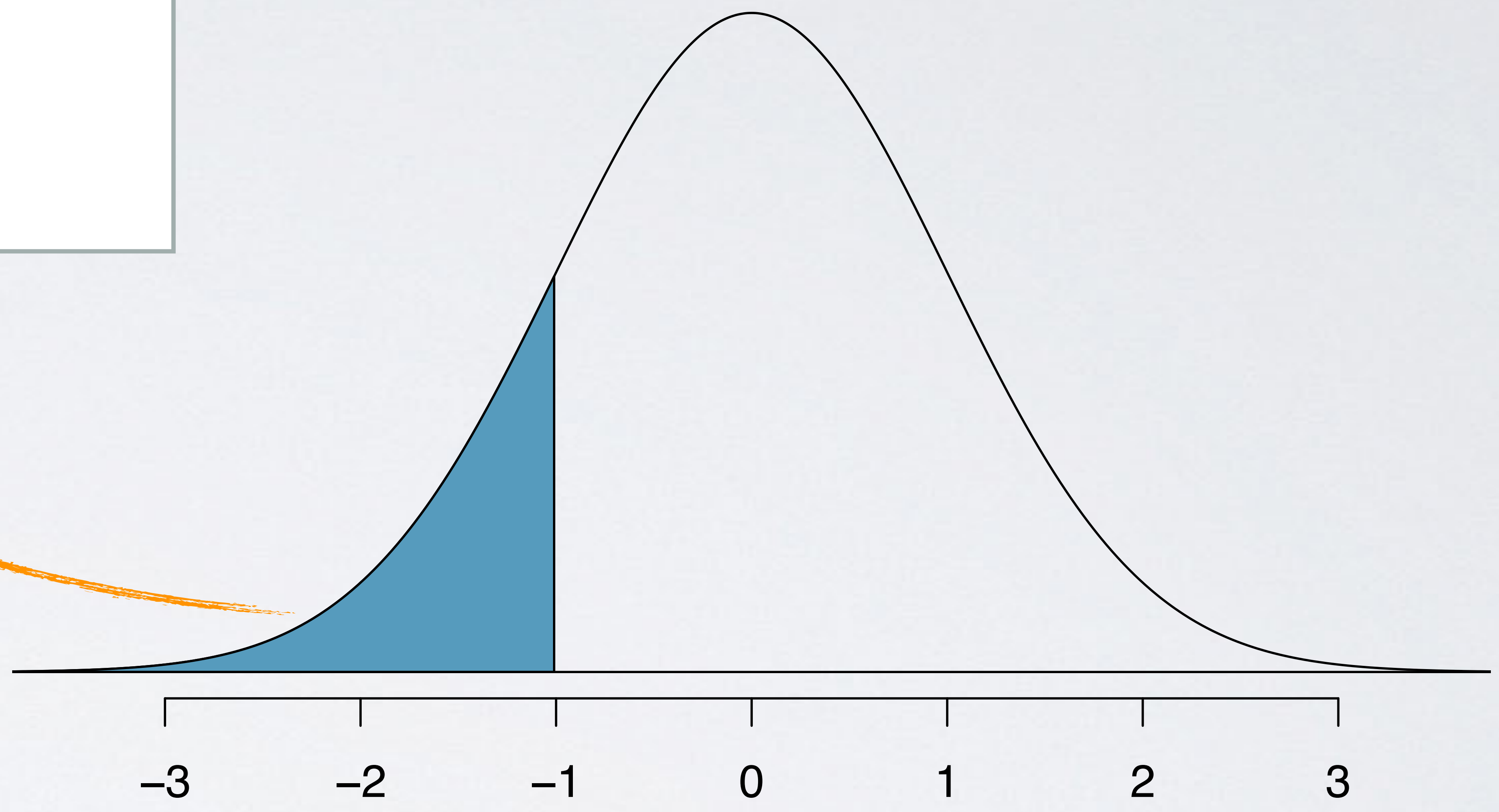
- when the distribution is normal, Z scores can be used to calculate percentiles
- percentile is the percentage of observations that fall below a given data point
- graphically, percentile is the area below the probability distribution curve to the left of that observation.



computing percentiles - using R

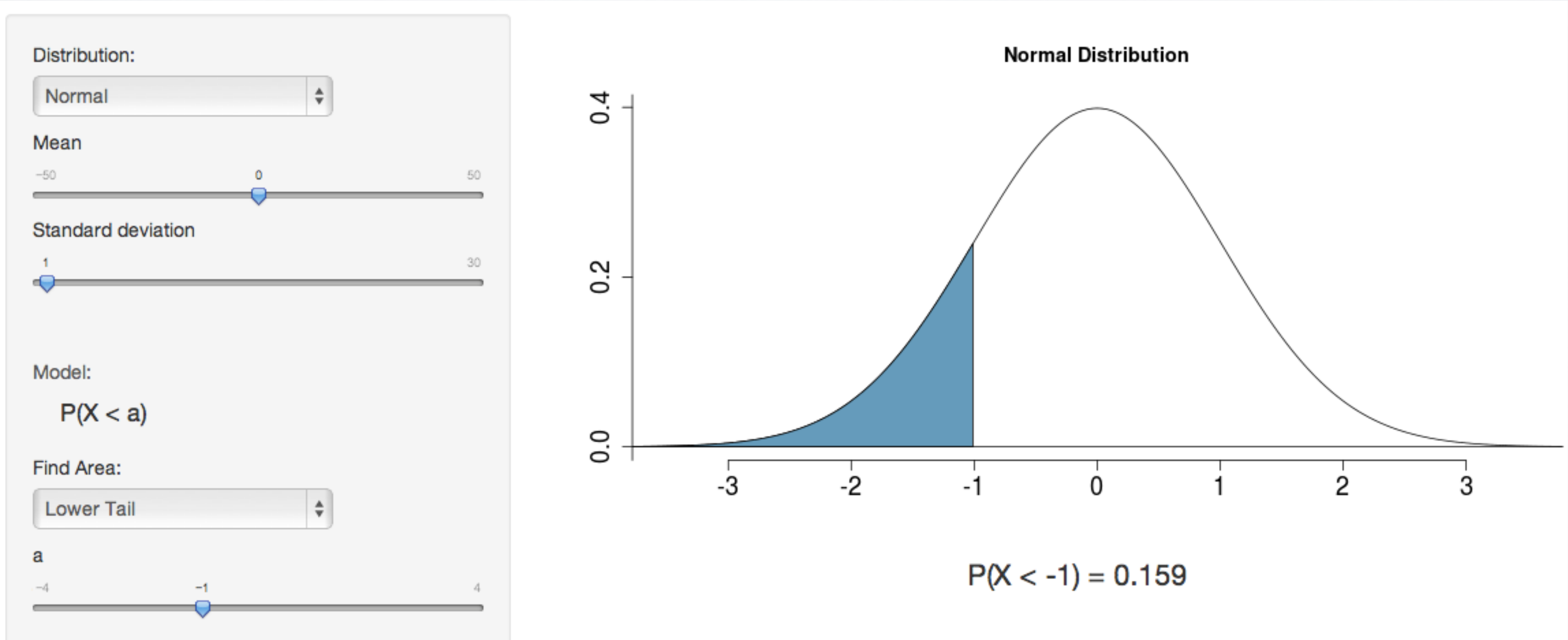
R

```
> pnorm(-1, mean = 0, sd = 1)
[1] 0.1586553
```



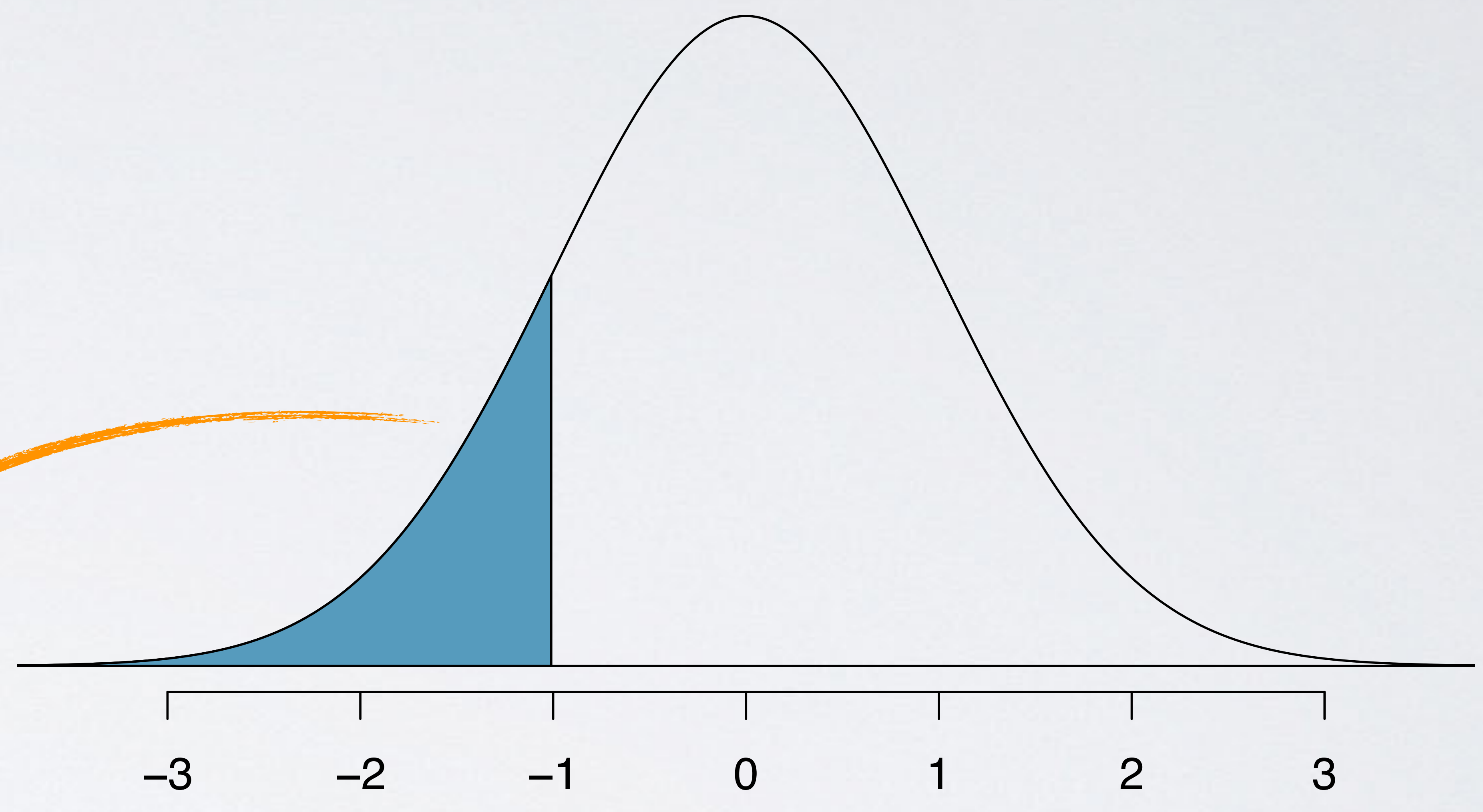
computing percentiles - using the applet

http://bitly.com/dist_calc



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

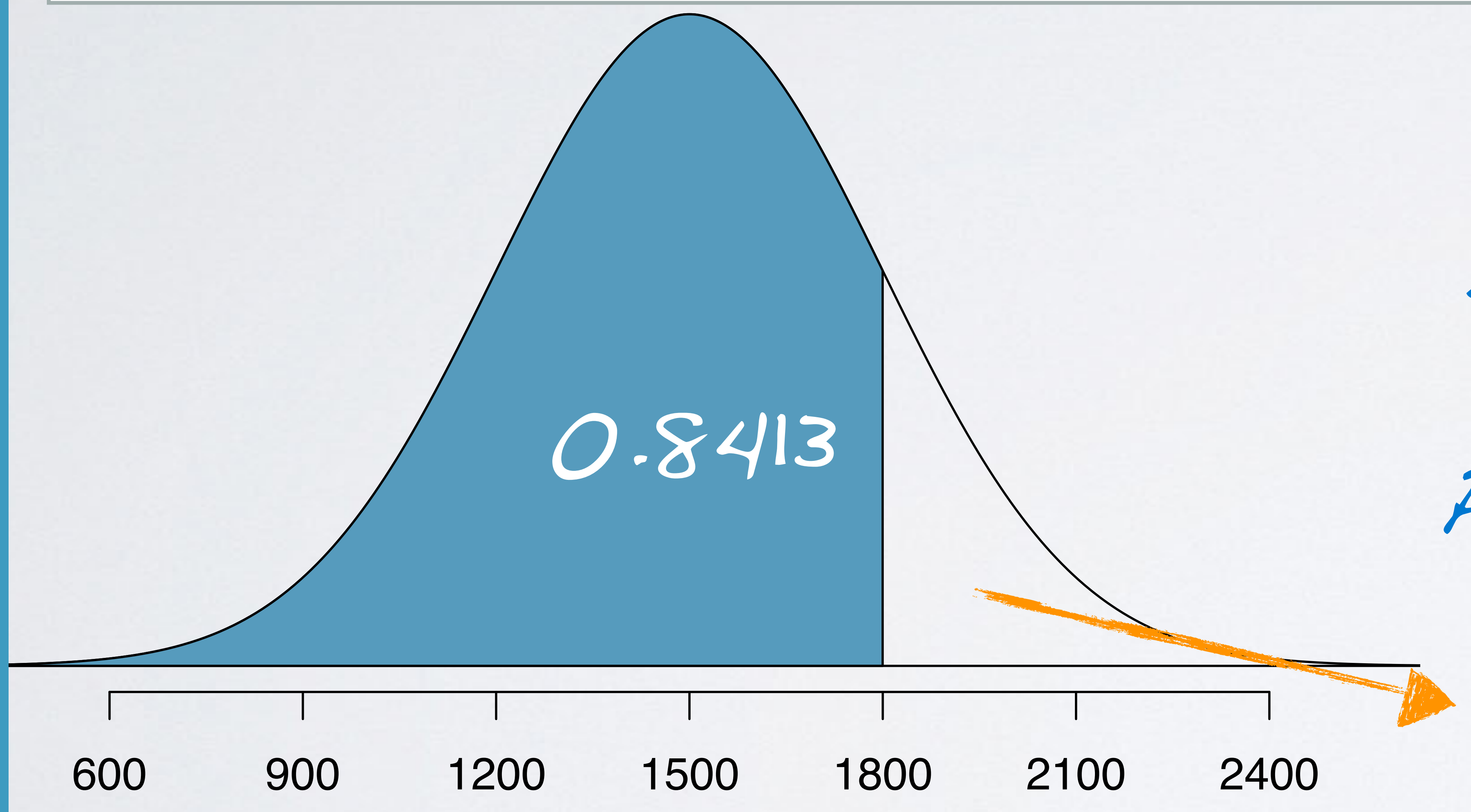


practice

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

R

```
> pnorm(1800, mean = 1500, sd = 300)
[1] 0.8413
```



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

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

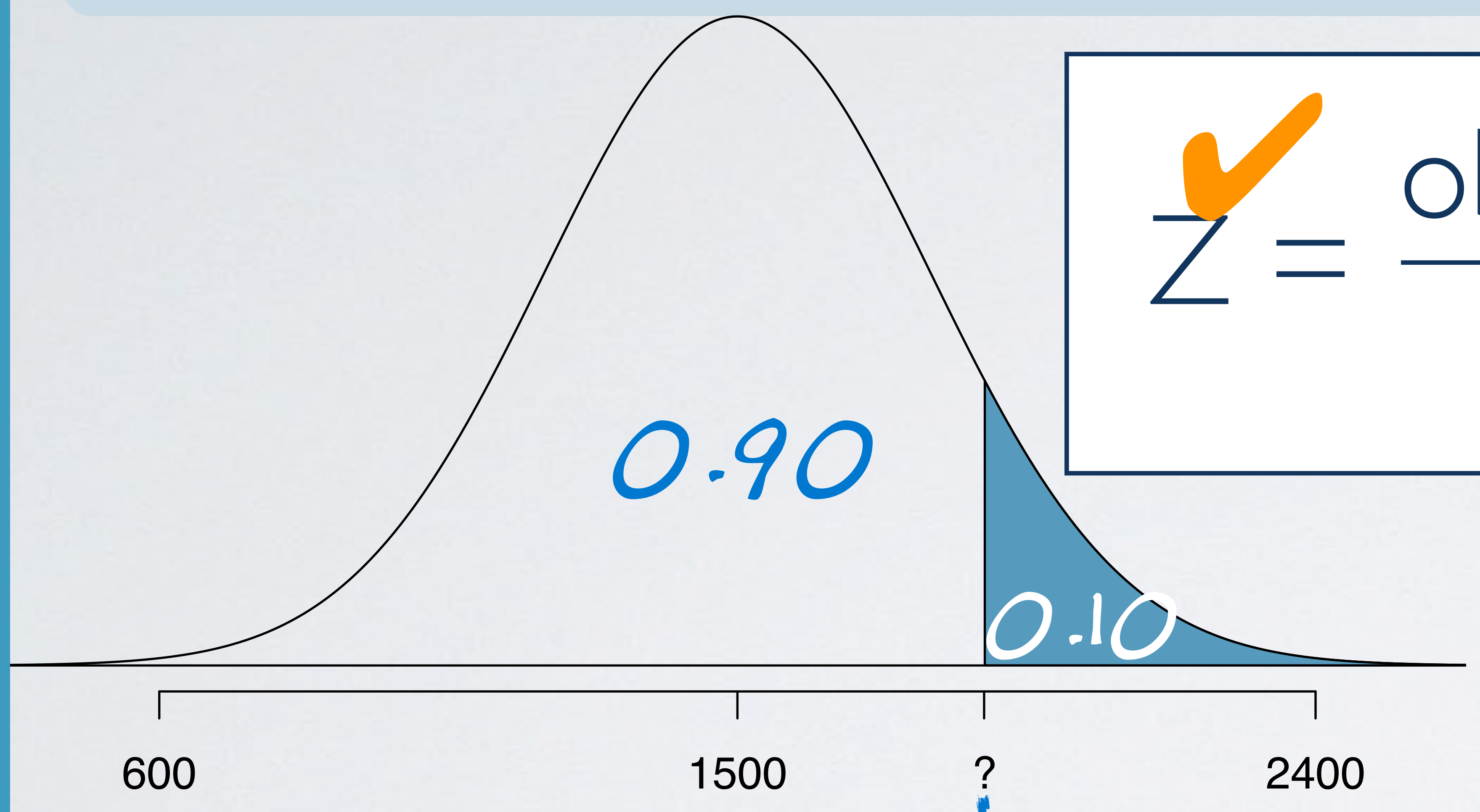
$$1 - 0.8413 = 0.1587$$

Z	Second decimal place of 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.

practice

A friend of yours tells you that she scored in the top 10% on the SAT. What is the lowest possible score she could have gotten?

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
			0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
			0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
			0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
			0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
			0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
			0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319



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

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

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

R

```
> qnorm(0.90, 1500, 300)
[1] 1884.465
```