Chapter 3: Normal distribution

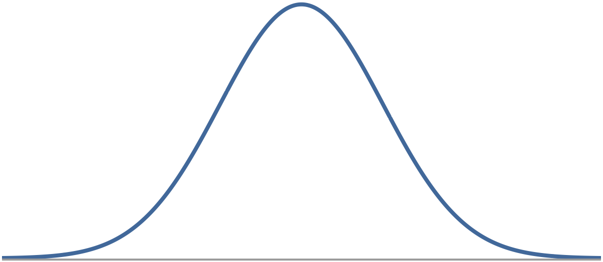
Yingbo Li

Southern Methodist University

STAT 2331

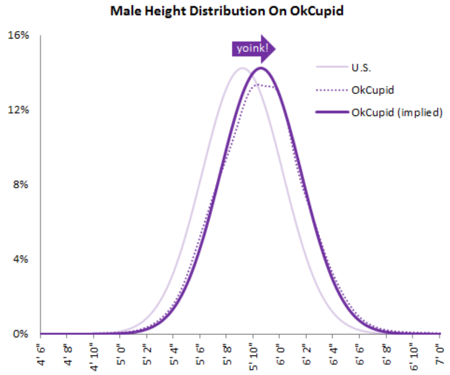
Normal distribution

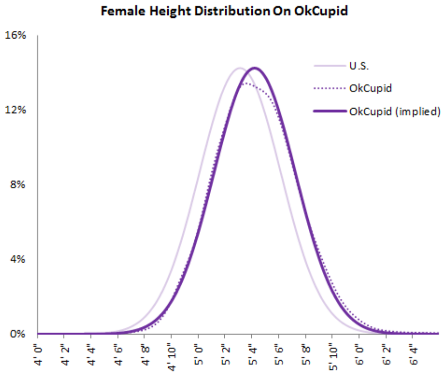
- Unimodal and symmetric, bell shaped curve
- Most variables are nearly normal, but none are exactly normal
- Denoted as $N(\mu, \sigma)$ $\rightarrow$ Normal with mean μ and standard deviation σ



Get yourself a 10 Deutsche Mark bill

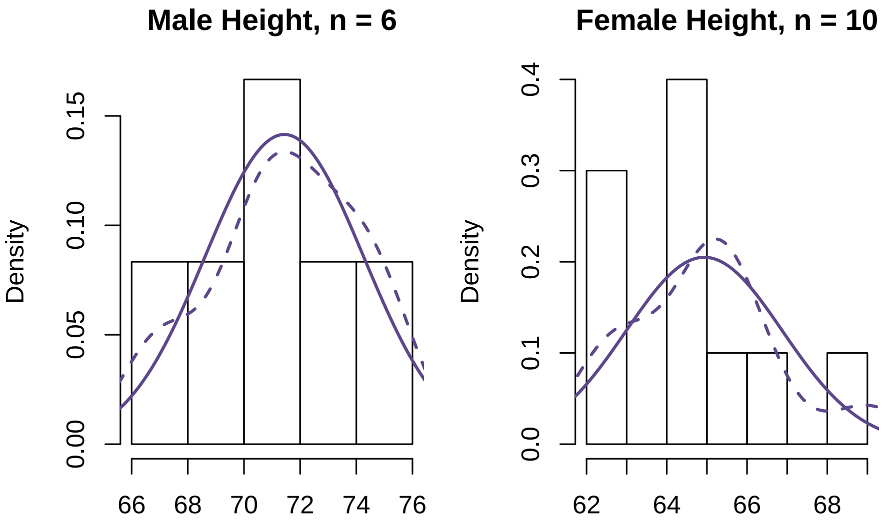
Heights of males and females





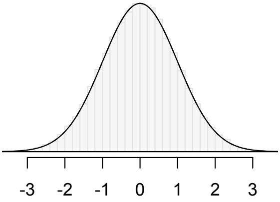
http://blog.okcupid.com/index.php/the-biggest-lies-in-online-dating/

Heights of students in our class

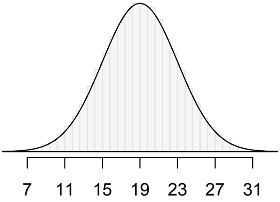


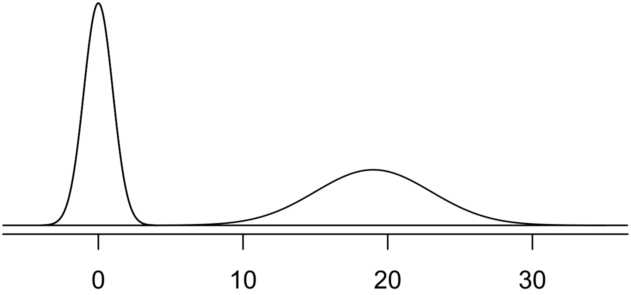
Normal distributions with different parameters

$N(\mu = 0, \sigma = 1)$



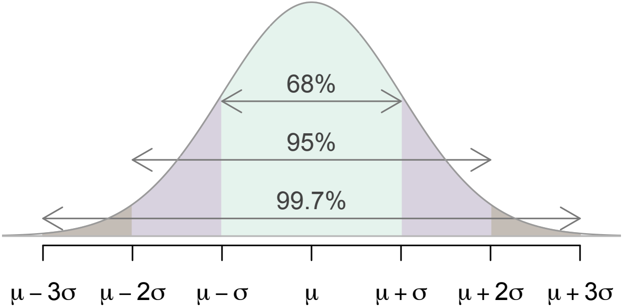
$N(\mu = 19, \sigma = 4)$





68-95-99.7 Rule

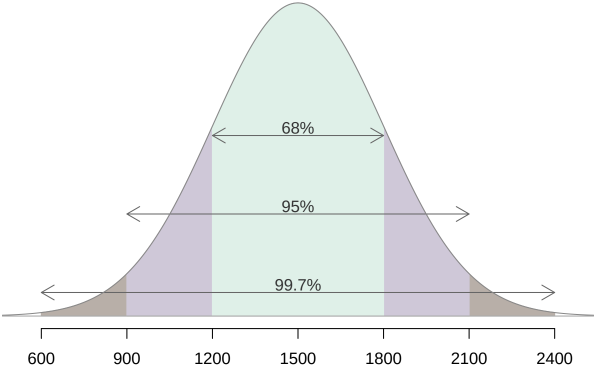
- For nearly normally distributed data,
  - about 68% falls within 1 SD of the mean,
  - about 95% falls within 2 SD of the mean,
  - about 99.7% falls within 3 SD of the mean.
- It is possible for observations to fall 4, 5, or more standard deviations away from the mean, but these occurrences are very rare if the data are nearly normal.



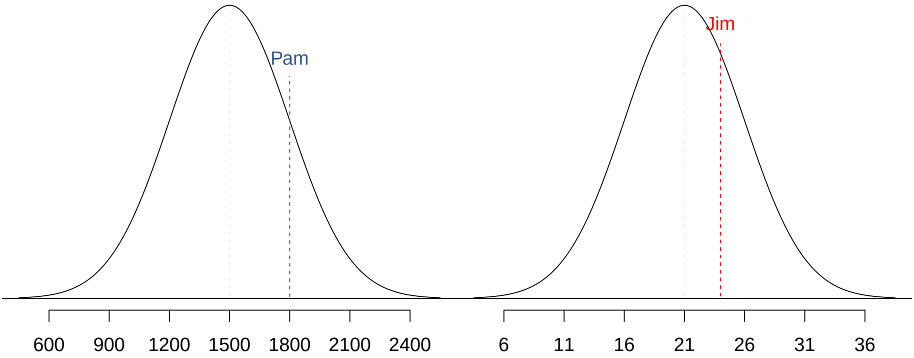
Describing variability using the 68-95-99.7 Rule

SAT scores are distributed nearly normally with mean 1500 and standard deviation 300.

- $\sim$68% of students score between 1200 and 1800 on the SAT.
- $\sim$95% of students score between 900 and 2100 on the SAT.
- $\sim$99.7% of students score between 600 and 2400 on the SAT.



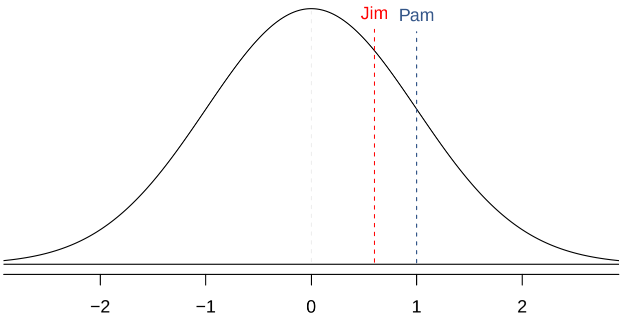
SAT scores are distributed nearly normally with mean 1500 and standard deviation 300. ACT scores are distributed nearly normally with mean 21 and standard deviation 5. A college admissions officer wants to determine which of the two applicants scored better on their standardized test with respect to the other test takers: Pam, who earned an 1800 on her SAT, or Jim, who scored a 24 on his ACT?



Standardizing with Z scores

Since we cannot just compare these two raw scores, we instead compare how many standard deviations beyond the mean each observation is.

- Pam's score is $\frac{1800-1500}{300} = 1$ standard deviation above the mean.
- Jim's score is $\frac{24-21}{5} = 0.6$ standard deviations above the mean.



Standardizing with Z scores (cont.)

- These are called *standardized* scores, or *Z scores*.
- Z score of an observation is the number of standard deviations it falls above or below the mean.

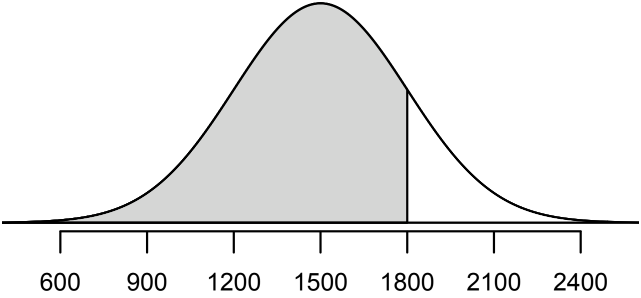
Z scores

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

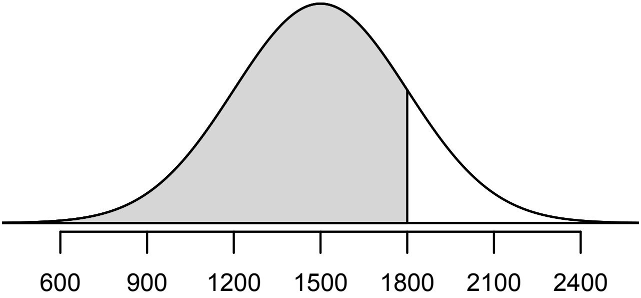
Note: Z scores can be used to describe observations from distributions of any shape (not just normal) but only when the distribution is normal can we use Z scores to calculate percentiles.

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.



Approximately what percent of students score below 1800 on the SAT? (Hint: Use the 68-95-99.7% rule.)



$$100 - 68 = 32\%$$
$$32/2 = 16\%$$
$$68 + 16 = 84\%$$

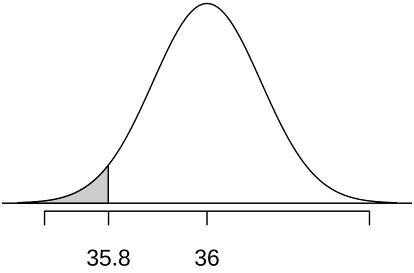
Calculating percentiles using normal tables

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	*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

Note: The normal tables are available on textbook page 428-429.

At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle will fails the quality control inspection. What's the probability that the amount of ketchup in a randomly selected bottle is less than 35.8 ounces?

Let X = amount of ketchup in a bottle: $X \sim N(\mu = 36, \sigma = 0.11)$



$$Z = \frac{x-\mu}{\sigma} = \frac{35.8 - 36}{0.11} = -1.82$$

$$P(X < 35.8) = P(Z < -1.82) = 0.0344$$

Second decimal place of Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	-2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	-2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	*0.0344*	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5

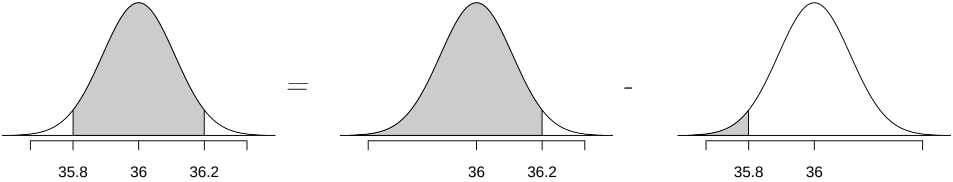
Question

1. At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of the bottle goes below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection.

What percent of bottles pass the quality control inspection?

(a) 1.82%
(b) 3.44%
(c) 6.88%
(d) *93.12%*
(e) 96.56%

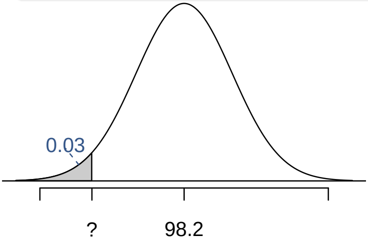
$$P(35.8 < X < 36.2) = ?$$



$$Z_{35.8} = \frac{35.8 - 36}{0.11} = -1.82$$
$$Z_{36.2} = \frac{36.2 - 36}{0.11} = 1.82$$

$$\begin{aligned} P(35.8 < X < 36.2) &= P(-1.82 < Z < 1.82) \\ &= P(Z < 1.82) - P(Z < -1.82) \\ &= 0.9656 - 0.0344 = 0.9312 \end{aligned}$$

Body temperatures of healthy humans are distributed nearly normally with mean 98.2°F and standard deviation 0.73°F. What is the cutoff for the lowest 3% of human body temperatures?



0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	-1.9
0.0294	*0.0301*	0.0307	0.0314	0.0322	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	-1.7

$$P(X < x) = 0.03 \to P(Z < \textit{-1.88}) = 0.03$$

$$Z = \frac{x - \mu}{\sigma} \to \frac{x - 98.2}{0.73} = -1.88$$

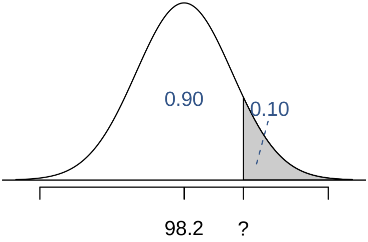
$$x = (-1.88 \times 0.73) + 98.2 = 96.8$$

Question

2. Body temperatures of healthy humans are distributed nearly normally with mean 98.2°F and standard deviation 0.73°F.

What is the cutoff for the highest 10% of human body temperatures?

(a) *99.1*
(b) 97.3
(c) 99.4
(d) 99.6



Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

$$P(X > x) = 0.10 \rightarrow P(Z < 1.28) = 0.90$$

$$Z = \frac{x - \mu}{\sigma} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

$$x = (1.28 \times 0.73) + 98.2 = 99.1$$