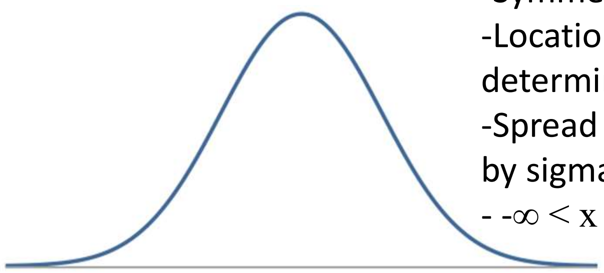
ECO239

Week 12

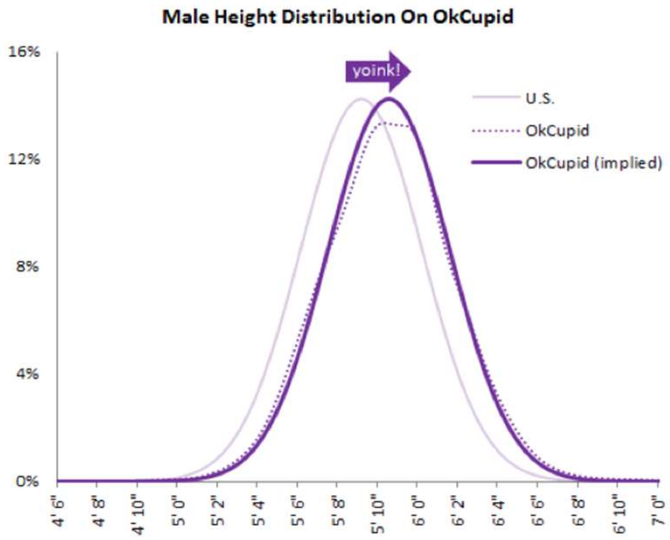
Normal Distribution (Gaussian)

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

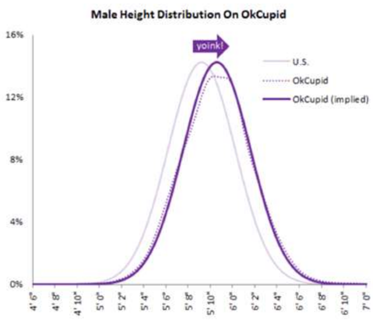


-Bell shaped
-Symmetric
-Location is determined by mean
-Spread is determined by sigma.
- $-\infty < x < \infty$

Heights of males



Heights of males



“The male heights on OkCupid very nearly follow the expected normal distribution -- except the whole thing is shifted to the right of where it should be. Almost universally guys like to add a couple inches.”

“You can also see a more subtle vanity at work: starting at roughly 5' 8", the top of the dotted curve tilts even further rightward. This means that guys as they get closer to six feet round up a bit more than usual, stretching for that coveted psychological benchmark.”

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

PDF (Probability Density Function) and CDF (Cumulative Distribution Function) for Normal Distribution

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

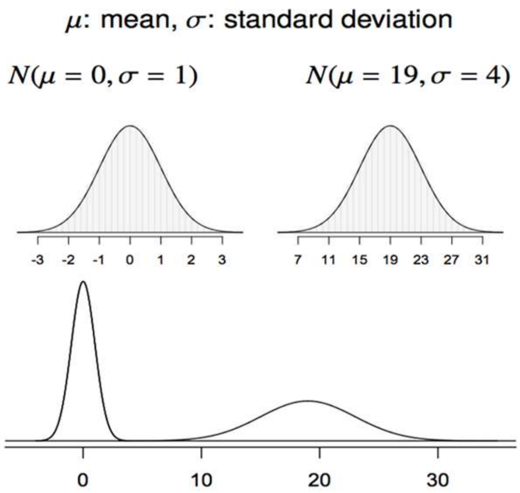
$$F(x_0) = P(X \leq x_0)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x_0} e^{-(x-\mu)^2/2\sigma^2} dx$$

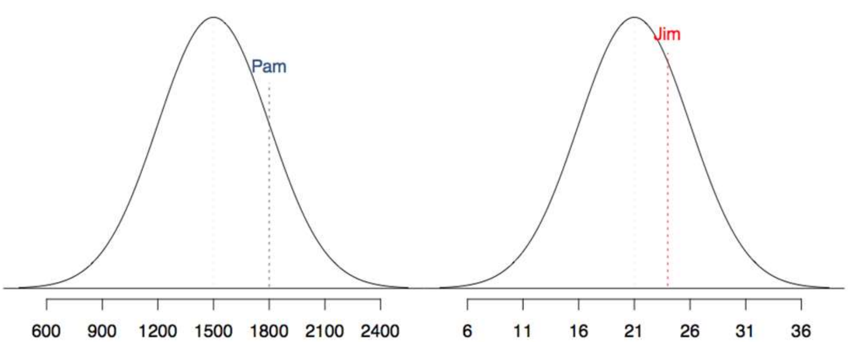
- μ : mean
- σ2 : variance
- π : 3.14159
- e : 2.71828

- Illustrations of pdf and cdf, finding normal probabilities:
- P(a<x<b)=F(b)-F(a) will be discussed on board.

Normal distributions with different parameters



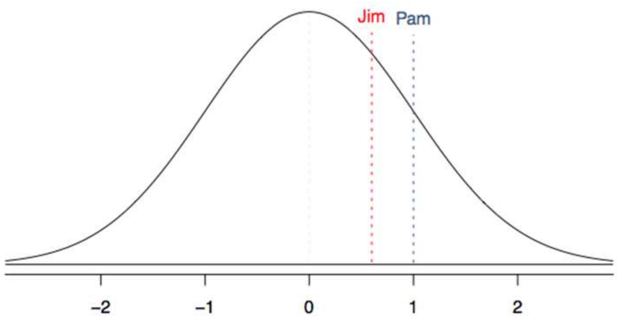
SAT scores are distributed nearly normally with mean 1500 and standard deviation 300. ACT scores are distributed nearly normally with mean 21 and standard deviation 5. A college admissions officer wants to determine which of the two applicants scored better on their standardized test with respect to 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 (1800 - 1500) / 300 = 1 standard deviation above the mean.
- Jim's score is (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.

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

- Z scores are defined for distributions of any shape, but only when the distribution is normal can we use Z scores to calculate percentiles.
- Observations that are more than 2 SD away from the mean (|Z| > 2) are usually considered unusual.

Standard Normal Distribution

$$z \sim N(0,1)$$

Standardized X, z value is following normal distribution with mean 0 and standard deviation 1.

$$PDF: f(z|0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$CDF: F(z_0) = P(Z \leq z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{-\frac{z^2}{2}} dz$$

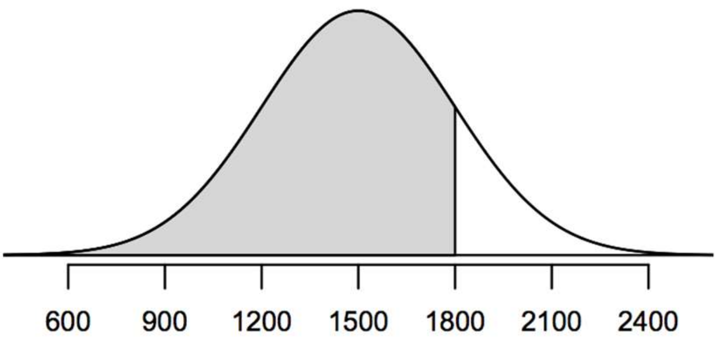
$$\mu = 0, \sigma = 1$$

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

$$F(x_0) = P(X \leq x_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x_0} e^{-(x-\mu)^2/2\sigma^2} dx$$

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.

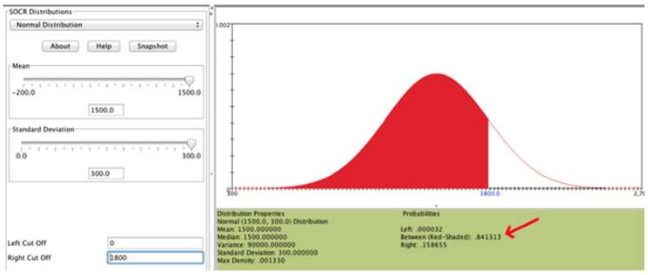


Calculating percentiles - using computation

There are many ways to compute percentiles/areas under the curve. R:

```
> pnorm(1800, mean = 1500, sd = 300)
[1] 0.8413447
```

Applet: www.socr.ucla.edu/htmls/SOCR_Distributions.html



Practice: Relationship between normal distribution and standard normal distribution

- X=200 => Z = ?
- μx= 100 => μz = ?
- σx = 50 =>σz = ?

=> Illustration/comparison is discussed on board.

Finding Normal Probabilities

$$P(a < x < b)$$
$$= P\left(\frac{a-\mu}{\sigma} < z < \frac{b-\mu}{\sigma}\right)$$
$$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

Practice

X~ N(15,4)
Find P(X > 18).

P(X>18)=P(Z> (18-15)/4) = P(Z>3/4)
=P(Z>0.75) = 1- P(Z<0.75)=1-0.7734 = 0.2266.

Negative z-values

- z = -2.
- Find the probability P(z < -2)
- HOW?

=> Use the fact that normal distribution is symmetric.

P(z < - 2) = P(z > 2) = 1 – P(z<2)
= 1-0.9772 (using z-table)
= 0.0228

Calculating percentiles - using tables

	Second decimal place of Z									
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

Practice

- $X \sim N(8, 5)$

(i) Find $P(X<8.6)$
(ii) Find $P(X>8.6)$
(iii) $P(8.6<x<10)$
(iv) $P(7<x<8.6)$
(v) $P(6<x<7)$

Answers: $X \sim N(8, 5)$

(i) Find $P(x<8.6)$
$= P(z<(8.6-8)/5)=P(z<0.12)=0.5478$
(ii)Find $P(X>8.6)$
$=1-P(x<8.6)=1-0.5478=0.4522$
(iii)$P(8.6<x<10)$
$=P((8.6-8)/5<z<(10-8)/5)=P(0.12<z<0.4)$
$=F(0.4)-F(0.12)=0.6554-0.5478=0.1076$
(iv) $P(7<x<8.6)$
$=P((7-8)/5 <z<(8.6-8)/5)=P(-0.2<z<0.12)$
$=F(0.12)-F(-0.2)=F(0.12)-[1-F(0.2)]$
$=0.5478-[1-0.5793]=0.1271$
(v)$P(6<x<7)$
$=P((6-8)/5<z<(7-8)/5)=P(-0.4<z<-0.2)$
$=P(0.2<z<0.4)=F(0.4)-F(0.2)$
$= 0.6654-0.5793 = 0.0761$

Quality control

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 fails the quality control inspection. What percent of bottles have less than 35.8 ounces of ketchup?

Quality control

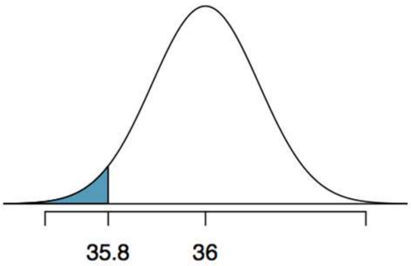
At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle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

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

Quality control

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 fails the quality control inspection. What percent of bottles have less than 35.8 ounces of ketchup?

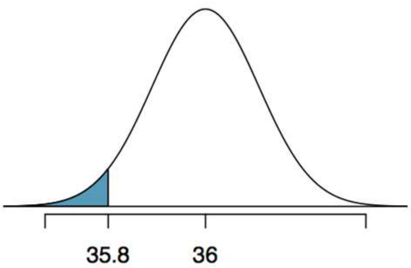
- *Let X = amount of ketchup in a bottle: X ~ N(μ = 36, σ = 0.11)*



Quality control

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 fails the quality control inspection. What percent of bottles have less than 35.8 ounces of ketchup?

- *Let X = amount of ketchup in a bottle: X ~ N(μ = 36, σ = 0.11)*



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

=P(z<-1.82)=P(z>1.82)=1-P(z<1.82)
=1-0.9656=0.0344

Practice

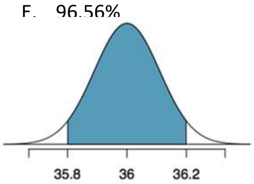
What percent of bottles pass the quality control inspection?

A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%

Practice

What percent of bottles pass the quality control inspection?

A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%

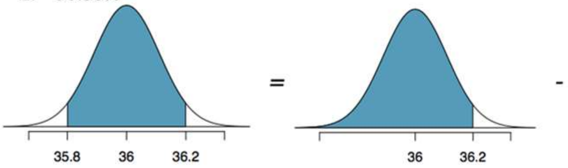


=

Practice

What percent of bottles pass the quality control inspection?

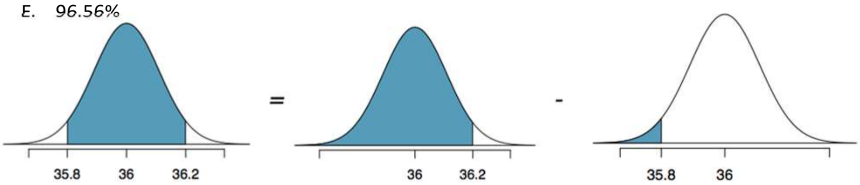
A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%



Practice

What percent of bottles pass the quality control inspection?

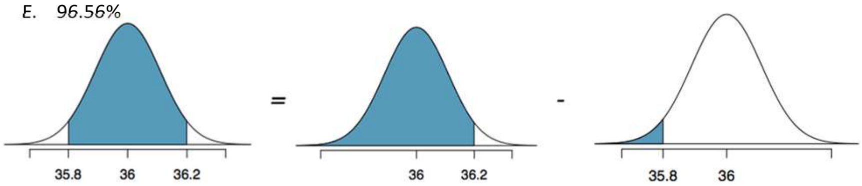
A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%



Practice

What percent of bottles pass the quality control inspection?

A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%

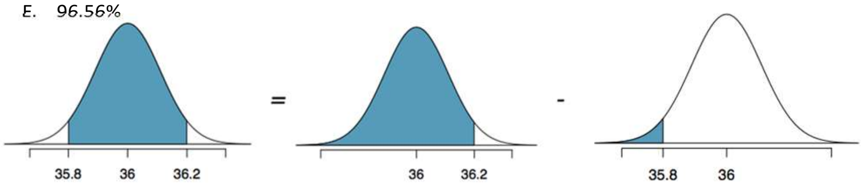


$$Z_{35.8} = \frac{35.8 - 36}{0.11} = -1.82$$

Practice

What percent of bottles pass the quality control inspection?

A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%



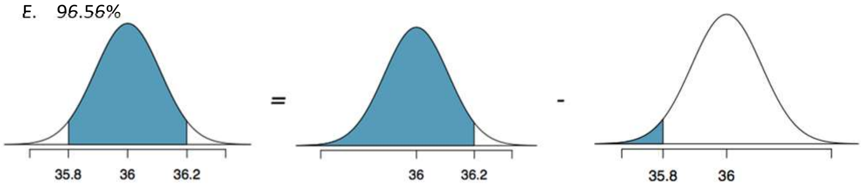
$$Z_{35.8} = \frac{35.8 - 36}{0.11} = -1.82$$

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

Practice

What percent of bottles pass the quality control inspection?

A. 1.82%
B. 3.44%
C. 6.88%
D. 93.12%
E. 96.56%



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

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

$$P(35.8 < X < 36.2) = P(-1.82 < Z < 1.82) = 0.9656 - 0.0344 = 0.9312$$

Finding the value of x for a known probability

Step1: Find z value for known probability from the table.

Step2: convert z to x units using

$$x = \mu + z\sigma$$

Practice

- X~N(8,5)
- Find the value of x so that only 20% of all values are below this x value.

⇒Since z0 value for P(z<z0)=0.2 is not on the table, find instead P(z<z0')=0.8. z0' = -(z0). [will be illustrated on board].
⇒Find the nearest value F(z)=0.7995 for z=0.84, F(z)=0.8023 for z = 0.85. Since F(z)=0.7995 is closer to F(z)=0.8, choose z = 0.84.
⇒Since z0= -(z0')=-0.84, x = 8+(-0.84)*5=3.8.

Practice

$$x = \mu + z\sigma$$

- X~N(60, 15)
- Find the cutoff point for the top 10% of all observation.

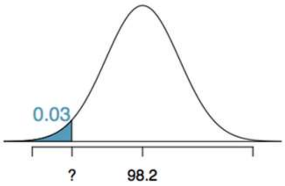
⇒P(z<z0)=0.9. Look for F(z) value which is the closest to 0.9. z0=1.28.
⇒x=60+1.28*15=79.2.

Finding cutoff points

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

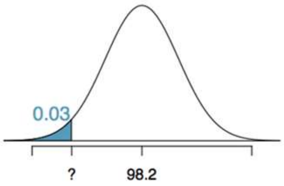
Finding cutoff points

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



Finding cutoff points

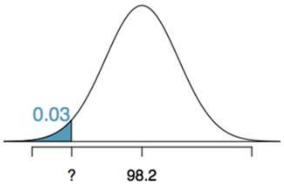
Body temperatures of healthy humans are distributed nearly normally with mean 98.2°F and standard deviation 0.73°F. What is the cutoff for the lowest 3% of human body temperatures?



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

Finding cutoff points

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?

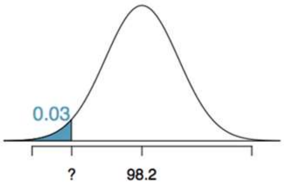


0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	−1.7

$$P(X < x) = 0.03 \rightarrow P(Z < -1.88) = 0.03$$

Finding cutoff points

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



P(X<x)=P(Z<z0)=0.03. P(Z>-z0)=0.03.
P(Z<z0)=0.97. Look for F(z) = 0.97 or closest.
-z0=1.88. z0=-1.88. X=98.2-1.88*0.73=
96.8276

F=>C conversion (extra content)
= (96.8276-32)/1.8= 36.01.

$$P(X < x) = 0.03 \rightarrow P(Z < -1.88) = 0.03$$

$$Z = \frac{obs - mean}{SD} \rightarrow \frac{x - 98.2}{0.73} = -1.88$$

$$x = 98.2 - 1.88 * 0.73 = 96.83$$

Practice

Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF
B. 99.1ºF
C. 99.4ºF
D. 99.6ºF

Practice

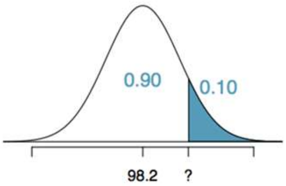
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

B. 99.1ºF

C. 99.4ºF

D. 99.6ºF



Practice

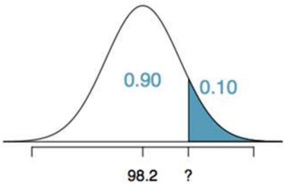
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

B. 99.1ºF

C. 99.4ºF

D. 99.6ºF



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

Practice

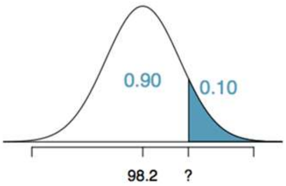
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

B. 99.1ºF

C. 99.4ºF

D. 99.6ºF



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

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

Practice

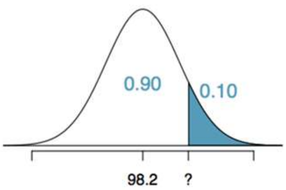
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

B. 99.1ºF

C. 99.4ºF

D. 99.6ºF



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

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

$$Z = \frac{obs - mean}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

Practice

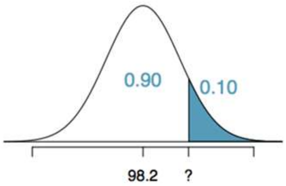
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3ºF

B. 99.1ºF

C. 99.4ºF

D. 99.6ºF



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

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

$$Z = \frac{obs - mean}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

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