Practice

Chapter 3: Distributions of Random Variables

Some slides developed by Mine Çetinkaya-Rundel of OpenIntro

Practice

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

A. 97.3^{o}F

B. 99.1^{o}F

C. 99.4^{o}F

D. 99.6^{o}F

Practice

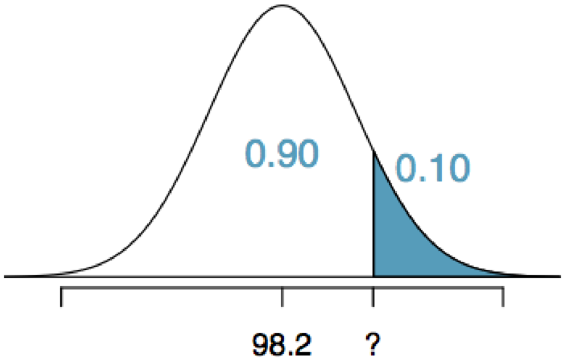
Body temperatures of healthy humans are distributed nearly normally with mean 98.2^{o}F and standard deviation 0.73^{o}F. What is the cutoff for the highest 10% of human body temperatures?

A. 97.3^{o}F

B. 99.1^{o}F

C. 99.4^{o}F

D. 99.6^{o}F



Practice

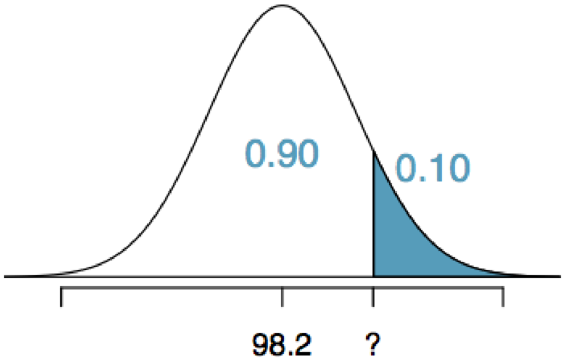
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

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

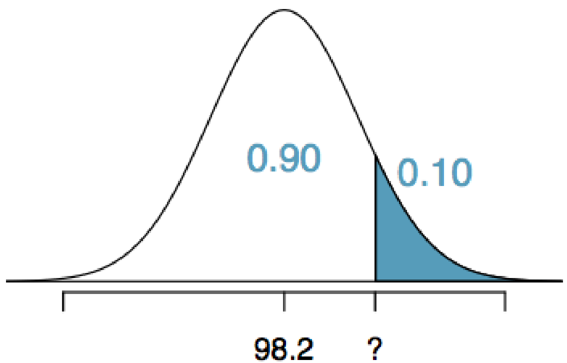
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

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

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

Practice

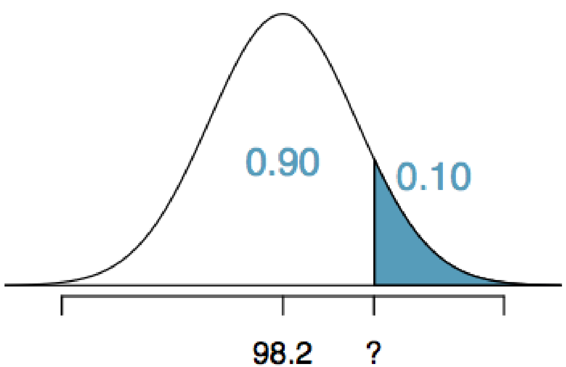
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

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

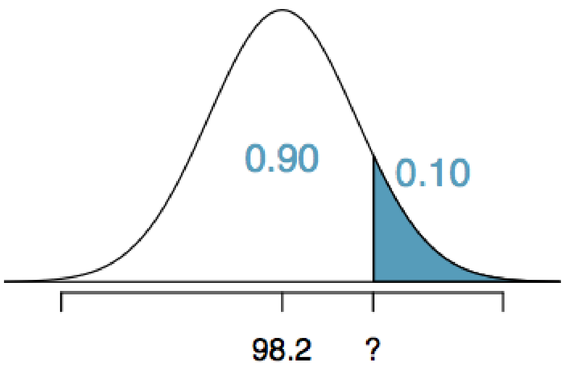
Body temperatures of healthy humans are distributed nearly normally with mean 98.2ºF and standard deviation 0.73ºF. What is the cutoff for the highest 10% of human body temperatures?

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

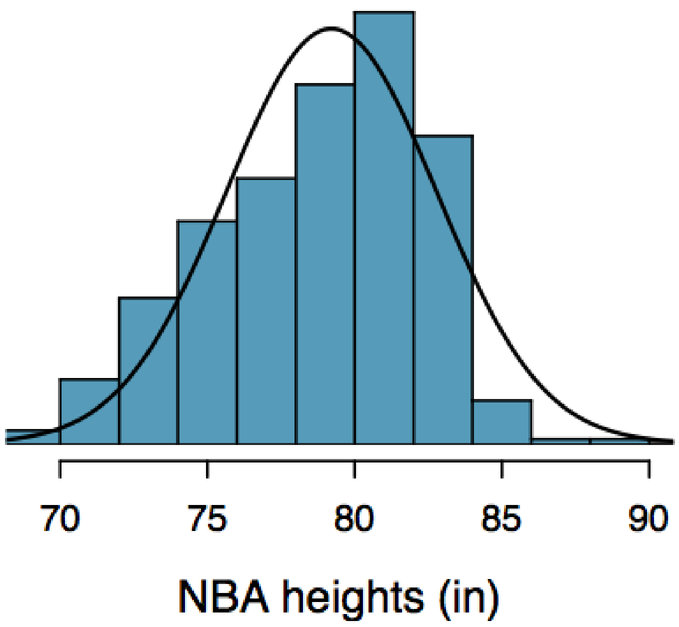
$$P(X > x) = 0.10 \rightarrow P(Z < 1.28) = 0.90$$

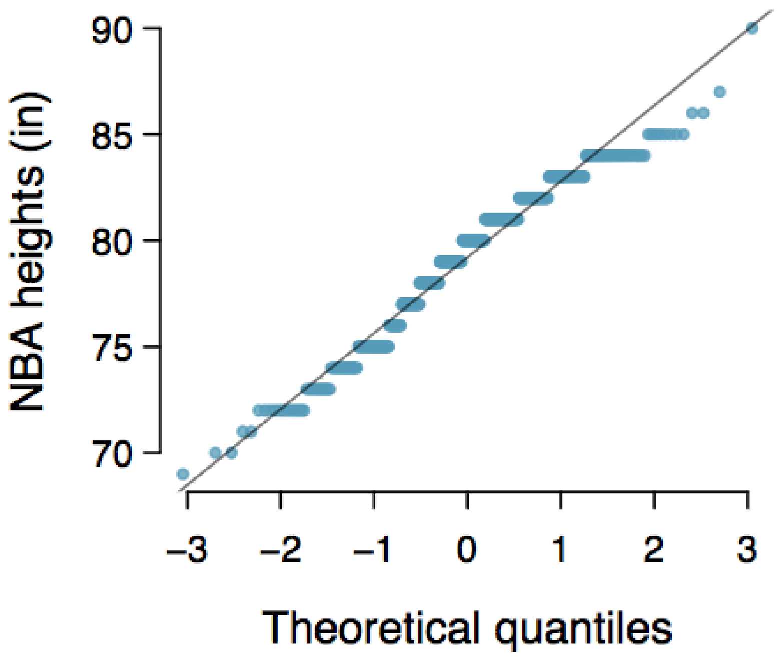
$$Z = \frac{obs - mean}{SD} \rightarrow \frac{x - 98.2}{0.73} = 1.28$$

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

Practice

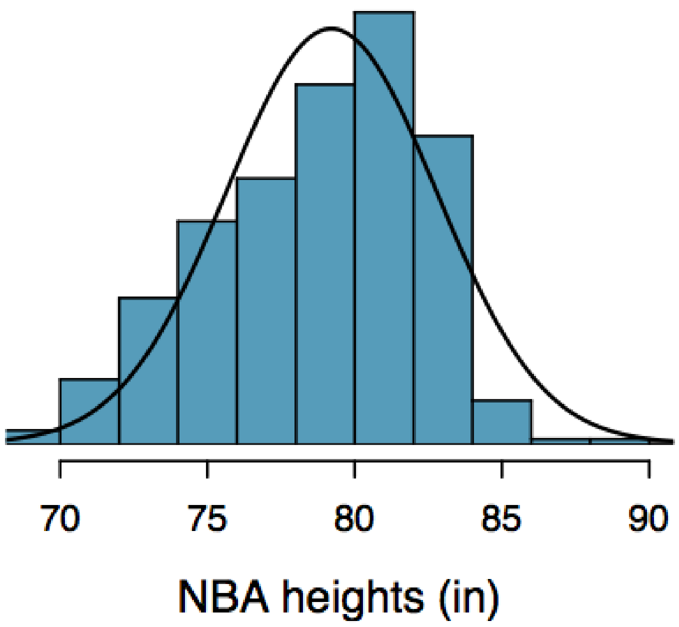
Below is a histogram and normal probability plot for the NBA heights from the 2008-2009 season. Do these data appear to follow a normal distribution?

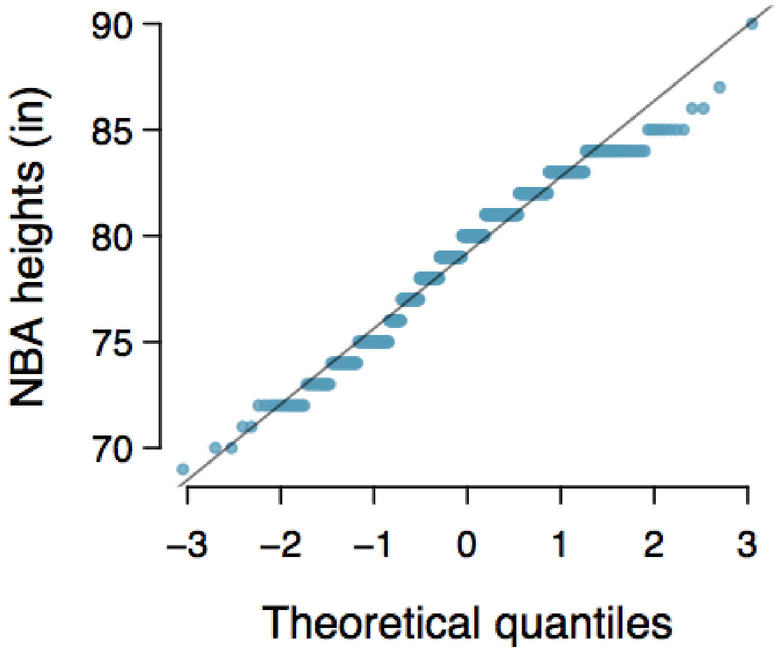




Practice

Below is a histogram and normal probability plot for the NBA heights from the 2008-2009 season. Do these data appear to follow a normal distribution?





Why do the points on the normal probability have jumps?

Practice

Can we calculate the probability of rolling a 6 for the first time on the 6^{th} roll of a die using the geometric distribution? Note that what was a success (rolling a 6) and what was a failure (not rolling a 6) are clearly defined and one or the other must happen for each trial.

1. no, on the roll of a die there are more than 2 possible outcomes
2. yes, why not

Practice

Can we calculate the probability of rolling a 6 for the first time on the 6th roll of a die using the geometric distribution? Note that what was a success (rolling a 6) and what was a failure (not rolling a 6) are clearly defined and one or the other must happen for each trial.

A. no, on the roll of a die there are more than 2 possible outcomes

B. *yes, why not*

$$P(6 \text{ on the } 6^{th} \text{ roll}) = \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right) \approx 0.067$$

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) There are $n - 1$ ways of getting $n - 1$ successes in n trials, $\binom{n}{n-1} = n - 1$.

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) *There are $n - 1$ ways of getting $n - 1$ successes in n trials, $\binom{n}{n-1} = n - 1$.*

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

A. the trials must be independent
B. the number of trials, *n*, must be fixed
C. each trial outcome must be classified as a *success* or a *failure*
D. the number of desired successes, *k*, must be greater than the number of trials
E. the probability of success, *p*, must be the same for each trial

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

A. the trials must be independent
B. the number of trials, *n*, must be fixed
C. each trial outcome must be classified as a *success* or a *failure*
D. *the number of desired successes, k, must be greater than the number of trials*
E. the probability of success, *p*, must be the same for each trial

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

A. pretty high
B. pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

A. pretty high
B. pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$

(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$

(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2$

(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$

(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$

(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2 = 45 \times 0.262^8 \times 0.738^2 = 0.0005$

(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes

B. No

	Excellent	Good	Only fair	Poor	Total excellent/ good
	%	%	%	%	%
Independent private school	31	47	13	2	78
Parochial or church-related schools	21	48	18	5	69
Charter schools	17	43	23	5	60
Home schooling	13	33	30	14	46
Public schools	5	32	42	19	37

Gallup, Aug. 9-12, 2012

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Method 1: *Range of usual observations:* $130 \pm 2 \times 10.6 = (108.8, 151.2)$
100 is outside this range, so would be considered unusual.

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Method 1: *Range of usual observations:* $130 \pm 2 \times 10.6 = (108.8, 151.2)$
100 is outside this range, so would be considered unusual.

Method 2: *Z-score of observation:* $Z = \frac{x - mean}{SD} = \frac{100-130}{10.6} = -2.83$
100 is more than 2 SD below the mean, so would be considered unusual.

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

A. $n = 100$, $p = 0.95$
B. $n = 25$, $p = 0.45$
C. $n = 150$, $p = 0.05$
D. $n = 500$, $p = 0.015$

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

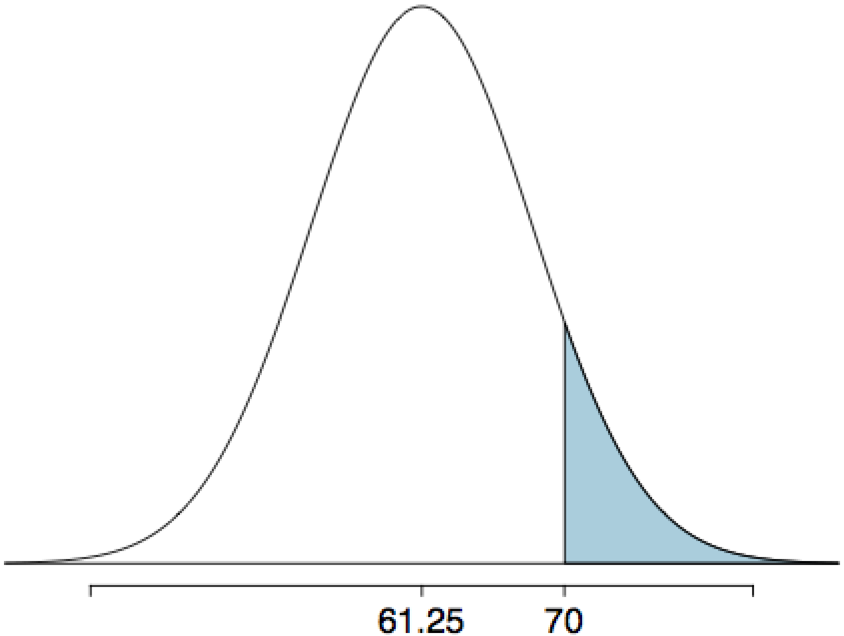
$$P(X \geq 70) = P(K = 70 \text{ or } K = 71 \text{ or } K = 72 \text{ or } \dots \text{ or } K = 245)$$
$$= P(K = 70) + P(K = 71) + P(K = 72) + \dots + P(K = 245)$$

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

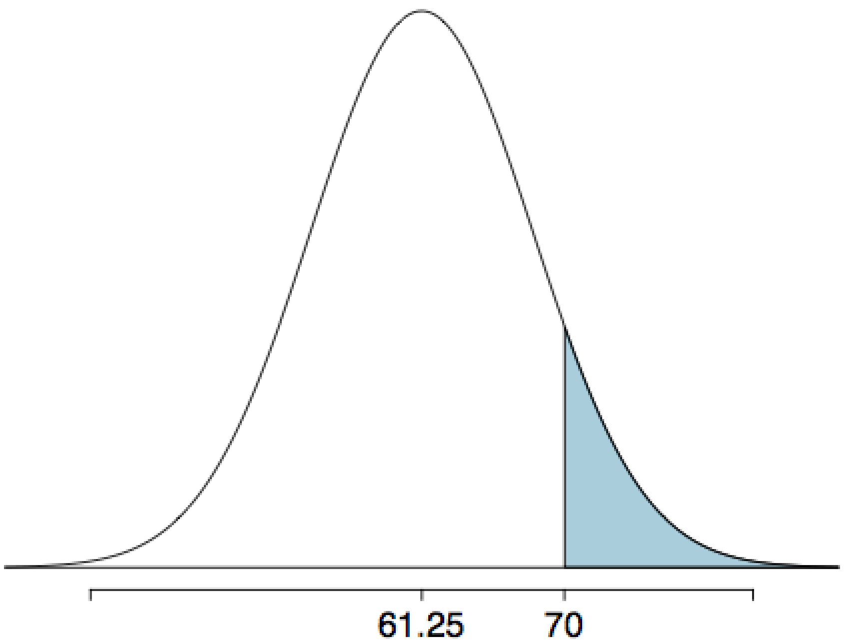
Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



Practice

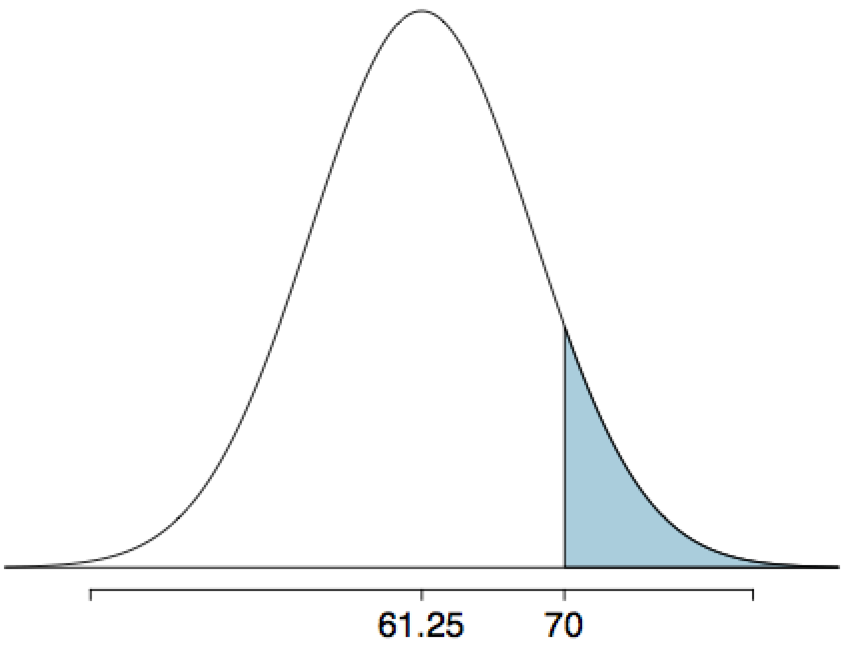
What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

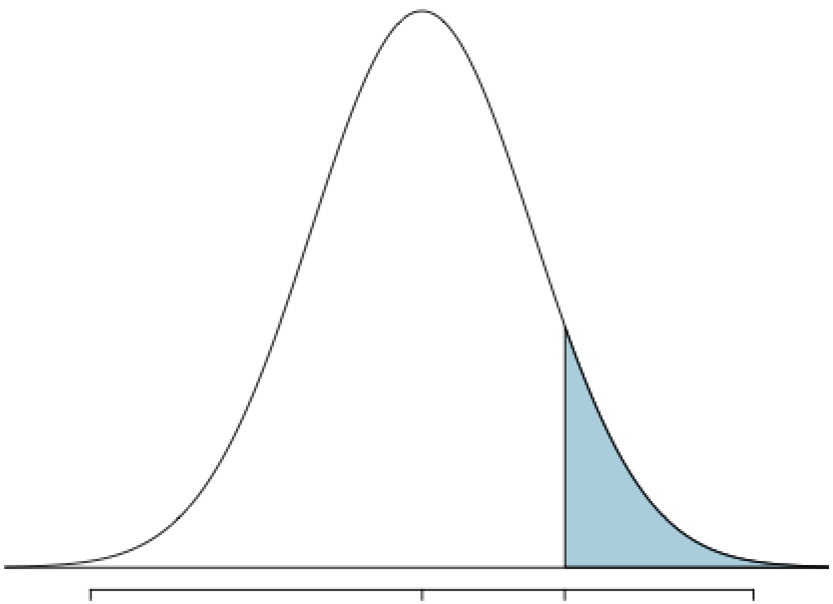


$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

	Second decimal place of Z				
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*

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

	Second decimal place of Z				
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

$$P(Z > 1.29) = 1 - 0.9015 = 0.0985$$

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

$$P(X \geq 70) = P(K = 70 \text{ or } K = 71 \text{ or } K = 72 \text{ or } \dots \text{ or } K = 245)$$
$$= P(K = 70) + P(K = 71) + P(K = 72) + \dots + P(K = 245)$$

This seems like an awful lot of work...

Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

A. $n = 100$, $p = 0.95$
B. $n = 25$, $p = 0.45 \rightarrow 25 \times 0.45 = 11.25$, $25 \times 0.55 = 13.75$
C. $n = 150$, $p = 0.05$
D. $n = 500$, $p = 0.015$

The birthday problem

What is the probability that 2 randomly chosen people share a birthday?

Pretty low, $1 / 365 \approx 0.0027$

What is the probability that at least 2 people out of 366 people share a birthday?

Exactly 1! (Excluding the possibility of a leap year birthday.)

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$
\begin{aligned}
P(no\ matches) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\
&= \frac{365!}{365^{121} \times (365 - 121)!} \\
&= \frac{121! \times \binom{365}{121}}{365^{121}} \approx 0
\end{aligned}
$$

$$P(at\ least\ 1\ match) \approx 1$$

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese.
Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.
- Or more formally, $\mu = np$ = 100 x 0.262 = 26.2.
- But this doesn't mean in every random sample of 100 people exactly 26.2 will be obese. In fact, that's not even possible. In some samples this value will be less, and in others more. How much would we expect this value to vary?

Shapes of binomial distributions

For this activity you will use a web applet. Go to http://socr.stat.ucla.edu/htmls/SOCR_Experiments.html and choose Binomial coin experiment in the drop down menu on the left.

- Set the number of trials to 20 and the probability of success to 0.15. Describe the shape of the distribution of number of successes.
- Keeping *p* constant at 0.15, determine the minimum sample size required to obtain a unimodal and symmetric distribution of number of successes. Please submit only one response per team.
- Further considerations:
  - What happens to the shape of the distribution as *n* stays constant and *p* changes?
  - What happens to the shape of the distribution as *p* stays constant and *n* changes?

Six sigma

The term *six sigma process* comes from the notion that if one has six standard deviations between the process mean and the nearest specification limit, as shown in the graph, practically no items will fail to meet specifications.

$$6\sigma$$

http://en.wikipedia.org/wiki/Six_Sigma

An analysis of Facebook users

A recent study found that “Facebook users get more than they give”. For example:

1. 40% of Facebook users in our sample made a friend request, but 63% received at least one request
2. Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content ``liked'' an average of 20 times
3. Users sent 9 personal messages, but received 12
4. 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx

An analysis of Facebook users

A recent study found that “Facebook users get more than they give”. For example:

1. 40% of Facebook users in our sample made a friend request, but 63% received at least one request
2. Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content ``liked'' an average of 20 times
3. Users sent 9 personal messages, but received 12
4. 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

Power users contribute much more content than the typical user.

http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx