Chapter 18 – Sampling Distribution Models

1. **Send money.**

   All of the histograms are centered around $p = 0.05$. As n gets larger, the shape of the histograms get more unimodal and symmetric, approaching a Normal model, while the variability in the sample proportions decreases.

2. **Character recognition.**

   All of the histograms are centered around $p = 0.85$. As n gets larger, the shapes of the histograms get more unimodal and symmetric, approaching a Normal model, while the variability in the sample proportions decreases.

3. **Send money, again.**

   a)

n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical standard deviation
20	0.0497	0.05	0.0479	$\sqrt{(0.05)(0.95)/20} \approx 0.0487$
50	0.0516	0.05	0.0309	$\sqrt{(0.05)(0.95)/50} \approx 0.0308$
100	0.0497	0.05	0.0215	$\sqrt{(0.05)(0.95)/100} \approx 0.0218$
200	0.0501	0.05	0.0152	$\sqrt{(0.05)(0.95)/200} \approx 0.0154$

   b) All of the values seem very close to what we would expect from theory.

   c) The histogram for $n = 200$ looks quite unimodal and symmetric. We should be able to use the Normal model.

   d) The success/failure condition requires np and nq to both be at least 10, which is not satisfied until $n = 200$ for $p = 0.05$. The theory supports the choice in part c.

4. **Character recognition, again.**

   a)

n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical standard deviation
20	0.8481	0.85	0.0803	$\sqrt{(0.85)(0.15)/20} \approx 0.0799$
50	0.8507	0.85	0.0509	$\sqrt{(0.85)(0.15)/50} \approx 0.0505$
75	0.8481	0.85	0.0406	$\sqrt{(0.85)(0.15)/75} \approx 0.0412$
100	0.8488	0.85	0.0354	$\sqrt{(0.85)(0.15)/100} \approx 0.0357$

   b) All of the values seem very close to what we would expect from theory.

   c) Certainly, the histogram for $n = 100$ is unimodal and symmetric, but the histogram for $n = 75$ looks nearly Normal, too. We should be able to use the Normal model for either.

d) The success/failure condition requires np and nq to both be at least 10, which is satisfied for both $n = 75$ and $n = 100$ when $p = 0.85$. The theory supports the choice in part c.

5. **Coin tosses.**

a) The histogram of these proportions is expected to be symmetric, but **not** because of the Central Limit Theorem. The sample of 16 coin flips is not large. The distribution of these proportions is expected to be symmetric because the probability that the coin lands heads is the same as the probability that the coin lands tails.

b) The histogram is expected to have its center at 0.5, the probability that the coin lands heads.

c) The standard deviation of data displayed in this histogram should be approximately equal to the standard deviation of the sampling distribution model, $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{16}} = 0.125$.

d) The expected number of heads, $np = 16(0.5) = 8$, which is less than 10. The Success/Failure condition is not met. The Normal model is not appropriate in this case.

6. **M&M's.**

a) The histogram of the proportions of green candies in the bags would probably be skewed slightly to the right, for the simple reason that the proportion of green M&M's could never fall below 0 on the left, but has the potential to be higher on the right.

b) The Normal model cannot be used to approximate the histogram, since the expected number of green M&M's is $np = 50(0.10) = 5$, which is less than 10. The Success/Failure condition is not met.

c) The histogram should be centered around the expected proportion of green M&M's, at about 0.10.

d) The proportion should have standard deviation $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.1)(0.9)}{50}} \approx 0.042$.

7. **More coins.**

a) $\mu_{\hat{p}} = p = 0.5$ and $\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{25}} = 0.1$

About 68% of the sample proportions are expected to be between 0.4 and 0.6, about 95% are expected to be between 0.3 and 0.7, and about 99.7% are expected to be between 0.2 and 0.8.
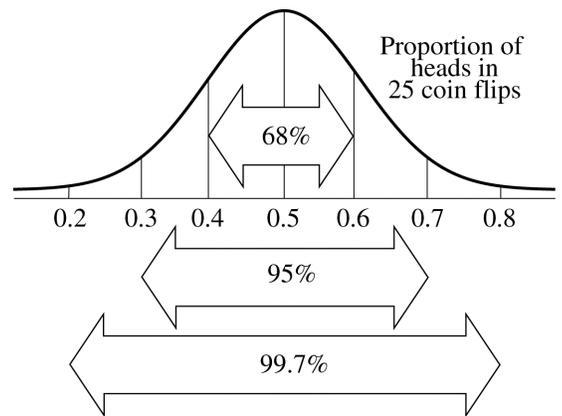

b) First of all, coin flips are independent of one another. There is no need to check the 10% Condition. Second, $np = nq = 12.5$, so both are greater than 10. The Success/Failure condition is met, so the sampling distribution model is $N(0.5, 0.1)$.

c) $\mu_{\hat{p}} = p = 0.5$ and $\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{64}} = 0.0625$

About 68% of the sample proportions are expected to be between 0.4375 and 0.5625, about 95% are expected to be between 0.375 and 0.625, and about 99.7% are expected to be between 0.3125 and 0.6875.

Coin flips are independent of one another, and $np = nq = 32$, so both are greater than 10. The Success/Failure condition is met, so the sampling distribution model is $N(0.5, 0.0625)$.
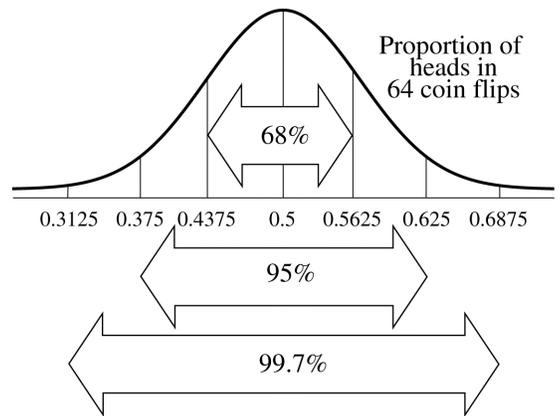
d) As the number of tosses increases, the sampling distribution model will still be Normal and centered at 0.5, but the standard deviation will decrease. The sampling distribution model will be less spread out.

8. **Bigger bag.**

a) **Randomization condition:** The 200 M&M's in the bag can be considered representative of all M&M's, and they are thoroughly mixed.

10% condition: 200 is certainly less than 10% of all M&M's.

Success/Failure condition: $np = 20$ and $nq = 180$ are both greater than 10.

b) The sampling distribution model is Normal, with:

$\mu_{\hat{p}} = p = 0.1$ and $\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.1)(0.9)}{200}} \approx 0.021$

About 68% of the sample proportions are expected to be between 0.079 and 0.121, about 95% are expected to be between 0.058 and 0.142, and about 99.7% are expected to be between 0.037 and 0.163.
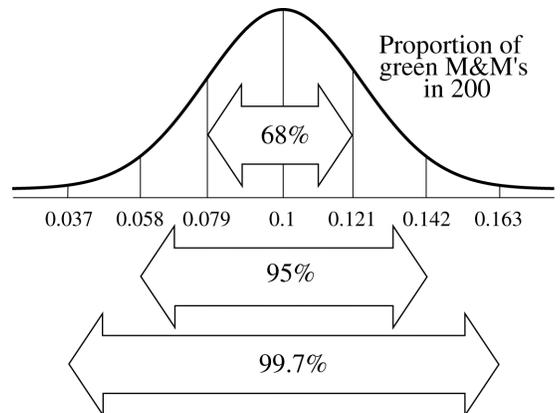
c) If the bags contained more candies, the sampling distribution model would still be Normal and centered at 0.1, but the standard deviation would decrease. The sampling distribution model would be less spread out.

9. **Just (un)lucky.**

For 200 flips, the sampling distribution model is Normal with $\mu_{\hat{p}} = p = 0.5$ and

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.5)(0.5)}{200}} \approx 0.0354$. Her sample proportion of $\hat{p} = 0.42$ is about 2.26 standard deviations below the expected proportion, which is unusual, but not extraordinary. According to the Normal model, we expect sample proportions this low or lower about 1.2% of the time.

10. Too many green ones?

For 500 candies, the sampling distribution model is Normal with $\mu_{\hat{p}} = p = 0.1$ and

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.1)(0.9)}{500}} \approx 0.01342$. The sample proportion of $\hat{p} = 0.12$ is about 1.49 standard deviations above the expected proportion, which is not at all unusual. According to the Normal model, we expect sample proportions this high or higher about 6.8% of the time.

11. Speeding.

a) $\mu_{\hat{p}} = p = 0.70$

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.7)(0.3)}{80}} \approx 0.051$.

About 68% of the sample proportions are expected to be between 0.649 and 0.751, about 95% are expected to be between 0.598 and 0.802, and about 99.7% are expected to be between 0.547 and 0.853.
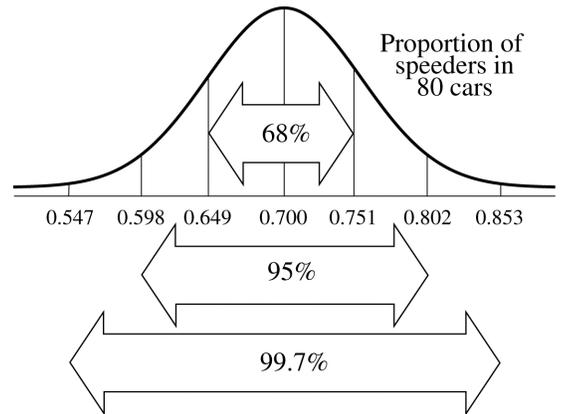
b) **Randomization condition:** The sample may not be representative. If the flow of traffic is very fast, the speed of the other cars around may have some effect on the speed of each driver. Likewise, if traffic is slow, the police may find a smaller proportion of speeders than they expect.

10% condition: 80 cars represent less than 10% of all cars

Success/Failure condition: $np = 56$ and $nq = 24$ are both greater than 10.

The Normal model may not be appropriate. Use caution. (And don't speed!)

12. Smoking.

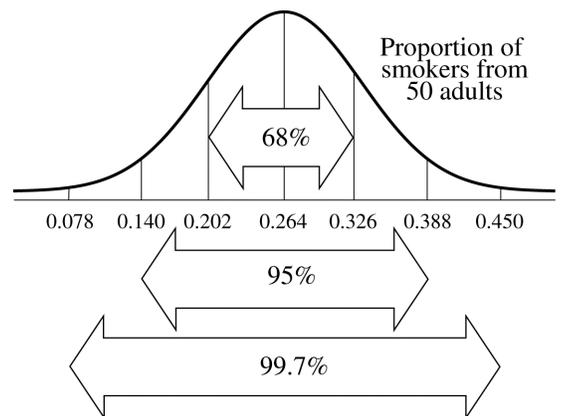**Randomization condition:** 50 people are selected at random

10% condition: 50 is less than 10% of all people.

Success/Failure condition: $np = 13.2$ and $nq = 36.8$ are both greater than 10.

The sampling distribution model is Normal, with:
$\mu_{\hat{p}} = p = 0.264$

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.264)(0.736)}{50}} \approx 0.062$

There is an approximate chance of 68% that between 20.2% and 32.6% of 50 people are smokers, an approximate chance of 95% that between 14.0% and 38.8% are smokers, and an approximate chance of 99.7% that between 7.8% and 45.0%are smokers.
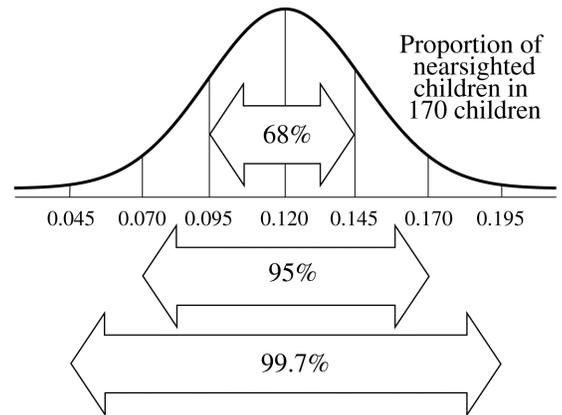
13. Vision.

a) **Randomization condition:** Assume that the 170 children are a representative sample of all children.

Proportion of nearsighted children in 170 children

10% condition: A sample of this size is less than 10% of all children.

Success/Failure condition: $np = 20.4$ and $nq = 149.6$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of 170 children who are nearsighted is $N(0.12, 0.025)$.

b) The Normal model is to the right.

c) They might expect that the proportion of nearsighted students to be within 2 standard deviations of the mean. According to the Normal model, this means they might expect between 7% and 17% of the students to be nearsighted, or between about 12 and 29 students.
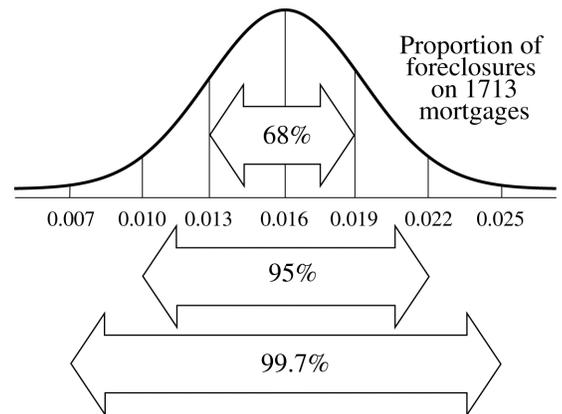
14. Mortgages.

a) **Randomization condition:** Assume that the 1713 mortgages are a representative sample of all mortgages.

Proportion of foreclosures on 1713 mortgages

10% condition: A sample of this size is less than 10% of all mortgages.

Success/Failure condition: $np = 27.7$ and $nq = 1703.3$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of foreclosures on 1713 mortgages is $N(0.016, 0.003)$.

b) The Normal model is to the right.

c) They might expect that the proportion of mortgage foreclosures to be within 2 standard deviations of the mean. According to the Normal model, this means they might expect between 1% and 2.2% of the mortgages to undergo foreclosure, or between about 17 and 38 foreclosures.

15. Loans.

a) $\mu_{\hat{p}} = p = 7\%$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.07)(0.93)}{200}} \approx 1.8\%$$

b) **Randomization condition:** Assume that the 200 people are a representative sample of all loan recipients.

   10% condition: A sample of this size is less than 10% of all loan recipients.

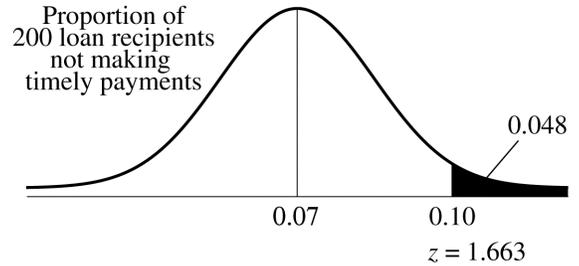   **Success/Failure condition:** np = 14 and nq = 186 are both greater than 10.

   Therefore, the sampling distribution model for the proportion of 200 loan recipients who will not make payments on time is $N(0.07, 0.018)$.

c) According to the Normal model, the probability that over 10% of these clients will not make timely payments is approximately 0.048.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\dfrac{pq}{n}}}$$

$$z = \frac{0.10 - 0.07}{\sqrt{\dfrac{(0.07)(0.93)}{200}}}$$

$$z \approx 1.663$$

Proportion of 200 loan recipients not making timely payments

0.048

0.07   0.10

$z = 1.663$

16. Contacts.

a) **Randomization condition:** 100 students are selected at random.

   10% condition: 100 is less than 10% of all of the students at the university, provided the university has more than 1000 students.
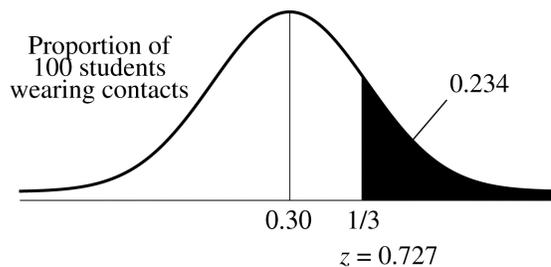
   Success/Failure condition: np = 30 and nq = 70 are both greater than 10.

   Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.30$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.30)(0.70)}{100}} \approx 0.046$$

b) According to the Normal model, the probability that more than one-third of the students in this sample wear contacts is approximately 0.234.

Proportion of 100 students wearing contacts

0.234

0.30   1/3

$z = 0.727$

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\dfrac{pq}{n}}}$$

$$z = \frac{\frac{1}{3} - 0.30}{\sqrt{\dfrac{(0.30)(0.70)}{100}}}$$

$$z \approx 0.727$$

17. Back to school?

   Randomization condition: We are considering random samples of 400 students who took the ACT.

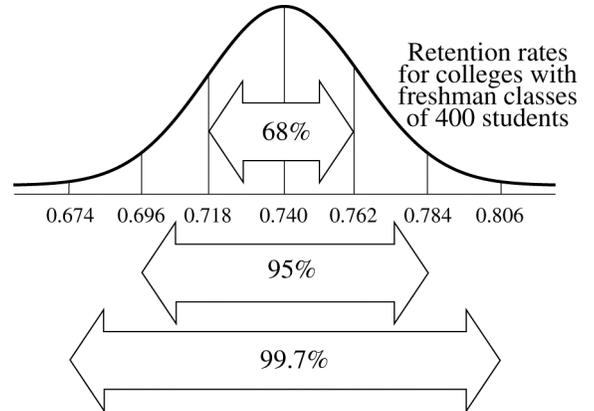   **10% Condition:** 400 students is less than 10% of all college students.

   Success/Failure condition: np = 296 and nq = 104 are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.74$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.74)(0.26)}{400}} \approx 0.022$$

According to the sampling distribution model, about 68% of the colleges are expected to have retention rates between 0.718 and 0.762, about 95% of the colleges are expected to have retention rates between 0.696 and 0.784, and about 99.7% of the colleges are expected to have retention rates between 0.674 and 0.806. However, the conditions for the use of this model may not be met. We should be cautious about making any conclusions based on this model.

18. Binge drinking.

Randomization condition: The students were selected randomly.

**10% condition:** 200 college students are less than 10% of all college students.

Success/Failure condition: $np = 88$ and $nq = 112$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.44$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.44)(0.56)}{200}} \approx 0.035$$

According to the sampling distribution model, about 68% of samples of 200 students are expected to have binge drinking proportions between 0.405 and 0.475, about 95% between 0.370 and 0.510, and about 99.7% between 0.335 and 0.545.
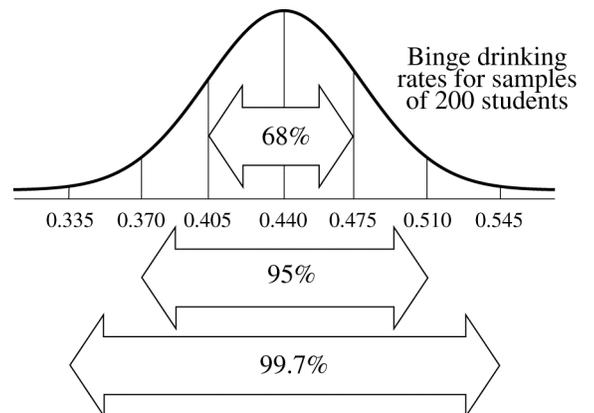
19. Back to school, again.

Provided that the students at this college are typical, the sampling distribution model for the retention rate, $\hat{p}$, is Normal with $\mu_{\hat{p}} = p = 0.74$ and standard deviation

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.74)(0.26)}{603}} \approx 0.018$$

This college has a right to brag about their retention rate. $522/603 = 86.6\%$ is over 7 standard deviations above the expected rate of 74%.

20. Binge sample.

Since the sample is random and the Success/Failure condition is met, the sampling distribution model for the binge drinking rate, $\hat{p}$, is Normal with $\mu_{\hat{p}} = p = 0.44$ and

standard deviation $\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.44)(0.56)}{244}} \approx 0.032$

The binge drinking rate at this college is lower than the national result, but not extremely low. $96/244 = 39.3\%$ is only about 1.5 standard deviations below the national rate of 44%.

21. Polling.

Randomization condition: We must assume that the 400 voters were polled randomly.

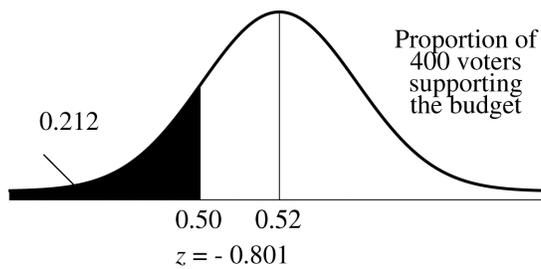**10% condition:** 400 voters polled represent less than 10% of potential voters.

Success/Failure condition: $np = 208$ and $nq = 192$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.52$

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.52)(0.48)}{400}} \approx 0.025$

According to the Normal model, the probability that the newspaper's sample will lead them to predict defeat (that is, predict budget support below 50%) is approximately 0.212.

Proportion of 400 voters supporting the budget

0.212

0.50   0.52

$z = -0.801$

$z = \dfrac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\dfrac{pq}{n}}}$

$z = \dfrac{0.50 - 0.52}{\sqrt{\dfrac{(0.52)(0.48)}{400}}}$

$z \approx -0.801$

22. Seeds.

Randomization condition: We must assume that the 160 seeds in a pack are a random sample. Since seeds in each pack may *not* be a random sample, proceed with caution.

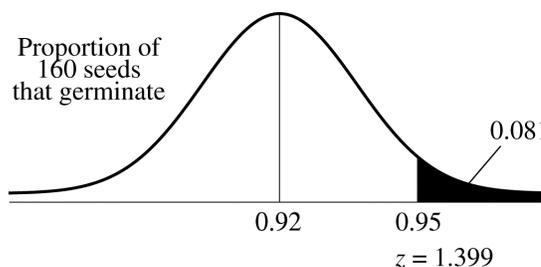**10% condition:** The 160 seeds represent less than 10% of all seeds.

Success/Failure condition: $np = 147.2$ and $nq = 12.8$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.92$

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.92)(0.08)}{160}} \approx 0.0215$

According to the Normal model, the probability that more than 95% of the seeds will germinate is approximately 0.081.

Proportion of 160 seeds that germinate

0.081

0.92   0.95

$z = 1.399$

$z = \dfrac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\dfrac{pq}{n}}}$

$z = \dfrac{0.95 - 0.92}{\sqrt{\dfrac{(0.92)(0.08)}{160}}}$

$z \approx 1.399$

23. Apples.

Randomization condition: A random sample of 150 apples is taken from each truck.

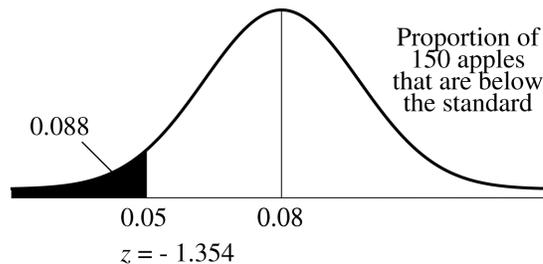**10% condition:** 150 is less than 10% of all apples.

Success/Failure Condition: $np = 12$ and $nq = 138$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.08$

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.08)(0.92)}{150}} \approx 0.0222$

According to the Normal model, the probability that less than 5% of the apples in the sample are unsatisfactory is approximately 0.088.

$z = \dfrac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\dfrac{pq}{n}}}$

$z = \dfrac{0.05 - 0.08}{\sqrt{\dfrac{(0.08)(0.92)}{150}}}$

$z \approx -1.354$

Proportion of 150 apples that are below the standard

0.088

0.05   0.08

$z = -1.354$

24. Genetic Defect.

Randomization condition: We will assume that the 732 newborns are representative of all newborns.

10% condition: These 732 newborns certainly represent less than 10% of all newborns.

Success/Failure condition: $np = 29.28$ and $nq = 702.72$ are both greater than 10.

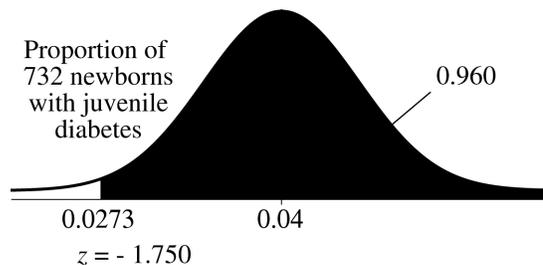Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$\mu_{\hat{p}} = p = 0.04$

$\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.04)(0.96)}{732}} \approx 0.0072$

In order to get the 20 newborns for the study, the researchers hope to find at least

$\hat{p} = \dfrac{20}{732} \approx 0.0273$ as the proportion of newborns in the sample with

juvenile diabetes.

According to the Normal model, the probability that the researchers find at least 20 newborns with juvenile diabetes is approximately 0.960.

$z = \dfrac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\dfrac{pq}{n}}}$

$z = \dfrac{^{20}/_{732} - 0.04}{\sqrt{\dfrac{(0.04)(0.96)}{732}}}$

$z \approx -1.750$

Proportion of 732 newborns with juvenile diabetes

0.960

0.0273   0.04

$z = -1.750$

25. Nonsmokers.

Randomization condition: We will assume that the 120 customers (to fill the restaurant to capacity) are representative of all customers.

10% condition: 120 customers represent less than 10% of all potential customers.

Success/Failure condition: $np = 72$ and $nq = 48$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.60$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.60)(0.40)}{120}} \approx 0.0447$$

Answers may vary. We will use 3 standard deviations above the expected proportion of customers who demand nonsmoking seats to be "very sure".

$$\mu_{\hat{p}} + 3\left(\sqrt{\frac{pq}{n}}\right) \approx 0.60 + 3(0.0447) \approx 0.734$$

Since $120(0.734) = 88.08$, the restaurant needs at least 89 seats in the nonsmoking section.

26. Meals.

Randomization condition: We will assume that the 180 customers are representative of all customers.

10% condition: 180 customers represent less than 10% of all potential customers.

Success/Failure condition: $np = 36$ and $nq = 144$ are both greater than 10.

Therefore, the sampling distribution model for $\hat{p}$ is Normal, with:

$$\mu_{\hat{p}} = p = 0.20$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.20)(0.80)}{180}} \approx 0.0298$$

Answers may vary. We will use 2 standard deviations above the expected proportion of customers who order the steak special to be "pretty sure".

$$\mu_{\hat{p}} + 2\left(\sqrt{\frac{pq}{n}}\right) \approx 0.20 + 2(0.0298) \approx 0.2596$$

Since $180(0.2596) = 46.728$, the chef needs at least 47 steaks on hand.

27. Sampling.

a) The sampling distribution model for the sample mean is $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

b) If we choose a larger sample, the mean of the sampling distribution model will remain the same, but the standard deviation will be smaller.

28. Sampling, part II.

a) The sampling distribution model for the sample mean will be skewed to the left as well, centered at μ, with standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

b) When the sample size is increased, the sampling distribution model becomes more Normal in shape, centered at μ, with standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

c) As we make the sample larger the distribution of data in the sample should more closely resemble the shape, center, and spread of the population.

29. Waist size.

a) The distribution of waist size of 250 men in Utah is unimodal and slightly skewed to the right. A typical waist size is approximately 36 inches, and the standard deviation in waist sizes is approximately 4 inches.

b) All of the histograms show distributions of sample means centered near 36 inches. As n gets larger the histograms approach the Normal model in shape, and the variability in the sample means decreases. The histograms are fairly Normal by the time the sample reaches size 5.

30. CEO compensation.

a) The distribution of total compensation for the CEOs for the 800 largest U.S. companies is unimodal, but strongly skewed to the right with several large outliers.

b) All of the histograms are centered near $10,000,000. As n gets larger, the variability in sample means decreases, and histograms approach the Normal shape. However, they are still visibly skewed to the right, with the possible exception of the histogram for $n = 200$.

c) This rule of thumb doesn't seem to be true for highly skewed distributions.

31. Waist size revisited.

a)

n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical standard deviation
2	36.314	36.33	2.855	$4.019 / \sqrt{2} \approx 2.842$
5	36.314	36.33	1.805	$4.019 / \sqrt{5} \approx 1.797$
10	36.341	36.33	1.276	$4.019 / \sqrt{10} \approx 1.271$
20	36.339	36.33	0.895	$4.019 / \sqrt{20} \approx 0.897$

b) The observed values are all very close to the theoretical values.

c) For samples as small as 5, the sampling distribution of sample means is unimodal and symmetric. The Normal model would be appropriate.

d) The distribution of the original data is nearly unimodal and symmetric, so it doesn't take a very large sample size for the distribution of sample means to be approximately Normal.

32. CEOs revisited.

a)

n	Observed mean	Theoretical mean	Observed st. dev.	Theoretical standard deviation
30	10251.73	10307.31	3359.64	$17964.62 / \sqrt{30} \approx 3279.88$
50	10343.93	10307.31	2483.84	$17964.62 / \sqrt{50} \approx 2540.58$
100	10329.94	10307.31	1779.18	$17964.62 / \sqrt{100} \approx 1796.46$
200	10340.37	10307.31	1260.79	$17964.62 / \sqrt{200} \approx 1270.29$

b) The observed values are all very close to the theoretical values.

c) All the sampling distributions are still quite skewed, with the possible exception of the sampling distribution for $n = 200$, which is still somewhat skewed. The Normal model would not be appropriate.

d) The distribution of the original data is strongly skewed, so it will take a very large sample size before the distribution of sample means is approximately Normal.

33. GPAs.

Randomization condition: Assume that the students are randomly assigned to seminars.

Independence assumption: It is reasonable to think that GPAs for randomly selected students are mutually independent.

10% condition: The 25 students in the seminar certainly represent less than 10% of the population of students.

Large Enough Sample condition: The distribution of GPAs is roughly unimodal and symmetric, so the sample of 25 students is large enough.
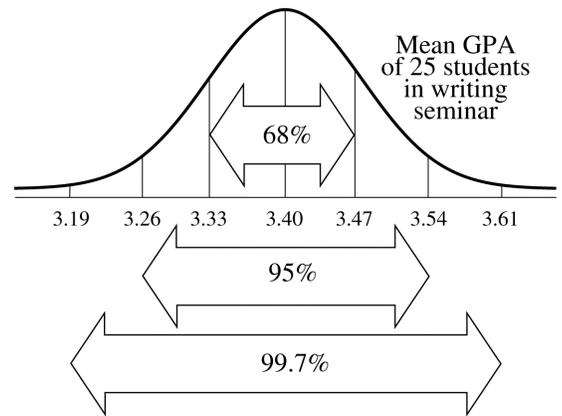
The mean GPA for the freshmen was $\mu = 3.4$, with standard deviation $\sigma = 0.35$. Since the conditions are met, the Central Limit Theorem tells us that we can model the sampling distribution of the mean GPA with a Normal model, with $\mu_{\bar{y}} = 3.4$ and standard deviation

$$\sigma(\bar{y}) = \frac{0.35}{\sqrt{25}} \approx 0.07.$$

The sampling distribution model for the sample mean GPA is approximately $N(3.4, 0.07)$.

34. Home values.

Randomization condition: Homes were selected at random.

Independence assumption: It is reasonable to think that assessments for randomly selected homes are mutually independent.

10% condition: The 100 homes in the sample certainly represent less than 10% of the population of all homes in the city. A small city will likely have more than 1,000 homes.

Large Enough Sample condition: A sample of 100 homes is large enough.

The mean home value was $\mu = \$140,000$, with standard deviation $\sigma = \$60,000$. Since the conditions are met, the Central Limit Theorem tells us that we can model the sampling distribution of the mean home value with a Normal model, with $\mu_{\bar{y}} = \$140,000$ and standard deviation

$$\sigma(\bar{y}) = \frac{60,000}{\sqrt{100}} = \$6000.$$

The sampling distribution model for the sample mean home values is approximately $N(140000,6000)$.
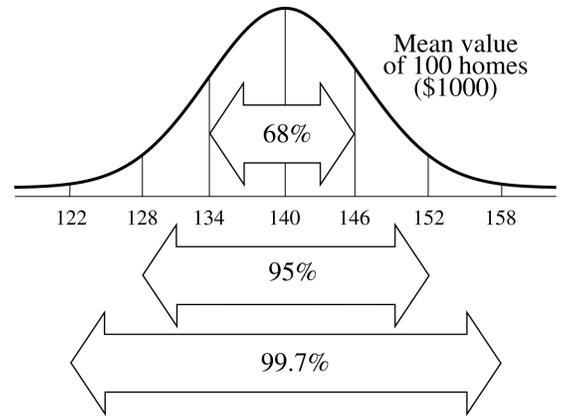
35. Lucky spot?

a) Smaller outlets have more variability than the larger outlets, just as the Central Limit Theorem predicts.

b) If the lottery is truly random, all outlets are equally likely to sell winning tickets.

36. Safe cities.

The standard deviation of the sampling model for the mean is $\dfrac{\sigma}{\sqrt{n}}$. So, cities in which the average is based on a smaller number of drivers will have greater variation in their averages and will be more likely to be both safest and least safe.
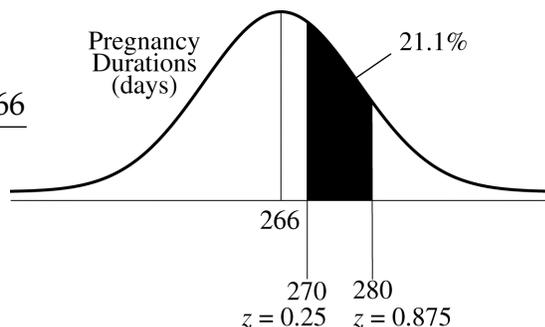
37. Pregnancy.

a)

$$z = \frac{y - \mu}{\sigma}$$

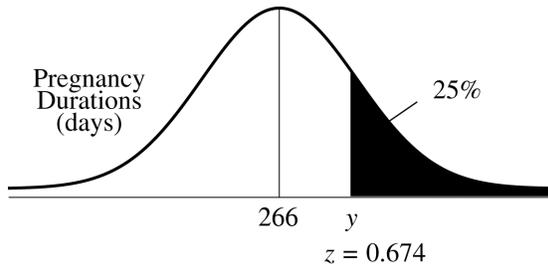$$z = \frac{270 - 266}{16}$$

$$z = 0.25$$

$$z = \frac{y - \mu}{\sigma}$$

$$z = \frac{280 - 266}{16}$$

$$z = 0.875$$

According to the Normal model, approximately 21.1% of all pregnancies are expected to last between 270 and 280 days.

b)

$$z = \frac{y - \mu}{\sigma}$$

$$0.674 = \frac{y - 266}{16}$$

$$y \approx 276.8 \text{ days}$$

According to the Normal model, the longest 25% of pregnancies are expected to last approximately 276.8 days or more.

c) **Randomization condition:** Assume that the 60 women the doctor is treating can be considered a representative sample of all pregnant women.

Independence assumption: It is reasonable to think that the durations of the patients' pregnancies are mutually independent.

10% condition: The 60 women that the doctor is treating certainly represent less than 10% of the population of all women.

Large Enough Sample condition: The sample of 60 women is large enough. In this case, any sample would be large enough, since the distribution of pregnancies is Normal.

The mean duration of the pregnancies was $\mu = 266$ days, with standard deviation $\sigma = 16$ days. Since the distribution of pregnancy durations is Normal, we can model the sampling distribution of the mean pregnancy duration with a Normal model, with
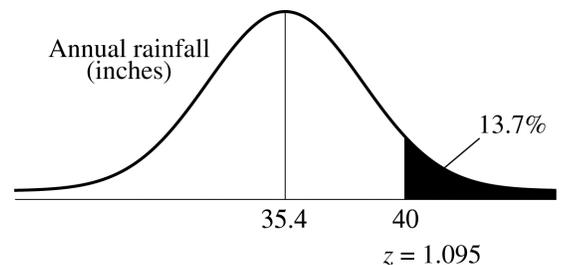
$\mu_{\bar{y}} = 266$ days and standard deviation $\sigma(\bar{y}) = \dfrac{16}{\sqrt{60}} \approx 2.07$ days.

d) According to the Normal model, the probability that the mean pregnancy duration is less than 260 days is 0.002.
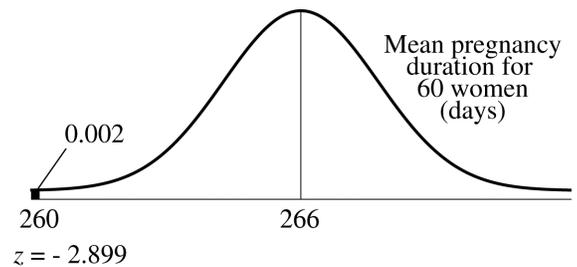
38. Rainfall.

a) According to the Normal model, Ithaca is expected to get more than 40 inches of rain in approximately 13.7% of years.
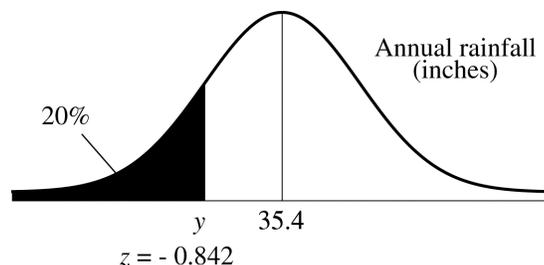
b) According to the Normal model, Ithaca is expected to get less than 31.9 inches of rain in driest 20% of years.

$$z = \frac{y - \mu}{\sigma}$$

$$-0.842 = \frac{y - 35.4}{4.2}$$

$$y \approx 31.9$$

Annual rainfall (inches)

20%

y 35.4

$z = -0.842$

c) **Randomization condition:** Assume that the 4 years in which the student was in Ithaca can be considered a representative sample of all years.

Independence assumption: It is reasonable to think that the rainfall totals for the years are mutually independent.

10% condition: The 4 years in which the student was in Ithaca certainly represent less than 10% of all years.

Large enough sample condition: The sample of 4 years is large enough. In this case, any sample would be large enough, since the distribution of annual rainfall is Normal.

The mean rainfall was $\mu = 35.4$ inches, with standard deviation $\sigma = 4.2$ inches. Since the distribution of yearly rainfall is Normal, we can model the sampling distribution of the mean annual rainfall with a Normal model, with $\mu_{\bar{y}} = 35.4$ inches and standard deviation

$$\sigma(\bar{y}) = \frac{4.2}{\sqrt{4}} = 2.1 \text{ inches}.$$

d) According to the Normal model, the probability that those four years averaged less than 30 inches of rain is 0.005.

Mean annual rainfall for 4 years (inches)

0.005

30 35.4

$z = -2.571$
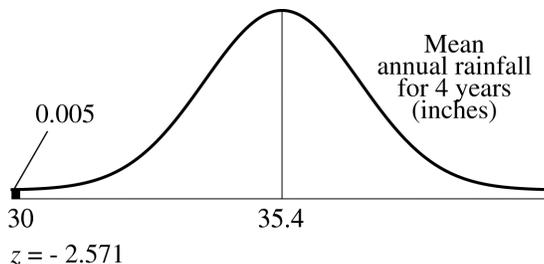
39. Pregnant again.

a) The distribution of pregnancy durations may be skewed to the left since there are more premature births than very long pregnancies. Modern practice of medicine stops pregnancies at about 2 weeks past normal due date by inducing labor or performing a Caesarean section.

b) We can no longer answer the questions posed in parts a and b. The Normal model is not appropriate for skewed distributions. The answer to part c is still valid. The Central Limit Theorem guarantees that the sampling distribution model is Normal when the sample size is large.

40. At work.

a) The distribution of length of time people work at a job is likely to be skewed to the right, because some people stay at the same job for much longer than the mean plus two or three standard deviations. Additionally, the left tail cannot be long, because a person cannot work at a job for less than 0 years.

b) The Central Limit Theorem guarantees that the distribution of the mean time is Normally distributed for large sample sizes, as long as the assumptions and conditions are satisfied. The CLT doesn't help us with the distribution of individual times.

41. Dice and dollars.

a) Let X = the number of dollars won in one play.

$$\mu = E(X) = 0\left(\frac{3}{6}\right) + 1\left(\frac{2}{6}\right) + 10\left(\frac{1}{6}\right) = \$2$$

$$\sigma^2 = Var(X) = (0-2)^2\left(\frac{3}{6}\right) + (1-2)^2\left(\frac{2}{6}\right) + (10-2)^2\left(\frac{1}{6}\right) = 13$$

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{13} \approx \$3.61$$

b) $X + X$ = the total winnings for two plays.

$$\mu = E(X + X) = E(X) + E(X) = 2 + 2 = \$4$$

$$\sigma = SD(X + X) = \sqrt{Var(X) + Var(X)}$$
$$= \sqrt{13 + 13} \approx \$5.10$$

c) In order to win at least $100 in 40 plays, you must average at least $\dfrac{100}{40} = \$2.50$ per play.

The expected value of the winnings is $\mu = \$2$, with standard deviation $\sigma = \$3.61$. Rolling a die is random and the outcomes are mutually independent, so the Central Limit Theorem guarantees that the sampling distribution model is Normal with $\mu_{\bar{x}} = \$2$ and standard deviation $\sigma(\bar{x}) = \dfrac{\$3.61}{\sqrt{40}} \approx \$0.571$.

According to the Normal model, the probability that you win at least $100 in 40 plays is approximately 0.191.

Mean winnings from 40 plays ($)

0.191

2   2.50

$z = 0.876$

(This is equivalent to using $N(80, 22.83)$ to model your total winnings.)
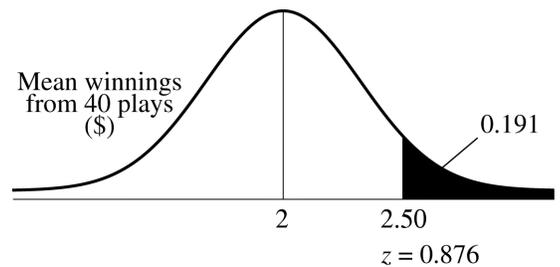
42. New game.

a) Let X = the amount of money won.

X	$40	$0	– $10
$P(X)$	$\dfrac{1}{6}$	$\left(\dfrac{5}{6}\right)\left(\dfrac{1}{6}\right) = \dfrac{5}{36}$	$\left(\dfrac{5}{6}\right)\left(\dfrac{5}{6}\right) = \dfrac{25}{36}$

b) $\mu = E(X) = 40\left(\dfrac{1}{6}\right) + 0\left(\dfrac{5}{36}\right) - 10\left(\dfrac{25}{36}\right) \approx -\$0.28$

$\sigma^2 = Var(X) = (40-(-0.28))^2\left(\dfrac{1}{6}\right) + (0-(-0.28))^2\left(\dfrac{5}{36}\right) + (-10-(-0.28))^2\left(\dfrac{25}{36}\right) \approx 336.034$

$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{336.034} \approx \$18.33$

c) $\mu = E(X + X + X + X + X) = 5E(X) = 5(-0.28) = -\$1.40$

$\sigma = SD(X + X + X + X + X) = \sqrt{5(Var(X))} = \sqrt{5(336.034)} \approx \$40.99$

d) In order for the person running the game to make a profit, the average winnings of the 100 people must be less than $0.

The expected value of the winnings is $\mu = -\$0.28$, with standard deviation $\sigma = \$18.33$. Rolling a die is random and the outcomes are mutually independent, so the Central Limit Theorem guarantees that the sampling distribution model is Normal with $\mu_{\bar{x}} = -\$0.28$ and standard deviation $\sigma(\bar{x}) = \dfrac{18.33}{\sqrt{100}} \approx \$1.833$.

According to the Normal model, the probability that the person running the game makes a profit is approximately 0.561.

0.561

Mean winnings
from 100 plays
($)

$0 $z = 0.153$

$-\$0.28$
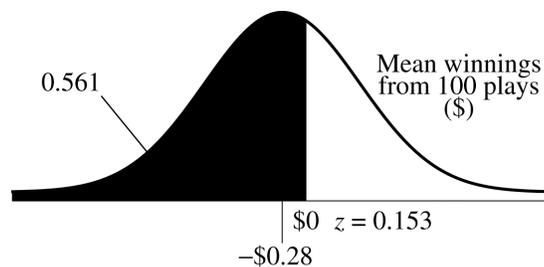
43. AP Stats 2006.

a) $\mu = 5(0.126) + 4(0.222) + 3(0.253) + 2(0.183) + 1(0.216) \approx 2.859$

$$\sigma = \sqrt{\begin{array}{c}(5 - 2.859)^2(0.126) + (4 - 2.859)^2(0.222) + (3 - 2.859)^2(0.253) \\ + (2 - 2.859)^2(0.183) + (1 - 2.859)^2(0.216)\end{array}} \approx 1.324$$

The calculation for standard deviation is based on a rounded mean. Use technology to calculate the mean and standard deviation to avoid inaccuracy.

b) The distribution of scores for 40 randomly selected students would not follow a Normal model. The distribution would resemble the population, mostly uniform for scores 1 – 4, with about half as many 5s.

c) **Randomization condition:** The scores are selected randomly.

Independence assumption: It is reasonable to think that the randomly selected scores are independent of one another.

10% condition: The 40 scores represent less than 10% of all scores.

Large Enough Sample condition: A sample of 40 scores is large enough.

Since the conditions are satisfied, the sampling distribution model for the mean of 40 randomly selected AP Stat scores is Normal, with $\mu_{\bar{y}} = \mu \approx 2.859$ and standard deviation

$\sigma(\bar{y}) = \dfrac{\sigma}{\sqrt{n}} = \dfrac{1.324}{\sqrt{40}} \approx 0.2093$.

44. Museum membership.

a) $\mu = 50(0.41) + 100(0.37) + 250(0.14) + 500(0.07) + 1000(0.01) \approx \$137.50$

$$\sigma = \sqrt{\begin{array}{l}(50-137.50)^2(0.41) + (100-137.50)^2(0.37) + (250-137.50)^2(0.14) \\ \quad + (500-137.50)^2(0.07) + (1000-137.50)^2(0.01)\end{array}} \approx \$148.56$$

The calculation for standard deviation is based on a rounded mean. Use technology to calculate the mean and standard deviation to avoid inaccuracy.

b) The distribution of donations for 50 new members would not follow a Normal model. The new members would probably make donations typical of the current member populations, so the distribution would resemble the population, skewed to the right.

c) **Randomization condition:** The members are not selected randomly. They are simply the new members that day. However, the donations they make are probably typical of the donations made by current members.

Independence assumption: It is reasonable to think that the donations for the new members would not affect one another.

10% condition: The 50 donations represent less than 10% of all donations.

Large Enough Sample condition: The sample of 50 donations is large enough.

Since the conditions are satisfied, the sampling distribution model for the mean of 50 donations is Normal, with $\mu_{\bar{y}} = \mu \approx \$137.50$ and standard deviation $\sigma(\bar{y}) = \dfrac{\sigma}{\sqrt{n}} = \dfrac{148.56}{\sqrt{50}}$.

45. AP Stats 2006, again.

Since the teacher considers his 63 students "typical", and 63 is less than 10% of all students, the sampling distribution model for the mean AP Stat score for 63 students is Normal, with

mean $\mu_{\bar{y}} = \mu \approx 2.859$ and standard deviation $\sigma(\bar{y}) = \dfrac{\sigma}{\sqrt{n}} = \dfrac{1.324}{\sqrt{63}} \approx 0.167$.
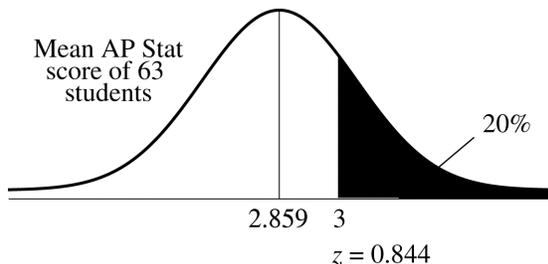
$z = \dfrac{\bar{y} - \mu_{\bar{y}}}{\sigma(\bar{y})}$

$z = \dfrac{3 - 2.859}{0.167}$

$z = 0.844$

According to the sampling distribution model, the probability that the class of 63 students achieves an average of 3 on the AP Stat exam is about 20%.

46. Joining the museum.

If the new members can be considered a random sample of all museum members, and the 80 new members are less than 10% of all members, then the sampling distribution model for the mean donation of 80 members is Normal, with $\mu_{\bar{y}} = \mu \approx \$137.50$ and standard deviation $\sigma(\bar{y}) = \dfrac{148.56}{\sqrt{80}} = \$16.61$.
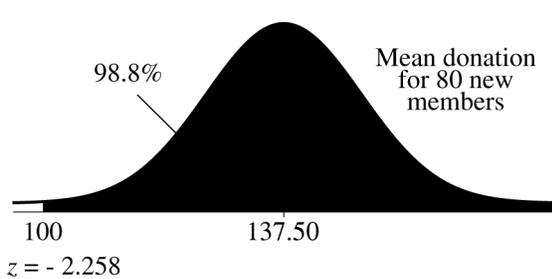
$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma(\bar{y})}$$

$$z = \frac{100 - 137.50}{16.61}$$

$$z = -2.258$$

According to the sampling distribution model, there is a 98.8% probability that the average donation for 80 new members is at least $100.

47. Pollution.

a) **Randomization condition:** Assume that the 80 cars can be considered a representative sample of all cars of this type.

   Independence assumption: It is reasonable to think that the CO emissions for these cars are mutually independent.

   10% condition: The 80 cars in the fleet certainly represent less than 10% of all cars of this type.

   Large Enough Sample condition: A sample of 80 cars is large enough.

   The mean CO level was $\mu = 2.9$ gm/mi, with standard deviation $\sigma = 0.4$ gm/mi. Since the conditions are met, the CLT allows us to model the sampling distribution of the $\bar{y}$ with a Normal model, with $\mu_{\bar{y}} = 2.9$ gm/mi and standard deviation $\sigma(\bar{y}) = \dfrac{0.4}{\sqrt{80}} = 0.045$ gm/mi.

b) According to the Normal model, the probability that $\bar{y}$ is between 3.0 and 3.1 gm/mi is approximately 0.0131.
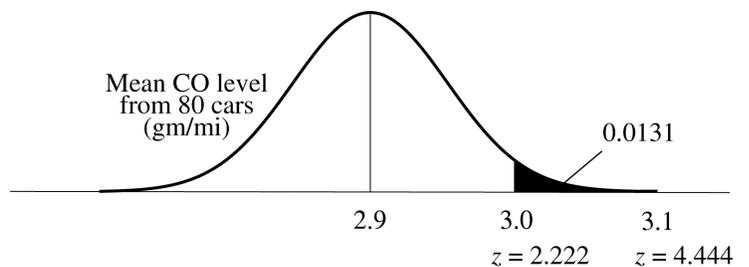
c) According to the Normal model, there is only a 5% chance that the fleet's mean CO level is greater than approximately 2.97 gm/mi.
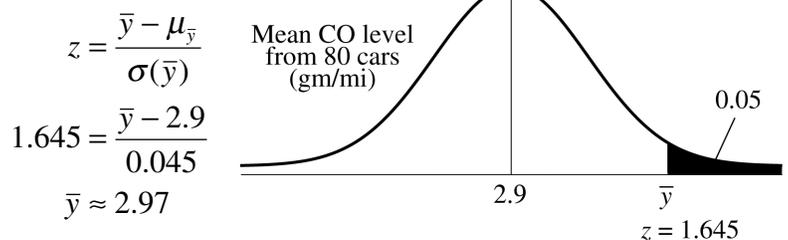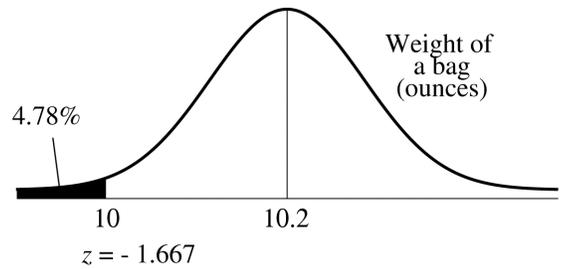
$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma(\bar{y})}$$

$$1.645 = \frac{\bar{y} - 2.9}{0.045}$$

$$\bar{y} \approx 2.97$$

48. Potato chips.

a) According to the Normal model, only about 4.78% of the bags sold are underweight.

b) P(none of the 3 bags are underweight)
$= (1 - 0.0478)^3 \approx 0.863$.

c) **Randomization condition:** Assume that the 3 bags can be considered a representative sample of all bags.

Independence assumption: It is reasonable to think that the weights of these bags are mutually independent.

10% condition: The 3 bags certainly represent less than 10% of all bags.

Large Enough Sample condition: Since the distribution of bag weights is believed to be Normal, the sample of 3 bags is large enough.

The mean weight is $\mu = 10.2$ ounces, with standard deviation $\sigma = 0.12$ ounces. Since the conditions are met, we can model the sampling distribution of $\bar{y}$ with a Normal model, with $\mu_{\bar{y}} = 10.2$ ounces and standard deviation

$\sigma(\bar{y}) = \dfrac{0.12}{\sqrt{3}} \approx 0.069$ ounces.

According to the Normal model, the probability that the mean weight of the 3 bags is less than 10 ounces is approximately 0.0019.
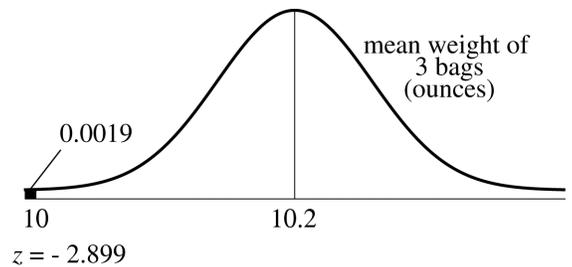
d) For 24 bags, the standard deviation of the sampling distribution model is
$\sigma(\bar{y}) = \dfrac{0.12}{\sqrt{24}} \approx 0.024$ ounces. Now, an average of 10 ounces is over 8 standard deviations below the mean of the sampling distribution model. This is extremely unlikely.

49. Tips.

a) Since the distribution of tips is skewed to the right, we can't use the Normal model to determine the probability that a given party will tip at least $20.

b) No. A sample of 4 parties is probably not a large enough sample for the CLT to allow us to use the Normal model to estimate the distribution of averages.

c) A sample of 10 parties may not be large enough to allow the use of a Normal model to describe the distribution of averages. It would be risky to attempt to estimate the probability that his next 10 parties tip an average of $15. However, since the distribution of tips has $\mu = \$9.60$, with standard deviation $\sigma = \$5.40$, we still know that the mean of the sampling distribution model is $\mu_{\bar{y}} = \$9.60$ with standard deviation $\sigma(\bar{y}) = \dfrac{5.40}{\sqrt{10}} \approx \$1.71$.

We don't know the exact shape of the distribution, but we can still assess the likelihood of specific means. A mean tip of $15 is over 3 standard deviations above the expected mean tip for 10 parties. That's not very likely to happen.

50. Groceries.

a) Since the distribution of Sunday purchases is skewed, we can't use the Normal model to determine the probability that a given purchase is at least $40.

b) A sample of 10 customers may not be large enough for the CLT to allow the use of a Normal model for the sampling distribution model. If the distribution of Sunday purchases is only slightly skewed, the sample may be large enough, but if the distribution is heavily skewed, it would be very risky to attempt to determine the probability.

c) **Randomization condition:** Assume that the 50 Sunday purchases can be considered a representative sample of all purchases.

Independence assumption: It is reasonable to think that the Sunday purchases are mutually independent, unless there is a sale or other incentive to purchase more.
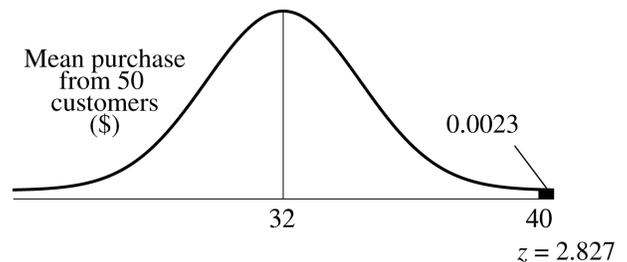
10% condition: The 50 purchases certainly represent less than 10% of all purchases.

Large Enough Sample condition: The sample of 50 purchases is large enough.

The mean Sunday purchase is $\mu = \$32$, with standard deviation $\sigma = \$20$. Since the conditions are met, the CLT allows us to model the sampling distribution of $\bar{y}$ with a Normal model, with $\mu_{\bar{y}} = \$32$ and standard

deviation $\sigma(\bar{y}) = \dfrac{20}{\sqrt{50}} \approx \$2.83$.

According to the Normal model, the probability the mean Sunday purchase of 50 customers is at least $40 is about 0.0023.

Mean purchase from 50 customers ($)

0.0023

32   40

$z = 2.827$

51. More tips.

a) **Randomization condition:** Assume that the tips from 40 parties can be considered a representative sample of all tips.

Independence assumption: It is reasonable to think that the tips are mutually independent, unless the service is particularly good or bad during this weekend.

10% condition: The tips of 40 parties certainly represent less than 10% of all tips.

Large Enough Sample condition: The sample of 40 parties is large enough.

The mean tip is $\mu = \$9.60$, with standard deviation $\sigma = \$5.40$. Since the conditions are satisfied, the CLT allows us to model the sampling distribution of $\bar{y}$ with a Normal model,
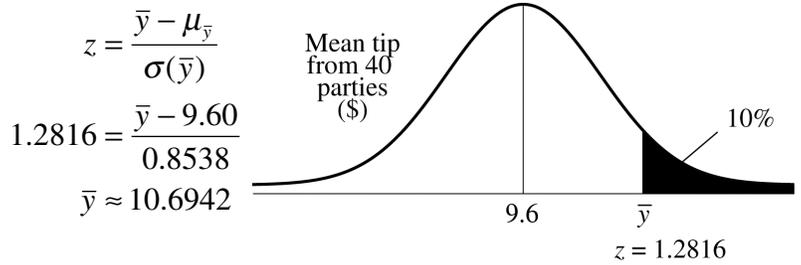
with $\mu_{\bar{y}} = \$9.60$ and standard deviation $\sigma(\bar{y}) = \dfrac{5.40}{\sqrt{40}} \approx \$0.8538$.

In order to earn at least $500, the waiter would

have to average $\dfrac{500}{40} = \$12.50$ per party.

According to the Normal model, the probability that the waiter earns at least $500 in tips in a weekend is approximately 0.0003.

Mean tip from 40 parties ($)

0.0003

9.60   12.50

$z = 3.396$
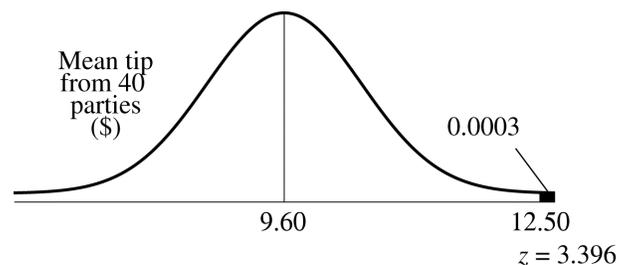
b) According to the Normal model, the waiter can expect to have a mean tip of about $10.6942, which corresponds to about $427.77 for 40 parties, in the best 10% of such weekends.

$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma(\bar{y})}$$

$$1.2816 = \frac{\bar{y} - 9.60}{0.8538}$$

$$\bar{y} \approx 10.6942$$

Mean tip from 40 parties ($)

10%

9.6 $\bar{y}$

$z = 1.2816$

52. More groceries.

a) Assumptions and conditions for the use of the CLT were verified in a previous exercise.

The mean purchase is $\mu = \$32$, with standard deviation $\sigma = \$20$. Since the sample is large, the CLT allows us to model the sampling distribution of $\bar{y}$ with a Normal model, with

$\mu_{\bar{y}} = \$32$ and standard deviation $\sigma(\bar{y}) = \dfrac{20}{\sqrt{312}} \approx \$1.1323$.

In order to have revenues of at least $10,000, the mean Sunday purchase must be at least

$\dfrac{10,000}{312} = \$32.0513$.

According to the Normal model, the probability of having a mean Sunday purchase at least that high (and therefore at total revenue of at least $10,000) is 0.482.
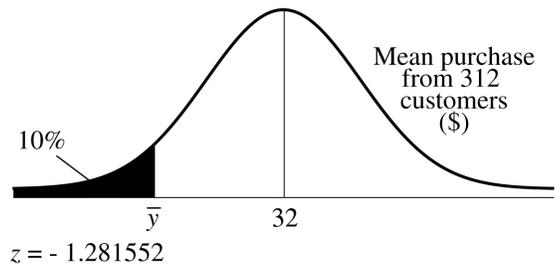
Mean purchase from 312 customers ($)

0.482

32.0513

32   $z = 0.0453$
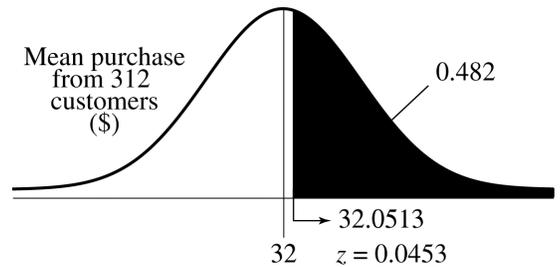
b) According to the Normal model, the mean Sunday purchase on the worst 10% of such days is approximately $30.548928, so 312 customers are expected to spend about $9531.27.
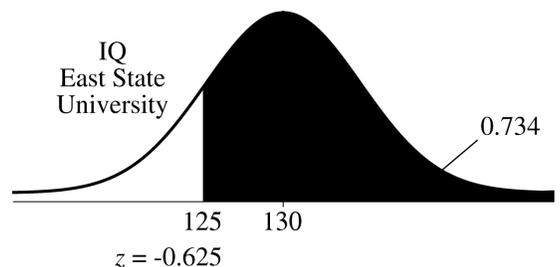
$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma(\bar{y})}$$

$$-1.281552 = \frac{\bar{y} - 32}{\dfrac{20}{\sqrt{312}}}$$

$$\bar{y} \approx 30.548928$$

10%

$\bar{y}$ 32

$z = -1.281552$

Mean purchase from 312 customers ($)

53. IQs.

a) According to the Normal model, the probability that the IQ of a student from East State is at least 125 is approximately 0.734.

IQ East State University

0.734

125   130

$z = -0.625$

b) First, we will need to generate a model for the difference in IQ between the two schools. Since we are choosing at random, it is reasonable to believe that the students' IQs are independent, which allows us to calculate the standard deviation of the difference.

$\mu = E(E - W) = E(E) - E(W) = 130 - 120 = 10$

$\sigma = SD(E - W) = \sqrt{Var(E) + Var(W)}$

$\qquad = \sqrt{8^2 + 10^2} \approx 12.806$
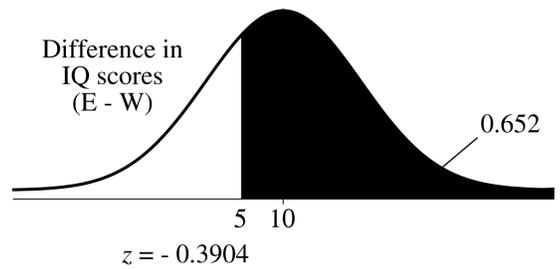
Since both distributions are Normal, the distribution of the difference is $N(10, 12.806)$.

According to the Normal model, the probability that the IQ of a student at ESU is at least 5 points higher than a student at WSU is approximately 0.652.

c) **Randomization condition:** Students are randomly sampled from WSU.

Independence assumption: It is reasonable to think that the IQs are mutually independent.

10% condition: The 3 students certainly represent less than 10% of students.

Large Enough Sample condition: The distribution of IQs is Normal, so the distribution of sample means of samples of any size will be Normal, so a sample of 3 students is large enough.

The mean IQ is $\mu_w = 120$, with standard deviation $\sigma_w = 10$. Since the distribution IQs is Normal, we can model the sampling distribution of $\overline{w}$ with a Normal model, with $\mu_{\overline{w}} = 120$ with standard deviation $\sigma(\overline{w}) = \dfrac{10}{\sqrt{3}} \approx 5.7735$.
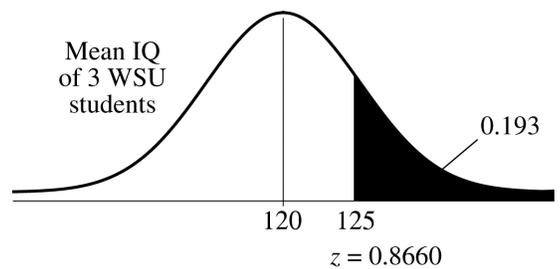
According to the Normal model, the probability that the mean IQ of the 3 WSU students is above 125 is approximately 0.193.

d) As in part c, the sampling distribution of $\overline{e}$, the mean IQ of 3 ESU students, can be modeled with a Normal model, with $\mu_{\overline{e}} = 130$ with standard deviation $\sigma(\overline{e}) = \dfrac{8}{\sqrt{3}} \approx 4.6188$.

The distribution of the difference in mean IQ is Normal, with the following parameters:

$\mu_{\overline{e}-\overline{w}} = E(\overline{e} - \overline{w}) = E(\overline{e}) - E(\overline{w}) = 130 - 120 = 10$

$\sigma_{\overline{e}-\overline{w}} = SD(\overline{e} - \overline{w}) = \sqrt{Var(\overline{e}) + Var(\overline{w})}$

$\qquad = \sqrt{\left(\dfrac{10}{\sqrt{3}}\right)^2 + \left(\dfrac{8}{\sqrt{3}}\right)^2} \approx 7.3937$

According to the Normal model, the probability that the mean IQ of 3 ESU students is at least 5 points higher than the mean IQ of 3 WSU students is approximately 0.751.

Difference of mean IQ scores for 3 students at each university (E - W)

0.751

5   10

$z = -0.6763$
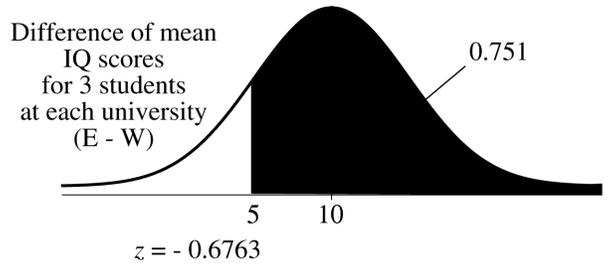
54. Milk.

a) According to the Normal model, the probability that an Ayrshire averages more than 50 pounds of milk per day is approximately 0.309.

Ayrshire daily milk production (pounds)

0.309

47   50

$z = 0.5$
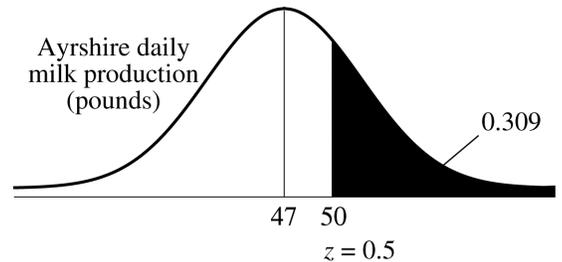
b) First, we will need to generate a model for the difference in milk production between the two cows. Since we are choosing at random, it is reasonable to believe that the cows' milk productions are independent, which allows us to calculate the standard deviation of the difference.

$$\mu = E(A - J) = E(A) - E(J) = 47 - 43 = 4 \text{ pounds}$$

$$\sigma = SD(A - J) = \sqrt{Var(A) + Var(J)}$$

$$= \sqrt{6^2 + 5^2} \approx 7.8102 \text{ pounds}$$

Since both distributions are Normal, the distribution of the difference is $N(4, 7.8102)$.

According to the Normal model, the probability that the Ayrshire gives more milk than the Jersey is approximately 0.696.

Difference in daily milk prodcution (pounds) (A - J)

0.696

0   4

$z = -0.5122$
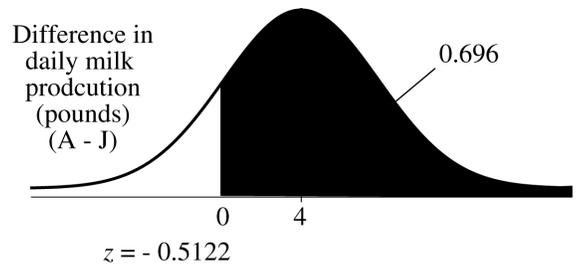
c) **Randomization condition:** Assume that the farmer's 20 Jerseys are a representative sample of all Jerseys.

Independence assumption: It is reasonable to think that the cows have mutually independent milk production.

10% condition: The 20 cows certainly represent less than 10% of cows.

Large Enough Sample condition: Since the distribution of daily milk production is Normal, the sample means of samples of any size are Normally distributed, so the sample of 20 cows is certainly large enough.

The mean milk production is $\mu_j = 43$ pounds, with standard deviation $\sigma_j = 5$. Since the distribution of milk production is Normal, we can model the sampling distribution of $\bar{j}$ with a Normal model, with $\mu_{\bar{j}} = 43$ pounds with

standard deviation $\sigma(\bar{j}) = \dfrac{5}{\sqrt{20}} \approx 1.1180$.

According to the Normal model, the probability that the mean milk production of the 20 Jerseys is above 45 pounds of milk per day is approximately 0.037.

Mean milk production (pounds) 20 Jerseys

0.037

43   45

$z = 1.789$
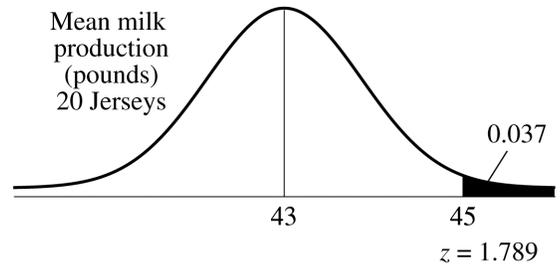
d) As in part c, the sampling distribution of $\bar{a}$, the mean milk production of 10 Ayrshires, can be modeled with a Normal model, with $\mu_{\bar{a}} = 47$ pounds with standard deviation

$\sigma(\bar{a}) = \dfrac{6}{\sqrt{10}} \approx 1.8974$ pounds.

The distribution of the difference in mean milk production is Normal, with the following parameters:

$\mu_{\bar{a}-\bar{j}} = E(\bar{a} - \bar{j}) = E(\bar{a}) - E(\bar{j}) = 47 - 43 = 4$ pounds

$\sigma_{\bar{a}-\bar{j}} = SD(\bar{a} - \bar{j}) = \sqrt{Var(\bar{a}) + Var(\bar{j})}$

$$= \sqrt{\left(\dfrac{6}{\sqrt{10}}\right)^2 + \left(\dfrac{5}{\sqrt{20}}\right)^2} \approx 2.2023 \text{ pounds}$$

According to the Normal model, the probability that the mean milk production of 10 Ayrshires is at least 5 pounds higher than the mean milk production of 20 Jerseys is approximately 0.325.

Difference in mean milk production (pounds) 10 Ayrshires 20 Jerseys

0.325

4 5

$z = 0.454$