Section 5.4

Exercise 1. Let's keep rolling a die until we get 5. How long, on average, does it take to get a 5? Take a guess.

We have different guesses:

We will find out the right answer shortly.

Exercise 2. What is the probability that it takes exactly four rolls of the die to get a 5 (i.e. you get 5 for the first time on the fourth roll of the die)?

If you got a 5 for the first time on the fourth roll, that means you didn't a 5 on the first three rolls.

Remark: Suppose X represents a random variable where we are waiting for an outcome to occur. If the probability of that outcome occurring is p, $P(X = x) = (1 - p)^{x-1}p$ for $x = 1, 2, 3, \ldots$.

Definition: A random variable X is a **geometric random variable** if its probability density function is $P(X = x) = (1 - p)^{x-1}p$ for $x = 1, 2, 3, \ldots$ and for some p between 0 and 1.

Notation: If X is a geometric random variable with parameter p, $X \sim G(p)$. (G stands for geometric.)

Example: The following are examples of geometric random variables:
Y = number of coin flips until first tail
Z = number of days until first rainy day (assuming that the probability it rains is the same every day)

Formula: Using mathematics beyond the scope of this course, it can be shown that $\mu = 1/p$ if $X \sim G(p)$.

Exercise 3. Suppose that there is a probability of 0.1 that a certain baseball player will hit a home run when he comes up to bat. Let X = number of times at bat it takes for the baseball player to hit a home run.
(a) Is X a geometric random variable?

(b) If so, what is p?

since the probability of hitting a home run is . ($X \sim G($ $)$).

Exercise 4. What is the expected number of times at bat it will take for the baseball player to hit a home run?

From the formula preceding the last exercise,

Class Exercise 1. Bob is a recent law school graduate who intends to take the state bar exam. According to the National Conference on Bar Examiners, about 57% of all people who take the state bar exam pass (Source: *The Book of Odds* by Shook and Shook, Signet). Let $n = 1, 2, 3,$ .... represent the number of times a person takes the bar exam until the *first* pass. (#10)
(a) Write out a formula for the probability distribution of the random variable n.
(b) What is the probability that Bob first passes the bar exam on the second try ($n = 2$)?
(c) What is the probability that Bob needs three attempts to pass the bar exam?
(d) What is the probability that Bob needs more than three attempts to pass the bar exam?
(e) What is the expected number of attempts at the state bar exam Bob must make for his (first) pass?

Class Exercise 2. Approximately 3.6% of all (untreated) Jonathan apples had bitter pit in a study conducted by the botanists Ratkowsky and Martin (Source: *Australian Journal of Agricultural Research*, Vol. 25, pp. 783-790). (Bitter pit is a disease of apples resulting in a soggy core, which can be caused either by over-watering the apple tree or by a calcium deficiency in the soil.) Let n be a random variable that represents the first Jonathan apple chosen at random that has bitter pit. (#12)
(a) Write out a formula for the probability distribution of the random variable n.
(b) Find the probabilities that $n = 3$, $n = 5$, and $n = 12$.
(c) Find the probability that $n \geq 5$.
(d) What is the expected number of apples that must be examined to find the first one with bitter pit?

Homework

C problems

Section 5.4: 9(a), 9(b), 9(c), 9(d), 11(a), 11(b), 11(c) 11(d), 13(a), 13(b)

B problems

Section 5.4: 1

A problems

Section 5.4: 3

Section 6.1

Definition: The **normal distribution** is a continuous distribution whose density curve is given by the following equation:

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

for some constants μ and σ where $\sigma > 0$.

Remark: You don't need to memorize the above formula for the exam.

Definition: A **normal random variable** is a random variable whose density is given by the above equation.

Properties of a Normal Curve
1. The curve is bell-shaped, with the highest point over the mean μ.
2. The curve is symmetric about a vertical line through μ.

3. The curve approaches the horizontal axis but never touches or crosses it.
4. The inflection (transition) points between cupping upward and downward occur above $\mu + \sigma$ and $\mu - \sigma$.
5. The area under the entire curve is 1.

Exercise 5. Sketch the normal curve with $\mu = 3$ and $\sigma = 4$.

Remark: Notice how the concavity of the curve changes at 7 and -1. For any normal curve, the concavity changes at $\mu + \sigma$ and μ - σ.

Exercise 6. Sketch the normal curve with $\mu = 2$ and $\sigma = 3$.

Exercise 7. Sketch the normal curve with $\mu = 0$ and $\sigma = 1$.

Remark: It may not be clear from the illustrations, but the curve never touches the x-axis. In other words, the curve extends out to $-\infty$ and ∞. (It doesn't stop at 4 and -4).

Empirical Rule

For a distribution that is symmetric and bell-shaped (in particular, for a normal distribution):
Approximately 68% of the data values will lie within 1 standard deviation on each side of the mean.
Approximately 95% of the data values will lie within 2 standard deviations on each side of the mean.
Approximately 99.7% (or almost all) of the data values will lie within 3 standard deviations on each side of the mean.

Exercise 8. What percentage of the area under the normal curve lies
(a) to the right of μ?

(b) between $\mu - 2\sigma$ and $\mu + 2\sigma$?

(c) to the right of $\mu + 3\sigma$?

Class Exercise 3. The incubation time for Rhode Island Red chicks is normally distributed with a mean of 21 days and standard deviation of approximately 1 day. If 1000 eggs are being incubated, how many do we expect will hatch
(a) in 19 to 23 days?
(b) in 20 to 22 days?
(c) in 21 days or fewer?
(d) in 18 to 24 days?

Class Exercise 4. A vending machine automatically pours soft drinks into cups. The amount of soft drink dispensed into a cup is normally distributed with a mean of 7.6 ounces and standard deviation of 0.4 ounce.
(a) Estimate the probability that the machine will overflow an 8-ounce cup.
(b) Estimate the probability that the machine will not overflow an 8-ounce cup.
(c) The machine has just been loaded with 850 cups. How many of these do you expect will overflow when served?

Homework

C problems

Section 6.1: 5 (all parts), 7 (all parts), 9(b), 11(a), 11(b), 13(a), 13(b)

B problems

Section 6.1: 1

A problems

Section 6.1: 3

Section 6.2

<u>**Definition**</u>: A <u>**z-score**</u> (or standardized value) is the number of standard deviations that a given value x is above or below the mean. For a sample, the z-score is

$$z = \frac{x-\bar{x}}{s}.$$

For a population, the z-score is

$$z = \frac{x-\mu}{\sigma}.$$

Exercise 9. Use z-scores to compare the above two test scores: 90 on an exam where the mean is 80 and the standard deviation is 2 and 95 on an exam where the mean is 80 and the standard deviation is 5.

Remark: A value with a z-score between -2 and 2 is an ordinary value; a z-score that is below -2 or above 2 is an unusual value. Both test scores in the above exercise are unusual.

Exercise 10. Compute the z-score for a value of 5 if $\mu = 10$ and $\sigma = 5$.

The value of 5 is one standard deviation below the mean.

Remark: If a data value is less than the mean, its corresponding z-score is negative.

Class Exercise 5. Eleanor scores 680 on the mathematics part of the SAT. The distribution of SAT scores in a reference population is symmetric and single-peaked with mean 500 and standard deviation 100. Gerald takes the American College Testing (ACT) mathematics test and scores 27. ACT scores also follow a symmetric, single-peaked distribution but with mean 18 and standard deviation 6. Find the standardized scores for both students. Assuming that both tests measure the same kind of ability, who has the higher score?

Class Exercise 6. Three landmarks of baseball achievement are Ty Cobb's batting average of .420 in 1911, Ted Williams's .406 in 1941, and George Brett's .390 in 1980. These batting averages cannot be compared directly because the distribution of major league batting averages has changed over the years. The distributions are quite symmetric, except for outliers such as Cobb, Williams, and Brett. While the mean batting average has been held roughly constant by rule changes and the balance between hitting and pitching, the standard deviation has dropped over time. Here are the facts:

In the 1910's, the mean batting average was .266 and the standard deviation was 0.0371.

In the 1940's, the mean batting average was .267 and the standard deviation was 0.0326.

In the 1970's, the mean batting average was .261 and the standard deviation was 0.0317.

Compute the standardized batting averages for Cobb, Williams, and Brett to compare how far each stood above his peers.

Class Exercise 7. Sketch the normal curve with $\mu = 0$ and $\sigma = 1$.

Definition: The **standard normal distribution** is a normal probability distribution with $\mu = 0$ and $\sigma = 1$. The total area under its density curve is equal to 1.

Exercise 11. Find the area of the shaded region. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

To find the area under the curve, we will use the normalcdf function on the calculator. normalcdf(b_1,b_2) gives the area under the standard normal curve between b_1 and b_2. The lower bound for the area is $-\infty$. We enter -10 instead of $-\infty$ because we cannot enter $-\infty$. (The area between $-\infty$ and -10 is extremely small so replacing $-\infty$ with -10 does not affect the value.)

Class Exercise 8. Find the area of the shaded region. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

Class Exercise 9. Find the area of the shaded region. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

Exercise 12. Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

For this question, we will utilize the invNorm() function on your calculator.

Definition: Suppose y = invNorm(b). Then, the area under the standard normal curve and to the left of y is b. For example, invNorm(0.5) = 0.

.

Exercise 13. Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

The area to the left of the line is

Class Exercise 10. Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

Class Exercise 11. Find the indicated z score. The graph depicts the standard normal distribution with mean 0 and standard deviation 1.

Homework

C problems

Section 6.2: 5, 9, 11(a), (b), (c), (d), (e), (f), 13 - 49 (every 4th)

B problems

Section 6.2: 1, 3

A problems

Section 6.2: 7