Section 3 Answer Key:

1) Suppose you wish to generate a random real number between 5 and 21. Write down an appropriate probability density curve and use it to compute the probability that the real number generated is

The probability density curve is a rectangle from 5 to 21, which is 21 –5 = 16 long. Since the area must be 1, (16)h = 1, and the height is 1/16 = 0.0625. If X is the number we get, the probabilities are represented by areas above the intervals.

(a) greater than 18:

$X > 18$: The rectangle above (18,21) is 3 long and 1/16 high, so 3/16.

(b) between 15 and 20, inclusive:

$15 \leq X \leq 20$: The rectangle above [15,20] is 5 long and 1/16 high, so 5/16.

(c) between 15 and 20, exclusive:

$15 < X < 20$: The rectangle above (15,20) is 5 long and 1/16 high, so 5/16. The presence or absence of exactly 15 and/or exactly 20 contributes no area.

(d) 13:

There is no area above this point; 0.

(e) less than 10 or greater than 17:

$X < 10$ or $X > 17$: A point makes this true if it's in either of the intervals [5,10) or (17,21]. The total area above these two intervals is the two rectangles which are 10-5=5 and 21-17=4 long respectively. The two areas are 5/16 and 4/16, so the total area is 9/16.

(f) less than 17 or greater than 10:

$X < 17$ or $X > 10$: A point makes this true if it's in either of the intervals [5,17) or (10,21]; that is, every point makes this true. The total area above these two intervals is the entire rectangle, which has area 1. Making both conditions true (e.g., X is 14) doesn't make it any more true; no area is double-counted.

(g) Find the mean, median, and quartiles of this distribution (standard deviation is more complicated and we haven't talked about it):

The mean is the "balance line" of the rectangle, which is obviously its midpoint: (5+21)/2=13. The median is the line where half the area is on either side, which is (for a rectangle) obviously the same place. The quartiles are the lines where a quarter and three-quarters of the area is on either side. For a rectangle, this is obviously one-quarter and three-quarters of the way across, so two more midpoints (5+13)/2=9 and (13+21)=17. Alternately, we're dividing the original length 16 into four equal intervals of length 16/4=4, so add it to 5, add again, add again.

2) If we average together two random real numbers between 0 and 1, the probability density curve is an isosceles triangle above the interval [0,1]. Sketch this and use it to find the probability that the average is less than ¼.

The isosceles triangle above [0,1] has length 2 and area 1. Its height h can then be found: ½*b*h= ½ * (1)*h=1, so h=2. The probability that Y is less than ¼ is the area above the interval [0,1/4]. This is a smaller triangle which is similar to the larger triangle above [0,1/2], a right triangle with length ½ and height 2. Setting up a proportion for similar triangles, the height h of the small triangle satisfies:

$$\frac{\frac{1}{4}}{\frac{1}{2}} = \frac{h}{2} \text{ or } \frac{0.25}{0.50} = \frac{h}{2} \text{ so } h = 0.25 * \frac{2}{0.5} = 1$$

Alternately, if we halve the base, we are going to halve the height.
Since this triangle has length 0.25 and height 1, the area is ½*0.25*1=0.125 or 1/8.
Note that for a rectangle (1), a quarter of the horizontal distance gives a quarter of the area above. For a triangle (or anything that peaks in the middle), this won't be the case. The smallest quarter [0,1/4] occurs 1/8 of the time. Likewise for the biggest quarter [3/4,1], by symmetry. The horizontal middle [1/2, ¾] thus occurs the remaining 6/8 of the time. That is, the number is three times more likely to be in the middle half (6/8) than it is to be in the outlying half (2/8).
Numerically, we can see that a small "average of two spins" is possible, but unlikely; for Y to be less than ¼, we must get two small spins in a row. This is possible, but not just as likely as everything else. It's more likely that one small spin and one large spin will average up to a "central" Y.

3) The length of human pregnancies is approximately normally distributed with mean 266 days and standard deviation 16 days.

$$X = N(266,16)$$

(a) Using the 68-95-99.7% rule, between what two lengths do the most typical 68% of all pregnancies fall? 95%? 99.7%?

The middle 68% of all pregnancies last between 266-16 and 266+16 days, 250 to 282. The middle 95% of all pregnancies last between 266-2*16 and 266+2*16 days, 234 to 298 (for future reference, note that this "rule" is rounded somewhat compared to the charts). The middle 99.7% of all pregnancies last between 266-3*16 and 266+3*16 days, 218 to 314.

(b) What percentage of pregnancies last less than 241 days?

$$X < 241 \quad Z < \frac{241 - 266}{16} \quad Z < -1.56 \;\; 0.0594 \quad 5.95\%$$

What percentage of pregnancies last between 241 and 286 days?

$$241 < X < 286 \quad \frac{241 - 266}{16} < Z < \frac{286 - 266}{16} \quad -1.56 < Z < 1.25 \;\; 0.8944 - 0.0594 = 0.8350 \;\; 83.5\%$$

What percentage of pregnancies last more than 286 days?

$$X > 286 \quad Z > \frac{286 - 266}{16} \quad Z > 1.25 \;\; 1 - 0.8944 = 0.1056 \quad 10.56\%$$

What percentage of pregnancies last more than 333 days?

$$X > 333 \quad Z > \frac{333 - 266}{16} \quad Z > 4.19 \;\; 1 - (\text{approx. } 0) = \text{approx. } 1 \text{ or } 100\%$$

(c) What length cuts off the shortest 2.5% of pregnancies? The longest 2.5% of pregnancies?

We wish to find the tick-mark which separates 2.5% of the area (0.0250) to the left. The closest entry (exact, in fact) is z=−1.96:

$$X = \sigma Z + \mu = 16(-1.96) + 266 = 234.54$$

We wish to find the tick-mark which separates 2.5% of the area (0.0250) to the right, or 0.9750 to the left. The closest entry (exact, in fact) is z=+1.96:

$$X = \sigma Z + \mu = 16(1.96) + 266 = 297.36$$

In passing, notice that this is another way of asking for the tick-marks that separate out the middle 95%. Compare with (a); these answers are slightly more accurate than the “rule”.

(d) Find the quartiles for pregnancy length.

We wish to find the tick-mark which separates 25% of the area (0.2500) to the left. The closest entry is z=−0.67.

$$X = \sigma Z + \mu = 16(-0.67) + 266 = 255.28$$

We wish to find the tick-mark which separates 25% of the area (0.2500) to the right, or 0.7500 to the left. The closest entry is $z = 0.67$.

$$X = \sigma Z + \mu = 16(+0.67) + 266 = 276.72$$

(e) Between what two lengths are the most typical 72% of all pregnancies?

We wish to find the tick-mark which separates 14% of the area (0.1400) to the left. The closest entry is $z = -1.08$. Multiply by 16 and add 266 to get 248.72. We wish to find the tick-mark which separates 14% of the area (0.1400) to the right, or 0.8600 to the left. The closest entry is $z = 1.08$. Multiply by 16 and add 266 to get 283.28.

4. Gerald scores 680 on the math-SAT (which is approximately normal with mean 500 and standard deviation 100). Eleanor scores 27 on the math-ACT (which is approximately normal with mean 18 and standard deviation 6). Without using the conversion formula in 2.6 above, how can we compare these scores? Who has the higher math ability?

$$S = N(500{,}100) \quad \text{and} \quad A = N(18{,}6)$$

Standardizing their scores:

$$\frac{680 - 500}{100} \text{ vs. } \frac{27 - 18}{6} \quad \text{are Z's of 1.80 vs. 1.50.}$$

Gerald has the higher z-score.

5. On the math-SAT, scores of 800 or higher are reported as 800, so a perfect paper is not required to score 800 on the SAT. What percent of students who take the SAT score 800?

$$S = N(500{,}100)$$

$$X \geq 800 \quad Z \geq \frac{800 - 500}{100} \quad Z \geq 3.00 \quad 1 - 0.9987 = 0.0013 \quad 0.13\%$$

6. The psychology department of a university finds that the scores of its applicants on the GRE exam are approximately normal with mean 544 and standard deviation 103.

$$X = N(544{,}103)$$

(a) Find the relative frequency of applicants with score above 700.

$$X > 700 \quad Z > \frac{700 - 544}{103} \quad Z > 1.51 \quad 1 - 0.9345 = 0.0655$$

(b) What percentage of applicants score below 500?

$$X < 700 \quad Z < \frac{500 - 544}{103} \quad Z < -0.43 \quad 0.3336 \quad 33.36\%$$

(c) If we randomly choose one applicant, what is the probability that his score is between 500 and 800?

$$500 < X < 800 \quad \frac{500 - 544}{103} < Z < \frac{800 - 544}{103} \quad -0.43 < Z < 2.49 \quad 0.9936 - 0.3336 = 0.6600$$

(d) Between what two scores are the middle 80% of all applicants?

We wish to find the tick-mark which separates 10% of the area (0.1000) to the left. The closest entry is $z = -1.28$. We wish to find the tick-mark which separates 10% of the area (0.1000) to the right, or 0.9000 to the left. The closest entry is $z = +1.28$.

$$X = \sigma Z + \mu = 103(\pm 1.28) + 544 = 412.16, 675.84$$

7. A patient is said to be hypokalemic (low potassium in the blood) if the measured level of potassium is 3.5 or less. However, an individual's potassium level varies from day to day, and the method of measuring potassium level varies as well. Suppose the overall variation follows a normal distribution. If Judy's mean potassium level is 3.8 with standard deviation 0.2 and she's measured on many days, on what proportion of days will the measurement report that that Judy is hypokalemic?

$$X = N(3.8, 0.2)$$

$$X < 3.5 \quad Z < \frac{3.5 - 3.8}{0.2} \quad Z < -1.50 \quad 0.0668$$

8. (a) Batting averages: Ty Cobb's = .420 in 1911, Ted Williams = .406 in 1941, George Brett = .390 in 1980. Which player stood the highest above his peers?

Decade	mean	standard deviation
1910's	.266	.0371
1940's	.267	.0326
1970's	.261	.0317

Assuming batting averages are normally distributed:

$$\frac{.420-.266}{0.0371} \text{ vs. } \frac{.406-.267}{0.0326} \text{ vs. } \frac{.390-.261}{0.0317} \text{ are z's of } 4.15 \text{ vs. } 4.26 \text{ vs. } 4.07$$

While all are exceptionally good (off the chart), 4.26 sweeps out the most area to the left.

(b) In the 1940's, what percentage of players had a batting average above 0.317?

$$X > 0.317 \quad Z > \frac{.317-.267}{0.0326} \quad Z \geq 1.53 \quad 1-0.9370 = 0.0630 \quad 6.30\%$$

(c) In the 1970's, how high did a batting average have to be for the player to be among the top 5%?

We wish to find the tick-mark which separates 95% of the area (0.9500) to the left. The closest entries are z=1.64 or 1.65.

$$X = \sigma Z + \mu = 0.0317\,(\pm 1.64 \text{ or } 1.65) + 0.261 \approx 0.313$$

(d) In the 1910's, between what numbers lay the most common 40% of batting averages?

We wish to find the tick-marks which separate 30% and 70% of the area (0.3000 and 0.7000) to the left. The closest entries are z= -0.52 and +0.52.

$$X = \sigma Z + \mu = 0.0371(\pm 0.52) + 0.266 = 0.247 \text{ and } 0.285$$

9. The distribution of a critical dimension on auto engine crankshafts is approximately normal with mean 224 mm and standard deviation 0.03 mm. Crankshafts with dimensions between 223.92 and 224.08 mm are acceptable. What percent of all crankshafts produced are acceptable?

$$X = N(224, 0.03)$$

$$223.92 < X < 224.08 \qquad \frac{223.92-224}{0.03} < Z < \frac{224.08-224}{0.3}$$

$$-2.67 < Z < 2.67 \qquad 0.9962 - 0.0038 = 0.9924 \qquad 99.24\%$$

10. Cholesterol levels of adult American women are approximately normally distributed with mean 188 (mg/dL) and standard deviation 24.
(a) What percent of adult women have a cholesterol level over 200?

$$X = N(188,24)$$

$$X > 200 \qquad Z > \frac{200-188}{24} \qquad Z \geq 0.50 \qquad 1 - 0.6915 = 0.3085 \qquad 30.85\%$$

(b) What percent of adult women have a cholesterol level between 150 and 170?

$$150 < X < 170 \qquad \frac{150-188}{24} < Z < \frac{170-188}{24} \qquad -1.58 < Z < -0.75$$

$$0.2266 - 0.0571 = 0.1695 \qquad 16.95\%$$

(c) How high does a cholesterol level have to be to be among the top 15%
We wish to find the tick-mark which separates 85% of the area (0.8500) to the left. The closest entry is z= 1.04.

$$X = \sigma Z + \mu = 24(1.04) + 188 = 212.96$$

11. Snow tires manufactured by a certain company have treadlife that is approximately normally distributed with mean 32,000 miles and standard deviation 2500 miles.

$$X = N(32000,2500)$$

(a) If you buy a set of tires, would it be reasonable to hope that they'll last for 40,000 miles?

What is the probability of a single tire lasting at least 40,000 miles?

$$X > 40000 \quad Z > \frac{40000-32000}{2500} \quad Z > 3.20 \quad 1 - 0.9993 = 0.0007$$

The probability of a particular tire lasting this long is very small; the probability of getting four such exceptional tires in a row must be very, very small (see section 4 for details).

(b) What percent of tires will last less than 30,000 miles?

$$X < 30000 \quad Z < \frac{30000-32000}{2500} \quad Z < -0.80 \quad 0.2119 \quad 21.19\%$$

(c) What is the probability that a randomly chosen tire will last between 30,000 and 35,000 miles?

$$30000 < X < 35000 \quad \frac{30000 - 32000}{2500} < Z < \frac{35000 - 32000}{2500}$$

$$-0.8 < Z < 1.20 \quad 0.8849 - 0.2119 = 0.6730$$

(d) If a dealer wants to give refunds to no more than 1 out of every 25 customers, for what mileage can he guarantee the tires to last?

We wish to separate out the worst 1/25 (shortest treadlives), that is, to find the tick-mark which separates 4% of the area (0.0400) to the left. The closest entry is $z = -1.75$.

$$X = \sigma Z + \mu = 2500(-1.75) + 32{,}000 = 27{,}625 \text{ miles}$$

12. Heights of kindergarten children vary normally with mean 38.2" and standard deviation 1.8".

$$X = N(38.2, 1.8)$$

(a) What is the probability that a randomly chosen kindergartner will be less than 3 feet tall?

$$X < 36 \text{ inches} \quad Z < \frac{36 - 38.2}{1.8} \quad Z < -1.22 \quad 0.1112$$

(b) The middle 80% of kindergartners span what interval of heights?

We wish to find the tick-marks which separate 10% and 90% of the area (0.1000 and 0.9000) to the left. The closest entries are z= −1.28 and +1.28. Multiply by 1.8 and add 38 to get.

$$X = \sigma Z + \mu = 1.8(\pm 1.28) + 38 = 35.696 \text{ and } 40.304 \text{ inches}$$

(c) How tall does a kindergartner have to be to be among the 10% tallest?

This is the same question as (b), so 40.304 inches.

13. The average body temperature is now believed to be 98.2 (degrees F), normally distributed with a standard deviation 0.7 (degrees F).

$$X = N(98.2, 0.7)$$

(a) What percent of people have a body temperature above 98.6?

$$X > 98.6 \quad Z > \frac{98.6 - 98.2}{0.7} \quad Z > 0.57 \quad 1 - 0.7157 = 0.2843 \quad 28.43\%$$

(b) The coolest 20% of body temperatures are below what number?

We wish to find the tick-mark which separates 20% (0.2000) to the left. The closest entry is $z = -0.84$.

$$X = \sigma Z + \mu = 0.7(-0.84) + 98.2 = 97.612 \text{ degrees}$$

(c) What percentage of people have a body temperature between 97 and 98.6?

$$97 < X < 98.6 \quad \frac{97 - 98.2}{0.7} < Z < \frac{98.6 - 98.2}{0.7} \quad -1.71 < Z < 0.57 \quad 0.7157 - 0.0436 = 0.6721 \quad 67.21\%$$

(d) The most common 30% of body temperatures span what interval?

We wish to find the tick-marks which separate 35% and 65% of the area (0.3500 and 0.6500) to the left. The closest entries are $z = -0.39$ and $+0.39$.

$$X = \sigma Z + \mu = 0.7(\pm 0.39) + 98.2 = 97.927 \text{ and } 98.473 \text{ degrees}$$