Assignment 2 Solutions

1. (a) Correlation is only defined for two quantitative variables. Since gender is a categorical variable, this is not a valid value of r.

   (b) The reaction time it takes to brake when driving is likely **positively** associated with amount of alcohol consumed. While it is true that amount of alcohol consumed will make you **slower** to react, slower to react means a **longer** reaction time. So, $r = -0.7$ is not a reasonable value of r.

   (c) People who wear larger shoes are neither more nor less likely to have a higher IQ score, so a value of $r = 0$ (indicating no association) is reasonable.

   (d) As distance from the equator for North American cities increases, it gets colder, in general, so average January temperature **decreases**. The sign of the correlation coefficient should be **negative**.

   (e) While there is likely to be a strong positive association between temperature and ice cream sales, a value of $r = 1.00$ implies a perfect linear relationship. To see why this is unreasonable, imagine the increase in ice cream sales when the temperature increases from $21°C$ to $24°C$. Now when the temperature increases by the same amount from $24°C$ to $27°C$, we must see the **exact same** increase in ice cream sales, and this must also be true for **any** $3°C$ increase in temperature.

2. (a) We are given that 79.43% of the variation in gas mileage is explained by its regression on horsepower. So $r^2 = 0.7943$. The correlation between horsepower and gas mileage is

$$r = -\sqrt{r^2} = -\sqrt{0.7943} = -0.8912$$

   Note that we take the negative square root of r^2, as the relationship between horsepower and gas mileage is negative. The sign of the correlation must be the same as the sign of the slope of the least-squares regression line (the slope is -0.11).

   (b) The residual for Car #5 is

$$y_5 - \hat{y}_5 = 15.5 - (35.42 - 0.11(150)) = 15.5 - 18.92 = -3.42$$

   The negative residual tells us that the point for Car #5 falls below the least-squares regression line, i.e., the gas mileage for Car #5 is lower than we would have predicted it to be, based on its horsepower.

   (c) For every 1 additional horsepower, the predicted gas mileage decreases by 0.11 miles per gallon.

3. (a) The first property of density curves is that all density curves must lie above the x-axis, which is true in this case.

   The second property of density curves is that the area underneath the density curve must be 1. We can see that the area under the density curve can be split up into a rectangle and two triangles. So the area underneath the density curve is simply the sum of the areas of the rectangle and the two triangles.

   The rectangle has a width of 0.75 - 0.25 = 0.5 and the height of the rectangle is 4/3. So the area of the rectangle is (0.5)(4/3) = 2/3.

   The two triangles have the same area as each other since the length of their bases and their heights are the same. The area of one of these triangles is (1/2)(0.25)(4/3) = 1/6. So the area of the two triangles combined is 2(1/6) = 1/3.

   Finally, the total area underneath the density curve is 2/3 + 1/3 = 1.

   (b) The proportion of values of X that are between 0.25 and 0.75 is the area of the region (the rectangle) between 0.25 and 0.75 divided by the area underneath the entire density curve.

   The area of the region between 0.25 and 0.75 is (0.5)(4/3) = 2/3. The area underneath the entire density curve is 1. So the proportion of values of X that are between 0.25 and 0.75 is $\frac{2/3}{1} = 2/3$.

   (c) The proportion of values of X that are less than 0.6 is the area of the region under the density curve to the left of 0.6 divided by the area under the entire density curve.

   The region under the density curve to the left of 0.6 consists of the left triangle in the diagram plus the rectangle that extends from 0.25 to 0.6. We have found above that the left triangle has an area of 1/6. The area of the rectangle that extends from 0.25 to 0.6 is $(0.6 - 0.25)(4/3) = 7/15$. So the total area under the density curve to the left of 0.6 is 1/6 + 7/15 = 0.6333. Finally, the proportion of values of X that are less than 0.6 is $\frac{0.6333}{1} = 0.6333$.

   (d) The area under the entire density curve is 1. If the proportion to the left of 0.6 is 0.6333, then the proportion to the right of 0.6 is $1 - 0.6333 = 0.3667$.

   (e) We know that the area of the middle rectangle (between 0.25 and 0.75) is 2/3 = 0.667, as found above. So, if we want an area of 0.5 in the region between 0.25 and some value x, the value of x that we're looking for must be to the left of 0.75.

   So, the region of interest that has an area of 0.5 is simply a rectangle. We want the area to be 0.5, we know the height of the rectangle is 4/3, so the base of the rectangle must have length $\frac{0.5}{4/3} = 0.375$. In order for the base to have a length of 0.375, the value of x must be 0.25 + 0.375 = 0.625.

4. (a) We know that $P(X > 200) = 0.3085$. We first find the value z such that $P(Z > z) = 0.3085 \Rightarrow P(Z < z) = 1 - 0.3085 = 0.6915$. From Table 2, we find this value to be $z = 0.50$. We now solve for σ as follows:

$$z = \frac{x-\mu}{\sigma} \Rightarrow \sigma = \frac{x-\mu}{z} = \frac{200-194}{0.50} = 12$$

   (b) $P(190 < X < 205) = P\left(\frac{190-194}{12} < Z < \frac{205-194}{12}\right) = P(-0.33 < Z < 0.92)$

   $= P(Z < 0.92) - P(Z < -0.33) = 0.8212 - 0.3707 = 0.4505$

5. (a) We know that $P(X > 85) = 0.0062$. We first find the value z such that $P(Z > z) = 0.0062 \Rightarrow P(Z < z) = 1 - 0.0062 = 0.9938$. From Table 2, we find this value to be $z = 2.50$. We now solve for μ as follows:

$$z = \frac{x-\mu}{\sigma} \Rightarrow \mu = x - z\sigma = 85 - (2.50)(8) = 65$$

   (b) $P(X > 73) = P\left(Z > \frac{73-65}{8}\right) = P(Z > 1.00) = 1 - P(Z < 1.00) = 1 - 0.8413$
   $= 0.1587$

   (c) We want to find the value x such that $P(X > x) = 0.05$. We first find the value z such that $P(Z > z) = 0.05 \Rightarrow P(Z < z) = 1 - 0.05 = 0.9500$. From Table 2, we find this value to be $z = 1.645$. (Since the proportion I'm looking for is exactly halfway between two of the proportions in Table 2, I am taking the average of the z values 1.64 and 1.65.) We solve for x as follows:

$$z = \frac{x-\mu}{\sigma} \Rightarrow x = \mu + z\sigma = 65 + (1.645)(8) = 78.16$$

6. (a) A completely randomized design should be used.

   (b) The experimental units are the 120 subjects.

   (c) There are two factors in this experiment – frequency and dose.

   Frequency has two factor levels – once per day, twice per day. Dose has three factor levels – 325 mg, 500 mg, 650 mg.

   (d) There are $2 \times 3 = 6$ treatments – once per day/325 mg, once per day/500 mg, once per day/650 mg, twice per day/325 mg, twice per day/500 mg, twice per day/650 mg.

   We allocate an equal number of experimental units to each treatment. That is, we allocate $120/6 = 20$ experimental units to each treatment.

   (e) The response variable is change in body temperature.

   (f) No, there is no blocking variable in this experiment.

7. (a) A randomized block design should be used.
   (b) The experimental units are the 30 cars.
   (c) There is only factor in this experiment – type of engine analyzer.
   Type of engine analyzer has two factor levels – computerized, electronic.
   (d) Sine there is only one factor, the treatments are the same as the factor levels – computerized, electronic.
   (e) The response variable is time needed to complete a minor engine tune-up.
   (f) Yes, the blocking variable is car size. (There are three blocks – compact cars, intermediate size cars, and full-sized cars.)