1 Quiz 1 (Feb 4) solutions

- 1. How are standard deviation and variance related?

If s is the standard deviation, then $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the variance.

- 2. Give the range, and mode of the following data set:

$$\begin{Vmatrix} 44 & 27 & 24 & 24 & 36 & 36 & 44 & 120 & 29 & 36 \\ 36 & 36 \end{Vmatrix}$$

The range is the largest minus the smallest value in the data set: x max – x min. Thus we have range = 120-24 = 96

The mode is the most frequently occurring value (or values; there may be more than one) in a data set. The mode here is then 36, occurring 5 times.

- 3. If a data set has 50 scores on a test e.g. 19, 25, 22, 23, ... (out of 30 points)

(a) What is an appropriate number of classes for a histogram of scores?

Using the rule of thumb #classes $\leq \frac{\#\text{data points}}{4} = \frac{50}{4} = 12.5$, we could use 12 classes.

(b) (bonus) If the range of the data set is 20, then using your answer in part (a), what is an appropriate width of a class?

Using part (a), we divide the range of the data by the # of classes (rounded up) to obtain an appropriate width of a class: $\frac{\text{range}}{\#\text{classes}} = \frac{20}{12} = 1.666..$ so rounded up, 2 is an appropriate width of a class.

2 Quiz 2 (Feb 11) solutions

Suppose that we are interested in the relation between carbon monoxide (CO) concentrations and the density of cars in some geographical area. The hundreds of cars per hour to the nearest 500 cars and the concentration of CO in parts per million (ppm) at a particular street corner are measured. The results are as follows:

$$\begin{Vmatrix} \text{Cars/hour} & 1 & 1 & 2 & 3 & 3 \\ \text{CO} & 9.0 & 7.7 & 12.3 & 20.7 & 21.6 \end{Vmatrix}$$

- 1. What should we expect to be the dependent measurement? The independent?

  We might should expect the CO levels to depend on the Cars/hour in the area, making CO in ppm the dependent measurement. The independent measurement would then be the Cars/hour.

- 2. What is $\bar{x}$?

  We calculate $\bar{x} = \frac{1+1+2+3+3}{5} = \frac{10}{5} = 2$.

- 3. Assuming that $\bar{y} = 14.3$, write out a formula for $\frac{S_{xy}}{S_{xx}} = \bar{m}$, the slope of the regression line using the data and $\bar{x}$.

Note that $\frac{S_{xy}}{S_{xx}} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$

so $\frac{S_{xy}}{S_{xx}} = \frac{(1-2)(9-14.3)+(1-2)(7.7-14.3)+(2-2)(12.3-14.3)+(3-2)(20.7-14.3)+(3-2)(21.6-14.3)}{(1-2)^2+(1-2)^2+(2-2)^2+(3-2)^2+(3-2)^2} = 6.4 = \bar{m}$

- 4. What is the y-intercept of the regression line, assuming we find that $\bar{m} = 2$?

  Assuming $\bar{m} = 2$, the y-intercept would be $\bar{b} = \bar{y} - \bar{m}\bar{x} = 14.3 - 2 * 2 = 10.3$

- 5. If we knew that our coefficient of determination $\rho^2 = 0.83$ what percentage of the variance in the data is accounted for by the regression line?

  Since $\rho^2 = 0.83$, 83% of the variance in the data in the data is accounted for by the regression line.

- 6. What does a correlation coefficient of 1 ($\rho = 1$) indicate? What if $\rho = 0$ or $\rho = -1$?

  A correlation coefficient of 1 indicates that the data all lie on a line of positive slope, a coefficient of 0 indicates the data are uncorrelated (this doesn't mean the data are not related), and a coefficient of -1 indicates the data all lie on a line of negative slope.

Bonus : What's the difference between interpolation and extrapolation?

Interpolation is to find the y-value predicted by the linear fit for an x- value that falls in the range of the x-values in your data set, while extrapolation is to do the same except for an x-value that falls outside the range of the x-values in your data.

3 Worksheet (Feb 11) solutions

(A) In a study of a free-living population of the snake Vipera bertis, researchers caught and measured nine adult females. Their body lengths x and weights y are shown in the table below:

length (cm) (x axis)	60	69	66	64	54	67	59	65	63
weight (g) (y axis)	136	198	194	140	93	172	116	174	145

- 1. Compute the regression line and the correlation coefficient. What does it indicate?

  To compute the regression line, we first compute $\bar{x} = \frac{60+69+66+64+54+67+59+65+63}{9} = 63$ and $\bar{y} = \frac{136+198+194+140+93+172+116+174+145}{9} = 152$.

  Now, to find $\bar{m}$, in order to find $\bar{b}$: $\bar{m} = \frac{S_{xy}}{S_{xx}}$
  $= \frac{(60-63)(136-152)+(69-63)(198-152)+(66-63)(194-152)+(64-63)(140-152)+(54-63)(93-152)+(67-63)(172-152)+(59-63)(116-152)+(65-$
  $\frac{}{(60-63)^2+(69-63)^2+(66-63)^2+(64-63)^2+(54-63)^2+(67-63)^2+(59-63)^2+(65-63)^2+(63-63)^2}$

  $= \frac{1237}{172} = 7.19$ where we found $S_{xy} = 1237$ and $S_{xx} = 172$.

  Then $\bar{b} = \bar{y} - \bar{m}\bar{x} = 152 - 7.19 * 63 = -300.97$, so the regression line is $y = \bar{m}x + \bar{b} = 7.19x - 300.97$.

  To calculate the correlation coefficient $\rho = \frac{S_{xy}}{\sqrt{S_{xx} * S_{yy}}}$, we need to find $S_{yy} = (136-152)^2+(198-152)^2+$
  $(194-152)^2+(140-152)^2+(93-152)^2+(172-152)^2+(116-152)^2+(174-152)^2+(145-152)^2 = 9990$.

Thus $\rho = \frac{1237}{\sqrt{172*9990}} = 0.94$. Since ρ is close to 1, we may conclude that the data are highly positively correlated.

- 2. Estimate the y-values corresponding to x-values $x = 69$ and $x = 70$. Which correspond with performing an interpolation and which with an extrapolation?

  We simply plug in $x = 69$ and $x = 70$ to our regression line: $7.19 * 69 - 300.97 = 195.14 \approx 195$ corresponds to an interpolation since $x = 69$ is in our data set. For $x = 70$, we perform an extrapolation since 70 is not an x-value in our data set: $7.19 * 70 - 300.97 = 202.33 \approx 202$

(B) Repeat step (A)(1) for the data below. It gives the infant mortality rate (MR) per 1000 live births in the United States for a period of 1960-1979:

$$\begin{Vmatrix} \text{Year} & '60 & '65 & '70 & '71 & '72 & '73 & '74 & '76 & '77 \\ \text{MR} & 26.0 & 24.7 & 20.0 & 19.1 & 18.5 & 17.7 & 16.7 & 15.2 & 14.1 \end{Vmatrix}$$

$$\begin{Vmatrix} \text{Year} & '78 & '79 \\ \text{MR} & 13.8 & 13.0 \end{Vmatrix}$$

Again, we calculate $\bar{x}, \bar{y}, S_{xx}, S_{xy}$, and S_{yy}:

$$\bar{x} = \frac{1960 + 1965 + 1970 + 1971 + 1972 + 1973 + 1974 + 1976 + 1977 + 1978 + 1979}{11} = 1972.27 \approx 1972$$

$$\bar{y} = \frac{26.0 + 24.7 + 20.0 + 19.1 + 18.5 + 17.7 + 16.7 + 15.2 + 14.1 + 13.8 + 13.0}{11} = 18.07 \approx 18.1$$

$S_{xx} = \sum (x_i - \bar{x})^2$

$= (1960 - 1972)^2 + (1965 - 1972)^2 + (1970 - 1972)^2 + (1971 - 1972)^2 + (1972 - 1972)^2 + (1973 - 1972)^2 + (1974 - 1972)^2 + (1976 - 1972)^2 + (1977 - 1972)^2 + (1978 - 1972)^2 + (1979 - 1972)^2 = 329$

$S_{yy} = \sum (y_i - \bar{y})^2$

$= (26.0 - 18.1)^2 + (24.7 - 18.1)^2 + (20.0 - 18.1)^2 + (19.1 - 18.1)^2 + (18.5 - 18.1)^2 + (17.7 - 18.1)^2 + (16.7 - 18.1)^2 + (15.2 - 18.1)^2 + (14.1 - 18.1)^2 + (13.8 - 18.1)^2 + (13.0 - 18.1)^2 = 181.77 \approx 181.8$

$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$

$= (26.0 - 18.1)(1960 - 1972) + (24.7 - 18.1)(1965 - 1972) + (20.0 - 18.1)(1970 - 1972) + (19.1 - 18.1)(1971 - 1972) + (18.5 - 18.1)(1972 - 1972) + (17.7 - 18.1)(1973 - 1972) + (16.7 - 18.1)(1974 - 1972) + (15.2 - 18.1)(1976 - 1972) + (14.1 - 18.1)(1977 - 1972) + (13.8 - 18.1)(1978 - 1972) + (13.0 - 18.1)(1979 - 1972) = -242.1$

Then $\bar{m} = \frac{S_{xy}}{S_{xx}} = \frac{-242.1}{329} = -0.735 \approx -0.74$, so $\bar{b} = \bar{y} - \bar{m}\bar{x} = 18.1 - (-0.74)(1972) = 1477.28 \approx 1477$. Then our regression line is $y = \bar{m}x + \bar{b} = -0.74x + 1477$.

To compute the correlation coefficient, we compute $\rho = S_{xy}/\sqrt{S_{xx}S_{yy}} = -242.1/\sqrt{329 * 181.8} = -0.9899... \approx -1$. Since ρ is close to -1, our data almost lies on a line of negative slope.

- 3. Is the TSS ($= S_{yy}$) close to the SSR the sum of squares of the regression?

  Since $SSR = \sum(\hat{y_i} - \bar{y})^2 = (-0.74(1960) + 1477 - 18.1)^2 + (-0.74(1965) + 1477 - 18.1)^2 + (-0.74(1970) + 1477 - 18.1)^2 + (-0.74(1971) + 1477 - 18.1)^2 + (-0.74(1972) + 1477 - 18.1)^2 + (-0.74(1973) + 1477 - 18.1)^2 + (-0.74(1974) + 1477 - 18.1)^2 + (-0.74(1976) + 1477 - 18.1)^2 + (-0.74(1977) + 1477 - 18.1)^2 + (-0.74(1978) + 1477 - 18.1)^2 + (-0.74(1979) + 1477 - 18.1)^2 = 183.4$

  Since the TSS=181.8 and the SSR is 183.4 are very similar, the data points are very close to the regression line.

4 Quiz 3 (Feb 18th) solutions

- 1. Solve $\log_4(4x) = 1$ for x.

  Take 4 raised to each side: if $\log_4(4x) = 1$, then $4^{\log_4(4x)} = 4^1$. Since $4^{\log_4(4x)} = 4x = 4 = 4^1$, we must have $x = 1$.

- 2. If $\log_a(x) = 3$ and $\log_a(y) = 4$, compute $\log_a(x^2 y^{1/2})$.

  We use the logarithm laws $\log_a(xy) = \log_a(x) + \log_a(y)$ and $\log_a(x^k) = k \log_a(x)$ of the logarithm: $\log_a(x^2 y^{1/2}) = \log_a(x^2) + \log_a(y^{1/2}) = 2 \log_a(x) + \frac{1}{2} \log_a(y) = 2 * 3 + \frac{4}{2} = 6 + 2 = 8$.

- 3. If we have data that follows an exponential trend so it was plotted as a semi-log plot as $Y = Mx + B$ where $Y = \log_{10}(y), M = \log_{10}(a), B = \log_{10}(b)$, write the equation of the exponential model y= .. (in terms of a, b, and x)

  Raise 10 to each side: since $Y = Mx + B$, $10^Y = 10^{Mx+B}$, and $10^Y = 10^{\log_{10}(y)} = y$. Also, $10^{Mx+B} = 10^{Mx} * 10^B = (10^M)^x * 10^B = (10^{\log_{10}(a)})^x * 10^{\log_{10}(b)} = a^x * b$. Then $y = b * a^x$.

- (bonus) Solve the following for x: $\log_2(\log_3(2x)) = 4$

  To solve, first raise 2 to each side: $\log_2(\log_3(2x)) = 4$ implies $2^{\log_2(\log_3(2x))} = 2^4 = 16$. Then our new equation is $\log_3(2x) = 16$. Now raise 3 to each side: $3^{\log_3(2x)} = 3^{16}$. Then $2x = 3^{16}$, so $x = \frac{3^{16}}{2}$.

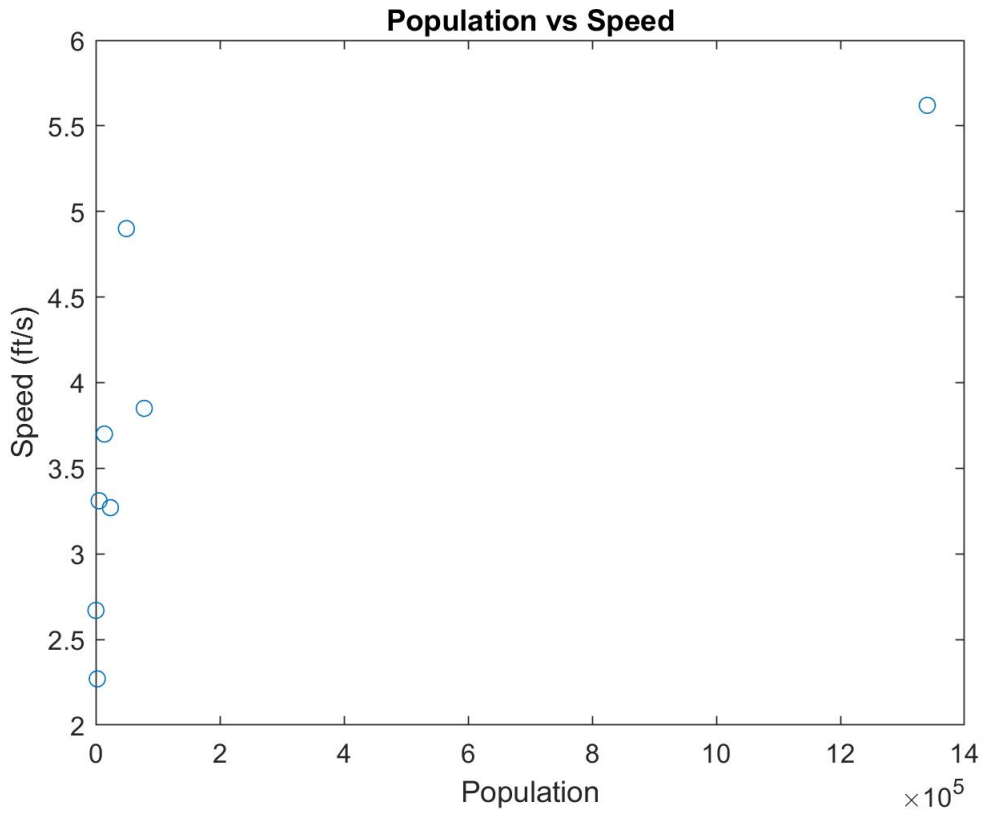
5 Worksheet (Feb 18) solutions

In an attempt to measure how the pace of city life is related to the size of the city, two researchers measured the mean speed of pedestrians in 15 cities by measuring the mean time it took them to walk 50 feet:

City	Population (x)	Speed (ft/s) (y)
Safed, Israel	14,000	3.7
Dimona, Israel	23,700	3.27
Corte, France	5491	3.31
Bastia, France	49,375	4.90
Munich, Germany	1,340,000	5.62
Psychro, Greece	365	2.67
Itea, Greece	2,500	2.27
Iraklion, Greece	78,200	3.85

With a graphing calculator,

- 1. Plot (x,y):



- 2. Compute the coefficient of determination $R^2 = \rho^2$ for the data:

We calculate $\bar{x}, \bar{y}, S_{xx}, S_{xy}$, and S_{yy}:

$\bar{x} = \frac{14{,}000+23{,}700+5{,}491+49{,}375+1{,}340{,}000+365+2{,}500+78{,}200}{8} = 189203.8 \approx 189,205$
$\bar{y} = \frac{3.70+3.27+3.31+4.90+5.62+2.67+2.27+3.85}{8} = 3.69875 \approx 3.70$
$S_{xx} = \sum(x_i - \bar{x})^2 = (14,000 - 189205)^2 + (23,700 - 189205)^2 + (5,491 - 189205)^2 + (49,375 - 189205)^2 + (1,340,000 - 189205)^2 + (365 - 189205)^2 + (2,500 - 189205)^2 + (78,200 - 189205)^2 = 1.5185625e+12 \approx 1.5 * 10^{12}$

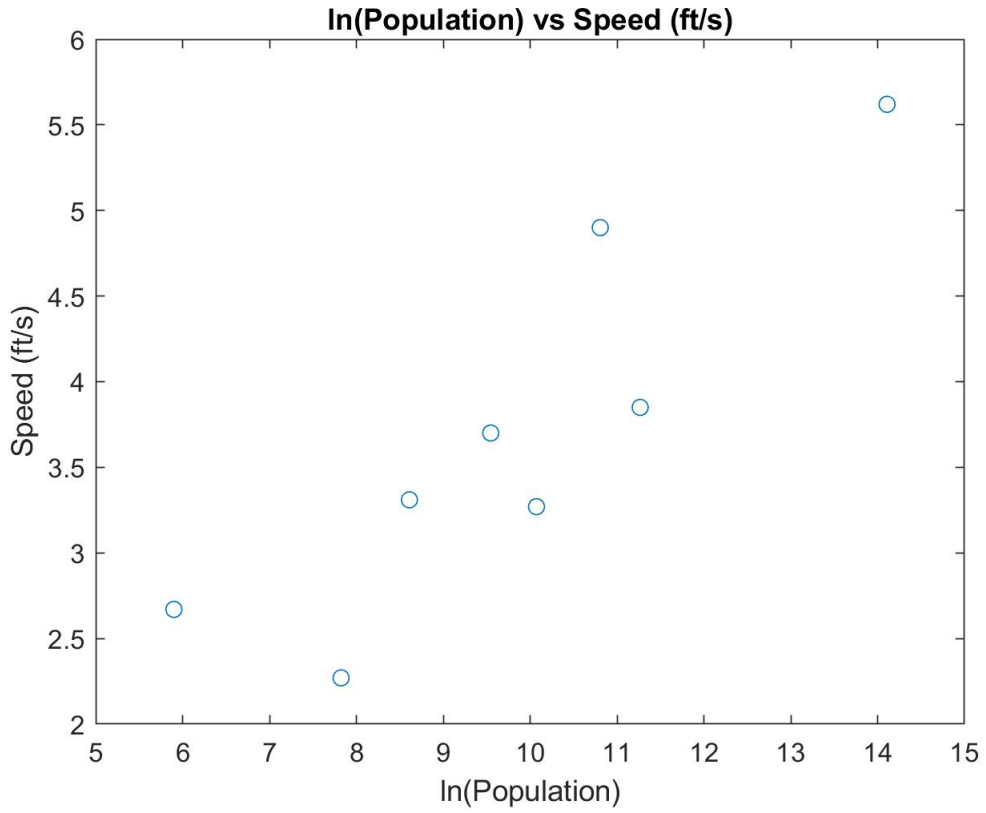
$S_{yy} = \sum(y_i - \bar{y})^2 = (3.70 - 3.70)^2 + (3.27 - 3.70)^2 + (3.31 - 3.70)^2 + (4.90 - 3.70)^2 + (5.62 - 3.70)^2 + (2.67 - 3.70)^2 + (2.27 - 3.70)^2 + (3.85 - 3.70)^2 = 8.5917 \approx 8.59$

$S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = (14,000 - 189205)(3.70 - 3.70) + (23,700 - 189205)(3.27 - 3.70) + (5,491 - 189205)(3.31 - 3.70) + (49,375 - 189205)(4.90 - 3.70) + (1,340,000 - 189205)(5.62 - 3.70) + (365 - 189205)(2.67 - 3.70) + (2,500 - 189205)(2.27 - 3.70) + (78,200 - 189205)(3.85 - 3.70) = 2,629,388.61 \approx 2,600,000$

Then $R^2 = \rho^2 = \frac{S_{xy}^2}{S_{xx} * S_{yy}} = \frac{2600000^2}{(1.5*10^{12})(8.59)} \approx 0.52$.

This indicates that the regression line only accounts for 52% of the variance in the data.

- 3. Repeat 1 & 2 for $(\ln x, y)$. Is R^2 closer to 1? What does this indicate? Find the regression line.



Now, if $X = \ln x$, we now need $\bar{X} = \overline{\ln x}$, S_{XX} and S_{Xy}:

$\bar{X} = \frac{\ln(14{,}000)+\ln(23{,}700)+\ln(5{,}491)+\ln(49{,}375)+\ln(1{,}340{,}000)+\ln(365)+\ln(2{,}500)+\ln(78{,}200)}{8} \approx 9.77$

$S_{XX} = (\ln(14,000) - 9.77)^2 + (\ln(23,700) - 9.77)^2 + (\ln(5,491) - 9.77)^2 + (\ln(49,375) - 9.77)^2 + (\ln(1,340,000) - 9.77)^2 + (\ln(365) - 9.77)^2 + (\ln(2,500) - 9.77)^2 + (\ln(78,200) - 9.77)^2 \approx 42.3$

$S_{Xy} = (\ln(14,000) - 9.77)(3.70 - 3.70) + (\ln(23,700) - 9.77)(3.27 - 3.70) + (\ln(5,491) - 9.77)(3.31 - 3.70) + (\ln(49,375) - 9.77)(4.90 - 3.70) + (\ln(1,340,000) - 9.77)(5.62 - 3.70) + (\ln(365) - 9.77)(2.67 - 3.70) + (\ln(2,500) - 9.77)(2.27 - 3.70) + (\ln(78,200) - 9.77)(3.85 - 3.70) \approx 16.9$

Thus $R^2 = \rho^2 = \frac{S_{Xy}^2}{S_{XX} * S_{yy}} = \frac{16.9^2}{42.3 * 8.59} \approx 0.79$. Therefore the regression line accounts for 79% of the variance in the data, much better than before.

The regression line is $y = \bar{m}x + \bar{b}$, where $\bar{m} = \frac{S_{Xy}}{S_{XX}}$ and $\bar{b} = \bar{y} - \bar{m} * \bar{X}$:

$\bar{m} = \frac{S_{Xy}}{S_{XX}} = \frac{16.9}{42.3} \approx 0.4$

$\bar{b} = \bar{y} - \bar{m} * \bar{X} = 3.70 - 0.4 * 42.3 = -13.22 \approx -13.2.$

Thus the regression line has equation $y = 0.4x - 13.2$.

6 Quiz 4 (Feb 25) solutions

- Find the first 5 terms of the sequence defined by $a_{n+1} - \frac{1}{4}a_n = 0$, $a_0 = 1$. Does $\lim_{n\to\infty} a_n$ exist? Why or why not?

  If $a_{n+1} - \frac{1}{4}a_n = 0$, then $a_{n+1} = \frac{1}{4}a_n$, so $a_1 = \frac{1}{4}a_0 = \frac{1}{4}$, $a_2 = \frac{1}{4}a_1 = \frac{1}{4^2}$, $a_3 = \frac{1}{4}a_2 = \frac{1}{4^3}$, $a_4 = \frac{1}{4}a_3 = \frac{1}{4^4}$. Since $4^n \to \infty$ as $n \to \infty$, $\frac{1}{4^n} \to 0$. Thus the limit of the sequence is 0.

- The probability that an adult elk survives the year is 90%. Suppose that a herd starts with 50 adult individuals and let x_n be the number of original adult elk still alive after n years. Assume that no individuals are added to the population.

  (a) Write the difference equation relating x_{n+1} to x_n.

  At the end of the first year, 90% of the elk survived. This is $x_1 = 0.9 * x_0$. At the end of the second year, 90% of those surviving elk, survived again. This is $x_2 = 0.9 * x_1$. Continuing in this pattern, the difference equation is $x_{n+1} = 0.9 * x_n$.

  (b) Find the number of original adult elk in the herd after 4 years.

  Note that $x_4 = 0.9*x_3 = 0.9*(0.9*x_2) = 0.9*(0.9*(0.9*x_1)) = 0.9*(0.9*(0.9*(0.9*x_0))) = (0.9)^4*50$.

  (bonus) What is $\lim_{n\to\infty} x_n$?
  Since we might have noticed that $x_n = (0.9)^n * 50$, and $(0.9)^n \to 0$ as $n \to \infty$, (the first few terms are $\{1, 0.9, 0.81, 0.729, ..\}$ which are decreasing) we have that x_n must also go to zero. This makes sense, eventually all the original elk will die after enough years.

7 Worksheet (Feb 25) solutions

Using the following data:

Year	'70	'75	'80	'85	'90	'00	'05	'10
No. of Eastern Blue Birds	200	300	125	250	425	575	675	500

- 1. Calculate the mean, median, range, standard deviation, and variance.

  The mean is $\bar{y} = \frac{200+300+125+250+425+575+675+500}{8} \approx 380$, the median is $\frac{425+300}{2} = 362.5 \approx 365$, and the range is $675 - 125 = 550$. The standard deviation is $\sigma = \sqrt{\frac{\sum (y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{(200-380)^2+(300-380)^2+(125-380)^2+(250-380)^2+(425-380)^2+(575-380)^2+(675-380)^2+(500-380)^2}{7}} \approx 194$ so the variance is $\sigma^2 \approx 37636$.

- 2. How many classes are appropriate for a histogram? What should be the class width for this number of classes?

  Since the data set is small, 4-6 classes is an appropriate number of classes (since typically we should have 10-20 classes, the rule of thumb fails here). We'll stick with 4 since our data set is so small. Then, an appropriate class width for this number of classes would be $\frac{range}{No.\ of\ classes} = \frac{550}{4}$ $= 137.5$(*rounded up to have the same precision as the measurements*) $= 140$.

- 3. Find the regression line and correlation coefficient. What percentage of the variance in the data is accounted for by the regression line?

  First, we find $\bar{x}$, S_{xx}, S_{yy}, and S_{xy}:

  $\bar{x} = \frac{1970+1975+1980+1985+1990+2000+2005+2010}{8} = 1989.37 \approx 1989$

  $S_{xx} = (1970-1989)^2 + (1975-1989)^2 + (1980-1989)^2 + (1985-1989)^2 + (1990-1989)^2 + (2000-1989)^2 + (2005-1989)^2 + (2010-1989)^2 = 1473$

  $S_{yy} = (200-380)^2 + (300-380)^2 + (125-380)^2 + (250-380)^2 + (425-380)^2 + (575-380)^2 + (675-380)^2 + (500-380)^2 = 262,200$

  $S_{xy} = (1970-1989)(200-380) + (1975-1989)(300-380) + (1980-1989)(125-380) + (1985-1989)(250-380) + (1990-1989)(425-380) + (2000-1989)(575-380) + (2005-1989)(675-380) + (2010-1989)(500-380) = 16,785.$

  The regression line is $y = \bar{m}x + \bar{b}$, where $\bar{m} = \frac{S_{xy}}{S_{xx}}$ and $\bar{b} = \bar{y} - \bar{m}\bar{x}$:
  $\bar{m} = \frac{S_{xy}}{S_{xx}} = \frac{16,785}{1473} \approx 11.4$

  $\bar{b} = \bar{y} - \bar{m} * \bar{X} = 380 - 11.4 * 1989 \approx -22,300.$
  Thus the regression line has equation $y = 11.4x - 22300$.
  Our correlation coefficient is $\rho = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{16,785}{\sqrt{(1473)(262,200)}} = 0.854$. This is close to 1, so our data are close to our regression line.

- 4. Now find the regression line for the $(\ln(x), y)$ data, and the correlation coefficient. Is this a better fit?
  If $X = \ln x$, we need to find $\bar{X} = \overline{\ln x_i}$, S_{Xy} and S_{XX}:

  $\bar{X} = \overline{\ln x_i} = \frac{\ln(1970)+\ln(1975)+\ln(1980)+\ln(1985)+\ln(1990)+\ln(2000)+\ln(2005)+\ln(2010)}{8} = 45.6$

  $S_{XX} = (\ln(1970)-45.6)^2 + (\ln(1975)-45.6)^2 + (\ln(1980)-45.6)^2 + (\ln(1985)-45.6)^2 + (\ln(1990)-45.6)^2 + (\ln(2000)-45.6)^2 + (\ln(2005)-45.6)^2 + (\ln(2010)-45.6)^2 \approx 8780$

  $S_{Xy} = (\ln(1970)-45.6)(200-380) + (\ln(1975)-45.6)(300-380) + (\ln(1980)-45.6)(125-380) + (\ln(1985)-45.6)(250-380) + (\ln(1990)-45.6)(425-380) + (\ln(2000)-45.6)(575-380) + (\ln(2005)-45.6)(675-380) + (\ln(2010)-45.6)(500-380) \approx -370$

  Thus the regression line is $y = \bar{m}x + \bar{b} = \frac{S_{Xy}}{S_{XX}}x + (\bar{y} - \frac{S_{Xy}}{S_{XX}}\bar{X}) = \frac{-370}{8780}x + 380 - \frac{-370}{8780} * 45.6 \approx -0.042x + 382$

  Our correlation coefficient is $\rho = \frac{S_{Xy}}{\sqrt{S_{XX}S_{yy}}} = \frac{-370}{\sqrt{8,780 * 262,200}} = -0.0077$ which is approximately 0. This indicates that our data are uncorrelated. Thus our $(\ln x, y)$ regression line is a much worse fit than the regression line from #3.