

1. Open the Minitab worksheet Homes.mtw in Student 9.
 - a) First, consider the first column, C1: Price.
 - (i) What is the shape of the distribution? Does it have a long tail? If so, on which side?

SOLUTION: Selecting **Histogram** from the graph menu, we see the distribution has a long right tail.

- (ii) How many values, if any, satisfy the rule for determining outliers? Are they at the bottom end of the distribution, the top end, or both?

SOLUTION: We select **Basic Statistics: Display Descriptive Statistics** from the stat menu, and specify **C1: Price** in the window it opens. The statistics appear in the Session window.

The interquartile range, or IQR, is $Q3 - Q1$, or 52725. An outlier below is a value less than $Q1 - 1.5IQR = 47412.5$. There are no outliers below in this distribution, since this value is less than the min.

An outlier above is a value more than $Q3 + 1.5IQR = 258312.5$. We can see from the max that there are outliers above. To find them, we need to sort the data. We do this by selecting **sort** from the manip menu, being sure to list C1 in both the **Sort Column** and **Sort by Column** slots.

Checking the upper end of the sorted data, we see there are 6 values above 258312.5, hence six outliers above.

- (iii) What is the percentile value of 200,000?

SOLUTION: We again use the sorted data. We see there are 133 values less than or equal to 200000. The total number of values in the sample is 150, so the percentile value of 200000 is $133/150 = 88.67\%$.

- b) Now consider the scatter diagram for the columns C1 and C2. What is the correlation coefficient, r ? Please give your answer to six decimal places.

Exam 1 Solutions

SOLUTION: This time, we select **Calculator** from the calc menu. First, we calculate the expression

$$(C1 - \text{mean}(C1))/\text{stdev}(C1)) * (C2 - \text{mean}(C2))/\text{stdev}(C2))$$

and store it in the first unused column, which we'll say is C7. The value of r is the average of this new column, which we calculate by using the **Calculator** again: The expression, this time, is

$$\text{sum}(C7)/N(C7),$$

and we store it in the constant k1. To find the value of k1, we select **info** from the windows menu, and use the scroll bar to get to the bottom of the **info** window. We see that $r = .692182$.

2. Suppose given a binormal distribution with

$$\begin{aligned} \bar{x} &= 420 & \bar{y} &= 345 \\ \sigma_x &= 50 & \sigma_y &= 40 & r &= -.7 \end{aligned}$$

- a) What is the slope-intercept formula for the regression line? (I.e., write it as $y = mx + b$.)

SOLUTION: The point-slope formula is

$$\begin{aligned} \frac{y - \bar{y}}{x - \bar{x}} &= r \frac{\sigma_y}{\sigma_x}, \quad \text{or} \\ \frac{y - 345}{x - 420} &= -.7 \frac{40}{50} = -.56, \quad \text{so} \\ y - 345 &= -.56x + 235.2, \quad \text{and} \\ y &= -.56x + 580.2 \end{aligned}$$

- b) What is the slope-intercept formula for the SD line?

SOLUTION: This time, the point-slope formula is

$$\begin{aligned} \frac{y - \bar{y}}{x - \bar{x}} &= \text{sign}(r) \frac{\sigma_y}{\sigma_x}, \quad \text{or} \\ \frac{y - 345}{x - 420} &= -\frac{40}{50} = -.8, \quad \text{so} \\ y - 345 &= -.8x + 336 \quad \text{and} \\ y &= -.8x + 681. \end{aligned}$$

- c) What is the regression estimate for $x = 500$?

SOLUTION: Substituting $x = 500$ in the formula for the regression line, we obtain $y = 300.2$.

- d) Suppose the x -value is in the 20-th percentile. What is the percentile value of its regression estimate?

SOLUTION: We use the SU formula:

$$\begin{aligned} \text{SU}_y(\text{reg. est.}) &= r\text{SU}_x(\text{20-th percentile of the } x\text{-dist.}) \\ &\approx (-.7)(-.84) \quad (\text{from normal table}) \\ &\approx .588 \end{aligned}$$

From the normal table, we see the percentile value of the regression estimate is approximately 72.24%.

- e) Suppose $y = 300$. What is the regression estimate for predicting the value of x from that of y ?

SOLUTION: Again, we use the SU formula, this time with the variables reversed:

$$\begin{aligned} \text{SU}_x(\text{reg. est.}) &= r\text{SU}_y(300) \\ &= -.7 \frac{300 - 345}{40} = .7875 \quad \text{so} \\ \text{reg. est.} &= .7875\sigma_x + \bar{x} = .7875 \cdot 50 + 420 = 459.375 \end{aligned}$$

3. Suppose given a binormal distribution with

$$\begin{aligned} \bar{x} &= 200 & \bar{y} &= 150 \\ \sigma_x &= 50 & \sigma_y &= 25 & r &= .8 \end{aligned}$$

Control for $x = 275$.

- a) What is the average of the controlled y -distribution?

SOLUTION: The average after controlling is the regression estimate for $x = 275$:

$$\begin{aligned} \text{SU}_y(\text{reg. est.}) &= r\text{SU}_x(275) \\ &= .8 \frac{275 - 200}{50} = 1.2 \quad \text{so} \\ \text{reg. est.} &= 1.2\sigma_y + \bar{y} \\ &= 1.2 \cdot 25 + 150 = 180 \end{aligned}$$

- b) What is the standard deviation of the controlled y -distribution?

SOLUTION:

$$\begin{aligned} \sigma_{\text{cont}} &= \sqrt{1 - r^2} \cdot \sigma_y \\ &= \sqrt{1 - .64} \cdot 25 \\ &= .6 \cdot 25 = 15 \end{aligned}$$

Exam 1 Solutions

- c) What is the 42-nd percentile of the controlled y -distribution?

SOLUTION: From the normal table, we see the SU value of the 42-nd percentile of a normal distribution is approximately $-.2$. The value in the controlled y -distribution whose SU value is $-.2$ is

$$-.2 * \sigma_{\text{cont}} + \text{reg. est.} = -.2 * 15 + 180 = 177.$$

- d) What is the percentile value of 150 in the controlled y -distribution?

SOLUTION:

$$\text{SU}_{\text{cont}}(150) = \frac{150 - 180}{15} = -2$$

From the normal table, we see this corresponds to a percentile value of 2.28%.

- e) What percent of the controlled y -values are between 150 and 200?

SOLUTION: We also need the percentile value of 200:

$$\text{SU}_{\text{cont}}(200) = \frac{200 - 180}{15} \approx 1.33$$

From the normal table, this corresponds to a percentile value of 90.82%. So the percent of the controlled y -values between 150 and 200 is $90.82\% - 2.28\% = 88.54\%$.