# Matt 1372 HW2

2.1 Mammal life spans. Data were collected on life spans (in years) and gestation lengths (in days) for 62 mammals. A scatterplot of life span versus length of gestation is shown below.[15]

(a) What type of an association is apparent between life span and length of gestation?

(b) What type of an association would you expect to see if the axes of the plot were reversed, i.e. if we plotted length of gestation versus life span?

(c) Are life span and length of gestation independent? Explain your reasoning.
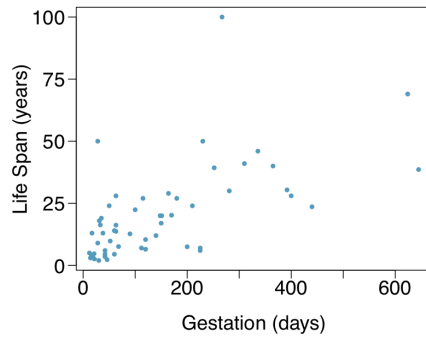
(a) positive association

(b) positive association

(c) NO, mammals with longer gestions tend to live longer as well based on the data.

2.2 Associations. Indicate which of the plots show (a) a positive association, (b) a negative association, or (c) no association. Also determine if the positive and negative associations are linear or nonlinear. Each part may refer to more than one plot.


(1) (2) (3) (4)
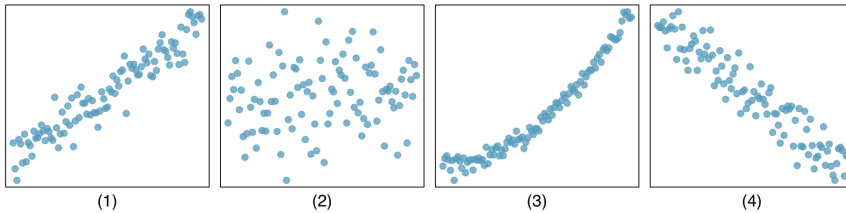
                    linear   nonlinear
(a) positive: (1) , (3)
(b) negative: (4) linear
(c) no association: (2)

2.3 Reproducing bacteria. Suppose that there is only sufficient space and nutrients to support one million bacterial cells in a petri dish. You place a few bacterial cells in this petri dish, allow them to reproduce freely, and record the number of bacterial cells in the dish over time. Sketch a plot representing the relationship between number of bacterial cells and time.
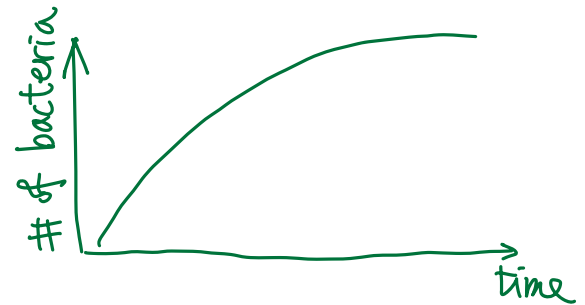
2.4 Office productivity. Office productivity is relatively low when the employees feel no stress about their work or job security. However, high levels of stress can also lead to reduced employee productivity. Sketch a plot to represent the relationship between stress and productivity.
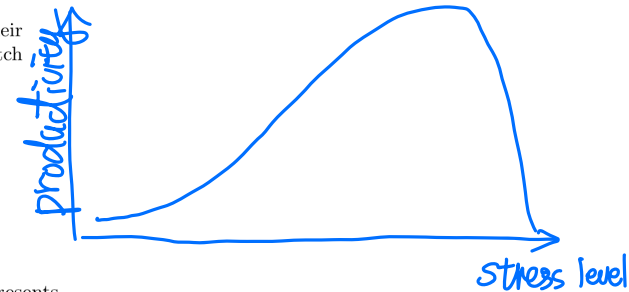
2.5 Parameters and statistics. Identify which value represents the sample mean and which value represents the claimed population mean.

(a) American households spent an average of about $52 in 2007 on Halloween merchandise such as costumes, decorations and candy. To see if this number had changed, researchers conducted a new survey in 2008 before industry numbers were reported. The survey included 1,500 households and found that average Halloween spending was $58 per household.

(b) The average GPA of students in 2001 at a private university was 3.37. A survey on a sample of 203 students from this university yielded an average GPA of 3.59 a decade later.

(a) population mean $\mu = \$52$
Sample mean $\bar{x} = \$58$

(b) population mean $\mu = 3.07$
Sample mean $\bar{x} = 3.59$

2.6 Sleeping in college. A recent article in a college newspaper stated that college students get an average of 5.5 hrs of sleep each night. A student who was skeptical about this value decided to conduct a survey by randomly sampling 25 students. On average, the sampled students slept 6.25 hours per night. Identify which value represents the sample mean and which value represents the claimed population mean.

Population mean $\mu = 5.5$ hrs
Sample mean $\bar{x} = 6.25$ hrs

2.7 Days off at a mining plant. Workers at a particular mining site receive an average of 35 days paid vacation, which is lower than the national average. The manager of this plant is under pressure from a local union to increase the amount of paid time off. However, he does not want to give more days off to the workers because that would be costly. Instead he decides he should fire 10 employees in such a way as to raise the average number of days off that are reported by his employees. In order to achieve this goal, should he fire employees who have the most number of days off, least number of days off, or those who have about the average number of days off?

Assume n employees in this mining site who have d_i days paid off. $i = 1, \cdots, n$

Since the manager doesn't want to give more day off to each employee, then $d_1, d_2, \cdots, d_n$ will not be changed.

Given the current mean $35 = \dfrac{d_1 + d_2 + \cdots + d_n}{n}$.

In order to increase the mean without changing any one of the d_i.

firing the employee who has paid day off between the mean and the minimum will be able to achieve this goal.

However, firing those employees who have less paid day off (and have longer work days) will hit the productivity the most and it is not very wise.

2.8 Medians and IQRs. For each part, compare distributions (1) and (2) based on their medians and IQRs. You do not need to calculate these statistics; simply state how the medians and IQRs compare. Make sure to explain your reasoning.

(a) (1) 3, 5, 6, 7, 9
    (2) 3, 5, 6, 7, 20

(b) (1) 3, 5, 6, 7, 9
    (2) 3, 5, 7, 8, 9

(c) (1) 1, 2, 3, 4, 5
    (2) 6, 7, 8, 9, 10

(d) (1) 0, 10, 50, 60, 100
    (2) 0, 100, 500, 600, 1000

(a) their medians are the same, but (2)'s IQR is higher than (1)'s.

(b) Both median and IQR from (2) is higher than those from (1)

(c) Their IQRs are the same, but (2)'s median is higher than (1)'s

(d) Both median and IQR from (2) is higher than those from (1)

2.9 Means and SDs. For each part, compare distributions (1) and (2) based on their means and standard deviations. You do not need to calculate these statistics; simply state how the means and the standard deviations compare. Make sure to explain your reasoning. *Hint:* It may be useful to sketch dot plots of the distributions.

(a) (1) 3, 5, 5, 5, 8, 11, 11, 11, 13
    (2) 3, 5, 5, 5, 8, 11, 11, 11, 20

(c) (1) 0, 2, 4, 6, 8, 10
    (2) 20, 22, 24, 26, 28, 30

(b) (1) -20, 0, 0, 0, 15, 25, 30, 30
    (2) -40, 0, 0, 0, 15, 25, 30, 30

(d) (1) 100, 200, 300, 400, 500
   (2) 0, 50, 300, 550, 600

(a) mean from (2) is higher than mean from (1) since 20 > 13.
So is standard deviation since 20 is further from the rest numbers than 13

(b) mean from (1) is higher than mean from (2) since -20 > -40
But (2)'s standard deviation is higher since -40 is further from the rest numbers than -20

(c) mean from (2) is higher than mean from (1) since each number from (2) is 20 more comparing with those from (1), but the standard deviations are the same since they are equally variable around their respective means.

(d) They have the same mean but data (2) has higher standard deviation since the data are further from the mean than in data (1).

2.10 Mix-and-match. Describe the distribution in the histograms below and match them to the box plots.
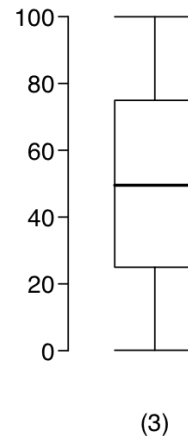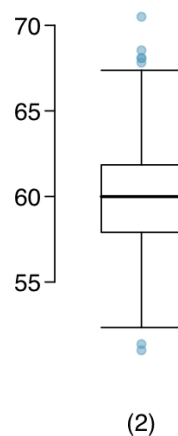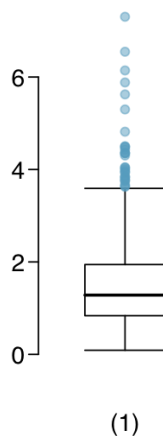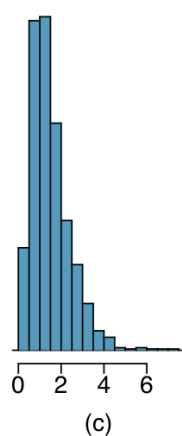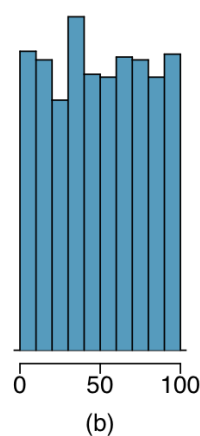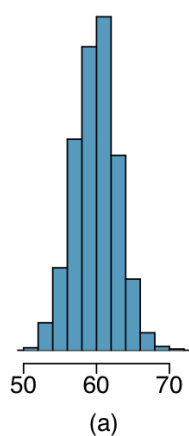
(a) It is unimodal and symmetric, and about 95% of the data falls within about 7 units of the center, so standard deviation will be about 3 or 5, this matches (2)

(b) It is uniformly distributed from 0 to 100 with median close to 50. this matches (3)

(c) It is right-skewed and unimodal which matches (1).