

STAT 201: Elementary Statistics Session 13 & 14

Exam 2

Instructor: Haigang Liu

1. (30 points) Toluca is a city in Mexico that makes replacement parts for refrigerators. We consider one particular part, which is manufactured in different lot sizes. In our study, researchers want to relate work hours to lot sizes. $n = 25$ pairs of observations are obtained.

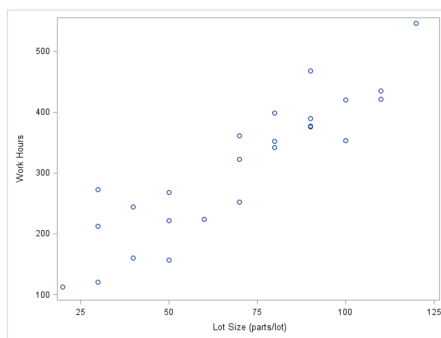


Figure 1: The scatter plot of lot size v.s. work time

1) Read the scatterplot in Figure 1 and comment on strength, direction of the correlation between lot sizes (X) and work hours (Y). Is it justifiable to apply a linear model in this scenario? Explain.

Root MSE	48.82331	R-Square	0.8215
Dependent Mean	312.28000	Adj R-Sq	0.8138
Coeff Var	15.63447		

Figure 2: Software Output 1

2) Calculate r , the correlation coefficient, based on Figure 2.

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	62.36586	26.17743	2.38	0.0259
size	Lot Size (parts/lot)	1	3.57020	0.34697	10.29	<.0001

Figure 3: Software Output 2

- 3) Write down the linear regression model based on the output given by Figure 3.
- 4) Interpret the slope of the regression equation you obtain in 3) *in the context of this problem*.
- 5) Apply this equation to predict the work hours for a lot size = 80.
- 6) Suppose the true value of work hour for lot size = 80 is 342. Find the residual of your prediction.

2. (20 points) Three PhD students in statistics department with no particular background in driver's education decide to take the permit exam in South Carolina. Each exam is graded as a Pass or Fail. Answer the following questions.

- 1) How many outcomes are possible for the grades received by the three friends together? Write down the sample space.
- 2) We assume that all possible outcomes are equally likely. (Or, equivalently, a randomly selected person has 50% chance to pass.) Calculate the probability of all three passing the exam.
- 3) In practice, the state-wide 80% of adults pass the exam. If the three students are a random sample of South Carolina residents, find the probability of any two of them passing the exam. *Hint:* You can use binomial distribution if you wish. The formula is given as follows:

$$\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x},$$

where n is the number of trials, and p is the probability of a success.

- 4) In practice, we may observe that probabilities that apply to a random sample are not likely to be valid for this sample of three friends. Provide a possible explanation.

3. (25 points) Suppose a Starbucks manger wants to study the relation between caffè macchiato and brownie sales. The manager recorded orders for 1000 customers of a certain week. The results are below.

	B	B ^c
C	300	500
C ^c	150	50

Table 1: The contingency table of Coffee and Brownie

Note that B denotes a person that orders Brownie, and B^c denotes a person that does not order Brownie. Similarly, C denotes a person that orders caffè macchiato, and C^c denotes a person that does not place the order. An alternative ways of displaying Table 1 is Figure 4. Read either Table 1 or Figure 4, and answer the following questions. *Hint:* The formula of conditional probability is $P(A|B) = \frac{P(A \cap B)}{P(B)}$, or equivalently, $P(A \cap B) = P(A)P(A|B)$.

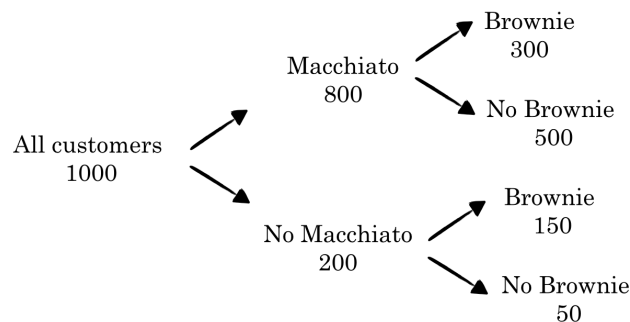


Figure 4: Tree diagram for Brownie and caffè macchiato

- 1) For this sample, what is the probability that a randomly selected customer ordered caffè macchiato, i.e., $P(C)$?
- 2) For this sample, given that someone ordered caffè macchiato, what is the probability that he or she has also ordered a Brownie, i.e. $P(B|C)$?
- 3) For this sample, what is the probability of a randomly selected person ordering both caffè macchiato and Brownie, i.e. $P(B \cap C)$?
- 4) Suppose $P(B) = 0.45$. Are choosing caffè macchiato and Brownie independent events? Justify your answer by comparing $P(B)$ with one of the previous quantities.
- 5) True or False: If B and C are independent, they can also be disjoint as well. Explain your claim.

4. (15 points) How to tell the difference between fabricated data and data from real world? This might be a question posed after the accounting scandals with Enron, a large energy company. In fact, there is a way named Benford's Law to examine accounting books to determine whether they had been "doctored" or used fabricated data. The Benford's law states that, under a variety of circumstances, numbers of population in small towns, figures in newspapers and magazines, tax returns and other business records are more likely to begin with 1. More precisely, the distribution of beginning digits of these numbers of is given in Table 2. Read the table and answer the following questions.

x	P(x)
1	0.30
2	0.18
3	0.12
4	0.10
5	0.08
6	0.07
7	0.06
8	0.05
9	0.04

Table 2: The distribution of starting digits given by Benford's Law.

- 1) Verify that this is a valid probability distribution.
 - 2) Find the expectation of X , i.e, EX .
 - 3) When you randomly select a digit with a random number table, the probability of obtaining 1-9 as starting digit are equally likely. That explains why people are more likely to pick 5 or 6 as starting digits when they fabricate numbers by themselves. Find and compare the probability of 5 or 6 as starting digits by i) random selection and ii) Benford's Law.
5. (10 points) When you buy Mac Pro from Apple's website, there are usually many different configurations. The first choice to make is quad-core or 6 core GPU. After that, you will be required to pick a processor from the following candidates.

- 3.5GHz 6-core with 12MB of L3 cache
- 3.0GHz 8-core with 25MB of L3 cache [Add \$1,500.00]
- 2.7GHz 12-core with 30MB of L3 cache [Add \$3,000.00]

Figure 5: Possible Processors

Choices of memories are given in Figure 6.

- 16GB (4x4GB) of 1866MHz DDR3 ECC
- 32GB (4x8GB) of 1866MHz DDR3 ECC [Add \$400.00]
- 64GB (4x16GB) of 1866MHz DDR3 ECC [Add \$1,200.00]

Figure 6: Possible Memories

Finally, there are 3 choices for storage.

- 256GB PCIe-based flash storage
- 512GB PCIe-based flash storage [Add \$300.00]
- 1TB PCIe-based flash storage [Add \$800.00]

Figure 7: Possible Storage

- 1) How many different configuration do you have when buying a Mac Pro?
- 2) If all configurations are equally likely, find the probability of getting a Mac Pro with 16GB memory and 1TB storage (whatever other configurations).

ANSWER SHEET PAGE ONE

First Name:

Last Name:

ANSWER SHEET PAGE TWO

First Name:

Last Name:

ANSWER SHEET PAGE THREE

First Name:

Last Name:

ANSWER SHEET PAGE FOUR

First Name:

Last Name:

ANSWER SHEET PAGE FIVE

First Name:

Last Name: