

---

*SHOW ALL WORK**In most calculations keep four decimal places.*

---

1. [10] Find the **mean** and the **standard deviation** of the sample below.

-2      1      4      3      4      2

2. An SRS of size 21 is as follow:

6   10   9   10   14   13   11   13   15   15   18   17   18   18   19   19   23   23   24   27   36

- 1) [10] Find the **five-number summary**.
- 2) [6] Is there any outlier in this data set? If any, which value? Use 1.5 IQR rule.
3. In the United States, 40% of the population have brown eyes. If a sample of 14 people are randomly selected, let  $x$  be the number of people having brown eyes in the sample, then  $x$  has binomial distribution with  $n=14$ ,  $p=0.4$ .
- 1) [6] What is the probability that exactly 2 people have brown eyes in the sample?
- 2) [6] What is the probability that at least one people have brown eyes in the sample?

4. Random variable  $X$  has binomial distribution  $B(100, 0.3)$ .
- 1) [6] Find the **mean** and **standard deviation** of  $X$ .
  - 2) [5] Use the normal approximation to find the probability  $P(15 \leq X \leq 34)$
5. Men's heights are Normal distributed with mean 69.0 in, and standard deviation 2.8 in.
- 1) [5] An airplane has a doorway height of 61.6 in. What percentage of adult men can fit through the doorway without bending?
  - 2) [5] The U.S. Marine Corps requires that men have heights between 58 in and 80 in. Find the percentage of men who meet the height requirements.
  - 3) [4] How short must a man be to be in the shortest 3%?
  - 4) [6] If randomly select 16 men what is the probability that their average height is greater than 70 in?

6. Suppose we use  $x$ , the cost of an airplane ticket purchased 30 days in advance to predict  $y$ , the cost of the ticket if purchased one day in advance. The correlation  $r=0.782$ . A summary of a random sample is:

Variables	Mean	standard deviation
Cost (in dollars) purchased 30 days in advance, $x$	$\bar{x} = 272.6$	$s_x = 23.34$
Cost (in dollars) purchased one day in advance, $y$	$\bar{y} = 690.3$	$s_y = 233$

- 1) [8] Find the **intercept** and **slope** of the least square regression line.
  
  - 2) [3] Write the equation of the regression line.
  
  - 3) [3] Predict the cost of a ticket if purchased one day in advance given that the cost of the ticket is \$300 purchased 30 days in advance.
7. Chose an American household at random. The table below shows the probability of number of cars the household owns.

Number of cars	0	1	2	3	4	5	6 or more
probability	0.09	0.29	0.38	0.16	0.05	0.02	0.01

- 1) [5] Is the table above a probability distribution? Why?
  
- 2) [5] What percent of household having more than 2 cars?

8. Suppose you applied two jobs. Defined the events:

A= You will get the first job.

B= you will get the second job.

Given that  $P(A) = 0.5$ ,  $P(B) = 0.7$ , and  $P(A \text{ and } B) = 0.3$ .

1) [5] Find  $P(A \text{ or } B)$

2) [5] Given you got the first job offer, what is the condition probability that you will get the second job.

9. It is desired to test whether there is any relationship between seat belt use and cigarette smoking. The two way table below summaries the data from a simple random sample of 459 people.

Seat belt use status		Number of Cigarettes Smoked per Day				total
		0	1-14	15-34	35 and over	
Wear seat belt	Observed	175	20	42	6	243
	Expected	(171.53)	(19.59)	(43.94)	(7.94)	
	Cell chisq	[ 0.07 ]	[0.009]	[0.086]	[0.489]	
Don't wear seat belt	Observed	149	17	41	9	216
	Expected	(152.47)	(14.56)	( )	(7.06)	
	Cell chisq	[ 0.079 ]	[ ]	[ ]	[ 0.533 ]	
total		324	37	83	15	459

1) [6] Find the value of the **expected** count that is not given in the table. Find the **2 cells chi square** contribution that need to be computed. Show work.

2) [12] Do a 4 - step hypotheses test. Show how the appropriate table is used ( $\alpha = 0.05$ ).

10. A class survey in a large class for first – year college students asked, “About how many minutes do you study on a typical weeknight?” The mean response of random selected 400 students was  $\bar{x} = 115$  minutes. Suppose we know that the study time follows a Normal distribution with mean  $\mu$  and standard deviation  $\sigma = 60$  minutes in the population of all first – year students.

- 1) [2] To make a confidence interval for  $\mu$  do you use T-procedure or Z-procedure? (Circle one)
  - a) T-procedure
  - b) Z-procedure
- 2) [8] Use the survey result to give a 95% confidence interval for  $\mu$ .

- 3) [5] If you want to make a 99% confidence interval with margin of error within 4 minutes, how large should sample size  $n$  be?

11. Bags of a certain brand of tortilla chips claim to have a net weight of 16 ounces. Net weights actually vary slightly from bag to bag and are Normally distributed with mean  $\mu$ . A representative of a consumer advocate group wishes to see if there is any evidence that the mean net weight is less than advertised.

To do this, he selects 24 bags of this brand at random and determines the net weight of each. He finds the sample mean to be  $\bar{x} = 15.68$  ounces and the sample standard deviation to be  $s = 0.64$  ounces.

- 1) [2] To do a hypotheses test for  $\mu$  do you use T-procedure or Z-procedure? (Circle one)
  - a) T- procedure
  - b) Z-procedure
- 2) [12] Do an appropriate four-step hypotheses test. (  $\alpha = 0.05$  )
  - a) Write the hypotheses:

$H_0$ :

$H_a$ :

- b) Calculate the test statistic.

- c) Find the p-value.

- d) Conclude:

12. A researcher suspect that lean people spent more time on daily activity ( $\mu_1$ ) than obese people do ( $\mu_2$ ). 10 lean people and 10 obese people were randomly selected from normal distributions. Attach sensors that monitor the subjects' every move for 10days. The times (in minutes per day ) that each subject spent on standing and walking were recorded. The MINITAB outputs below are based on the data collected from the subjects.

Let Mud =  $\mu_1 - \mu_2$

test	alternative	T-value	p-value	95% CI
matched pairs :	mud not = 0	3.814	0.0041	( 62.04, 242.96)
matched pairs :	mud > 0	3.814	0.0021	( 62.04, 242.96)
matched pairs :	mud < 0	3.814	0.9979	( 62.04, 242.96)
2 sample t :	mud not = 0	3.807	0.0017	( 67.2, 237.8)
2 sample t :	mud > 0	3.807	0.0008	( 67.2, 237.8)
2 sample t :	mud < 0	3.807	0.9992	( 67.2, 237.8)

- 1) [4] To test the researcher's suspect, which procedure should be used? ( Circle one)
  - a) One sample z-test
  - b) two sample t-test
  - c) matched-pairs t-test
  
- 2) [10] Using the output above do a 4- step hypotheses test. (  $\alpha=0.05$ )

13. Each person in a random sample of 2000 "likely voters" (as defined by a professional polling organization) was questioned about his or her political views. Of those surveyed, 1308 felt that "the economy's state" was the most urgent national concern.

- 1) [2] Calculate  $\hat{p}$ , the sample proportion of likely voters that felt the economy's state was the most urgent national concern.
  
- 2) [4] What is the margin of error of the 95% confidence for the proportion  $p$  of all likely voters that feel the economy's state is the most urgent national concern?
  
- 3) [4] Make a 95% confidence interval for  $p$ .

14. [6] Make a stemplot for the data below.

5 3 10 11 11 20 24 29 28 41 42 44 47 61

15. [20] Circle the correct answer

1. How large a sample would you need to estimate population proportion  $p$  with margin of error 0.03 with 99% confidence? Assume that you don't know anything about the value of  $p$ .  
A) 600      B) 1843      C) 22      D) 1844
2. Event A and event B = 'A does not occur' are independent.  
A) True      B) False
3. If there is an extreme high outlier in your data, which summary would you use?  
A) Mean and standard deviation    B) five-number summary
4. Which correlation value indicates the weakest linear relationship?  
A)  $r = -0.95$     B)  $r = 0.81$     C)  $r = -0.56$     D)  $r = 0.4$
5. If the correlation between two variables is close to 0, you can conclude that a scatterplot show  
A) A strong straight-line pattern.  
B) A cloud of points with no visible pattern.  
C) No straight-line pattern, but there might be a strong pattern of another form.
6. Let  $x$  be the number of lottery you buy until you win. Then  $X$  has a binomial distribution.  
A) True      B) False
7. Suppose the 95% confidence interval for mean salary of SIU graduates is (\$50,000, \$55,000). Then you will have a 95% chance to get a job with salary between \$50,000 and \$55,000.  
A) True      B) False
8. To study how college students feel about the Social Security system. A questionnaire was mailed to 750 college students selected at random. Only 600 questionnaires were returned. The 750 students are:  
A) Population    B) sample      c) not population and not sample.
9. Take a simple random sample of size 25 from a normal distribution  $N(4, 20)$ . The distribution of the sample mean  $\bar{X}$  is  
A)  $N(0,1)$       B)  $N(4,4)$       C)  $N(4, 5)$
10. A student stands in front of student center collecting student opinions. Identify the type of sampling is used.  
A) Stratified      B) convenience      C) voluntary response sampling.