

Math 13
Statistics
Problem Sets

Fall 2009
Section #5117

Instructor
Greg Perkins

I. EXPERIMENTAL APPROACH

Two Coins. Flip two coins ten times. Record the results below, then find the probabilities asked for.

$P(H H)=$

$P(T H)=$

$P(H T)=$

$P(T T)=$

Three Coins. Flip 3 coins ten times. Record the results below, then find the probabilities asked for.

$P(H H H)=$

$P(2 \text{ heads})=$

$P(2 \text{ tails})=$

$P(T T T)=$

Class Averages:

$P(H H) =$

$P(H H H) =$

II. CLASSICAL APPROACH

Two Coins. Write down all possible head/tail combinations when two coins are flipped. Fill in the probabilities.

$P(H H)=$

$P(T H)=$

$P(H T)=$

$P(T T)=$

Three Coins. Write down all possible head/tail combinations when 3 coins are flipped. Fill in the probabilities.

$P(H H H)=$

$P(2 \text{ heads})=$

$P(2 \text{ tails})=$

$P(T T T)=$

III. CONCLUSIONS

- 1) Why are your answers for $P(H H)$ different for the experimental and classical approach?
- 2) Which is a better approximation for $P(H H)$, your experimental answer or the class average? Explain!

1) Does the order in which the objects are chosen matter? Answer yes or no.

_____ Runners are chosen for a cross-country team.

_____ Three club members are chosen to work on committee.

_____ Three club members are chosen to serve as president, vice-president and treasurer.

_____ Answers to a 20-question true/false test.

_____ Ten different calculators are chosen from a batch of 100 calculators.

_____ Three numbers are chosen for a combination lock.

_____ Names are chosen for 3 newborn babies.

2) According to the US Census Bureau, 64.1% of the people living in Salinas are Latino. Suppose that 10 people living in Salinas were chosen at random for our survey.

a) What is the probability that all 10 would be Latino?

b) Find the probability that at least one of the 10 people would be Latino.

3) In California, car license plates are made up of one digit, three letters and then 3 digits. For example, 3 GJW 832. It's not true, but assume that all numbers and letters could be used.

a) How many possible license plates are there in California?

b) A witness to a crime saw the license plate of the getaway car. She said it started out 3 TKY
How many cars in California can have this? What percentage of all license plates in California start with 3 TKY?

4) Cards are drawn from a well-shuffled deck of 52. Find the probability for each event below.

a) Drawing a King or a heart.

b) Drawing a Jack or a seven.

c) Drawing two Kings, assuming cards are not replaced.

d) [Extra Credit] Drawing a Queen and then a heart.

5) It is a fact that Allan Iverson makes 81.6% of his freethrows. Find the following probabilities.

a) He makes 4 freethrows out of 4 attempts.

b) He makes at least one shot when he attempts 2 shots.

c) He makes 5 shots out of 6 attempts.

Mega Millions Pick 5 numbers from the numbers 1 to 56 and one MEGA number from 1 to 46.

- 1) Let X = the number of correct numbers on one lotto ticket.
 - a) Construct the probability distribution table for this situation.
 - b) Find the mean for the random variable X . EXPLAIN what this number represents.
 - c) Calculate the probability of winning a cash prize.
 - d) Suppose that a person bought 1 million of these lotto tickets. How many would you expect to be cash winners?
- 2) We will now turn this into a binomial situation by considering it a success if a person wins any cash prize. See part c above for the value of p .
 - a) Suppose that a person bought 10 lotto tickets. How many cash winners would you expect?
 - b) Find the probability that out of 10 tickets purchased, none will be cash winners.
 - c) How many tickets must you purchase in order to expect one cash winner?
- 3) Let X = the amount of profit from one lotto ticket. Listed below are the winning amounts for the drawing that took place July 22, 2008.

**Matching
Numbers**

Prize Value

5 + Mega	\$ 170,000,000
5	\$ 250,000
4 + Mega	\$ 13,171
4	\$ 163
3 + Mega	\$ 180
3	\$ 7
2 + Mega	\$ 11
1 + Mega	\$ 3
Mega	\$ 2

- a) Construct the probability distribution table for this situation. Do not forget the cases when the person loses!! You can find the probabilities you need in #1 part a.
 - b) Find the mean for the random variable X . EXPLAIN what this number represents.
- 4) Is it worthwhile to play the lottery? Which of the statistics above influenced your decision the most?

Observe the changes in the confidence intervals as the parameters change.

1) An older Americans organization believes that a certain corporation is discriminating against the elderly in their hiring practices. They were able to obtain the ages of people hired in the last year.

(a) 40 people hired $\bar{x} = 35$ $s = 4.5$ 90% CI _____

(b) 40 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(c) 40 people hired $\bar{x} = 35$ $s = 4.5$ 99% CI _____

(d) Describe what happens to the size of the confidence interval as the level of confidence increased.

Why does this make sense?

(e) 40 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(f) 400 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(g) 4000 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(h) Describe what happens to the size of the confidence interval as the sample size increased.

Why does this make sense?

(i) 40 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(j) 40 people hired $\bar{x} = 35$ $s = 2.5$ 95% CI _____

(k) 40 people hired $\bar{x} = 35$ $s = 0.5$ 95% CI _____

(l) Describe what happens to the size of the confidence interval as the standard deviation decreased.

Why does this make sense?

Now consider what happens when the sample size gets small.

(m) 32 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(n) 30 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(o) 20 people hired $\bar{x} = 35$ $s = 4.5$ 95% CI _____

(p) Describe what happens to the size of the confidence interval as the sample size decreased.
Why does this make sense?

2) A surgeon is trying to determine is a new procedure if as successful as the old method. The new method causes the patient less discomfort, but the old method is 92% successful. Consider the following test groups for the new procedure.

(a) 30 patients, 27 successes 95% CI _____

(b) 40 patients, 36 successes 95% CI _____

(c) 87 patients, 81 successes 95% CI _____

(d) 143 patients, 130 successes 95% CI _____

(e) What happens to the confidence interval as the sample size is increased?

(f) Which of the four test groups should the doctor use for the report to the medical journals? WHY??

a) Tell which estimate is closer to the actual answer.

b) Explain in your own words WHY one estimate is closer.

c) Explain WHY neither estimate is perfectly correct.

- 1) An educational testing company has designed a standard test of mechanical aptitude. Scores on this test are normally distributed with a mean of 75 and a standard deviation of 15.
- If a subject is randomly selected and tested, find the probability that the person will score between 70 and 90.
 - If a subject is randomly selected and tested, find the probability that the person will score at least 80.
 - Find the 33rd percentile for the distribution.
 - If 80 subjects are randomly chosen and tested, find the probability that their mean score is below 78.
- 2) In a study of causes of death, it was found that 52% of Americans die of heart disease.
- What would the random variable be in this situation?
 - Find the probability that in a group of 8 randomly selected Americans, at least one will die of heart disease.
 - Find the probability that in a group of 800 randomly selected Americans, at least 450 will die of heart disease.
- 3) Chips Ahoy! claims that there are at least 1000 chips in every bag of their chocolate chip cookies. A consumer advocate tested 458 cookies and found that there were an average of 24.7 chips in each cookie with $s = 12.3$. Assuming that there are usually 35 cookies in each bag, is the claim that Chips Ahoy! made true? Use a confidence interval to prove your point.
- 4) [Fred Hoiberg](#) of the [Minnesota Timberwolves](#) makes 48.3% of his 3-point attempts. Last season, [Allen Iverson](#) made 104 out of 338 of his 3-point shots. Is Iverson even comparable to Hoiberg when it comes to 3-point shots?
- Use a confidence interval to answer this question.
 - Find the probability that [Fred Hoiberg](#) would make at most 104 shots out of 338 attempts. Use this to answer the question comparing the two players.
- 5) A psychic runs the following ad in a magazine:

Expecting a baby? Renowned psychic will tell you the gender of the unborn child from any photograph of the mother. Cost \$10. Money back guarantee.

This may be a profitable con game. Suppose that the psychic replies "girl" to every inquiry. In the worst case, only those people that have a boy will demand their money back. Find the expected value of the psychic's profit from each customer.

MULTIPLE CHOICE:

- 6) According to the Central Limit Theorem, if you randomly select all samples of size 16 from a population, then the standard deviation of the sample means will be:
- (a) equal to σ (b) half of σ (c) one fourth of σ (d) one sixteenth of σ
- 7) For the probability distribution of a DISCRETE random variable X , the probability of any single value of X is always:
- (a) 0 (b) 1 (c) between 0 and 1 (d) none of these
- 8) For the probability distribution of a CONTINUOUS random variable X , the probability of any single value of X is always:
- (a) 0 (b) 1 (c) between 0 and 1 (d) none of these

For these tests you will need to use the Student-t test.

1) The claim is that the mean IQ score is 100. Test the claim using 98% confidence.

a) $n = 20$ $\bar{x} = 104$ $s = 10$

CV _____ TS _____ Conclusion _____

b) $n = 40$ $\bar{x} = 104$ $s = 10$

CV _____ TS _____ Conclusion _____

c) $n = 100$ $\bar{x} = 104$ $s = 10$

CV _____ TS _____ Conclusion _____

d) Explain what happens to the results of the test as the sample size increases. WHY?

2) The claim is that the mean IQ score is 100. Test the claim using 98% confidence.

a) $n = 40$ $\bar{x} = 104$ $s = 5$

CV _____ TS _____ Conclusion _____

b) $n = 40$ $\bar{x} = 104$ $s = 15$

CV _____ TS _____ Conclusion _____

c) $n = 40$ $\bar{x} = 104$ $s = 25$

CV _____ TS _____ Conclusion _____

d) Explain what happens to the results of the test as the standard deviation increases. WHY?

- 3) A claim has been made that men and women with BA's earn the same amount, \$28,590 per year. Three samples of women were taken. Which sample provides the **strongest** evidence against the claim? Use either the CV or the P-value method to support your answer.

Sample 1: $n = 32$ $\bar{x} = \$26,100$ $s = \$7,500$

Sample 2: $n = 45$ $\bar{x} = \$26,200$ $s = \$8,900$

Sample 3: $n = 53$ $\bar{x} = \$26,150$ $s = \$9,650$

- 4) A new IQ test has been introduced. The claim is the mean is the same as the old test (100). The new test was given to two sample groups. Test the claim using each sample.

Which sample provides **stronger** evidence against the claim? WHY?

Sample 1: $n = 30$ $\bar{x} = 104.76$ $s = 13.1$

CV _____ TS _____ Conclusion _____

Sample 2: $n = 31$ $\bar{x} = 104.76$ $s = 13.1$

CV _____ TS _____ Conclusion _____

On a separate page are the current homework and overall averages for THIS semester's Statistics students.

1) Test for a linear correlation between HW and OVERALL scores.

2) Predict the overall average of a person that scores 85% on homework.

3) Test the claim that the means for HW and OVERALL scores are equal.

4) Test the claim that the mean for OVERALL scores is at least 70.

5) Construct the CI for the proportion of people that are getting a passing (70 or better) average overall.

6) Divide the data into two groups: those that have a passing HW average and those that do not.

It would help to use 2 different colors to highlight people in each group.

a) Test the claim that the proportion of people with passing OVERALL score is the same for the two groups.

b) Construct the CI for the difference between the mean OVERALL scores for the two groups of students.

7) Explain what all of these tests say about the relationship between HW and OVERALL scores in this class.

The data below was collected from Hartnell College students. They were asked for their Hartnell GPA, the number of hours they work per week and the number of units they have.

GPA	WORK	UNITS
3.4	15	6
2.48	30	5
3.15	8	15
3.4	16	8
2.8	40	6
3.94	20	12
2.3	31	10
2.75	26	14
2.9	12	13
3.26	28	12
3.44	20	16
1.8	7	18
2.93	0	15
2.8	46	5
3.25	30	6.5
3.13	25	11

GPA	WORK	UNITS
2.78	22	14
2.55	24	8
3.6	15	16
3	10	10
3.1	19	11
3.5	40	8
2.85	35	9
2.4	55	5
2.8	4	13
3.2	26	10
3.47	32	12
3.33	9	15
2.5	18	16
3.4	16	19
3.95	0	21
2.88	29	10

1) Use a 2 X 2 contingency table to test for independence between GPA and WORK.

2) Use a 2 X 2 contingency table to test for independence between UNITS and WORK.

3) What does WORK have the biggest effect on?

Below are the grades for the third exam in a Trigonometry class I taught.

<u>Male</u>	<u>Female</u>
66	77
54	72
94	74
32	69
96	89
69	87
72	79
100	57
80	69
57	85
84	98
63	66
87	89
75	93
27	99
64	82
53	
57	
61	
78	

1) Use ANOVA to test the claim of equal means for men and women.

2) Use the t-test to test the claim of equal means.

3) Use the contingency table below to test the claim of independence between GRADE and GENDER.

		GRADE	
		<70	≥ 70
GENDER	Male		
	Female		

4) Explain why different tests should be used to reach a conclusion about one set of data.

1) Data is provided for 3 professional basketball teams. The scores listed are the total number of points scored by each team's top 10 players last season.

a) Use ANOVA to test the claim of equal means for the three teams.

b) Use the Kruskal-Wallis test to see if the three teams do equally well.

c) How do the results of (a) and (b) compare? Which of the two tests should be used in this situation?

Miami	Utah	LA
293	382	432
152	303	394
137	198	158
116	143	118
104	95	109
92	94	98
70	89	79
68	86	74
67	73	56
65	69	43

Church	Volunteer	
0	0	
0	1	
0	0	
0	5	
0	0	
0	1	
1	0	
1	0	
1	0	
1	1	
2	3	
2	2	
3	0	
4	2	
4	2	
4	2	
8	5	
8	8	
9	3	
10	3	

4) The paired data listed here is the number of times a person goes to church per month and the number of hours of volunteer work they do per month.

a) Use Rank Correlation to test for a correlation . b) Test for a liner correlation between the variables.

c) How do the results of (a) and (b) compare? Which of the two tests should be used in this situation?