Stat300: L.C.Larsen
Exam #1 Name: Solution
Spring 2013
Tue/Thu 7:00 - 9:05 p.m.

1. Use the data in the following table to answer parts (a) through (e).

Education Level	Employment Type for People in Sample				Total
	Farm	Factory	Health	Financial	
No Schooling	256	267	212	292	1027
Elementary	226	245	288	294	1053
Secondary	267	228	257	291	1043
College	266	208	219	250	943
Graduate	207	293	263	233	996
Total	1222	1241	1239	1360	5062

(1 point, 1 minute)
(a) What is the probability that a randomly selected person from this sample will have college-level education?

$\frac{943}{5062} = 0.186$

(3 points, 3 minutes)
(b) What is the probability that a person randomly selected from this sample will have no schooling given that the person works in the "Farm" sector of the economy?

$P(\text{No sch}|\text{Farm}) = \frac{P(\text{No school \& Farm})}{P(\text{Farm})} = \frac{(256/5062)}{(1222/5062)} = \frac{256}{1222} = 0.209$

(3 points, 3 minutes)
(c) What is the probability that a person randomly selected from this sample will have no schooling given that the person works in the "Financial" sector of the economy?

$P(\text{no school}|\text{Finance}) = \frac{292}{1360} = \frac{\text{No school \& Financial}}{\text{Financial}} = 0.215$

(3 points, 3 minutes)
(d) Based on your answers to parts (b) and (c), say whether "no schooling" is independent or dependent on the sector of the economy, and explain why.

Dependent because $P(\text{No school}|\text{Farm}) \neq P(\text{no school}|\text{Finance})$

(3 points, 3 minutes)
(e) What is the probability that a person randomly selected from this sample will have "Secondary" education or work in the "Health" sector of the economy?

$P(\text{secondary OR Health}) = P(\text{Secondary}) + P(\text{Health}) - P(\text{Both})$

$= \frac{1043}{5062} + \frac{1239}{5062} - \frac{257}{5062} = \frac{2025}{5062}$

Stat300: L.C.Larsen
Exam #1

Name: Solution

Summer 2013
Mon/Tue/Wed/Thu 5:30 - 8:15 p.m.

(5 points; 5 minutes)

2. The scores of two different tests have bell-shaped distributions. The scores for Test A have a mean of 110 points and a standard deviation of 12 points. The scores for Test B have a mean of 330 and a standard deviation of 60 points. Use this information to answer the two questions below.

What percent of the scores for Test A should be found between 86 and 134? Show why!

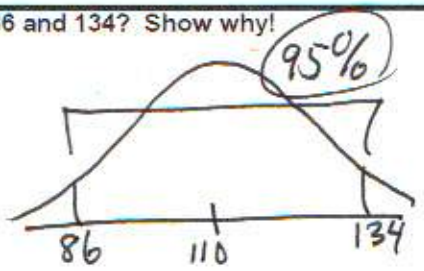
$\mu_A = 110$

$\sigma_A = 12$

$134 - 110 = 24 = 2(\sigma)$

$86 - 110 = -24 = -2\sigma$

Empirical Rule says ~95% of the values should be within $2*\sigma$ of μ.



Would a score of 220 on Test B be considered unusual? Show why!

$X = 220$

$\mu_B = 330$

$\sigma_B = 60$

$Z = \frac{220 - 330}{60} = \frac{-110}{60} = -1.83$

$|Z| < 2$ so 220 is Not unusual.

(10 points; 8 minutes)

3. Complete the columns in the "Frequency Distribution" table using the data values given below, and answer the two questions beoow the table.

Frequency Distribution

Class Limits Lower	Class Limits Upper	Tally	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
40	80	III	3	3/9	3	3/9
90	130	IIII	4	4/9	7	7/9
140	180	II	2	2/9	9	9/9 = 1

Class Midpoint	Class Boundary
60	
	85
110	
	135
160	

N = 9

Data: ~~75.9~~ 61.0 ~~91.4~~ ~~78.7~~ ~~145.7~~ ~~86.8~~ 101.1 ~~134.2~~ ~~150.1~~

Class Width
50

= 90 - 40

OR

= 135 - 85

What is the frequency for Class #2 ? 4

What is the upper limit for Class #1 ? 80

4. Give a short definition of statistics (1 point; 1 minute):

The art and science of making sense out of data.

(5 points; 5 minutes)

5. For each situation below, select the appropriate statistical term from the list provided and write it in the blank next to the description or situation. Choose the term that is best connected to the underlined text in the description or situation.

Terms:	1. randomization	4. blinding	7. experimental unit
	2. replication	5. placebo / control	8. treatment
	3. confounding	6. block	

Blinding (4)

Educators tested the effect of three different breakfasts on students diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). A stratified random sample of 2400 such kids included 300 boys and 300 girls in each of four age groups. Critics say race and culture should have been included in the structure of the stratified sampling. In each stratum, each child was assigned a normal, high fiber, or high protein breakfast with equal and independent probability, and the kids did not know which one.

(7)
Experimental Unit

Educators tested the effect of three different breakfasts on students diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). A stratified random sample of 2400 such kids included 300 boys and 300 girls in each of four age groups. Critics say race and culture should have been included in the structure of the stratified sampling. In each stratum, each child was assigned a normal, high fiber, or high protein breakfast with equal and independent probability, and the kids did not know which one.

(3)
Confounding

Educators tested the effect of three different breakfasts on students diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). A stratified random sample of 2400 such kids included 300 boys and 300 girls in each of four age groups. Critics say race and culture should have been included in the structure of the stratified sampling. In each stratum, each child was assigned a normal, high fiber, or high protein breakfast with equal and independent probability, and the kids did not know which one.

(8)
Treatment.
~~Experimental Unit~~

Educators tested the effect of three different breakfasts on students diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). A stratified random sample of 2400 such kids included 300 boys and 300 girls in each of four age groups. Critics say race and culture should have been included in the structure of the stratified sampling. In each stratum, each child was assigned a normal, high fiber, or high protein breakfast with equal and independent probability, and the kids did not know which one.

(1)
Randomization

Educators tested the effect of three different breakfasts on students diagnosed with Attention Deficit Hyperactivity Disorder (ADHD). A stratified random sample of 2400 such kids included 300 boys and 300 girls in each of four age groups. Critics say race and culture should have been included in the structure of the stratified sampling. In each stratum, each child was assigned a normal, high fiber, or high protein breakfast with equal and independent probability, and the kids did not know which one.

Stat300: L.C.Larsen
Exam #1

Name: Solution

Summer 2013
Mon/Tue/Wed/Thu 5:30 - 8:15 p.m.

(4 points; 4 minutes)

6. A subway train has six safety systems that will stop the train in an emergency. Each system is independent of the others and each system will work with probability 0.6 each time an emergency happens. What is the probability that exactly 5 of the systems will work the next time an emergency happens?

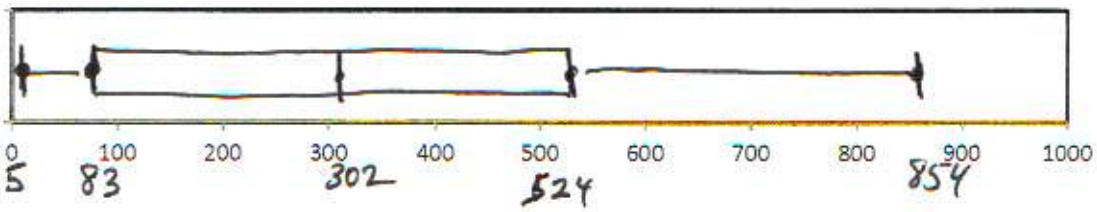
The 6 safety systems are 6 trials. They are independent. Each trial will work ("success") or not ("fail"). X is # of successes. "Binomial" $P(S) = 0.6 = p$ $P(F) = 0.4 = q$

$$P(5) = {}_nC_x (p^x)(q^{n-x}) = {}_6C_5 (0.6)^5 (0.4)^1$$

$$= 0.1866$$

(5 points and 4 points; 8 minutes)

7. (a) For the set of 77 values shown below in sorted order, prepare a Boxplot inside the rectangle that is above the number line.



5	5	19	23	23	27	27	27	32	42
47	47	47	52	58	58	70	70	70	83 = P_{25}
83	96	110	118	118	118	140	148	148	156
190	207	207	253	263	282	292	302 P_{50}	302	364
364	408	419	430	430	442	442	465	465	465
476	488	488	512	512	512	524 P_{75}	524	536	548
561	561	561	586	611	637	650	650	689	702
716	716	743	770	770	840	854			

Min = 5 Max = 854

P_{25} @ $L = \left(\frac{25}{100}\right)77 = 19.25 \uparrow 20$

P_{50} @ $L = \left(\frac{50}{100}\right)77 = 38.5 \uparrow 39$

P_{75} @ $L = \left(\frac{75}{100}\right)77 = 57.8 \uparrow 58$

(b) For the set of 77 values shown above in sorted order, what percentile is represented by the value 70?

$X = 70 = P_{(k)} = ?$

$$k = \left(\frac{\#\text{ of values} < 70}{\text{Total \# of values}}\right) * 100$$

$$= \left(\frac{16}{77}\right) 100 = 0.208$$

$70 = P_{21}$

(OR) $P_{20.8}$

(OR) $P_{20.78}$

(8 points; 10 minutes)

8. Answer parts (a), (b), and (c).
Use the columns in the table in any way you wish to use them.

(a) Is this distribution "proper" (circle "YES" or "NO")? YES NO

Why? $\sum P(x) = 1$

x	P(x)	$x \cdot P(x)$	$(x-\mu)^2 \cdot P(x)$
146	0.26	37.96	331.4
163	0.16	26.08	56.0
183	0.22	40.26	0.4
215	0.36	77.40	399.2

$\sum P(x) = 1.00$ $\quad \sum = 181.7 = \mu$ $\quad \sum = 787.0 = \sigma^2$

(b) Write the formulas for the mean, the variance, and the standard deviation of a discrete probability distribution.

$\mu =$ $\sum x \cdot P(x)$

$\sigma^2 =$ $\sum (x-\mu)^2 \cdot P(x)$

$\sigma =$ $\sqrt{\sigma^2} = \sqrt{\sum (x-\mu)^2 \cdot P(x)}$

(c) Write the values for the mean, the variance, and the standard deviation of this discrete probability distribution.

$\mu =$ 181.7

$\sigma^2 =$ 787.0

$\sigma =$ 28.1 $= \sqrt{787.0}$

(4 points and 2 points; 5 minutes)

9. A quiz has two "true" (T) or "false" (F) questions and one "Pass" or "Fail" essay question. What is the sample space of possible answers to the set of three questions?

(T T Pass) (F T Pass)
(T T Fail) (F T Fail)
(T F Pass) (F F Pass)
(T F Fail) (F F Fail)

False for 1st and Fail for essay

If an outcome is selected at random (all equally likely) from your sample space, what is the probability that your selection is "False" for the first problem and "Fail" for the essay?

$Prop = \frac{2}{8} = 0.25$

(4 points; 4 minutes)

10. A lock uses a touch pad with the letters A B and C and the digits 1 2 3 4 5 6 7 and 8. The unlock code begins with 2 of the letters and ends with 4 of the digits. Each letter and each digit can be used each time a letter or a digit is needed. Every possible sequence of letters and digits is a different unlock code. How many different unlock codes are possible?

L L D D D D

$3 \cdot 3 \cdot 8 \cdot 8 \cdot 8 \cdot 8$ Fundamental Counting Rule

= 36,864 possible codes

(3 points; 3 minutes)

11. You are one of twenty people that work in a company. Four of the 20 workers will be sent to New York for special training. How many sets of four workers are possible to form from the twenty workers, and what is the probability that you will not get to go to New York for the training?

$_{20}C_4$ = 4845 possible sets of 4 people

without you there are $_{19}C_4 = 3876$ possible sets

P(you do not get to go) = $\frac{3876}{4845} = 0.80$

(4 points; 6 minutes)

12. The Medicare administration wants to evaluate the standard deviation of the costs for a specific medical test this year to see if the standard deviation is acceptable under the Medicare rules. Because it is believed that variability is the same throughout each year, the Medicare staff uses the charges for the 229 times the medical test has been done so far this year and calculates the standard deviation of the 229 values. The result is that the calculated standard deviation is not acceptable.

Use the information in the "story" to answer the following:

(a) What is the population of interest to the Medicare administration?

All of the costs this year for a specific medical test

(b) What parameter was important for the Medicare administration to know?

The standard deviation of all the costs this year for a specific medical test

(c) What statistic did the Medicare staff use instead?

The standard deviation of the ~~229~~ costs for 229 of the medical tests done so far this year.

Stat300: L.C.Larsen
Exam #1

Name: Solution

Summer 2013
Mon/Tue/Wed/Thu 5:30 - 8:15 p.m.

(14 points; 8 minutes)

12. **Use the data below to determine the value of each statistic. Write an expression for each statistic or describe how it is calculated in principle** (do NOT describe how to use the calculator to determine the result).

Data		Expression or Description	Value of statistic
40 60 41 56 17 59 46 24 50	median	the value in the middle when data are in sorted order	46
	mode	the value that occurs most often (most frequent value)	No mode all values are unique
17 24 40 41 46 median 50 56 59 60	variance	$\frac{\sum (x-\bar{x})^2}{n-1} = s^2$	227.3
	range	max − min (largest − smallest)	60 − 17 = 43
	standard deviation	$s = \sqrt{s^2} = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$	15.1
	midrange	$\frac{\min + \max}{2}$	(17+60)/2 = 38.5
	mean	$\frac{\sum x}{n}$	43.7

Stat300: L.C.Larsen | Summer 2013
Exam #1 | Name: | Mon/Tue/Wed/Thu 5:30 - 8:15 p.m.

3 X 3

(5 points; 5 minutes)

13. Last year, the dropout rate for High School students was 35%. Using the "relative frequency" approach to probability, what is the probability that a random selection of five high school students at the beginning of this year will end up in the exact sequence shown below? (Show your work.)

$P(\text{Dropout}) = 0.35 \quad P(\text{Not Dropout}) = 0.65$

Student	Dropout
1	Yes
2	Yes
3	No
4	No
5	Yes

$P(\text{yes, yes, No, No, yes}) = (.35)(.35)(.65)(.65)(.35)$

$= 0.018$

(5 points; 6 minutes)

14. Last year, the dropout rate for High School students was 35%. Using the "relative frequency" approach to probability and assuming that decisions to drop out of school are independent, what are the mean and standard deviation for the number of students that will drop out this year at a high school with 3500 students?

Binomial

$n = 3500$

$p = 0.35 \quad q = 0.65$

$\mu = np = (3500)(0.35) = 1225$

$\sigma = \sqrt{npq} = \sqrt{3500(0.35)(0.65)} = 28.2$

(5 points; 5 minutes)

15. A gambler knows that playing "black" in a particular Roulette game will win with probability 0.475. The gambler bets \$10 on "black" for each spin of the wheel. The gambler will get \$10 if he wins and will lose \$10 if he loses. What is the expected value of the next two bets together?

Expected value = -\$1.00

Why? each bet has Expected Value of -\$0.50 so 2 bets would be double that.

X	P(X)	X · P(X)
-10	0.525	-5.25
+10	0.475	4.75
		-0.50
		X 2
		-1.00

(6 points; 6 minutes)

16. Circle the best answer for each situation.

An MTV program shows music videos and asks all viewers to call a free 800 number to rate each video on a scale of 1 to 10.
self selection

Simple Random	Systemmatic
Stratified Random	Cluster
Convenience	Census

A cable TV company rates the popularity of TV shows among its customers by constantly tracking the channel to which each of its cable connect boxes is tuned.
comprehensive sample

Simple Random	Systemmatic
Stratified Random	Cluster
Convenience	**Census**

A research company rates the popularity of TV shows by taking random samples of all Americans in each of 5 income groups within each of 5 age groups.

Simple Random	Systemmatic
Stratified Random	Cluster
Convenience	Census

The Department of Corrections (Prisons) selects a group of 5000 prisoners released in 2001 and studies key characteristics of their past lives to find out what types of decisions decrease the percent that return to prison at a later time.

retrospective	**observational study**
cross-sectional	experiment
prospective	

The Department of Corrections releases a group of 400 prisoners who share many key characteristics (race, education, family history, type of crime, etc.). Half of the 400 go into the army and the others do not, so the effect of military service can be studied.

retrospective	observational study
cross-sectional	**experiment**
prospective	

The Department of Corrections randomly selects 5000 prisoners in 2013 and conducts a detailed health exam on each one in order to study the present state of health in the prison population at that time.

retrospective	**observational study**
cross-sectional	experiment
prospective	