Statistics 300
Instructor: Lawrence C. Larsen

Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(5 points; 6 minutes)

1. A small ferry boat carries people and cars across a river. The boat can carry 10 people but only 2 cars. Four cars are waiting to cross the river -- 1 is Red, 1 is Green, 1 is Blue, and 1 is White. The car owners all claim to have arrived at the same time, so the ferry boat operator decides to pick one at random to get on the boat first and another at random to get on the boat second.

(a) List the sample space for the boat operators' procedure (e.g. G,W)

RG GR BR WR
RB GB BG WG
RW GW BW WB

(b) What is the probability that the Red car will be picked to get on the ferry boat.
(Note: All the elements in the sample space are equally likely.)

6 of the 12 items in the sample space include the Red car, so

$$P(\text{red gets on the Ferry Boat}) = \frac{6}{12}$$

(7 points; 7 minutes)

2. Given: X ~ Binomial(n = 2000, p = 0.72) and Y ~ Binomial(n = 800, p = 0.44)
Which would be more unusual, X = 1392 or Y = 378 ?

For unusual, the Z-score is the tool presently available. $Z = \frac{X-\mu}{\sigma}$, so we need the mean (μ) and the standard deviation (σ) for each of the Binomial Distributions.

$x = 1392$

$\mu_x = np = 2000(0.72) = 1440$

$\sigma_x = \sqrt{npq} \approx \sqrt{2000(.72)(.28)} = 20.08$

$Z_x = \frac{1392 - 1440}{20.08} = \boxed{-2.39}$

$y = 378$

$\mu_y = np = 800(0.44) = 352$

$\sigma_y = \sqrt{npq} = \sqrt{800(.44)(.56)} = 14.04$

$Z_y = \frac{378 - 352}{14.04} = \boxed{1.852}$

x more unusual than y

Statistics 300 | Name: Solution | Summer 2012
Instructor: Lawrence C. Larsen | M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(12 points; 10 minutes)

3. **Use the dataset at the bottom of this page to answer parts (a), (b) and (c). There are 180 values in the dataset, in rows of 10, sorted from the smallest at the top left to the largest at the bottom right.**

(a) What percentile is represented by the value 1087 ?

$1087 = P_{K} \rightarrow K = \left[\frac{\#\text{ of values} < 1087}{\#\text{ of all values}}\right] \times 100$

$= \left(\frac{40}{180}\right) 100 = 22.2 \qquad 1087 = P_{22.2}$

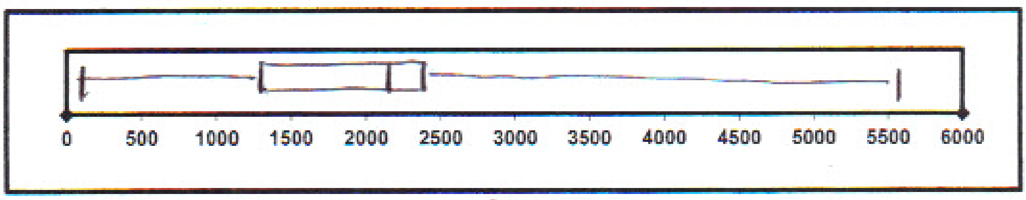
(b) What is the value of the 75th percentile, P_{75} ?

$k = 75 \quad \text{Location} = L = \left(\frac{k}{100}\right) N = \left(\frac{75}{100}\right) 180 = \boxed{135}$ (whole #, so...)

value at $L = \cancel{136}\ 135$ is 2348

$L = 136$ is 2354 → average = 2351 = P_{75}

(c) Using the number line below, make a Boxplot to represent the distribution of the dataset.



$Q_1 = P_{25}$ & $L = (25/100)\,180 = 45 \qquad (1278 + 1335)/2 = 1306.5$

$Q_2 = P_{50}$ & $L = (50/100)\,180 = 90 \qquad (2125 + 2128)/2 = 2126.5$

$Q_3 = P_{75}$: $L = (25/100)\,180 = 135 \qquad (2348 + 2354)/2 = 2351$

min

100	107	144	149	170	193	200	226	263	294
322	340	344	363	372	385	402	440	475	514
523	545	584	599	627	657	669	697	715	740
752	770	778	830	863	963	988	1015	1042	1070
1087	1087	1087	1223	1278	1335	1365	1430	1484	1521
1551	1564	1603	1613	1657	1727	1791	1798	1877	1904
1934	1948	1948	1954	1962	1966	1970	1980	1985	1989
1994	2004	2012	2020	2029	2032	2041	2047	2057	2063
2063	2068	2070	2080	2090	2098	2102	2112	2121	2125
2128	2131	2136	2137	2140	2142	2146	2155	2161	2161
2165	2174	2183	2193	2195	2204	2210	2216	2219	2222
2228	2229	2229	2231	2237	2243	2251	2255	2264	2266
2272	2279	2286	2289	2293	2301	2310	2315	2318	2324
2329	2337	2342	2345	2348	2354	2357	2360	2366	2368
2377	2382	2385	2391	2399	2404	2407	2471	2599	2769
2785	2895	2945	2969	3156	3296	3351	3531	3655	3819
3850	3852	3930	4026	4051	4231	4371	4509	4665	4795
4814	4836	4961	5023	5071	5257	5305	5416	5531	5593

130

max

Statistics 300
Instructor: Lawrence C. Larsen

Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

4

(8 points; 10 minutes)

4. **For each of the underlined segments in the situations below, select the appropriate 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	5. placebo	9. parameter
	2. replication	6. block	10. statistic
	3. confounding	7. experimental unit	11. population
	4. blinding	8. treatment	12. (no term)

(a.) An experiment is done to estimate the average of **the responses of all autistic children to large doses of vitamins**. The study involved 600 autistic children in each of 5 age groups. In each age group, 200 children were given a pill with no vitamins, 200 a pill with the standard dose, and 200 a pill with a large dose. Children stayed in their family homes, and each family believed their child was receiving the "large dose". Cameras in each home recorded the behavior of each child which was scored on a "20 point scale" for "severity of autism". Conclusions were based on the difference between the average score of the large dose group and the average score of the standard dose group. The study could not control for the possible effects of unique factors in each household that may also affect autism.

population (11)

(b.) An experiment is done to estimate the average of the responses of all autistic children to large doses of vitamins. The study involved 600 autistic children in each of 5 age groups. In each age group, 200 children were given a pill with no vitamins, 200 a pill with the standard dose, and 200 a pill with a large dose. Children stayed in their family homes, and each family believed their child was receiving the "large dose". Cameras in each home recorded the behavior of each child which was scored on a "20 point scale" for "severity of autism". Conclusions were based on **the difference between the average score of the large dose group and the average score of the standard dose group**. The study could not control for the possible effects of unique factors in each household that may also affect autism.

statistic (10)

(c.) An experiment is done to estimate the average of the responses of all autistic children to large doses of vitamins. The study involved 600 autistic children in each of **5 age groups**. In each age group, 200 children were given a pill with no vitamins, 200 a pill with the standard dose, and 200 a pill with a large dose. Children stayed in their family homes, and each family believed their child was receiving the "large dose". Cameras in each home recorded the behavior of each child which was scored on a "20 point scale" for "severity of autism". Conclusions were based on the difference between the average score of the large dose group and the average score of the standard dose group. The study could not control for the possible effects of unique factors in each household that may also affect autism.

blocks (6)

(d.) An experiment is done to estimate the average of the responses of all autistic children to large doses of vitamins. The study involved 600 autistic children in each of 5 age groups. In each age group, 200 children were given a pill with **no vitamins**, 200 a pill with **the standard dose**, and 200 a pill with **a large dose**. Children stayed in their family homes, and each family believed their child was receiving the "large dose". Cameras in each home recorded the behavior of each child which was scored on a "20 point scale" for "severity of autism". Conclusions were based on the difference between the average score of the large dose group and the average score of the standard dose group. The study could not control for the possible effects of unique factors in each household that may also affect autism.

treatments (8)

Statistics 300 Instructor: Lawrence C. Larsen	Name: Solution	Summer 2012 M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(8 points; 5 minutes)

6. **For each of the discrete probability distributions below, calculate the mean, variance, and standard deviation.**

(a)

X	P(X)	$x \cdot P(x)$	$(x-\mu)^2 \cdot P(x)$	
7	0.23	1.61	22.04	
19	0.66	12.54	3.22	
24	0.11	2.64	5.72	
		$\Sigma = \mu = 16.79$	$\Sigma = \sigma^2 = 30.98$	$\sigma = \sqrt{\sigma^2} = \sqrt{30.98} = 5.57$

$\Sigma P(x) = 1$

"proper distribution"

(b)

X	P(X)			
0	0.136			
1	0.279			
2	0.423			

$\Sigma P(x) = 0.838$

"improper" distribution

(9 points; 6 minutes)

7. **Use the data below to complete the frequency distribution .**

Boundaries	Class Limits Lower	Class Limits Upper	Tally	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
	10	20	||||	4	4/14	4	4/14
25	30	40	|||	3	3/14	7	7/14
45	50	60	||||	4	4/14	11	11/14
65	70	80	|||	3	3/14	14	1 = 14/14

N = 14

Data:						
	~~67~~ ~~42~~ ~~20~~	~~24~~ ~~57~~ ~~59~~	~~33~~ ~~46~~	~~12~~ ~~56~~	~~68~~ ~~24~~	~~38~~ ~~70~~

20 = the class width.

25 = boundary between class 1 and class 2.

30 = the lower class limit for class #2.

4 = the frequency of class #3.

Statistics 300
Instructor: Lawrence C. Larsen

Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(4 points; 6 minutes)

8. The American Association of Realtors wants to estimate the percentage of all single-family homes in the USA that have more than two bathrooms. For this purpose, the Association takes a random sample of 8000 addresses for single-family homes and determines the number of bathrooms in each of these homes. There were 976 homes with more than two bathrooms, so the estimate for all single family homes in the USA is 12.2%.

(a) What is the population of interest in this situation?

the Number of bathrooms in each of the single-family homes in the USA

(b) What is the parameter of interest in this situation?

the % of the single-family homes in the USA that have more than 2 bathrooms.

(c) What statistic was used in this situation?

the % of 8000 randomly selected homes ~~from~~ in the USA that have more than 2 bathrooms.

(d) Was a sample or a census used for this study, and why did you choose your answer?

Sample of 8000 homes is stated, and that number is much smaller than the population size.

Statistics 300 Instructor: Lawrence C. Larsen	Name: Solution	Summer 2012 M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(14 points; 10 minutes)

9. **For the sample of data given below, provide the formula (expression) or description of calculation** (not how to use the calculator) **for each statistic listed and also provide the value of each statistic. The mean and standard deviation must be calculated using your calculator's "statistics mode".**

Data:

~~16~~
~~20~~
~~16~~
~~13~~
~~17~~
~~19~~
~~10~~
~~16~~
~~18~~
~~17~~
~~10~~
~~17~~

Statistic	Formula / Description	Value
Standard deviation	$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = s$	3.194
Mode	The most frequently occurring value	16 and 17 3x each
Mean	$\Sigma x / n$	15.75
Median	Value in the center of data when in sorted order	$(16+17)/2 =$ 16.5
Range	max − min	$20 - 10 =$ 10
Mid-range	$(\max + \min)/2$	$(10+20)/2$ $= 15$
Variance	$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$	10.20

even number, so use avg. of 2 in the middle

10
10
13
16
mode 16
16 → median
17
mode 17
17
18
19
20

Statistics 300
Instructor: Lawrence C. Larsen

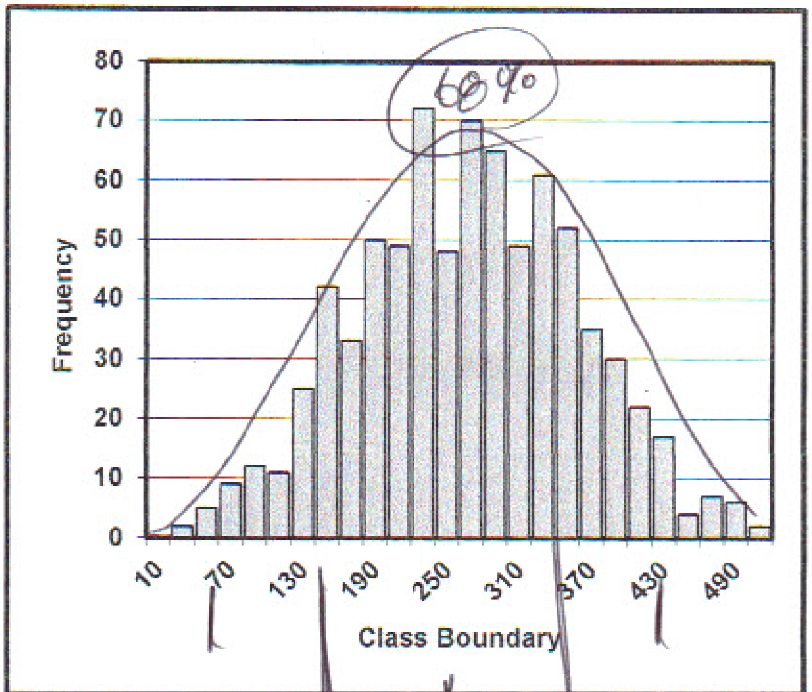
Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(6 points; 6 minutes)

9. A sample of 2500 data values were collected in a random sample. A graph of the distribution shows it is bell-shaped. Of the 2500 data values, 68% are between 160 and 340. Use these facts to estimate the percent of the data that are between 70 and 430.



70 160 σ σ 340 430

90 90

σ σ

μ or $\bar{x}$

180 $\cong 2\sigma$

because empirical rule says 68% of values (approximately) are contained between $(\mu-\sigma)$ and $(\mu+\sigma)$.

So, $\sigma \cong 90$

70 is 2σ below μ
430 is 2σ above μ
so
95% of the data should be contained between 70 and 430.

Statistics 300
Instructor: Lawrence C. Larsen

Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(3 points; 3 minutes)

10. A final exam in statistics must have 10 out of 16 possible problems. If the professor decides to choose the 10 problems at random and arrange them in a random order, how many different ways could the test turn out?

the order of the problems makes a difference, so

$_{16}P_{10} = \boxed{2.91 \times 10^{10}}$ different ways for the exam to turn out.

(3 points; 3 minutes)

11. A statistics exam will have 2 different versions so students will be discouraged from trying to cheat. If the class has 40 students and half will be assigned to each of the two versions, how many different ways could the professor divide the class into two groups?

Each group will have 20 students, so

$_{40}C_{20} = \boxed{1.38 \times 10^{11}}$ different ways to divide the class of 40 into 2 groups of 20 each.

(5 points; 4 minutes)

12. A different statistics professor likes multiple choice problems. That professor gives an exam that has 10 problems with 4 possible answers in each one.

(a) If a student decides to use the calculator's random number function to guess on each problem, what is the probability that the student will guess the correct answer on exactly 4 of the 10 problems? ON each problem, the student's guess is correct or incorrect. Random choices ⇒ independent. There are 10 problems ⇒ 10 trials.

Binomial

$p = 0.25$ correct

$q = 0.75$ wrong

$P(\text{exactly 4 out of 10}) = {}_{10}C_4 (0.25)^4 (0.75)^6 = \boxed{0.146}$

(b) If a student decides to use the calculator's random number function to guess on each problem, what is the probability that the student will guess the correct answer on at least one of the 10 problems?

5 pts.

$p(\text{correct}) = 0.25$

$q(\text{wrong}) = 0.75$

$P(\text{at least one correct}) = 1 - P(\text{all wrong}) = 1 - P(0)$

$n = 10$

Binomial

$= 1 - {}_{10}C_0 (0.25)^0 (0.75)^{10}$

$= 1 - 0.0563 = 0.9437$

other solution $P(\text{at least one correct}) = 1 - P(\text{all wrong})$

$= 1 - (0.75)(0.75) \cdots (0.75) = 1 - (0.75)^{10} = 1 - 0.0563$

$= 0.9437$

Statistics 300
Instructor: Lawrence C. Larsen

Name: ____________________

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(3 points; 4 minutes)

14. For the study described below, select the appropriate statistical terms from the list provided and write them in the blanks, choose the term that is best connected to the <u>underlined text</u>.

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

Write the selected term

a. ____________________ A total of 60 children were included in a study of a new medication. The study used <u>30 similar girls and 30 similar boys</u> that already used the standard medication every day. In the study, 10 boys and 10 girls were given a "medication" that had no effect at all, 10 boys and 10 girls were given the standard medication, and 10 boys and 10 girls were given the new medication. So, each child received one of the three types of medication. Before the study began, each child was equally likely to be assigned to each one of the medications. To prevent "bias", neither the children nor the experimenters knew which medication each child was being given.

b. ____________________ A total of 60 children were included in a study of a new medication. There were 30 girls and 30 boys in the study, who were already using the standard medication every day. In the study, <u>10 boys and 10 girls were given a "medication" that had no effect at all, 10 boys and 10 girls were given the standard medication, and 10 boys and 10 girls were given the new medication</u>. So, each child received one of the three types of medication. Before the study began, each child was equally likely to be assigned to each one of the medications. To prevent "bias", neither the children nor the experimenters knew which medication each child was being given.

c. ____________________ A total of 60 children were included in a study of a new medication. There were 30 girls and 30 boys in the study, who were already using the standard medication every day. In the study, 10 boys and 10 girls were given a "medication" that had no effect at all, 10 boys and 10 girls were given the standard medication, and 10 boys and 10 girls were given the new medication. So, <u>each child</u> received one of the three types of medication. Before the study began, each child was equally likely to be assigned to each one of the medications. To prevent "bias", neither the children nor the experimenters knew which medication each child was being given.

Statistics 300
Instructor: Lawrence C. Larsen

Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

6 ~~(8~~ points; 3 minutes)

15. Circle the correct choice in each box in relation to the underlined text.

	Are the data ... ?	Are the data ... ?
a. The ~~total gallons~~ pounds of jet fuel of all the gasoline used by Americans ~~to drive~~ who fly to work today.	Qualitative Quantitative and Discrete Quantitative and continuous	Nominal Interval Ordinal Ratio
b. The number of ~~"subcompact", "compact", "mid-size", and "standard" cars used by Americans to drive to work today.~~ classes of eggs: small, med, large, X-large, jumbo.	Qualitative Quantitative and Discrete Quantitative and continuous	Nominal Interval Ordinal Ratio
c. The total profit of all ~~the gasoline companies that sell gasoline to Americans who drive to work today.~~ # of Americans ~~that~~ that will use gasoline to commute to work today.	Qualitative Quantitative and Discrete Quantitative and continuous	Nominal Interval Ordinal Ratio

(3 points; 3 minutes)

16. A standard California license plate for a car has 4 numbers (digits) and 3 letters in the format "DLLLDDD". Each D can be a digit from 0 through 9 and each L can be any one of the 26 letters in our alphabet. How many different standard license plates are possible? if all the 3 letters must be the same.

$$\underset{D}{10}\ \underset{L\,L\,L}{26}\ \underset{D}{10}\ \underset{D}{10}\ \underset{D}{10} = 260{,}000$$

(3 points; 3 minutes)

17. A bowl contains 20 jelly beans. Five are "Cherry", 8 are "Orange", 2 are "Lemon" and 5 are "Grape". What is the probability ~~of getting the sequence "O,O,G"~~ that none of the lemon jelly beans will be used when if 3 jelly beans are taken out of the bowl (and not put back in between picks)?

18 are not lemon.
20 jelly beans in all.

P(1st not lemon and 2nd not lemon and 3rd not lemon) = ~~P(1st not lemon) P(not lemon | 1st not lemon)~~

$$= \left(\frac{18}{20}\right)\left(\frac{17}{19}\right)\left(\frac{16}{18}\right) = \cancel{0.805} = 0.7158$$

$$= \left({}_{18}C_3\right) / \left({}_{20}C_3\right) = 0.7158$$

Statistics 300
Instructor: Lawrence C. Larsen

Name: Solution

Summer 2012
M/Tu/W/Th 8:00 - 10:45 a.m.

Exam #1

(8 points; 8 minutes)

18. Use the information in the table to answer the probability questions (a) - (c).

Number of Years at Current Job	Teacher at a ... Element. School	Junior High	High School	College	Total
0 to 5 years	72	102	116	74	364
6 to 10 years	129	92	106	101	428
> 10 years	123	114	118	92	447
Total	324	308	340	267	1239

(2 pts)

(a) What is the probability that someone picked at random from the individuals in this table will be a teacher who has been at their current job for > 10 years?

$$\left(\frac{447}{1239}\right) = 0.3608$$

(3 pts)

(b) What is the probability that someone picked at random from the individuals in this table will will be a High School Teacher given that the person has been at their current job for 6 to10 years?

Use only the 6 to 10 years row: $\frac{106}{428} = 0.2477$

OR

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{106/1239}{428/1239} = \frac{106}{\cancel{1239}} \cdot \frac{\cancel{1239}}{428} = \frac{106}{428} = 0.2477$$

(3 pts)

(c) What is the probability that someone picked at random from the individuals in this table will be someone who teaches at an Elementary School or College or has been at their current job for 6 to10 years?

overlaps

$$= \underbrace{\left(\frac{324}{1239}\right)}_{\text{elem. school}} + \underbrace{\left(\frac{267}{1239}\right)}_{\text{college}} + \underbrace{\left(\frac{428}{1239}\right)}_{\text{6 to 10 yrs.}} - \underbrace{\left(\frac{129}{1239}\right)}_{\text{elem. and 6 to 10}} - \underbrace{\left(\frac{101}{1239}\right)}_{\text{college and 6 to 10}}$$

$$= \frac{789}{1239} = 0.6368$$