

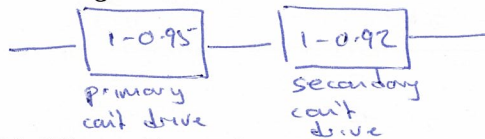
Name: _____ Section: (day/time) _____

AMS5 - MIDTERM
Thursday 30th April, 2009

A Normal Table is on the last page of this exam.
You must explain all answers and/or show working for full credit.

1. Instead of driving from Palo Alto to Santa Cruz, I now take the vanpool. The vanpool has two drivers, a primary and a backup. The primary driver is available to drive with probability 0.95, and the backup driver is available with chance 0.92.

- (a) What is the chance that, on any given day, the van will not run, due to neither driver being available?



chance that van won't run is
 $(1-0.95) \times (1-0.92)$.

- (b) What additional assumption did you make?

1 ~~the~~ chance for each driver is independent of the other.

Aside from the driver's seat, there are 11 passenger seats in the van. There are 14 people signed up to ride as passengers. The chance that any given passenger shows up on a particular day is 0.8.

- (c) What is the chance that exactly 11 passengers show up to ride the van on a particular day?

2

$$P(11 \text{ successes out of } 14) = \frac{14!}{11!(14-11)!} (0.8)^{11} (1-0.8)^{14-11}$$

- (d) What is the chance that on a given day one or more of the potential passengers who show up don't get a seat?

2

$$P(12 \text{ out of } 14) + P(13 \text{ out of } 14) + P(14 \text{ out of } 14)$$
$$= \frac{14!}{12! 2!} 0.8^{12} 0.2^2 + \frac{14!}{13! 1!} 0.8^{13} 0.2^1 + \frac{14!}{14! 0!} 0.8^{14} 0.2^0$$

- (e) One day I decide to work from home. What is the chance that exactly 11 people (not including the driver) ride the van on that day?

2

$$P(11 \text{ out of } 13) = \frac{13!}{11!(13-11)!} (0.8)^{11} (0.2)^{13-11}$$

[TURN OVER]

2. We recorded the number of chips in a batch of 161 chocolate chip cookies in class. The frequencies of different numbers of chips is summarized in the first two columns of the table below.

Class Interval (chips)	Frequency	percentage	percent-per-chip
9-12	1	0.6	0.2 (0.16)
13-16	8	5.0	1.2
17-20	39	24.2	6.1
21-24	51	31.7	7.9
25-28	44	27.3	6.8
29-32	15	9.3	2.3
33-36	3	1.9	0.5

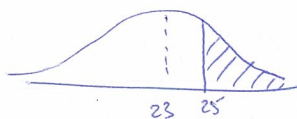
2
2
1

- (a) Complete the table.
 (b) Sketch the histogram on the graph at the top of page 3. Label the axes.
 (c) Does the histogram appear to follow the normal curve? Explain briefly.

Yes - reasonably symmetric
 peaked (unimodal) in the middle.

2

- (d) The mean number of chips was 23, and the SD was 4.38. The cookie manufacturer claims "25 chips per cookie". Using the Normal Approximation, what is the estimate of how many students had 25 or more chips in their cookie? How many students actually had 25 or more chips in their cookie?



in standard units 25 is $\frac{25-23}{4.38} \approx 0.46$

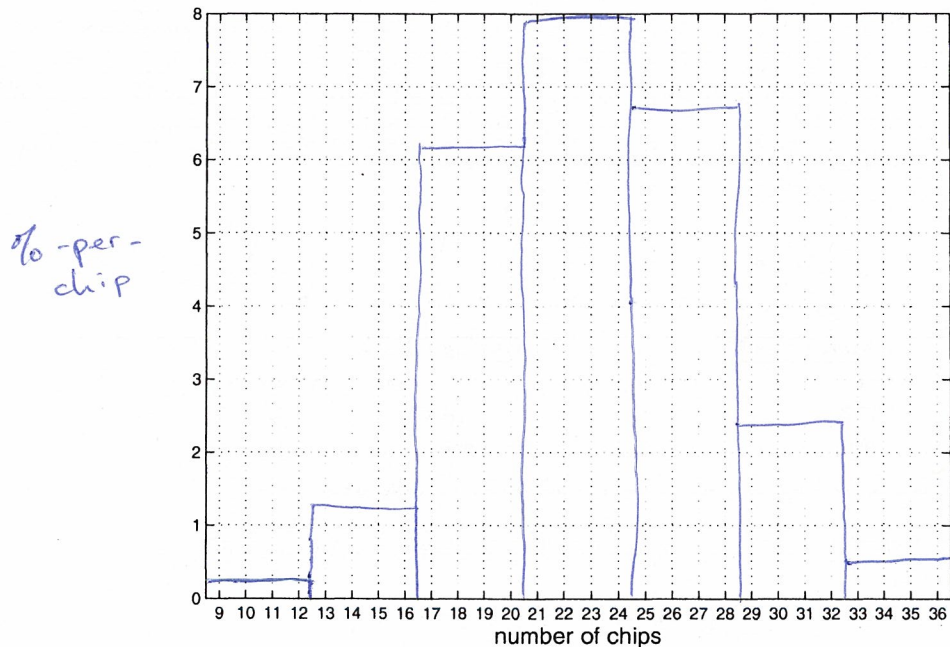
area is $\frac{1}{2}(100 - 35) = 32.5\%$

32.5% of 161 is $\approx \underline{\underline{52}}$

[CONTINUED]

2

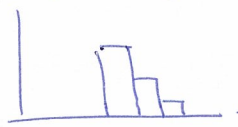
Actual number of students with ≥ 25 chips
 $= 44 + 15 + 3 = \underline{\underline{62}}$



1

(e) Considering only the cookies that had 25 or more chips, the mean was 27.6 and the SD was 2.4. Does the histogram for only these cookies follow the Normal Curve? Answer yes/no, and explain briefly.

No - it is the right end of the distribution above.



2

3. Cervical cancer is more common among women who have been exposed to the herpes virus, according to many observational studies. Is it fair to conclude that the virus causes cervical cancer? Explain your answer.

No. A potential confounding factor is sexual activity. The observational study shows association between herpes + cervical cancer.

[TURN OVER]

3

4. A (hypothetical) study is carried out to determine the effect of party affiliation on voting behavior in a certain city. The city is divided up into wards. In each ward, the percentage of registered Democrats who votes is higher than the percentage of registered Republicans who vote. True or false: for the city as a whole, the percentage of registered Democrats who vote must be higher than the percentage of registered Republicans who vote. If true, why? If false, give an example.

False - Simpson's paradox. The Democrats may be concentrated in wards with low turnouts.

	Democrats		Republicans	
	total #	# voting	total #	# voting
ward A	1000	100	100	5
ward B	100	60	1000	500

5. Which of the following are true? false? Explain or give examples.

(a) The median and the average of any list are always close together.

1

No. 1, 1, 1, 1, 100 median is 1 average is 20.8

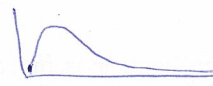
(b) Half a list is always below average.

1

No - same example 1, 1, 1, 1, 100

(c) With a large, representative sample, the histogram is bound to follow the normal curve quite closely.

1

No. incomes have long right tail 

(d) If two lists of numbers have exactly the same average of 50, and the same SD of 10, then the percentage of entries between 40 and 60 must be exactly the same for both lists.

1

False - the % within 1SD of the mean is approximately 68%

[TURN OVER]

6. Read the abstract of the paper "Beliefs About the Health Effects of "Thirdhand" Smoke and Home Smoking Bans" printed at the end of this exam paper.

1

(a) Was this a controlled experiment or an observational study? Explain briefly.

observational study. The participants self-assigned to smoker / non-smoker

1

(b) What was the investigators' objective?

Assess health beliefs of adults regarding third-hand smoke, and whether those beliefs differed between smokers and nonsmokers

2

(c) What method did the investigators use to gather their data? Give one important problem with this methodology.

Random-digit dialing

They can't call cellphones.

1

(d) The investigators say that the sample was weighted by race and gender using census data. Explain what this means.

They adjust their data to match the proportion of men/women given by the census. Ditto for race.

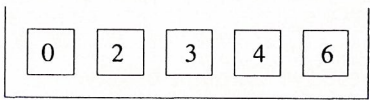
1

(e) Do the results say anything about the effect of third hand smoke on children's' health? Explain your answer briefly.

No. They say something about people's beliefs about the effect of third hand smoke.

[TURN OVER]

7. One hundred draws are going to be made at random with replacement from the box



True or false, and explain.

1

(a) The expected value for the sum of the draws is 300.

True. expected value = # draws × (average of box)
 $= 100 \times \frac{(0+2+3+4+6)}{5} = 100 \times 3 = 300.$

1

(b) The expected value for the sum of the draws is 300, give or take 20 or so.

False. There is no uncertainty in an expected value.

1

(c) The sum of the draws will be 300.

False. The sum of the draws will show fluctuations about the expected value.

1

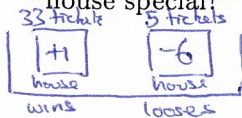
(d) The sum of the draws will be around 300, give or take 20 or so.

True. The standard error is $\sqrt{\# \text{ draws} \times \text{SD box}}$
 $= \sqrt{100} \times \sqrt{\frac{(0-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (6-3)^2}{5}} = 10 \times 2 = 20$

8. At Nevada roulette tables, the "house special" is a bet on the numbers 0, 00, 1, 2, 3. The bet pays 6 to 1 (i.e., if you bet \$1 and you win, you get your original dollar back, plus 6 more), and there are 5 chances in 38 to win.

1

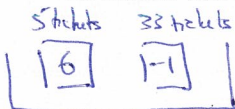
(a) For all the other bets at Nevada roulette tables, the house expects to make about 5 cents on every dollar put on the table. How much does it expect to make per dollar on the house special?



average of box = $\frac{33 \times 1 + 5 \times (-6)}{38} = 0.0789$
 or 7.89 cents for a \$1 bet.

[2 Bonus]

(b) Someone plays roulette 100 times, betting a dollar on the house special each time. Estimate the chance that this person comes out ahead.

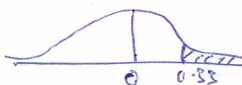


expected value = # draws × average of box
 $= 100 \times \frac{-3}{38} = -7.895$

SF = $\sqrt{\# \text{ draws}} \times \text{SD box} = \sqrt{100} \times (6 - (-1)) \sqrt{\frac{5}{38} \times \frac{33}{38}}$
 $= 23.7$ [TURN OVER]

comes out ahead \Rightarrow value > 0

0 is $\frac{7.9}{23.7}$ SD above mean ≈ 0.33 SE above mean



chance $\approx \frac{1}{2} (100 - 25) \approx 37.5\%$