27 + 4 bonus

Name:_____ Section: (day/time) _____

AMS5 - MIDTERM
Tuesday February 9th, 2016

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

5

1. To measure the effect of exercise on the risk of heart disease, investigators compared the incidence of this disease for two large groups of London Transport Authority busmen – drivers and conductors. The conductors got a lot more exercise as they walked around all day collecting fares.

   The age distributions for the two groups were very similar, and all the subjects had been on the job for 10 years or more. The incidence of heart disease was substantially lower among the conductors, and the investigators concluded that exercise prevents heart disease.

   Other investigators were sceptical. They went back and found that London Transport Authority had issued uniforms to drivers and conductors at the time of hire. A record had been kept of the sizes.

   (a) Why does it matter that the age distributions of the two groups were similar?

   (1) Because older people tend to have higher incidence of heart disease

   (b) Why does it matter that all the subjects had been on the job for 10 years or more?

   (1) To remove the effect of their starting health - after 10 years, the effect of the job should dominate.

   (c) Why did the first group of investigators compare the conductors to drivers, not to London Transport Authority executive staff?

   (1) Compare groups of similar socioeconomic status

   (d) Why might the second group of investigators have been sceptical?

   (1) It's possible that people in worse health would chose the sedentary (drivers) job rather than the active (conductor) job.

   (e) What would you do with the sizes of the uniforms?

   (1) check that the distribution of sizes was the same for the two groups.

[8]

2. Read the Summary of the article *"Bacterial ecology of hospital workers' facial hair: a cross-sectional study"* printed at the end of this exam.

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

(1) Observational study. The subjects self-assigned to facial hair / no facial hair.

(b) What was the investigators' hypothesis?

(1) That facial hair harbours bacteria at a different rate than clean-shaven faces.

(c) What does *"a cross-sectional study"* mean?

(1) The subjects were all chosen/studied at the same point in time. (ie not studying the same person with + without facial hair)

(d) What did the investigators conclude?

(1) overall colonization is similar; some bacteria species were more common in clean-shaven workers.

(e) Of the participants with facial hair, 103 had a full beard, 52 had a goatee, 25 had a moustache, and 19 had "other" facial hair. Assuming that the sample is representative of the population of hospital workers with facial hair, if we were to take another sample of size 199, we would expect to have

(1) + (1)

52 ± 6.2

goatees in our sample. (Show your working in the space below.)

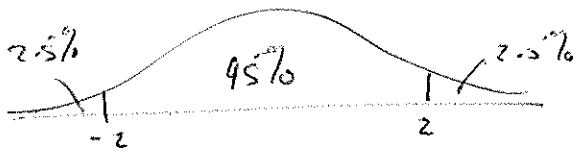
147 | 52 | not goatee | goatee

Expected value $= \frac{52}{199} \times 199 = 52.$

$SE = \sqrt{199} \times (1-0)\sqrt{\frac{52}{199} \times \frac{147}{199}} = 6.2$

(1) + (1) (f) What is the chance that our new sample will contain more than 64 people with goatees?

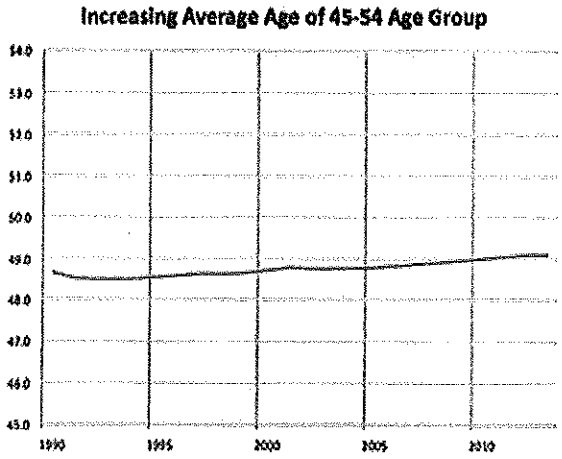
64 in standard units is ≈ 2 $\left(\frac{64-52}{6.2} = 1.94\right)$



chance is $\approx 2.5\%$

12

3. Using data from the census, it is possible to calculate the mean age of all US residents between the age of 45 and 54 (inclusive). The figure below shows how this mean age varied with time – from 1990 to 2010 the average age of US residents between the age of 45 and 54 has been increasing, with the increase being about 0.5 years.



How might you explain that, on average, US residents aged 45 to 54 have been getting older?

(2) Because of the Baby Boomers – in 2010 the ~~45-54~~ baby boomers were towards the old end of the 45-54 range, so there were more 54 year olds than 45 year olds, raising the average age.

13

4. There are 8 people in a club. One person makes up a list of all the possible committees with 2 members. Another person makes up a list of all the possible committees with 5 members. True or false: the second list is longer than the first. Explain your answer.

(3) 1st list contains

$$\binom{8}{2} = \frac{8!}{(8-2)!\,2!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{(6\times5\times4\times3\times2\times1)\times(2\times1)} = \frac{8\times7}{2} = 28 \text{ entries}$$

2nd list contains

$$\binom{8}{5} = \frac{8!}{3!\,5!} = \frac{8\times7\times6}{3\times2} = 8\times7 = 56$$

$\Rightarrow$ 2nd list is longer.

16
2 bonus

5. (Benford's Law). The first digit of a number is its leftmost digit (eg the first digit of 567 is 5). If we consider the first digits of all the adjusted gross incomes on US tax returns for a particular year, we find that they are *not* uniformly distributed. Instead of their being approximately the same number of each digit, we find the following distribution.

1st digit	1	2	3	4	5	6	7	8	9
%	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6

(a) Which digit is missing from the table? Why?

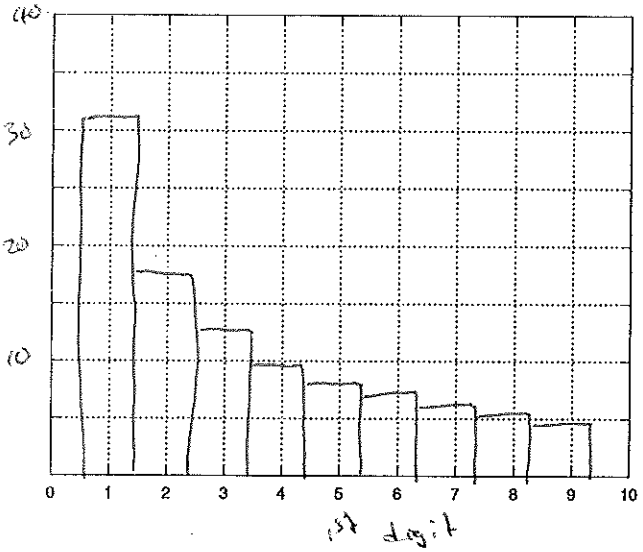
(1) Zero - it can't be a 1st digit

(b) Is "first digit" a categorical or numerical variable. Explain briefly.

(1) categorical - it doesn't make sense to do math on "first digits"

(c) Plot the distribution on the graph below. Label the axes. Is your plot a histogram or a bar chart?

(1)+(1)+(1)



Bar chart

(d) Does the distribution follow the normal curve? Why/why not?

(1) No. skewed/non-symmetric

also, categorical variables - order is meaningless

(e) [Bonus] Why do you think the first digits follow this distribution?

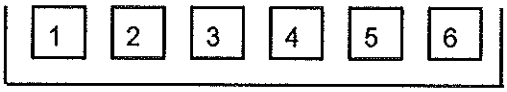
(2) uniform on log scale

think about % change needed to increase 1st digit

- it's larger if the 1st digit is a one than if the 1st digit is an 8.

[3] 6. One ticket is drawn at random from each of the two boxes below

(A)



(B)

Find the chance that

(1) (a) One of the numbers is 2 and the other is 5.

30 possibilities ; 2 are of interest

$\frac{2}{30} = \frac{1}{15}$

(1) (b) The sum of the numbers is 7.

30 possibilities ; 5 are of interest

$\frac{5}{30} = \frac{1}{6}$

(1) (c) One number is bigger than twice the other.

30 possibilities ;
10 are of interest

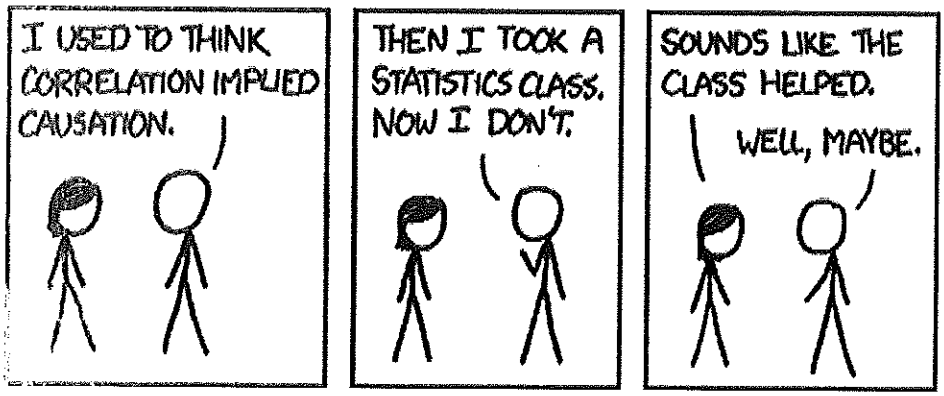
$\frac{10}{30} = \frac{1}{3}$

	1	2	3	4	5	6
1			(c)	(c)	(c)	(b) (c)
2					(a) (b) (c)	(c)
3	(c)			(b)		
4	(c)		(b)			
5	(c)	(a) (b) (c)				

BONUS QUESTION (You can get full marks on this exam without answering this questions. If you do a correct answer, you will get extra marks. You cannot score more than 100% on the exam, however.)

6. Why is this cartoon funny?

(2)



(xkcd.com)

Because she is claiming causation, where only correlation has been shown.