AMS7-Practice Midterm

You must explain all answers and/or show working for full credit.

This exam is closed book, but you may use one 8.5 by 11 piece of paper with notes, and a calculator. You may not confer with any other person (other than the one administering the examination).

1. [3 points] Data for the word lengths for articles in Popular Science is given in the table below.

<table>
<thead>
<tr>
<th>Length</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.6</td>
</tr>
<tr>
<td>2</td>
<td>14.8</td>
</tr>
<tr>
<td>3</td>
<td>18.7</td>
</tr>
<tr>
<td>4</td>
<td>16.0</td>
</tr>
<tr>
<td>5</td>
<td>12.5</td>
</tr>
<tr>
<td>6</td>
<td>8.2</td>
</tr>
<tr>
<td>7</td>
<td>8.1</td>
</tr>
<tr>
<td>8</td>
<td>5.9</td>
</tr>
<tr>
<td>9</td>
<td>4.4</td>
</tr>
<tr>
<td>10</td>
<td>3.6</td>
</tr>
<tr>
<td>11</td>
<td>2.1</td>
</tr>
<tr>
<td>12</td>
<td>0.9</td>
</tr>
<tr>
<td>13</td>
<td>0.6</td>
</tr>
<tr>
<td>14</td>
<td>0.4</td>
</tr>
<tr>
<td>15</td>
<td>0.2</td>
</tr>
</tbody>
</table>

(a) What is the median number of letters per word?

(b) Sketch the boxplot showing the distribution of the data. Is the distribution symmetrical or skewed?

2. [3 points] Read the abstract of the article “Tea drinking habits and oesophageal cancer in a high risk area in northern Iran: population based case-control study” printed at the back of this exam.

(a) Does this describe an experiment or an observational study?

(b) Was it cross-sectional, retrospective or prospective? Why?

(c) Does it show that drinking hot tea causes oesophageal cancer?
3. [8 points] A certain professor drives from Palo Alto to Santa Cruz each week. He records the
time it takes. So far, the times recorded are 57, 55, 60, 58 and 54 minutes.

(a) What is the mean time?

(b) What is the median time?

(c) What is the standard deviation of the times?

(d) What percentage of the times recorded so far are within 1 standard deviation of the
mean?

(e) Assuming the distribution of all journey times follows the Normal distribution, with
mean and standard deviation as computed in a) and c) above, how long are the longest
2.5% of all journeys?

(f) The professor’s longest drive time was 60 minutes. What is the corresponding z-score?

(g) Sketch a standard normal distribution, and show the region corresponding to drive times
longer than 60 minutes.

(h) Do you expect the distribution of all journey times to follow the normal distribution?
    Explain briefly.
4. [8 points] According to the registrar, in 2009 there were 14,662 undergraduate and 1425 graduate students at UCSC.

(a) If you take a simple random sample of 10 students, what is the probability that your sample will contain 2 graduate students?

(b) You take a randomly ordered list of students, and, going down the list, you ask each student if they are a graduate student or an undergraduate.

What’s the chance that the 10th person you ask is your second graduate student? [Hint: consider separately the first 9 people and the 10th person.]

(c) You randomly sample 200 students. Do you expect to be able to conclude that the proportion of graduate students is different from 10%.

(d) How large a sample do you need for a margin of error of 0.02 in the proportion?

(e) You mail out surveys to the number of students estimated in part d). Not all reply, so you keep mailing out surveys until you have the number of responses calculated in part d). Will these responses give you an estimate with the desired margin of error? Explain briefly.
5. [3 points] A friend recently had an ultrasound scan, which revealed that she was pregnant with twin boys. Her doctor told her that a third of all twins are identical twins. My friend isn’t so interested in the chances of twins in the entire population; she’s interested in *her* twins. What’s the chance that my friend’s twins are identical?

6. [3 points] Weights of newborn babies in the United States are normally distributed with a mean of 3420g and a standard deviation of 495g.

   (a) A baby is considered “at risk” if his or her birth weight is in the lowest 2%. What is the weight below which a baby is considered “at risk”?

   (b) If 49 newborn babies are randomly selected, find the probability that their mean weight is between 3300g and 3700g.
7. [6 points] A survey organization takes a simple random sample of 625 households from a city of 80,000 households. On the average, there are 2.30 persons per sample household, and from the census, the standard deviation of household size is 1.75. Say whether each of the following statements is true or false, and explain.

(a) The standard error for the sample mean is 0.07.

(b) A 95% confidence interval for the average household size in the sample is 2.16 to 2.44.

(c) A 95% confidence interval for the average household size in the city is 2.16 to 2.44.

(d) 95% of the households in the city contain between 2.16 and 2.44 persons.

(e) The 95% confidence level is about right because household size follows the normal distribution.

(f) The 95% confidence level is about right because, with a sample of size 625, the distribution for the sample mean is close to being normal.
Tea drinking habits and oesophageal cancer in a high risk area in northern Iran: population based case-control study

Farhad Islami, research fellow,1,2,3 Akram Poursahms, associate professor,1 Dariush Nasrollahzadeh, PhD student,4,5 Farin Kamangar, research fellow,5 Saman Fahimi, PhD student,1,6 Ramin Shakeri, research fellow,1 Behnoush Abedi-Ardekani, pathologist,1 Shahin Merat, associate professor,1 Homayoon Vahedi, associate professor,1 Shahryar Semnani, associate professor and director,7 Christian C Abnet, investigator,5 Paul Brennan, group head,2 Henrik Møller, professor and director,2 Farrokh Saidi, professor,1 Sanford M Dawsey, senior investigator,2 Reza Malekzadeh, professor and director,1 Paolo Boffetta, group head and cluster coordinator2

ABSTRACT
Objective To investigate the association between tea drinking habits in Golestan province, northern Iran, and risk of oesophageal squamous cell carcinoma.

Design Population based case-control study. In addition, patterns of tea drinking and temperature at which tea was drunk were measured among healthy participants in a cohort study.

Setting Golestan province, northern Iran, an area with a high incidence of oesophageal squamous cell carcinoma.

Participants 300 histologically proved cases of oesophageal squamous cell carcinoma and 571 matched neighbourhood controls in the case-control study and 48 582 participants in the cohort study.

Main outcome measure Odds ratio of oesophageal squamous cell carcinoma associated with drinking hot tea.

Results Nearly all (98%) of the cohort participants drank black tea regularly, with a mean volume consumed of over one litre a day. 39.0% of participants drank their tea at temperatures less than 60°C, 38.9% at 60-64°C, and 22.0% at 65°C or higher. A moderate agreement was found between reported tea drinking temperature and actual temperature measurements (weighted κ 0.49). The results of the case-control study showed that compared with drinking lukewarm or warm tea, drinking hot tea (odds ratio 2.07, 95% confidence interval 1.28 to 3.35) or very hot tea (8.16, 3.93 to 16.9) was associated with an increased risk of oesophageal cancer. Likewise, compared with drinking tea four or more minutes after being poured, drinking tea 2-3 minutes after pouring (2.49, 1.62 to 3.83) or less than two minutes after pouring (5.41, 2.63 to 11.1) was associated with a significantly increased risk. A strong agreement was found between responses to the questions on temperature at which tea was drunk and interval from tea being poured to being drunk (weighted κ 0.68).

Conclusion Drinking hot tea, a habit common in Golestan province, was strongly associated with a higher risk of oesophageal cancer.

INTRODUCTION
Oesophageal squamous cell carcinoma constitutes most of the oesophageal cancers worldwide.1,2 In Europe and America it is mainly caused by tobacco and alcohol use and is more common in men than in women.3 However, in Golestan province of Iran, which has one of the highest incidence rates for oesophageal squamous cell carcinoma in the world,4 this cancer has relatively unique epidemiological features: smoking and alcohol consumption are not major risk factors,5,6 and women are as likely to have a diagnosis of oesophageal cancer as men.7 This epidemiological pattern resembles that observed in Linxian, China, another area with a high incidence of oesophageal squamous cell carcinoma.8 Earlier studies in Golestan have suggested that low intake of fresh fruits and vegetables, low socioeconomic status, and opium consumption are associated with a higher risk of oesophageal cancer.9-11 In addition, studies have pointed towards the possible role of drinking very hot tea.9

An association between drinking hot beverages and risk of oesophageal cancer has been reported in several studies from different parts of the world.3 10-14 In Golestan, tea and water are the only drinks commonly consumed, with comparable average intake.15 An ecological study showed that inhabitants of Golestan drank more tea and at a higher temperature than people living in a nearby area with a low incidence of oesophageal cancer.16 A case-control study carried out in areas of both low and high incidence in northern Iran in the 1970s showed about a twofold increase in risk of oesophageal cancer among people who drank hot tea, but there was no trend in risk according to quantity of tea consumed.5

In contrast with some other risk factors, such as opium consumption, which may affect only a fraction of the general population or may have dissimilar distributions among men and women, tea drinking in Golestan is widespread in both sexes, usually starts at

BMJ | ONLINE FIRST | bmj.com