UNIVERSITY OF TEXAS AT AUSTIN, DEPARTMENT OF MATHEMATICS
M378K - Introduction to Mathematical Statistics

IN-TERM EXAM I

Formulas.

If X has the binomial distrubution with parameters n and p, then $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, \ldots, n$, $\mathbb{E}[X] = np$, $\mathrm{Var}[X] = np(1-p)$.

If X has the standard normal distrubution, then its mean is zero, its variance is one, and its density equals

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad x \in \mathbb{R}.$$

True/false questions.

Problem 1.1. (2 points) The following statement makes sense:

The correlation between eye color and height equals 0.05.

True or false?

Problem 1.2. (2 points)

We know that the distribution of wealth in the United States is right skewed. Thus, the majority of people in the US have more than the average (i.e. mean) level of wealth. *True or false?*

Free-response problems.

Problem 1.3. (15 points)

The Alarming Studies Weekly reported on a study suggesting that

"frequently 'heading' the ball in soccer lowers players' IQs".

A psychologist tested 60 male soccer players, ages $14 - 29$, who played soccer up to 5 times a week. Players who averaged 10 or more "headers" a game had an average IQ of 103, while players who "headed" 1 or fewer times per game had an average IQ of 112.

(a) (2 points) Describe the population of interest to the psychologist.

(b) (2 points) Describe the sample.

(c) (4 points) Identify the variables of interest, i.e., the explanatory and response variables.

(d) (2 points) Identify the type of the variables in (c). Are they categorical or quantitative?

(e) (5 points)

What do you think the inference (i.e. conclusions) made by the psychologist is? Discuss in short (two or three sentences) the possible reasons why this inference might be misleading!

Problem 1.4. (12 points)

The population of Midsomer University students can roughly be divided into two subgroups - coffee addicts and tea addicts. GPAs of both populations are approximately normally distributed - $N(mean = 3.6, sd = 0.25)$ for coffee addicts and $N(mean = 3.4, sd = 0.1)$ for tea addicts.

(a) (5 points) What is the probability that a coffee addict chosen at random has a GPA higher than 3.7?

(b) (5 points) What is the range of GPAs for the top 15% of the tea addicts?

(c) (2 points) Would it be correct to say that the GPAs of the population of **all** Midsomer University students follow the normal distribution (given the assumptions of the problem)?

Problem 1.5. (17 points)

Suppose that the thumb sizes of the US males follow a normal distribution with an unknown mean μ and standard deviation $\sigma = 20$ on the GPI - scale (*Grey's Pollex Index - GPI - from* 50 *to* 280). The US Department of Thumbs and Toes (DTT) reports that the mean thumb size in the country is $\mu = 150$.

Being the chairman of the Faculty of Thumbs of the local university you see an excellent opportunity here and decide to conduct your own study of the size of the average American thumb. After collecting a SRS of 100 American thumbs you obtain the following sample average $\bar{x} = 153$.

i (5 pts) Construct a 95%-confidence interval for the unknown parameter μ based on your study.

ii. (7 pts)

Assess the strength of evidence your study carries against the DTT findings. In other words: state the hypotheses, carry out the hypothesis test, and report your findings.

(iii) (5 pts) You dream of achieving fame and fortune by being the first person ever to estimate the mean thumb size up to the margin of error equal to ± 0.1. How large a sample size do you need for that?

Problem 1.6. Don't mess with Texas!

The *Anti-Littering League* wishes to gauge the success of the ingenious *Don't mess with Texas!* campaign.

Realizing the obvious problems with conducting a survey which outright asks the questions: "Are you or have you ever been a litterer?", they resort to the randomized-response method.

They prompt a computer to display the question

"Have you ever littered?"

with probability 0.6. The rest of the time, a virtual fair coin is flipped on the screen and the subject is asked

"Is the outcome heads or tails?"

In both cases, the subject is prompted to click the button with **Yes** or **No**. The interviewer did not know the actual question asked, just the ultimate response. So, there was no real reason for the subject to lie, and we assume that the subjects responded truthfully.

i. (5 points)

It turned out that 50% of the subjects answered "yes". Give an estimate of the proportion of *litterers* in this population.

ii. (5 points) What percentage of "yes" answers would you have obtained in an ideal world in which nobody ever litters?

Multiple-choice problems.

Problem 1.7. (5 points) A medical researcher thinks that adding calcium to the diet will help reduce blood pressure. She believes that the effect is different for men and women. 20 men and 20 women are willing to participate in the study. The researcher chooses 10 of the men and 10 of the women at random. These chosen 20 men and women take a calcium pill every day. The other 20 men and women take a placebo. This is a ...

a.: stratified random sample design.
b.: simple random sample design.
c.: randomized block experimental design.
d.: completely randomized experimental design.
e.: None of the above is correct.

Problem 1.8. (5 points)

In a hypothesis testing problem, $p-$value $= 3\%$ means that ...

a.: Null hypothesis has a 3% chance to be wrong.

b.: If the null hypothesis is true, the probability of observing as extreme or more extreme than what was observed is 3%.

c.: Alternative hypothesis has a 3% chance to be wrong.

d.: If we repeat the procedure a lot times, approximately 3% of the tests will be significant.

e.: None of the above.

Problem 1.9. (5 points)

A new headache remedy is given to a group of 250 patients who suffer severe headaches. Of these patients, 200 report that the remedy is very helpful in treating their headaches. From this information you conclude

a.: The remedy is effective for the treatment of headaches.

b.: Nothing, because the sample size is too small.

c.: The new treatment is better than aspirin.

d.: Nothing, because there is no control group.

e.: None of the above.

Problem 1.10. (5 points)

To estimate a population mean, our resident statistician Martyn Rivera plans to pick two simple random samples, each of size 100, from the population. He also plans to calculate the confidence interval with level C for each sample. What is the probability that at least one of his confidence intervals will cover the population mean?

a.: C^2

b.: $1 - C^2$

c.: $2C$

d.: $1 - (1 - C)^2$

e.: None of the above