STOR 155: SPRING 2024
Midterm Two, March 26, 2024

YOUR NAME (PRINT).....

PID.....................................

Please sign the **pledge**: "On my honor, I have neither given nor received unauthorized aid on this examination."

Sign.....................................

Open book in-class exam: time limit 75 minutes. Numbers in **boldface** indicate the number of points per part-question (100 points total for the whoe exam).

1. The following is a (very much simplified) model for a homeowners' insurance company. An insurance company offers a policy for which it charges \$1,250 as an annual premium. Potential payouts from this policy are assessed as either "partial loss of the property" for which the payout is \$50,000, or "total loss of the property" for which the payout is \$500,000. In any given year, for each policy, the company assesses the probability of a partial loss as 0.005 (1 in 200) and the probability of a total loss as 0.001 (1 in 1000). Assume that these are disjoint events (i.e. a policyholder does not claim for a total loss and a partial loss in the same year) and if there is no loss in a given year, the payout is 0.

   (a) Complete the following table showing the possible payouts x (in thousands of dollars) and their probabilities p. **[6 points.]**

Event	No loss	Partial loss	Total loss
x	0	50	
p			0.001

   (b) What are the mean μ, and the standard deviation σ, for the payout to one customer in one year? Assume the units are thousands of dollars. **[10 points.]**

(Question continues on next page)

(c) Now suppose the company has 10,000 policy holders. What are the mean and standard deviation of the *total* payout by the company in a given year? **[8 points.]**

(d) The company is receiving premiums at a rate of 1.25 (thousands of dollars) per policy holder. Ignoring other costs such as administration, what is the probability that the total payout in a given year is greater than the revenue received from premiums? (Assume that the distribution of the total payout is approximately normal.) **[10 points.]**

2. A student club consists of 15 members who identify as male and 30 who identify as female. They have to choose a committee of 5 students. Although females outnumber males in the club, the members are concerned to make sure the committee is approximately gender-balanced. To achieve this, they may try different methods for selecting the committee. Assume that every member of the club identifies as one of male or female.

   (a) Suppose the number of males in the selected committee followed a binomial distribution with $n = 5$ and $p = \frac{1}{3}$. What is the probability that you would end up with (i) a committee of two males and three females, (ii) a committee of less than two males? **[10 points.]**

   (b) Now assume a more realistic model for the selection process: the selection is *without replacement*, and every possible subset of five individuals from the 45 club members is equally likely to be chosen. Under this model, what is the probability of (i) a committee of two males and three females, (ii) a committee of less than two males? **[11 points.]**

   (c) Now suppose the club is much larger, with 150 members who identify as male and 300 female. A group of 40 is chosen to represent the club at an out-of-town event. If the selection is random, what is the probability that the group contains (i) at least 15 who identify as male, (ii) more than 15 who identify as male? State carefully any assumptions you are making here. **[12 points.]**

3. The following statement is from a recent survey: "The share of all adults who said they had tried marijuana in their lifetimes reached 49 percent in 2021, the highest number measured in half a century of Gallup polling."

   A new survey sampled 341 adults and asked them "Have you tried marijuana in your lifetime?". 147 said yes. Answer the following questions:

   (a) Calculate $\hat{p}$, the proportion of adults in the sample who have tried marijuana. Also calculate the standard error of $\hat{p}$. [**9 points.**]

   (b) Let p denote the proportion of adults in the population who have tried marijuana. Assuming the sample is random, calculate (i) a 95% confidence interval, (ii) a 99% confidence interval, for p. [**12 points.**]

   (c) Does this sample support the statement made at the head of this question? Show how to formulate this as a hypothesis test, state the null and alternative hypotheses, and calculate the p-value. Summarize your conclusions. [**12 points.**]

SOLUTIONS

Note: You will not be penalized for small numerical errors due to rounding error. Large numerical errors may be penalized.

1. (a) See the following table:

Event	No loss	Partial loss	Total loss	Row sum
x	0	50	500	
p	0.994	0.005	0.001	1
$x \cdot p$	0	0.25	0.05	$\mu = 0.75$
$x - \mu$	–0.75	49.25	499.25	
$(x-\mu)^2$	0.5625	2425.5625	249250.5625	
$p \cdot (x-\mu)^2$	0.559125	12.127813	249.250563	$\sigma^2 = 261.9375$

(b) The mean is $\mu = 0.75$ and the standard deviation is $\sigma = \sqrt{261.9375} = 16.18$ to two decimal places.

(c) For 10,000 policy holders (assumed to be independent) the total payout has a mean of $10000 \times \mu = 7500$ and a standard deviation of $100 \times \sigma = 1618$ (all in thousands of dollars). [Explanation of the SD: for independent random variables, the variance of the sum is the sum of the variances, so $10000\sigma^2$. But then the standard deviation is the square root of this, in other words $\sqrt{10000}\sigma = 100\sigma$.]

(d) The total received in premium is 12500 (thousands of dollars). Since the total payout has the mean and SD given in part (c), the associated z score is $\frac{12500-7500}{1618} = 3.09$ to two decimal places. From the normal table with $z = 3.09$, the left hand tail probability is 0.9990 and therefore the right hand tail probability is 0.001. Therefore, there is about a 0.001 probability that the total payout will exceed the revenue received from premiums.

2. (a) (i) Probability of exactly two male members being chosen is $\binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \frac{10\times 8}{3^5} = \frac{80}{243} = 0.3292$ to four decimal places. (ii) The probabilities of exactly one and no male members are respectively $\binom{5}{1}\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^4 = \frac{5\times 16}{243} = 0.3292$ and $\left(\frac{2}{3}\right)^5 = \frac{32}{243} = 0.1317$. Adding the last two probabilities gives the answer 0.4609.

(b) The given scenario corresponds to a hypergeometric distribution, for which the probability of k male members being chosen for the committee is $\frac{\binom{15}{k}\binom{30}{5-k}}{\binom{45}{5}}$. Note that $\binom{45}{5} = \frac{45\times 44\times 43\times 42\times 41}{5\times 4\times 3\times 2\times 1} = 1221759$.

(i) For $k = 2$, the required probability is $\frac{105\times 4060}{1221759} = 0.3489$.

(ii) For $k = 1$ and 0, the corresponding probabilities are $\frac{15\times 27405}{1221759} = 0.3365$ and $\frac{142506}{1221759} = 0.1166$. Adding the last two probabilities, the answer is 0.4531.

(c) We treat this as a binomial distribution with continuity correction. First calculate the mean $\mu = 40 \times \frac{1}{3} = 13.3333$ and $\sigma = \sqrt{40 \times \frac{1}{3} \times \frac{2}{3}} = 2.9814$. With $x = 14.5,\ 15.5$ respectively, we calculate $z = \frac{x-\mu}{\sigma}$ which (to two decimal places) comes to 0.39 or 0.73. The corresponding left tail probabilities under the normal distribution are 0.6517 and

0.7673. Subtract from 1 to get the right tail probabilities: the (approximate) answer to (i) is 0.3483 and to (ii) is 0.2327.

[*Footnotes (this was not part of the answer you were expected to derive for yourself).* We can use Excel or R to derive the exact probabilities, which, assuming the binomial distribution, come to 0.3422 and 0.2312. Or alternatively, repeating the argument in part (b), one can note that the hypergeometric distribution is the most appropriate model, in which case the probabilities come to 0.3362 and 0.2213. In setting this question, I was assuming that with relatively much larger samples, the distinction between the binomial and hypergeometric distributions would be less important; but you can see that it does make some difference.]

3. (a) $\hat{p} = \frac{147}{341} = 0.4311$. The standard error is $SE = \sqrt{\frac{\hat{p}\times(1-\hat{p})}{341}} = \sqrt{\frac{147\times 194}{341^3}} = 0.0268$ to four decimal places.

   (b) (i) $\hat{p} \pm 1.96 \times SE = (0.3785, 0.4836)$; (ii) $\hat{p} \pm 2.58 \times SE = (0.3619, 0.5003)$.

   (c) The null and alternative hypotheses are $H_0 : \ p = 0.49, \ H_A : \ p \neq 0.49$. The standard error for the hypothesis test is $SE = \sqrt{\frac{0.49\times 0.51}{341}} = 0.0271$. In this case we calculate $z = \frac{\hat{p}-0.49}{SE} = -2.1763$. Rounding this off to -2.18, we calculate the left-hand tail probability (from the normal table) as 0.0146. Multiplying by 2 because it is a two-sided test, the two-sided p-value is 0.0292. This is < 0.05, so we *reject* the null hypothesis at significance level 0.05. In layman's language, the data are not consistent with the national result reported in the survey (which could be for a variety of reasons, including random variation in the survey itself, but you were not asked to discuss this).

   [*Some refinements.* Although the test is significant at level 0.05, it is not significant at 0.01. The level of significance matters: in this instance, we could note that the evidence against the null hypothesis is strong but not overwhelming. Another way to tackle the problem would be to state the z^* values associated with significance levels 0.05 and 0.01, which are $z^* = 1.96$ and $z^* = 2.58$ respectively. Taking absolute values, the observed value of $|z|$ is 2.18 which is between 1.96 and 2.58; this again implies that the test is significant at level 0.05 but not at level 0.01. Note, however, that the question did ask you to compute the p-value, which this version of the answer does not give you.]