Final Review, MAT1372

- Show all work and justify your answers
- wishing you success.
- Useful formulas:

Formulas

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\quad s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$ $\quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$\Pr(A^c) = 1 - \Pr(A)$ $\quad \Pr(\text{A and B}) = \Pr(\text{A}) \times \Pr(\text{B})$

$\Pr(\text{A or B}) = \Pr(\text{A}) + \Pr(\text{B}) - \Pr(\text{A and B})$ $\quad \Pr(A|B) = \frac{\Pr(\text{A and B})}{\Pr(\text{B})}$

$\mu = \sum_{i=1}^{n} x\Pr(x)$ $\quad z = \frac{x - \mu}{\sigma}$ $\quad x = \mu + z\sigma$

$\text{P}(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, ..., n$ $\quad \mu = np$ $\quad \sigma = \sqrt{np(1-p)}$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ $\quad Q_1 - 1.5 \times IQR,$ $\quad Q_3 + 1.5 \times IQR$

$\hat{p} \pm z_{score} \times se_{\hat{p}}$ $\quad se_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\quad z = \frac{\hat{p} - p_0}{se_0}$ $\quad se_0 = \sqrt{\frac{p_0(1-p_0)}{n}}$

$\bar{x} \pm t_{score} \times se_{\bar{x}}$ $\quad t = \frac{\bar{x} - \mu_0}{se_{\bar{x}}}$ $\quad se_{\bar{x}} = \frac{s}{\sqrt{n}}$ $\quad df = n - 1$

$(\hat{p}_1 - \hat{p}_2) \pm z_{score} \times se$ $\quad se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0}$ $\quad se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ $\quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$(\bar{x}_1 - \bar{x}_2) \pm t_{score} \times se$ $\quad t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se}$ $\quad se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $\quad df = \min(n_1 - 1, n_2 - 1)$

$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ $\quad se = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $\quad df = n_1 + n_2 - 2$

$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ $\quad \text{expected} = \frac{\text{row} \times \text{column}}{\text{total}}$ $\quad df = (r-1) \times (c-1)$

1. It is known that 41 percent of the U.S. population has type A blood, 9 percent has type B, 4 percent has type AB, and 46 percent has type O. We suspect that the blood type distribution of people suffering from stomach cancer is different from that of the overall population. Suppose that a random sample of 200 stomach cancer patients yielded 92 having blood type A, 20 having blood type B, 4 having blood type AB, and 84 having blood type O.

   Conduct a hypothesis test for the following question: Are these data significant enough, at the 5 percent level of significance, to enable us to reject the null hypothesis that the blood type distribution of stomach cancer sufferers is the same as that of the general population?

2. To determine whether accidents are more likely to occur on certain days of the week, data have been collected on all the accidents requiring medical attention that occurred over the last 12 months at an automobile plant in northern California. The data yielded a total of 250 accidents, with the number occurring on each day of the week being as follows:

Monday	Tuesday	Wednesday	Thursday	Friday
62	47	44	45	52

   Conduct a hypothesis test for the following question: Are these data significant enough, at the 5 percent level of significance, to enable us to reject the null hypothesis that an accident is equally likely to occur on any day of the week?

3. Among a clinic's patients having high blood cholesterol levels of at least 240 milliliters per deciliter of blood serum, volunteers were recruited to test a new drug designed to reduce blood cholesterol. A group of 40 volunteers were given the drug for 60 days, and the changes in their blood cholesterol levels were noted: The **average change** was a decrease of 6.8 with a sample standard deviation of 12.1.

   Conduct a hypothesis test for the following question: Are these data significant enough, at the 5 percent level of significance, to enable us to reject the null hypothesis that any changes in blood cholesterol levels were due purely to chance?

4. Historical data indicate that the mean acidity (pH) level of rain in a certain industrial region in West Virginia is 5.2. To see whether there has been any recent change in this value, the acidity levels of 12 rainstorms over the past year have been measured, with the following results:

   6.1, 5.4, 4.8, 5.8, 6.6, 5.3, 6.1, 4.4, 3.9, 6.8, 6.5, 6.3

   Are these data strong enough, at the 5 percent level of significance, for us to conclude that the acidity of the rain has changed from its historical value?

Please turn over and finish the rest of the question.

5. We are interested in estimating the proportion of students at a university who smoke. Out of a random sample of 200 students from this university, 40 students smoke.

(a) Calculate a 95% confidence interval for the proportion of students at this university who smoke, and interpret this interval in context. (Reminder: Check conditions.)

(b) If we wanted the margin of error to be no larger than 2% at a 95% confidence level for the proportion of students who smoke, how big of a sample would we need?

6. It is believed that nearsightedness affects about 8% of all children. In a random sample of 194 children, 21 are nearsighted. Conduct a hypothesis test for the following question: Do these data provide evidence that the 8% value is inaccurate at the 5 percent level of significance?

7. 400 students were randomly sampled from a large university, and 289 said they did not get enough sleep. Conduct a hypothesis test to check whether this represents a statistically significant difference from 50%, and use a significance level of 0.01.

8. Heights of 10 year olds, regardless of gender, closely follow a *normal* distribution with mean 55 inches and standard deviation 6 inches.

(a) What is the probability that a randomly chosen 10 year old is shorter than 48 inches?

(b) What is the probability that a randomly chosen 10 year old is between 60 and 65 inches?

(c) If the tallest 10% of the class is considered "very tall", what is the height cutoff for "very tall"?

9. In the following problems, state whether the random variable X is binomial or hypergeometric. Also give its parameters (n and p if it is binomial or r, n and p if it is hypergeometric).

(a) A lot of 200 items contains 18 defectives. Let X denote the number of defectives in a sample of 20 items.

(b) A restaurant knows from past experience that 15 percent of all reservations do not show. Twenty reservations are expected tonight. Let X denote the number that show.

(c) In one version of the game of lotto each player selects six of the numbers from 1 to 54. The organizers also randomly select six of these numbers. These latter six are called the winning numbers. Let X denote how many of a given player's six selections are winning numbers.

(d) Each new fuse produced is independently defective with probability 0.05. Let X denote the number of defective fuses in the last 100 produced.

10. If 65 percent of the population of a certain community is in favor of a proposed increase in school taxes, find the approximate probability that a random sample of 100 people will contain

(a) At least 45 who are in favor of the proposition.

(b) Fewer than 60 who are in favor.

(c) Between 55 and 75 who are in favor.

Please turn over and finish the rest of the question.

11. Of the registered voters in a certain community 54 percent are women and 46 percent are men. Sixty-eight percent of the registered women voters and 62 percent of the registered men voters voted in the last local election. If a registered voter from this community is randomly chosen, find the probability that this person is

(a) A woman who voted in the last election

(b) A man who did not vote in the last election

(c) What is the conditional probability that this person is a man given that this person voted in the last election?

End of this test.