1. (2 points) Pica is an eating disorder defined as the craving of nonfood items (for example: paper, metal, sand, rocks). A study has shown that 30% of pregnant women are affected by pica. If, at a given time, 2% of the female population is pregnant, find the probability that a randomly selected female is pregnant and has pica

2. A fondue set contains 10 fondue forks. The forks are numbered 1 through 10, and there are 2 forks of each color: blue, yellow, orange, green and red.

   (a) (2 points) How many different groups of two forks can be selected?

   (b) (2 points) What is the probability that, if you select two forks at random, you get two forks of the same color?

3. When getting up in the morning and preparing to leave for work, John needs to wear the following 6 items: pants, socks, shirt, watch, contact lenses and ring. He can put on these items in any order.

   (a) (2 points) In how many different ways can John prepare for work?

   (b) (2 points) If John randomly decides the order, what is the probability that he will wear his watch first, and his contact lenses last?

4. A report from the World Health Organization states that France is the "most depressed country". Suppose the table below shows percentages of the adult population of France categorized by depression and gender.

	Suffered from depression	Never suffered from depression	Total:
Men	7.6%	42.6%	50.2%
Women	12.4%	37.4%	49.8%
Total:	20%	80%	100%

   Find the probability that a French adult selected at random:

   (a) (2 points) is a man and has suffered depression.

   (b) (2 points) has never suffered depression given that she is a woman.

   (c) (2 points) is a man or has suffered depression, but not both.

   (d) (2 points) Are the events "woman" and "never suffered from depression" mutually exclusive? Justify.

   (e) (2 points) Are the events "woman" and "never suffered from depression" independent? Justify.

5. (4 points) The Florida Museum of Natural History has compiled statistics on shark attacks. It was found that 38% of shark attacks are committed by white sharks. Among the attacks from white sharks, 75% of the victims survived. Among the attacks committed by other types of sharks, 84% of the victims survived. If someone has just survived a shark attack, find the probability that the perpetrator was a white shark.

6. During a 3-hour final exam, suppose each student independently has a 4% chance of needing a trip to the washroom at some point during the exam. Consider a packed gym with a total of 720 students.

   (a) (2 points) Calculate the probability that exactly 25 students will need to go to the washroom during this exam.

   (b) (2 points) Find the mean (expected value) and the standard deviation for the number of students who will need to go to the washroom.

   (c) (4 points) Use a normal approximation with continuity correction to find the probability that at least 25 students, but less than 40 students will require a trip to the washroom during the exam. Verify explicitly that using a normal approximation is appropriate for this question.

7. When preparing an exam, teachers will sometimes make small typos (typographical errors). Consider x to be the number of typos in a given exam, and suppose the probability distribution of this random variable is as follows:

x	0	1	2	3	4
$P(x)$	0.25	0.35	0.25	0.12	0.03

   (a) (4 points) Calculate the mean (expected value) and standard deviation for this random variable.

   (b) (2 points) Find the probability that a given exam will have more typos than the expected value.

8. Suppose that the speed of moving vehicles on highway 40 is normally distributed with a mean of 105 km/h and a standard deviation of 7 km/h. (Assume that the speed can be measured very precisely, with several decimals.)

   (a) (3 points) Find the probability that a random vehicle has a speed between 100 km/h and 120 km/h.

   (b) (3 points) Suppose the police would like to give tickets to the fastest 2% of drivers on the highway. What is the precise speed, that if exceeded, will earn you a speeding ticket according to this rule?

   (c) (3 points) A police officer is planning to measure the speed of 25 randomly selected vehicles. Find the probability that the average speed of this sample will be greater than 103 km/h.

   (d) (3 points) Find the probability that a random vehicle has a speed below 100 km/h. Using this information as a probability of success, what is the probability that exactly 6 vehicles in a random sample of 25 have a speed below 100 km/h?

9. In a study relating physical attractiveness and mental disorders, 231 subjects were rated for attractiveness, and the sample average rating was 3.94. Assume that the population standard deviation of attractiveness rating is $\sigma = 0.75$.

   (a) (4 points) Construct a 93% confidence interval for the true average attractiveness rating in the population.

   (b) (2 points) Interpret your answer from part (a) in the context of this problem.

   (c) (3 points) Find the minimum sample size required to estimate the population average attractiveness rating (with a 93% confidence interval) so that the margin of error is at most 0.05.

10. (6 points) Two types of medication to treat hives are being tested. A random sample of 240 patients suffering from hives were given medication A. After 30 minutes, 216 of the patients no longer had hives. Another random sample of 350 patients with hives were given medication B. After 30 minutes, 329 of the patients no longer had hives.

   Test the claim that the percentages of effectiveness of the two medications are different. Use a 1% significance level and follow the method of your choice (rejection region or P-value).

11. Many studies and discussions have targeted the topic of "home advantage" in sports; do teams playing in their home city have a better chance of winning than when playing elsewhere? In a random sample of 16 hockey games during the playoffs, the home team won 10 of those games, and the visiting team won the other 6.

   (a) (5 points) Test the hypothesis that the proportion of all playoff games won by the home team is greater than 0.5. Use a 10% significance level and follow the P-value approach. Make sure to verify the requirements for the test you are applying.

   (b) (2 points) Interpret the conclusion of the test in part (a) in the context of this application.

   (c) (3 points) If you intend to construct a 95% confidence interval for the true proportion of all playoff games won by the home team, how many **more** games would you have to sample in order for the margin of error to be at most 0.05?

12. At a circus, there is a booth where a man claims that he can guess your age within a certain tolerance. Of course, to participate, people have to pay, and if the guess is not close to the correct age, the customer would win a small prize.

   (a) (7 points) A random sample of 6 people line up at this booth. An observer records the following data.

person	1	2	3	4	5	6
guess	31	48	19	53	52	37
correct age	35	47	21	53	50	42

   Test the claim that the man significantly underestimates people's age. Use a 10% significance level and follow the rejection region approach. (Assume that the difference between the guess and correct age is normally distributed.)

   (b) (2 points) What can you say about the P-value for the test in part (a)?

   (c) (5 points) A bus with 40 young men arrives at the circus. All of them are 18 years old; 23 of them have some facial hair, and the other 17 are clean-shaven. These young men all participate in the "guess your age" game, at different times of the day. For those with facial hair, the average age guessed was 18.6 with a standard deviation of 1.5. For those without facial hair, the average age guessed was 17.5 with a standard deviation of 1.1.

   Test the hypothesis that there is a significant difference between the average age guessed for men with or without facial hair. Use a 5% significance level and follow the rejection region approach. (Assume that the guesses are normally distributed among young men with and without facial hair.)

13. (7 points) The following data is inspired by a study analyzed by the famous statistician Karl Pearson in 1909. The table classifies a sample of 285 crimes by type, and by whether the criminal drinks alcohol or not.

	violence	stealing	fraud	total
drinker	38	94	15	147
abstainer	27	75	36	138
total	65	169	51	285

Test the hypothesis that the type of crime committed and the drinking habits of the criminal are dependent variables. Use a 1% significance level and follow the rejection region approach.

14. A brand of candy-coated chocolate comes in five different colors: brown (40%), yellow (20%), orange (20%), green (10%) and red (10%). A candy enthiusiast is conducting a test of hypothesis to check the validity of this distribution: a random sample of 580 pieces of candy contained 216 brown, 125 yellow, 136 orange, 42 green and 61 red candies.

(a) (2 points) Find the expected frequencies for this test.

(b) (2 points) Suppose a software reports the value of the test statistic to be 9.8190. Estimate the corresponding P-value.

ANSWERS:

1. 0.006

2. (a) 45   (b) 0.1111

3. (a) 720   (b) 0.0333

4. (a) 0.076   (b) 0.751   (c) 0.550

(d) No, P(Woman and Never suffered from depression) $= 0.374 \neq 0$.

(e) P(Never suffered from depression) $= 0.800 \neq 0.751 = P$(Never suffered from depression — Woman)

so the events are dependent.

5. 0.3537

6. (a) 0.0616   (b) $\mu = 28.8$ $\sigma = 5.258$ (c) 0.7727

7. (a) $\mu = 1.33$ $\sigma = 1.068$ (b) 0.4

8. (a) 0.7449   (b) 119.35   (c) 0.9236   (d) 0.1840

9. (a) (3.8507, 4.0293)

(b) We are 93% confident that the average rating for the population is between 3.85 and 4.03.

(c) The sample should consist of at least 738 persons.

10. H_0: $p_1 - p_2 = 0$ H_A: $p_1 - p_2 \neq 0$ test statistic: $z^* = -1.798$

P-value $= 0.0658$ fail to reject H_0

There is insufficient evidence to claim that the percentages of effectiveness of the two medications are different, at the 1% level of significance.

11. (a) $np = 8 > 5$ $nq = 8 > 5$ H_0: $p = 0.5$ H_A: $p > 0.5$

test statistic: $z^* = 1.00$ P-value $= 0.1587$ Fail to reject H_0

(b) There is insufficient evidence to conclude that the home team has an advantage, at the 10% level of significance.

(c) We would need 345 more games.

12. (a) H_0: $\mu_{\bar{d}} = 0$ H_A: $\mu_{\bar{d}} > 0$ (d = correct − guess)   Rejection region: $t > 1.476$

test statistic: $t^* = 1.164$ fail to reject H_0

There is insufficient evidence to support the claim that the man underestimates people's age, at the 10% level of significance.

(b) P-value between 0.125 and 0.250

(c) H_0: $\mu_1 - \mu_2 = 0$ H_A: $\mu_1 - \mu_2 \neq 0$ rejection region: $t < -2.120$ or $t > 2.120$

test statistic: $t^* = 2.676$ reject H_0.

There is enough evidence to claim that there is a difference between the average age guessed for men with or without facial hair, at the 5% level of significance.

13. H_0: the type of crime committed and the drinking habits are independent

H_A: the type of crime committed and the drinking habits are dependent

rejection region: $\chi^2 > 9.21$ test statistic: $\chi^2 = 12.3788$ reject H_0

The evidence suggests that the type of crime and drinking habits are dependent, at the 1% level of significance.

14. (a) Brown: 232 Yellow: 116 Orange: 116 Green: 58 Red: 58

(b) P-value between 0.025 and 0.05