1. A basketball player has a 0.8 probability of making a 1-point free throw, a 0.4 probability of making a 2-point shot, and a 0.2 probability of making a 3-point shot. All shots are independently likely to score.
   (a) If the player takes 1000 free throws, briefly explain why the total number of points scored is approximately normal, and find the mean and standard deviation.
   (b) Using the normal approximation with continuity correction, estimate the probability that she scores at least 820 points on 1000 free throws.
   (c) If the player takes 1000 free throws, 2000 2-point shots, and 500 3-point shots during the season, describe the approximate distribution of the total number of points she scores.
   (d) Using the normal approximation with continuity correction, estimate the probability that she scores between 2500 and 2700 points inclusive during the season described in part (c).

2. The weights of widgets are normally distributed with mean 10.05g and standard deviation 0.10g. Sample A consists of 10 widgets and sample B consists of 25 widgets.
   (a) Describe the distributions of the respective average weights of Sample A and Sample B.
   (b) Find the probability that a random widget has weight over 10.15g.
   (c) Find the probability that at least one widget in Sample A has weight over 10.15g.
   (d) Find the mean and standard deviation of the total weight of Sample A.
   (e) Find the probability that the total weight of Sample A exceeds 101g.
   (f) Find the probability that the average weight of Sample A is less than 9.95g.
   (g) Describe the distribution of the difference in the average weights of Sample A and Sample B.
   (h) Find the probability that the average weight of Sample A exceeds the average of B by 0.05g or more.

3. You wait for the Orange Line at Ruggles during rush hour: your wait time is uniformly distributed between 0min and 7min.
   (a) What is the probability that you will have to wait 5+ minutes (5 minutes or more)?
   (b) Find the expected value and standard deviation of your wait time.
   (c) If you take the train 75 times, describe the approximate distribution of your average wait time.
   (d) If you take the train 75 times, estimate the probability that your average wait time exceeds 3.57 minutes.
   (e) If you take the train 75 times, estimate the probability that you have to wait 5+ minutes at least 30 times.

4. On average, the Green Line through Northeastern has a service interruption 3 times per week.
   (a) If X is the random variable measuring the number of times there is a service interruption in one week, what type of distribution is X, and what is its pdf?
   (b) What is the probability that there is no service interruption this week?
   (c) What is the probability that there are at least 5 service interruptions this week?
   (d) What is the probability that there is at least 1 service interruption today?
   (e) What is the probability that there are exactly 15 service interruptions within the next 30 days?

5. You wait for the #28 bus at Ruggles during a snowstorm. Your expected wait time is 30 minutes, independent of the amount of time you have already been waiting.
   (a) If Y is the random variable measuring your wait time, what type of distribution is Y, and what is its pdf?
   (b) Find the probability that you wait at least 30 minutes.
   (c) Find the probability that the bus comes within the next 10 minutes.

6. Students in Math 3081 enjoy the course 2% of the time[1]. Assume there are 130 students in the course.
   (a) Find the exact probability that either 0 or 1 student really enjoys the course.
   (b) An approximation to the probability in (a) can be found using a Poisson model. What is the parameter for this model, and what is the probability estimate?
   (c) An approximation to the probability in (a) can also be found using a normal model. What are the parameters for this model, and what is the probability estimate?
   (d) Which estimate (Poisson or normal) is more accurate? Briefly explain why this should be the case.

---

7. For each independent sample of the given random variable, find (i) the likelihood function $L(\theta)$, (ii) the derivative of the log-likelihood $\frac{d}{d\theta}[\ln L(\theta)]$, and (iii) the maximum likelihood estimate $\hat{\theta}$:
   (a) $p_X(n;\theta) = \theta(1-\theta)^n$ for integers $n \geq 0$, with sample values 3, 4, 1, and 6.
   (b) $p_X(n;\theta) = \theta^n e^{-\theta}/n!$ for integers $n \geq 0$, with sample values 5, 3, and 6.
   (c) $p_X(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x \geq 0$, with sample values 1, 2, and 5.
   (d) $p_X(x;\theta) = e^{-(x-\theta)^2}/\sqrt{\pi}$ for all real x, with sample values 1, −2, 0.
   (e) $p_X(x;\theta) = 2(\theta - x)/\theta^2$ for $0 \leq x \leq \theta$, with sample values 2 and 5. (Note $\theta > 5$ here.)

---

8. A Poisson distribution with parameter λ is independently sampled four times, yielding values x_1, x_2, x_3, x_4. For each estimator (i) find its expected value and decide whether it is biased and (ii) find its variance. Then decide which of the unbiased estimators is the most efficient.
   (a) $\hat{\lambda}_1 = \frac{1}{2}(x_1 + x_3)$. (b) $\hat{\lambda}_2 = \frac{1}{4}(x_1+x_2+x_3+x_4)$ (c) $\hat{\lambda}_3 = x_1 - x_2 + 2x_3$. (d) $\hat{\lambda}_4 = \frac{1}{5}(x_1+x_2+2x_3+x_4)$

---

9. A normal distribution with mean μ and variance 4 is sampled once, yielding the value x. A normal distribution with mean 3μ and variance 3 is sampled once, yielding y. Consider estimators for μ of the form $\hat{\mu} = ax + by$.
   (a) Find $E(\hat{\mu})$ and $\mathrm{var}(\hat{\mu})$. (b) Find the values of a and b that give the most efficient unbiased estimator $\hat{\mu}$.

---

10. Researchers want to test whether hearing a certain type of music improves the intelligence of infants. 880 infants are randomly assigned to receive a treatment: 330 hear classical music, 350 hear thrash metal, and 200 hear a placebo (nature sounds) for two months. The infants are then assessed on a basic skills scale whose scores are normally distributed with standard deviation 1. The classical music group has an average score of 6.81, the thrash metal group has an average score of 6.90, and the placebo group has an average score of 6.83.
    (a) Test at the 4% level whether classical music improves the skills score relative to a placebo.
    (b) Test at the 4% level whether thrash metal improves the skills score relative to a placebo.
    (c) Test at the 4% level whether classical music and thrash metal have different skills scores.
    (d) If there is actually no difference among any of the three groups, what is the approximate probability of making a type I error in *at least one* of the hypothesis tests (a)-(c)?
    (e) Briefly explain why it would be improper statistical practice to report the result of only the tests performed in (a)-(c) that have a p-value below the 4% threshold.

---

11. Perform a hypothesis test for each situation: give the hypotheses, the significance level α, the test statistic and its observed value, the p-value, and the result of the test.
    (a) Researchers want to test whether crystal healing is more effective than a placebo. 134 patients are randomly assigned to receive a treatment: 67 receive crystal healing while the other 67 receive an equivalent placebo treatment with regular rocks. 22 of the crystal healing patients and 20 of the placebo patients report improved wellness. Test at the 4% level whether crystal healing was more effective than a placebo.
    Also, find 95% confidence intervals for the proportion of patients reporting improved wellness in the placebo and treatment groups.
    (b) The tenure-track faculty salaries at a university are approximately normally distributed with standard deviation \$30000. The university samples 100 male faculty and finds their average salary to be \$108591, while a sample of 25 female faculty finds their average salary to be \$91513. Test at the 3% significance level that the female faculty are paid less than the male faculty.
    Also, find 95% confidence intervals for the two populations' salaries.

[1]This is false; the actual value is 0%.

(c) Researchers want to test whether people with Black-sounding names are less likely to receive job interviews than people with White-sounding names[2]. They send 2,425 resumes with White-sounding names and receive 234 callbacks, and also 2,425 resumes with Black-sounding names and receive 157 callbacks. (The resumes are otherwise identical.) Test at the 0.1% significance level that the resumes with Black-sounding names were less likely to receive an interview than ones with a White-sounding name.
Also, find 95% confidence intervals for the two callback percentages.

---

12. Researchers want to test whether taking hydroxychloroquine lowers the rate of hospitalization in (unvaccinated) patients with COVID-19.[3] It is estimated that patients who do not take the drug require hospitalization within 15 days 15.6% of the time. 30 patients out of 159 who took hydroxychloroquine required hospitalization within 15 days.
    (a) Test whether hydroxychloroquine affects the hospitalization rate at the 20%, 9%, and 2% significance levels.
    (b) Test whether hydroxychloroquine lowers the hospitalization rate at the 20%, 9%, and 2% significance levels.
    (c) Test whether hydroxychloroquine increases the hospitalization rate at the 20%, 9%, and 2% levels.
    (d) Based on the results of (a)-(c), critique the accuracy of the following statements: (i) "the study proves that hydroxychloroquine is effective", (ii) "the study proves that hydroxychloroquine is not effective", (iii) "it is impossible to conclude from the study whether hydroxychloroquine is effective".
    (e) If the number of patients were tripled (but all the percentages stayed the same), qualitatively describe the effect on the power of the test. Would any of the answers in the previous parts change?

---

13. The math department wants to determine whether students actually enjoy Math 3081.[4] To determine this, they randomly poll 50 students in the course and find that 38% (19 students) enjoy it.
    (a) Find 80%, 90%, 95%, and 99% confidence intervals for the true proportion of students who enjoy the course.
    (b) If the department wanted to have confidence intervals that were 1/3 as wide in part (a), how many students would they need to poll?
    (c) Assuming the sample above is representative, if the department wanted a 95% margin of error of $\pm 2\%$, how many students would they need to poll? How would the answer change if the department had no idea of the true proportion of students who enjoy the course?
    (d) Test at the 11%, 3%, and 0.5% significance levels that exactly 30% of the students enjoy the course.
    (e) Test at the 11%, 3%, and 0.5% significance levels that less than half of students enjoy the course.
    (f) For each hypothesis test performed in (d)-(e), identify whether it would be a type I error, type II error, or correct decision if (i) the null hypothesis were true, or (ii) the true proportion was equal to 45%.
    (g) If the true proportion is 46%, find the type-II error probability and the power for testing the null hypothesis in (e) that less than half the students enjoy the course at the 11% significance level, with sample sizes of (i) 50 and (ii) 500.

---

14. A normal distribution with unknown mean μ and known standard deviation 4 is sampled five times, yielding values 0, -2, 3, 1, 6.
    (a) Find 80%, 90%, 95%, and 99.5% CIs for μ.
    (b) Test the null hypothesis $H_0 : \mu = 0$ against $H_a : \mu > 0$ at the 10%, 3%, and 1% significance levels.
    (c) If the true mean is 4, find the type-II error probability and the power of the test if a sample of size 5 is used for the hypothesis test in (b) at the 10%, 3%, and 1% significance levels.
    (d) If the true mean is 4, find the type-II error probability and the power of the test if a sample of size 20 is used for the hypothesis test in (b) at the 10%, 3%, and 1% significance levels.

---

[2] Bertrand and Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination", Amer. Econ. Rev. (2004), https://cos.gatech.edu/facultyres/Diversity_Studies/Bertrand_LakishaJamal.pdf .

[3] Cavalcanti *et al.*, Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19, NEJM (July 2020), https://www.nejm.org/doi/full/10.1056/NEJMoa2019014 .

[4] As noted in problem 6, the actual proportion of such students is 0%.