1. Last year, during a lecture, I asked my students: “How many text messages did you send today so far?” (It was about noon.) Here is a stem-and-leaf plot showing the data I obtained.

| | |
|---|---|
| 0 | 00000001133455777 |
| 1 | 000002555 |
| 2 | 056 |
| 3 | 005 |
| 4 | 0 |
| 7 | 00 |
| 8 | 0 |
| 10 | 0 |
| 11 | 3 |

In this sample of 38 students, the average number of text messages is $\overline{x} = 20.76$, with a standard deviation of 28.695.

[1] (a) What can you say about the symmetry in this distribution?

[4] (b) Find the outliers (if any) by using the quartiles, and draw a box-plot.

[2] 2. Explain in words what it means for two events to be **disjoint**, and give a concrete example.

[2] 3. Explain in words what it means for two events to be **independent**, and give a concrete example.

[3] 4. How many possible five-letter passwords are there if the first two letters are not allowed to be the same?

5. Texas hold ’em is a variation of the standard card game of poker. The game consists of two cards being dealt face down to each player, and then five community cards being placed face-up by the dealer. So each player sees 7 cards in total (2 in his hands, and 5 on the table).

[2] (a) How many different sets of 7 cards are possible?

[2] (b) What is the probability of getting 4 kings?

[3] (c) What is the probability of getting a flush, meaning that at least 5 of the 7 cards are of the same suit? (Careful not to count the same hand multiple times.)

[2] (d) Suppose you hold in your hands the ace of spades and the 2 of spades. What is the probability that the five cards on the table are revealed to be the 3, 4, 5, 6 and 7 of spades, in that specific order?

6. On the basis of a physical examination and symptoms, a physician assesses the probabilities that the patient has no tumor, a benign tumor, or a malignant tumor as 0.70, 0.20 and 0.10 respectively. A thermographic test is subsequently given to the patient. This test gives a negative result with probability 0.90 if there is no tumor, with probability 0.80 if there is a benign tumor, and with probability 0.20 if there is a malignant tumor.

[4] (a) Suppose the test gives a negative for this patient. What is the probability that this patient has (i) no tumor? (ii) a benign tumor? (iii) a malignant tumor?

[2] (b) Is the event “*test is negative*” independent from the event “*patient has no tumor*”? Justify mathematically.

7. Consider the following discrete **cumulative** probability distribution:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1/8 & \text{for } 0 \leq x < 1 \\ 1/2 & \text{for } 1 \leq x < 2 \\ 7/8 & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

[3] (a) Find the associated probability distribution $p(x)$ and check that it is a valid distribution.

[2] (b) Find $P(-5 \leq X < 2.5)$.

[2] (c) Find $P(X = 2 \mid X \neq 0)$.

[2] 8. Suppose $X$ is a discrete random variable with a Poisson distribution with parameter $\lambda$. Find $E(X^2)$.

9. A small lake contains a total of 100 fish, among these 40 are trouts. Three different researchers want to make some measurements on trouts only, but each person has a different approach.

- Researcher A waits until a trout swims close by ("an arrival"), which occurs on average 8 times per hour.
- Researcher B uses a fishing pole and catches 20 fish, putting the fish back in the lake each time. Note that for each fish being caught, the probability that it is a trout is 0.4.
- Researcher C takes a big fish net, and catches a sample of 20 fish at once.

Denote by $X_A$, $X_B$ and $X_C$ the number of trouts that each researcher sees during his experiment. Assume that these random variables are independent.

[2] (a) What well-known distributions should be used to model $X_A$, $X_B$ and $X_C$?

[5] (b) Find $P(X = 8)$ and $E(X)$ for each researcher.

[3] (c) Which researcher, $A$ or $B$, has a better chance of seeing more than 8 trouts?

[2] (d) Calculate the probability that researcher A sees 8 trouts **and** researcher B sees 8 trouts.

[2] (e) Calculate the probability that researcher A sees 8 trouts **or** researcher B sees 8 trouts.

[2] (f) Find $P(3 \leq X_A < 12)$.

[3] (g) Suppose each researcher is paid a base amount of 10\$, plus 5\$ for each trout studied. Calculate the expected pay and the variance of the pay for each researcher.

10. Consider the following continuous probability density function (pdf):

$$f(x) = \begin{cases} \dfrac{x^2}{9} & \text{for } 0 \leq x \leq 3, \\ 0 & \text{elsewhere.} \end{cases}$$

[2] (a) Calculate the mean of $X$ (also called the expected value).

[3] (b) Calculate the median of $X$ (also called the 50th percentile).

11. A machine fills bags with white sand. The actual weight of the sand in each bag follows a normal distribution with a standard deviation of 1.5 pounds. The mean of the distribution depends on the setting of the machine.

[3] (a) At what mean weight should the machine be set so that only 5% of the bags contain less than 50 pounds of sand?

[3] (b) Suppose we set the mean weight at $\mu = 50.5$ pounds, and that we fill up a sample of 40 bags of sand. What is the probability that the average weight of this sample is less than 50 pounds?

[2] 12. Explain in words the difference between a biased versus an unbiased estimator.

[3] 13. Consider constructing a 97% confidence interval for a proportion. Find how large the sample size $n$ should be in order to guarantee that the width of this interval is at most 0.01.
(It is known that a sample proportion of $\hat{p} = 0.5$ would yield the widest interval.)

[3] 14. Suppose you did a hypothesis test on a mean with the alternative $H_a : \mu \neq 10$, and found a $P$-value of 0.0062. Your sample mean and standard deviation were $\overline{x} = 11.3$ and $s = 4$, however the sample size $n$ was lost; all you remember is that it was larger than 30. Can you recover the value of $n$?

15. A test measuring math anxiety is administered to 25 males and 20 females. Among males, the average measure was 70.5 with standard deviation 13.2. Among females, the average measure was 75.7 with standard deviation 13.6.

[5] (a) Is there a significant difference in math anxiety between males and females? Use a 10% significance level and the $P$-value approach.

[3] (b) Construct a 99% confidence interval for the difference in the average math anxiety measure between males and females.

[2] (c) Interpret your answers from part (a) and (b).

16. On a junior basketball team, the number of personal fouls ($X$) and the number of points scored ($Y$) by each player was recorded.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| personal fouls (X) | 1 | 24 | 1 | 18 | 9 | 3 | 5 | 20 | 1 | 3 | 22 |
| points scored (Y) | 2 | 75 | 0 | 59 | 9 | 7 | 35 | 46 | 0 | 2 | 57 |

From this data, is was calculated that:

$$n = 11, \quad \sum x_i = 107, \quad \sum y_i = 292, \quad \sum x_i^2 = 1911, \quad \sum y_i^2 = 15834, \quad \sum x_i y_i = 5321$$

[3] (a) Find the correlation coefficient $r$ and interpret. What can you say about any *causation* effect?

[2] (b) Find the equation of the least-square regression line.

[3] (c) Find the probability that a player committing 10 personal fouls would score more than 40 points.

[4] (d) Test that there is a difference in variance between the number of personal fouls and the number of points scored. Use a 2% significance level and the rejection region approach.
(*Note that* $\dfrac{S_{xx}}{n-1}$ *is the sample variance for* $X$.)

17. Consider doing a test on a variance with the alternative hypothesis $H_a : \sigma^2 < 2$, and a 5% significance level. You are planning to use a sample of size $n = 31$.

[1] (a) Draw the rejection region for this test.

[3] (b) Calculate the probability of a type II error if the true population variance is 0.7268.

## ANSWERS:

1. (a) not symmetric, positively skewed
   (b) $\tilde{x} = 10, \quad Q_1 = 3, \quad Q_3 = 26, \quad \text{outliers} = 70, 70, 80, 100, 113$
2. disjoint events: mutually exclusive, cannot occur at the same time (ex: being right and wrong)
3. independent events: knowledge of one doesn't affect the other (ex: being tall and winning the lottery)
4. $26 \cdot 25 \cdot 26 \cdot 26 \cdot 26 = 11,424,400$
5. (a) $C_7^{52} = 133,784,560$
   (b) $\frac{C_3^{48}}{C_7^{52}} = 0.000129$
   (c) $4\left[\frac{C_5^{13} \cdot C_2^{39} + C_6^{13} \cdot C_1^{39} + C_7^{13}}{C_7^{52}}\right] = 0.0306$
   (d) $\frac{1}{P_5^{50}} = 3.9331 \times 10^{-9}$
6. (a) $P(\text{no tumor}|\text{test negative}) = 0.7778$
   $P(\text{benign tumor}|\text{test negative}) = 0.1975$
   $P(\text{malignant tumor}|\text{test negative}) = 0.0247$
   (b) not independent since $P(\text{no tumor}) \neq P(\text{no tumor}|\text{test negative})$
7. (a) $p(0) = 1/8, \quad p(1) = 3/8, \quad p(2) = 3/8, \quad p(3) = 1/8$
   (b) $7/8$
   (c) $3/7$
8. $E(X^2) = \lambda(1 + \lambda)$
9. (a) $X_A =$ Poisson with $\lambda = 8$
   $X_B =$ binomial with $n = 20, p = 0.4$
   $X_C =$ hypergeometric with $N = 100, M = 40, n = 20$
   (b) $P(X_A = 8) = 0.1396, \quad P(X_B = 8) = 0.1797, \quad P(X_C = 8) = 0.2008$
   the expected value is 8 in all cases
   (c) $P(X_A > 8) = 0.407, \quad P(X_B > 8) = 0.404$
   (d) 0.0251
   (e) 0.2942
   (f) 0.874
   (g) $Y =$ amount paid $= 10 + 5X$
   $E(Y) = 50$ (same for all researchers)
   $V(Y_A) = 200, \quad V(Y_B) = 120, \quad V(Y_C) = 96.9697$
10. (a) 2.25 (b) 2.3811
11. (a) 52.4675 (b) 0.0174
12. *unbiased:* on average, the estimator gives the true value
    *biased:* more often overshoots or undershoots the true value
13. $n \geq 47089$
14. $n \approx 71$
15. (a) $\nu = 40, \quad t^* = -1.29, \quad P - \text{value} \approx 0.2,$ fail to reject $H_0$
    (b) $(-16.0893, 5.6893)$
    (c) there is not sufficient evidence to claim there is a difference between math anxiety in men and women, and we are 99% confident that the difference in math anxiety is within the iterval shown in part (b)
16. (a) $r = 0.9354$, a strong positive linear relationship,
    but we cannot say that committing more fouls will cause you to score more often
    (b) $Y = -1.1841 + 2.8507X$
    (c) 0.1151
    (d) rejection region is $(-\infty, 0.2062) \cup (4.85, \infty)$
    test statistic $f^* = 0.1077$ (or $f^* = 9.2885$)
    reject $H_0$, there is a significant difference in the variances
17. (a) $\chi^2$ distribution with $\nu = 30$, rejection region is $(-\infty, 18.493)$
    (b) $\beta(0.7268) \approx 0.01$