Midterm I
Math 11, UCSD, Winter 2018
Tuesday, January 30th, 2pm–3:20pm
Instructor: Eddie Aamari

- Write your PID, Name and Section in the spaces provided above.
- Do not staple the pages.
- Write your solutions clearly in the spaces provided.
- Answers written outside the answer boxes will not be graded.
- You may use a calculator (any type is fine), but no other electronic devices.
- You may not use your cell phone, tablet, or computer as a calculator.
- One handwritten page of notes allowed. (both sides OK, 8.5” by 11”)
- Put away (and silence!) your cell phone and other devices that can be used for communication or can access the Internet.
- Show all of your work; no credit will be given for unsupported answers.
- You may either give exact answers (like 1/3) or round to three decimal places (like 0.333).

Do not turn page until instructed to do so
Exercise I (2 points)

1. (½ point each) Indicate whether the following statements are True or False. Make sure to blacken completely the circle of the answer you pick.

(a) A bimodal distribution is symmetric.  ○ True  ☐ False
(b) The IQR is a more appropriate measure of spread than the standard deviation in distributions with outliers.  ☐ True  ○ False
(c) For two events $A$ and $B$, if $P(A) = P(A|B) \cdot P(B)$, then $A$ and $B$ are independent.  ○ True  ☐ False
(d) If all the points of a scatterplot fall exactly on a line, then the correlation of the two variables must be 1.  ○ True  ☐ False

Exercise II (5 points)

Leah is flying from Boston to Denver with a connection in Chicago. The probability her first flight leaves on time is 0.15. If the flight is on time, the probability that her luggage will make the connecting flight in Chicago is 0.95, but if the first flight is delayed, the probability that the luggage will make it is only 0.65.

1. (2 point) Are the first flight leaving on time and the luggage making the connection independent events? Explain.

Let $A$ denote "first flight leaving on time"
B denotes "luggage making the connection"
then $P(B|A) = 0.95 \quad P(B|A^c) = 0.65$

since they are not equal, $A$ and $B$ are not independent
(or by definition of independence, whether first flight leaves on time affects whether luggage making the connection)

2. (3 points) What is the probability that her luggage arrives in Denver with her?

\[
P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) \\
= 0.95(0.15) + 0.65(1-0.15) \\
= 0.695
\]
Exercise III (8 points)

You land on a foreign planet that is home to two species of aliens. You decide to collect weight data (in pounds) for the creatures of this new world. Below is the histogram of 212 aliens you weighed:

1. (2 points) Describe the important visual features of this histogram. (The numbers on the x axis have been replaced with letters for part b of this question.)

   Bi-modal, asymmetric and outlier near H.

2. (2 points) Between what two consecutive letters would we find the mean for this histogram? Give a reason for your choice.

   C and D. The balance point is between the two humps and since there are more points in the D-E span we need to move the balance point in that direction (vs. toward B)
3. (2 points) Which measure of spread is more appropriate to use when describing this histogram and why?

Use the IQR given outliers and asymmetry.

4. (2 points) Give a possible theory for the strange shape of the above histogram and suggest how you might test to see if your theory is correct.

The two different alien species might be different weights or this might show a gender-based weight difference.
Exercise IV (3 points)

If an animal is able to lower its heart rate, it uses less oxygen to maintain its function. Because of this, researchers believe that animals should be able to stay underwater longer if they have lower heart rates. To test this, you study the dives of 125 penguins, testing their heart rates (beats per minute, bpm) and how long they can stay underwater (duration, minutes). You get the below scatterplot.

1. (1 point) Given the visible curve in the graph, what transformations are reasonable to try when attempting to straighten the scatterplot? Circle all those that are appropriate.

- $\ln y$ vs. $x$
- $y^2$ vs. $x$
- $\sqrt{y}$ vs. $x$
- $y^4$ vs. $x$

2. (2 points) After making a suitable transformation, you fit a linear model and have the computer display the residual plot. Why is it impossible to get the below residual plot when you create the line of best fit through the straightened scatterplot (which is not pictured)?

The residuals can't be all positive. $e = y - \hat{y} > 0$ means the line is below all the points ($y > \hat{y}$ for all points), we can find a better line by moving the line up some.
Exercise V (5 points)

After a statistics course, 83% of students know the formula for standard deviation. Among those who know this formula, 95% passed, but only 56% of those students who did not know the formula passed.

1. (2 points) Construct a tree diagram of this scenario

Let $A$ denote "student knows the formula".

Let $B$ denote "student pass the exam".

Tree diagram

- 83% $A$ 95% $B$
- 17% $A^c$ 5% $B^c$
- 56% $B$
- 44% $B^c$

2. (3 points) Compute the probability that a randomly picked student knows the formula for standard deviation if it is known that he/she passed.

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} = \frac{(95\%) \cdot (83\%)}{(95\%) \cdot (83\%) + (56\%) \cdot (17\%)}$$

$$\approx 89.23\%$$
Exercise VI (7 points)

You play the following game from a well-shuffled deck of 52 cards. If you draw a black card, you win $1. If you draw a heart, you win $4. For any diamond, you win $7, plus an additional $15 for the king or ace of diamonds.

1. (4 points) Create a probability model for the amount you win playing this game. Find the expected value and standard deviation for this model.

\[
\begin{array}{c|c|c|c|c}
X & 1 & 4 & 7 & 22 \\
P(X) & \frac{1}{2} & \frac{1}{4} & \frac{13}{4} & \frac{2}{13} \\
\end{array}
\]

\[
E(X) = \sum X P(X) = \frac{1}{2} + 4 \left(\frac{1}{4}\right) + 7 \left(\frac{13}{4}\right) + 22 \left(\frac{2}{13}\right) \\
\approx \$3.827
\]

\[
\text{Var}(X) = \sum \frac{(X-\mu)^2}{n} P(X) = (1-3.827)^2 \cdot \frac{1}{2} + (4-3.827)^2 \cdot \frac{1}{4} + (7-3.827)^2 \cdot \frac{13}{4} + (22-3.827)^2 \cdot \frac{2}{13} \\
\approx \$18.835 \quad \text{(18.354 if use } \mu = 199 \text{)}
\]

\[
\text{SD}(X) = \sqrt{\text{Var}(X)} \approx \$4.34
\]

2. (3 points) Assume there is no fee to play. If you play the game each day of the week (7 days/week), what do you expect your weekly earnings to be? What is the standard deviation of the weekly totals?

We are exploring \( X_1 + X_2 + \cdots + X_7 \).

\[
E(X_1 + X_2 + \cdots + X_7) = 7E(X) = \$26.789
\]

\[
\text{Var}(X_1 + X_2 + \cdots + X_7) = \text{Var}(X_1) + \cdots + \text{Var}(X_7) \quad \text{Given } X_1, \ldots, X_7 \text{ independent.}
\]

\[
= 7(18.835) = 131.845 \quad \$^2
\]

\[
\text{SD}(X_1 + X_2 + \cdots + X_7) = \sqrt{\text{Var}(X_1 + \cdots + X_7)} = \$11.482
\]