FEBRUARY REGIONAL STATISTICS TEAM ROUND--CONDENSED

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 1

The following information is known about two independent sets of data *a* and *b*:

$\mu_a = 69, \sigma_a = 8, \mu_b = 89, \sigma_b = 15$

Let A be the mean of $3a + b$

Let B be the standard deviation of $2a + 2b$

Let C be the variance of $4a - 3b$

Let D be the standard deviation of $5a + 2b$

Find A + B + C + D.

P. S. Pay attention to the hints in questions 2 - 4. You may find them helpful for later questions.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 2

Sophia and Liz are playing Scrabble®. Suppose that over the course of one game, which happens to last fifteen turns, Sophia scores an average of 20 points per turn with a standard deviation of 4.9 points, and Liz scores an average of 24 points per turn with a standard deviation of 6.6 points. Assume their points per turn are approximately normally distributed. What is the probability that Sophia's overall per-turn average (they play Scrabble® a lot) is higher than Liz's? Round to 3 decimal places.

By the way, the standard deviation of a random variable *X* can be written as $\sqrt{E\left(X^2\right) - \left(E(X)\right)^2}$. I'd save this, if I were you.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 3

Consider the following probabilities of possible outcomes for the discrete random variable X, where X represents the number of people in a group of eight who like Mr. Snow ☺.

X	1	2	3	4	5	6	7	8
P(X)	0.1	0.15	0.25	0.15	0.1	0.1	0.05	0.1

Let A be the square root of the weighted average of the square of the distance of each data point from the mean.

Let B be the probability of rolling a prime number on a fair dodecahedral die and an even number of people liking Mr. Snow. Assume these two events are, unbelievable though it may sound, independent of one another.

Let C be the value of A rounded to 2 decimal places.

Find the exact value of $\left|C - \frac{1}{B}\right|$.

By the way, $\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$, $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$, and $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. Save this.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 4

Let A be the sum of the *x*-coordinates of all possible points on the curve $a(x) = 2 - x^2$ such that the rectangular region with one vertex at the origin and one vertex on $a(x)$ in the first quadrant is a uniform probability density curve. **EXACT ANSWER REQUIRED**.

Let B and C be the values of k and n which will make $r(x)$ a **continuous** probability density curve given that

$$r(x) = \begin{cases} x & 0 < x < 1 \\ 2k - kx & 1 \le x < n \\ 0 & \text{otherwise} \end{cases}$$

Find A + B + C.

P.S. A triangular number is a number of the form $\frac{n(n+1)}{2}$, where *n* is a natural number.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 5

Let the standard deviations of two random variables M and N be σ_M and σ_N. Suppose it is known that $\sigma_{M+N} < 5$. Let $x = \sigma_M$ and $y = \sigma_N$ in the Cartesian plane. What is the area bound by the set of all possible variances σ^2_{M+N} and standard deviations σ_M and σ_N?

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 6

Suppose Kobe Bryant passes the ball to his teammates in 30% of games (the actual figure may be lower; who knows).

Let A be the probability that he passes the ball to his teammates in exactly 7 out of 10 games.
Let B be the probability that he passes the ball to his teammates in at least one game of 10.
Let C be the probability that he passes the ball to his teammates in 90% or higher of 10 games.
Let D be the probability that he passes the ball to his teammates in 40 to 60 percent (inclusive) of 10 games.

Find A + B + C + D to 3 decimal places.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 7

I roll a biased six-sided die k^2 times (where k is a positive integer). The probability of attaining an even number on any roll is twice the probability of attaining an odd number. If I roll an even number, I place half that number in a random empty position in a k x k matrix; if I roll an odd number, I place three times that number in a random empty position of this matrix. What is the expected value of the trace (sum of the elements in the main diagonal) of this matrix, in terms of k?

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 8

Consider a standard 52-card deck. In a game of poker, Sida, quite the corrupt dealer, rigs the deck such that his hand will consist of exactly two aces (no more, no less). Assume that only Sida and Lisa are left. What is the probability that both Sida and Lisa will get a full house in a five-card hand (a full house is 3 of one card and 2 of the other—i.e. 3 kings and 2 sixes)? Round your answer to 7 decimal places.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 9

Define a Raghunandan number to be a triangular number that's multiplied by 4 if it's square and by 2 otherwise (for example, $R_1 = 4$, $R_2 = 6$, $R_3 = 12$, and so on). Letting R_n be the nth Raghunandan number, consider sets $S = \{R_1, R_2, R_3, ..., R_{50}\}$ and $T = \{1, 2, 3, ..., 50\}$, where T is the set of all possible outcomes of a random variable V, with each outcome having the same probability of occurring. Hint: S has three square numbers. One of them comes from the 1st triangular number, one from $\{T_{6,} T_7, ..., T_{10}\}$, and one from $\{T_{41,} T_{41}, ..., T_{50}\}$

Let A be the mean of set S, to the nearest integer.
Let B be the median of set S.
Let C be the variance of the random variable V.
Let D be the expected value of the random variable V.

Find A + B + C + D^2, to the nearest integer.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 10

Let P(X) and P(Y) denote the probabilities of the five possible outcomes for values of the discrete random variables X and Y, respectively. Given the following tables:

X	1	2	3	4	5
P(X)	0.1	0.2	0.1	b	a

Y	1	2	3	4	5
P(Y)	0.1	a	b	$2b$	0.2

Find the value of $a - b$.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 11

We know that the standard error s is a decent approximation for the population standard deviation σ. Unfortunately, as we also know, s is not perfectly unbiased; there is, however, a way to get an unbiased estimator! There's a constant c_n which depends on the value of the sample size n, such that $s = c_n\sigma$. The formula for c_n is

$c_n = \left(\sqrt{\frac{2}{n-1}}\right)\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)$. Given that $\Gamma(4.5) \approx 11.63$ and $\Gamma(5) = 24$:

I ask ten people how many loaves of bread they buy weekly and get the responses 3, 7, 1, 1, 2, 3, 2, 8, 8, 3. Using this data, find an unbiased estimator for the population standard deviation. Round to 2 decimal places, using the given values of $\Gamma(4.5)$ and $\Gamma(5)$.

DON'T WASTE YOUR TIME; READ THIS AFTER YOU'VE SOLVED THE PROBLEM.
You don't need this to solve the problem, as I've given you the values of the gamma function you'll need, but for those of you with knowledge of calculus, the gamma function is defined as: $\Gamma(n) = \begin{cases} (n-1)! & x \in Z^+ \\ \int_0^\infty t^{n-1}e^{-t}\,dt & \text{otherwise} \end{cases}$

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 12

The information entropy is the measure of uncertainty associated with a random variable. For a discrete random variable X that can take possible values/outcomes of $\{x_1, x_2, ..., x_k\}$, the entropy associated with it is given by

$\text{I(X)} = \sum_{m=1}^{k}\left(P(x_m)\log_2\left(\frac{1}{P(x_m)}\right)\right)$

Let A be the information entropy of a roll of a die which is rigged such that prime numbers are twice as likely to show as composite numbers, to 2 decimal places.
Let B be the information entropy of the toss of a coin which has the tendency to land heads-up 60% of the time, to 2 decimal places.
Let the random variable Y represent the number of people in a group of five who will obey Mr. Snow's instructions. If the probability distribution for this variable is given by

Y	0	1	2	3	4	5
P(Y)	k +0.04	k + 0.03	k + 0.02	k + 0.01	k	0

then let C be the information entropy of Y, to 2 decimal places.

Find A + B + C.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 13

Chenyu thinks there's some type of relation between a person's amount of peaceful sleep at night and their calorie consumption at dinner. Letting x = peaceful sleep (in hours) and y = calorie consumption at dinner, consider the data points $(x, y) = (3,55)$, $(4,161)$, $(5,488)$, and $(6,1460)$.

Let A be the value corresponding to the transformation which would give the highest correlation coefficient, out of:
(2) x vs. $\sqrt[3]{y}$ (3) y vs. 3^x (4) x vs. ln(y) (5) y vs. $3x^2$

Let B and C be the slope and y-intercept, **respectively**, of the least-squares regression line generated for this set of data, each rounded to one decimal place then multiplied by 10. Assume x to be the explanatory variable.

Let D and E be the slope and y-intercept of the least-squares regression line generated from the transformation you picked in part (A), each rounded to one decimal place then multiplied by 10. Again, assume the transformed (or not, if you picked 2 or 5) x-values to be the explanatory variable.

Find $\frac{\log_2\left(A^{4B+C}\right)}{D+E}$ to the nearest integer.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 14

Let A be the variance of the first ten square numbers.
Let B be the variance of the first ten whole numbers.
Let C be the variance of the first ten odd positive numbers.
Let D be the variance of the first ten even positive numbers.
Find $\frac{AB}{CD}$ to 2 decimal places.

FEBRUARY REGIONAL STATISTICS TEAM QUESTION 15

Brandon conducts an experiment consisting of a measurement taken once a day for 11 days. Because it's summer and he's bored, he conducts it twice, and gets as his measurements $\{1,6,1,8,0,3,3,9,8,8,7\}$ and $\{3,1,4,1,5,9,2,6,5,3,5\}$ (assume these two samples are completely independent of one another). Unbeknownst to Brandon, Craig conducts this experiment and gets as his measurements $\{2,7,1,8,2,8,1,8,2,8,4\}$. (All three sets are arranged in the same order of days; that is, the k^{th} value of each set, for $1 \leq k \leq 11$, corresponds to the same day)
Define:
Set X to be the set of Brandon's measurements which are the first 11 digits of $\phi = \frac{1+\sqrt{5}}{2}$.
Set Y to be the set of Brandon's measurements which are the first 11 digits of π.
Set Z to be Craig's measurements.

Let A = the *p*-value of the hypothesis test $H_0:$ Set X = Set Z vs. $H_a:$ Set X < Set Z, to 2 decimal places.
Let B = the *p*-value of the hypothesis test $H_0:$ Set Y = Set Z vs. $H_a:$ Set Y $\neq$ Set Z, to 2 decimal places.

Find A + B.