October 7, 2008

Math 102 / Core 143 BX — Exam I

Show all work clearly for partial credit. An unsimplified answer like $12\sqrt{3.51} + 6/7$ is usually worth more than 23.3, because it is easier to understand where it came from. If your calculator has buttons for statistical functions like average or standard deviation, do not use them for the computations below.

1. *(20 points)* For a list of numbers consisting of 6 ones, 3 twos, and 1 eighteen, find the
   (a) average (b) standard deviation (c) median
   (d) 20-th percentile (e) IQR (interquartile range)

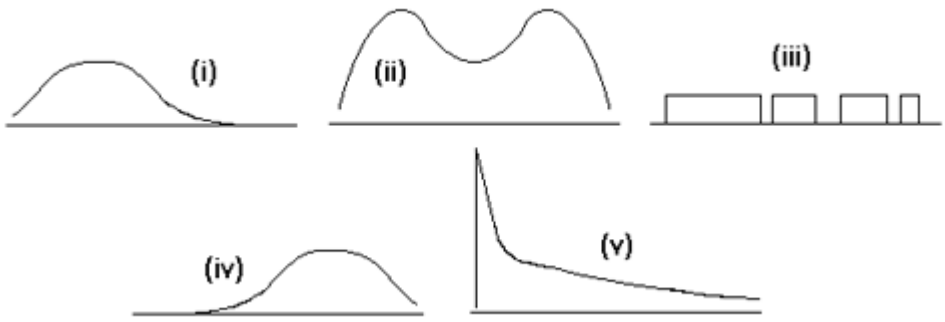
2. *(16 points)* Which of these histograms represent each of the following lists of data? You may use any histogram more than once or not at all. If you are not sure, or if you cannot find one that looks right, feel free to sketch your own and explain why it is better than the given histograms.

   (a) jersey numbers of Colgate football players in the 2005 season

   (b) heights of people at a party of Big Brothers/Sisters and their Little Brothers/Sisters

   (c) the numbers of shirts owned by various men

   (d) scores on an easy statistics exam



3. *(15 points)* The coordinates of a data point are described. If many data points of each kind are collected, would the correlation coefficient be closest to -1, -0.5, 0, 0.5, or 1?

   (a) A person's birthdate (number of days into the year) and his/her IQ.

   (b) The average number of hours per day a person spends watching TV and the hours he/she spends reading.

   (c) A child's height measured in inches and the same child's height measured in centimeters.

4. *(30 points)* A side effect of a certain drug is to change blood pressure. Dosage of the drug and systolic blood pressure are recorded for a large number of patients, and the following "summary data" is found:

	dose (mg)	pressure (mmHg)	
AV	40	120	$r = 0.8$
SD	20	15	

On the basis of this data, estimate:

   (a) the systolic blood pressure of a randomly chosen patient (with no information on his dosage).

   (b) the likely error in the estimate found in (a)

   (c) the systolic blood pressure of a patient whose dosage is 60 mg.

   (d) the dosage of a patience whose systolic blood pressure is measured as 110 mmHg.

   (e) the likely error in the estimate found in (d).

5. *(9 points)* Every 6th-grade class in an elementary school must take an achievement test. Here are the percentages of boys and girls who passed in each class:

	Ms. A	Mr. B	Ms. C	Mr. D
boys	20%	10%	60%	40%
girls	30%	20%	70%	50%

Which of the following statements must be true on the basis of this information? (Neither, one or both may be necessarily true.) Explain your answers.

   (a) The percentage of girls in that school who passed is larger than the percentage of boys who passed.

   (b) The percentage of girls who passed must be between 20% and 70%.

6. *(10 points)* According to the article "Coddling Human Guineas Pigs" by Sharon Begley, for what kind of research is it more difficult to get funding from the National Institutes of Health, and what two reasons are given for it?

Some possibly useful formulas:

$$\sqrt{\text{average of } (x-\overline{x})^2} \qquad \sigma_y \cdot \sqrt{1-r^2}$$

$$r = \text{average of } ((x \text{ in std units}) \cdot (y \text{ in std units}))$$

$$z = \frac{x-\overline{x}}{\sigma_x} \qquad x = z \cdot \sigma_x + \overline{x}$$

$$y - \overline{y} = (\text{sign of } r)\frac{\sigma_y}{\sigma_x}(x-\overline{x}) \qquad y - \overline{y} = r\,\frac{\sigma_y}{\sigma_x}(x-\overline{x})$$

Normal table (Area between $-z$ and z)									
z	Area(%)	z	Area(%)	z	Area(%)	z	Area(%)	z	Area(%)
0.0	0.0	0.9	63.19	1.8	92.81	2.7	99.31	3.6	99.968
0.05	3.99	0.95	65.79	1.85	93.57	2.75	99.4	3.65	99.974
0.1	7.97	1	68.27	1.9	94.26	2.8	99.49	3.7	99.978
0.15	11.92	1.05	70.63	1.95	94.88	2.85	99.56	3.75	99.982
0.2	15.85	1.1	72.87	2	95.45	2.9	99.63	3.8	99.986
0.25	19.74	1.15	74.99	2.05	95.96	2.95	99.68	3.85	99.988
0.3	23.58	1.2	76.99	2.1	96.43	3	99.73	3.9	99.99
0.35	27.37	1.25	78.87	2.15	96.84	3.05	99.771	3.95	99.992
0.4	31.08	1.3	80.64	2.2	97.22	3.1	99.806	4	99.9937
0.45	34.73	1.35	82.3	2.25	97.56	3.15	99.837	4.05	99.9949
0.5	38.29	1.4	83.85	2.3	97.86	3.2	99.863	4.1	99.9959
0.55	41.77	1.45	85.29	2.35	98.12	3.25	99.885	4.15	99.9967
0.6	45.15	1.5	86.64	2.4	98.36	3.3	99.903	4.2	99.9973
0.65	48.43	1.55	87.89	2.45	98.57	3.35	99.919	4.25	99.9979
0.7	51.61	1.6	89.04	2.5	98.76	3.4	99.933	4.3	99.9983
0.75	54.67	1.65	90.11	2.55	98.92	3.45	99.944	4.35	99.9986
0.8	57.63	1.7	91.09	2.6	99.07	3.5	99.953	4.4	99.9989
0.85	60.47	1.75	91.99	2.65	99.2	3.55	99.961	4.45	99.9991

Math 102 / Core 143 — Solutions to Exam I

1. (a) $(6(1) + 3(2) + 18)/(6 + 3 + 1) = 3$
   (b) $\sqrt{(6(1-3)^2 + 3(2-3)^2 + (18-3)^2)/10} \approx 5.02$
   (c) 1, because over half the values are less than or equal to it.
   (d) Also 1.
   (e) $2 - 1 = 1$

2. (a) (iii) — No jersey number is used more than once.
   (b) (ii) — The Big Brothers/Sisters have heights in the right high point, the Little Brothers/Sisters in the left one (though that one should really be broader and lower, because there is probably more variability in the children's heights).
   (c) (i) — Most men have a few shirts, where "few" can cover a broad range, say 10 to 30. But a few men have many more.
   (d) (iv) — Exam scores are usually between 70 and 90, with a few at the extremes. But there is less room for a right tail than for a left one, because scores cannot exceed 100%.

3. (a) 0 — There is no reason to assume an association between birthday and IQ.
   (b) $-.5$ — More time spent on TV leaves less time for reading; but the correlation is far from perfect, because there are so many other things on which time could be spent.
   (c) 1 — Except for small roundoff errors, the two measurements should give identical lists of numbers in standard units.

4. (a) 120 mmHg — guess the average.
   (b) The standard deviation of pressure, 15 mmHg.
   (c) Because his dosage is $(60 - 40)/20 = 1$ standard deviation above average, we project his blood pressure as $0.8(1)$ standard deviation above average: $120 + (0.8)(15) = 132$ mmHg.
   (d) Because his blood pressure is $(110 - 120)/15 = -2/3$ standard deviation above average (i.e., two-thirds of a standard deviation below average), we project his dosage as $0.8(-2/3)$ standard deviations above average, i.e., $40 + (0.8)(-2/3)(20) \approx 29.3$ mg.
   (e) The RMS error for regression: $20\sqrt{1 - (0.8)^2} = 12$ mg.

5. (a) This need not be true. Depending on the relative sizes of the classes and the numbers of boys and girls in each, Simpson's paradox may apply, giving a different result.
   (b) This must be true: Even a weighted average must lie between the highest and lowest of the values being averaged.

6. It is difficult to get funding for clinical research on human subjects, as compared to more "original" research on animals. The two reasons given in the article are: (1) Scientists who conduct original research on animals think that clinical studies on humans, which must be based on earlier studies on animals, are not "innovative enough" to receive funding. (2) "Internal review boards," who review research on human subjects at various universities and other institutions, are often so concerned about ill treatment of the subjects (despite Begley's opinion that they should expect some rough going as human lab rats), that the research is often limited or blocked. Knowing this, reviewers for external funding agencies often give lower ratings to proposals for research on humans, so that such research does not receive funding.