October 9, 2007

Math 102 / Core 143 DX — Exam I

Put your answers to all the problems on the numbered sheet, except for the essay (#4), which goes on the back. The blue book is for your scratch work and will NOT be graded.

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

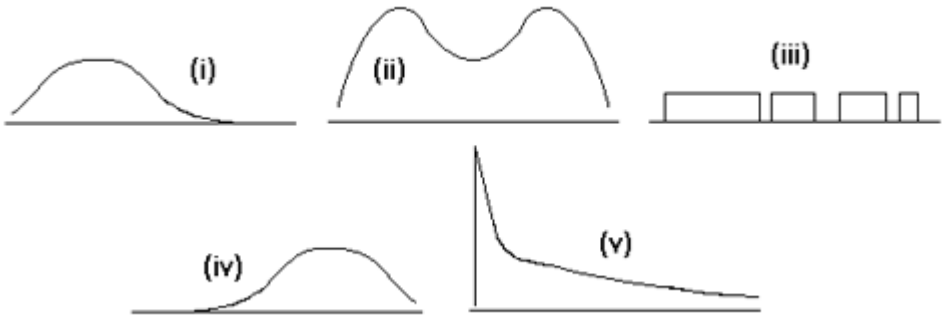
2. *(16 points)* Which of these histograms represent each of the following lists of data? You may use any histogram more than once or not at all. If you are not sure, or if you cannot find one that looks right, feel free to sketch your own and explain why it is better than the given histograms. (A new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

   (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) numbers of shirts owned by various men

   (d) scores on an easy statistics exam



3. *(18 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. *(10 points)* Relative the article, "After the Bell Curve", by David L. Kirp: How does the paragraph on adoptions between different social classes give evidence for Kirp's argument? How does the same data give evidence against his argument?

5. *(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)	
average	40	120	$r = 0.8$
standard deviation	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).

6. *(6 points)* Suppose that in every county in the State of New York the employment rate (fraction of potential workers who have jobs) is higher for women than for men. Are the following statements true of false?

   (a) The employment rate of men for the state as a whole must be between the highest and lowest county rates for men.

   (b) The employment rate of men for the state as a whole is lower than the employment rate of women.

Some possibly useful formulas:

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

$z = \dfrac{x-\overline{x}}{\sigma_x}$ $\qquad$ $x = z \cdot \sigma_x + \overline{x}$ $\qquad$ $y - \overline{y} = (\text{sign of } r)\dfrac{\sigma_y}{\sigma_x}(x-\overline{x})$ $\qquad$ $y - \overline{y} = r\,\dfrac{\sigma_y}{\sigma_x}(x-\overline{x})$

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

Name: _____

Answers to Exam I

1. (a)
   (b)
   (c)
   (d)
   (e)
2. (a)
   (b)
   (c)
   (d)
3. (a)
   (b)
   (c)
4. On back.
5. (a)
   (b)
   (c)
   (d)
   (e)
6. (a)
   (b)

4.

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
   (d) 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 reasonably small number of shirts, with some variation; but a few men have a very large number (and no one has a negative number).
   (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 probably no association between birthdate 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. Kirp's claim that home environment is crucial in determining IQ is supported by the fact that, no matter what kind of family they came from, children adopted by more well-to-do people had an average higher IQ than those adopted by working-class parents (for children born to well-to-do parents, 119.6 vs. 107.5, and for children born to working-class parents, 103.6 vs. 92.4 — but I don't expect you to quote figures). But the claim is rebutted by the fact that children born to working-class parents but adopted by well-to-do parents still had average IQ lower than children born to well-to-do parents and adopted by working-class parents (103.6 vs. 107.5).

5. (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 $0.8(-2/3)$ SD's 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.

6. (a) True: Even a weighted average (with nonnegative weights that add up to 1) must fall between the highest and lowest values being averaged.
   (b) False: Simpson's paradox may apply.