Final for Statistics 114
Elements of Probability and Statistics - Fall 1997
Material Covered: entire course
For: 17th December for Division 1; 16th December for Division 2

This is a 2 hour final, worth 34% and marked out of 34 points. The total possible points awarded for each question is given in square brackets at the beginning of each question. Anything that can fit on two sides of an $8\frac{1}{2}$ by 11 inch piece of paper may be used as a reference during this final. A calculator and appropriate statistical tables may also be used. No other aids are permitted.

1[1]**.** An experiment was carried out to determine the effect of providing free milk to school children in Lanarkshire, Scotland. The children were assigned at random either to receive or not to receive milk.

(a) [2] Assign a sample of eight children to either receiving or not receiving milk by filling in the following table by circling "yes" or "no" for the eight children. Use the *third* row of the random numbers table, start at column 1 and proceed left to right. Numbers $0, 1, 2, 3, 4$ represent a child receiving milk; numbers $5, 6, 7, 8, 9$ represent a child not receiving milk.

child	1	2	3	4	5	6	7	8
receives milk?	**yes/no**	**yes/no**	**yes/no**	**yes/no**	**yes/no**	**yes/no**	**yes/no**	**yes/no**

(b) [1] This is a (circle one) **controlled experiment / observed study**

because: _____.

(c) [1] If the children were allowed to decide whether they wanted milk or not, this would become a **controlled experiment / observed study**

because: _____.

[1]Freedman et. al, 5, p 23, 1993.

2. Consider the following three tables of data which concern whether or not patients improved or not using either drug 1 or drug 2. The first table summarizes the data combined from two different hospitals. The second and third tables summarizes this data separately for each of the two hospitals. For example, 5,450 people improved when taking drug 1 in the combined hospital study.

		drug 1	drug 2	subtotals
improved?	yes	5,450	8,500	13,950
	no	4,550	1,500	6,050
	subtotals	10000	10,000	20,000

	hospital A	drug 1	drug 2	subtotals
improved?	yes	950	8100	9050
	no	50	900	950
	subtotals	1,000	9,000	10,000

	hospital B	drug 1	drug 2	subtotals
improved?	yes	4,500	400	4,900
	no	4,500	600	5,100
	subtotals	9,000	1,000	10,000

(a) [1] The combined study tells us (circle one)
drug 1 is better / drug 2 is better / both drugs are equally as good.

(b) [1] The hospital A study tells us (circle one)
drug 1 is better / drug 2 is better / both drugs are equally as good.

3. Consider the histogram of the family income for a wealthy suburb below. Based on this histogram, answer the questions below.

(a) [2] The percentage of the families in this suburb that had incomes between \$90,000 and \$110,000 a year is closest to:

(i) 100% (ii) 27% (iii) 50% (iv) 21% (v) 0%

(b) [1] The percentage of the families in this suburb that had incomes exactly equal to \$100,000 a year is closest to:

(i) 100% (ii) 27% (iii) 50% (iv) 21% (v) 0%

(c) [1] The percentage of the families in this suburb that had incomes greater than \$50,000 a year is closest to:

(i) 100% (ii) 27% (iii) 50% (iv) 21% (v) 0%

4[2]**.** Answer true (T) or false (F):

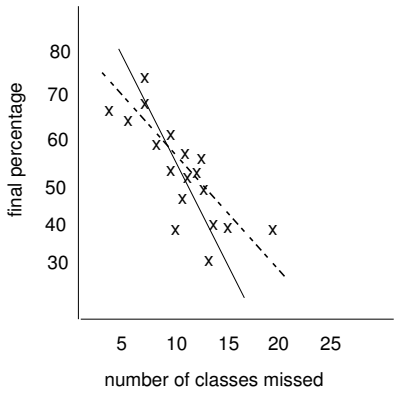
(a) [1] **T** / **F** The median and the average of a list of numbers are always close together.

(b) [1] **T** / **F** Half of a list of numbers is always below average.

(c) [1] **T** / **F** With a large representative sample, the histogram is bound to follow the normal curve quite closely.

---

5. Consider, below, the scatter plot of the percentage obtained on the final exam versus the number of classes missed for a group of statistics students. Assume "final percentage" is the dependent variable.



Based on this scatter plot, with the dotted line and solid line, answer the following questions.

(a) [1] The dotted line is the:

(i) SD line
(ii) regression line
(iii) $r \times$ least squares line
(iv) average of final percentage
(v) unknown, because we are not able to determine this from the graph

(b) [1] If $SD_y = 10$ and $SD_x = 2.5$, *approximate* slope of SD line is closest to:

(i) -10 (ii) -4 (iii) -0.2 (iv) -1 (v) not able to tell

(c) [1] The *approximate* slope of the regression line is closest to:

(i) -3 (ii) -1 (iii) -0.2 (iv) -0.5 (v) not able to tell

[2]Freedman et al., 11, p 88, 1993.

6. In a study of the sorghum yield, y (tons), versus the distance upslope the field is from water, x (feet), the scatter plot of data is found to be football shaped. The summary statistics are given by:

average sorghum yield, $y = 400$	SD ≈ 100	
average distance upslope, $x = 100$	SD ≈ 30	$r \approx -0.7$

(a) [1] The percentage of sorghum yield over 450 tons is closest to:

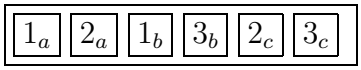
(i) 61% (ii) 31% (iii) 57% (iv) 0.5% (v) 38%

(b) [3] Of the sorghum fields located 100 feet upslope from water, the percentage of sorghum yield over 450 tons is closest to:

(i) 17% (ii) 31% (iii) 42% (iv) 24% (v) 5%

---

7. A box has six tickets. Each ticket possesses one of the numbers 1, 2 or 3 and one of three letter subscripts: a, b or c.



Two tickets are randomly drawn (with replacement and where order matters) from the box. [HINT: Use Venn diagrams.]

(a) [1] The chance that the "one of the two tickets is a 1 *and* one of the two tickets is a 2" is closest to:

(i) $\frac{20}{36}$ (ii) $\frac{8}{36}$ (iii) $\frac{9}{36}$ (iv) $\frac{10}{36}$ (v) $\frac{11}{36}$

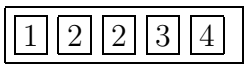
(b) [1] The chance that the "one of the two tickets is a 1 *or* one of the two tickets is a 2" is closest to:

(i) $\frac{9}{36}$ (ii) $\frac{3}{36}$ (iii) $\frac{16}{36}$ (iv) $\frac{17}{36}$ (v) $\frac{32}{36}$

(c) [1] The chance that the "one of the two tickets is a 1 *and* one of the two tickets is a 2", given that (or conditional on) "one of the two tickets is a 1 *or* one of the two tickets is a 2" is closest to:

(i) $\frac{8}{32}$ (ii) 1 (iii) $\frac{16}{36}$ (iv) $\frac{16}{20}$ (v) $\frac{32}{8}$

8. A box has five tickets, as follows.



(a) [1] Four tickets are drawn at random (with replacement and where order matters) from the box. Determine the chance that the "ticket 2 is drawn four times".

(b) [1] Four tickets are drawn at random (with replacement and where order matters) from the box. Determine the chance that the "ticket 2 is drawn at least once".

(c) [1] Four tickets are drawn at random (with*out* replacement and where order matters) from the box. Determine the chance that the "ticket 2 is drawn at least once".

9. In a town of 10,000 people, 35% are enrolled in college. A simple random sample of 250 such persons is drawn and it found 39 of them are enrolled in college.

(a) [1] The correction factor, under a sampling with replacement procedure, is about the same as the correction factor, under a sampling with*out* replacement procedure, because:

(i) number of draws equals number of tickets
(ii) 250 is so much smaller than 10,000
(iii) 35% is so much smaller than 100%
(iv) 39 is so much smaller than 250
(v) 39 is so much smaller than 10,000

(c) [3] The chance that exactly 45% of this sample are enrolled in college is closest to:

(i) 1% (ii) 0.99% (iii) 0.03% (iv) 0.08% (v) 0.47%

(c) [1] The 95% confidence interval for the percentage of people enrolled in college is closest to:

(i) $16\% \pm 3\%$ (ii) $16\% \pm 6\%$ (iii) $35\% \pm 3\%$ (iv) $3\% \pm 6\%$ (v) $2.5\% \pm 3\%$

---

10. [4] A paper on the amount of selenium content in orchard leaves reported the following five values:

0.072, 0.073, 0.080, 0.078 0.088

Use a t test at level 0.01 to test the hypothesis "null: average = 0.08" versus "alternative: average < 0.08".

(1) (a) **no, yes, no, no, yes, no, no, yes** (b) **controlled** (c) **observed**

(2) (a) **drug 2 is better** (b) **drug 1 is better**

(3) (a) **(ii) 25%** (b) **(v) 0%** (c) **(i) 100%**

(4) (a) **F** (b) **F** (c) **F**

(5) (a) **(ii) regression line** (b) **(ii)** $\mathbf{-4}$ (c) **(i)** $\mathbf{-3}$

(6) (a) **(ii) 31%** (b) **(iv) 24%**

(7) (a) **(ii)** $\frac{8}{36}$ (b) **(v)** $\frac{32}{36}$ (c) **(i)** $\frac{8}{32}$

(8) (a) $\left(\frac{2}{5}\right)^4 \approx 0.0256$ (b) $1 - \left(\frac{3}{5}\right)^4 \approx 0.8704$ (c) **1**

(9) (a) **(ii)** (b) **(iv) 0.09%** (c) **(ii)** $16\% \pm 6\%$

(10) $(25\%, 50\%)$