Math 10 Practice Final 2

June 3, 2011

1. (a) [5 points] The heights of 10 male Dartmouth undergraduates were collected and the results are shown below (you may assume these students were selected by a simple random sample of all male undergraduates). Find the average, median and standard deviation of the heights.

$$heights: \; 71 \quad 72 \quad 68 \quad 73 \quad 70 \quad 71 \quad 69 \quad 68 \quad 72 \quad 76$$

   (b) [5 points] Assuming there are 2000 male undergraduates, determine a 95% confidence interval for the average height of all male undergraduates.

(c) [5 points] The weights of the same subjects were also collected, and are given in the table below. Find the correlation coefficient for these heights and weights. (The average of the weights is 175 lbs. and the SD of the weights is approximately 18.79 lbs.)

heights :	71	72	68	73	70	71	69	68	72	76
weights :	177	181	153	189	169	173	161	149	185	213

(d) [5 points] Now find the equation for the regression line predicting height from weight and use it to predict the height of someone who is 165 lbs.

(e) [5 points] Oops! The scale was miscalibrated by 5 lbs.; each weight should actually be 5 lbs. less than what is recorded above.Without recalculating, how would you correct the average of the weights, their standard deviation and the correlation coefficient for the heights and weights?

(f) [5 points] 1 inch is approximately 2.54 cm and 1 lb. is approximately 0.45 kg. If we used these conversion factors to convert the heights and weights from part (c) into metric units, explain how the averages, standard deviations and correlation coefficient would change. (*Hint: You should not have to recompute all of these.*)

2. A simple random sample of 125 graduating Dartmouth students were surveyed. Each student was asked whether they were an undergraduate or graduate student and whether or not they planned to walk at graduation. The number of undergraduates was 100 and the number of graduate students was 25. Of these 125 graduates, 75 said they were planning to walk while 50 said they were not.

   (a) [5 points] We would like to determine whether there is any relationship between a student's status (grad or undergrad) and their decision on whether to walk at graduation. Formulate a null hypothesis and compute the expected values under this hypothesis. (Draw some sort of table, if that helps you.)

   (b) [5 points] If I tell you that 14 graduate students reported that they would *not* walk at graduation, is this enough information to determine whether there is a relationship? Why or why not?

(c) [6 points] Now, if I give you the full response numbers (64 undergraduates walking; 36 undergraduates not walking; 11 graduate students walking; and 14 graduate students not walking), determine the appropriate test and the corresponding p-value, and use this to decide whether there is a relationship between these variables. Show your work and indicate some reasoning behind your conclusion.

3. [15 points] A study of 5,000 babies looked at the relationship between their weight at birth and the age at which they first slept through the night. The birth weights averaged out to 90 oz. with an SD of 15 oz. The ages at which they first slept all night averaged 50 days with an SD of 10 days. The correlation between the two variables was -0.60. If we draw a sample of 25 babies who weighed 105 oz. at birth, what is the probability that the average age at which these 25 babies sleep through the night is less than 40? (You may assume that all distributions involved are normal and that the sample was a simple random sample.)

4. [6 points] The US Department of Justice made a study of civil jury cases in state courts in the nation's 75 largest counties. In these courts, during the year ending in June 1992, juries gave money damages to plaintiffs in 5,949 cases. The median amount was $50,000 and the average was $450,000. Percentiles were computed for this distribution. Consider the following intervals:

   (i) the interval between the 50th %-ile and the 0th %-ile

   (ii) the interval between the 100th %-ile and the 50th %-ile

   Which of the following statements is true? (choose one)

   (a) Interval (i) is bigger.

   (b) Interval (ii) is bigger.

   (c) Intervals (i) and (ii) are about the same size.

   (d) It is impossible to tell with the information we have been given.

5. [6 points] A large sample of children was followed over time. One investigator looked at all the children who were at the 90th percentile in height at age four. Some of these children turned out to be above the 90th percentile in height at age eighteen, and others were below. Was the number who were above

   (i) quite a bit smaller than

   (ii) about the same as

   (iii) quite a bit larger than

   the number who were below? Or is more information needed? Give a reason to support your answer. (You may assume the data is homoscedastic.)

6. A statistical analysis is made of the midterm and final scores in a large course with the following results (x gives midterm scores while y gives final) $\bar{x} = 60$, $SD_x = 15$, $\bar{y} = 65$, $SD_y = 20$ and $r = 0.50$. The scatter diagram is football shaped.

   (a) [4 points] About what percentage of students scored over 80 on the final?

   (b) [4 points] Of the students who scored 80 on the midterm, about what percentage scored over 80 on the final?

   (c) [3 points] For each student the final score was predicted from the midterm score using the regression line. For about 1/3 of the students, the prediction for the final score was off by more than ______ points. Options:

   6 9 12 15 25

7. [6 points] A hundred draws are made at random with replacement from a box with 5 tickets; three tickets are marked with a 0, one with a 1 and another with a 2. Estimate the chance that 1 turns up on exactly 20 draws. Show your work.

8. A casino offers the following game. The gambler stakes $1 on either spades, hearts, diamonds or clubs. Then a deck of cards is shuffled and the top card is turned over. If this card is of selected suit, the gambler wins $2 (so he gets $3 back, including his own). If it is not of the selected suit, he loses his $1. That ends the game. The card is put back in the deck, and the deck is shuffled for the next round. A gambler plays the game 75 times, betting on spades every time.

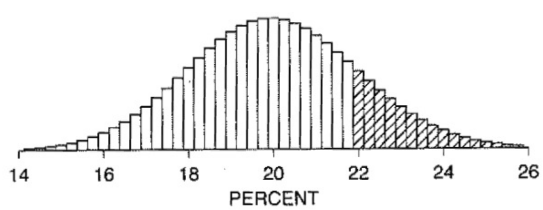
   (a) [4 points] After 75 plays, the net gain of the gambler will be around ______ give or take ______ or so. (fill in the blanks)

   (b) [4 points] The gambler will win $2 on around ______ of the 75 plays, give or take ______ or so. (i.e., how many; fill in the blanks)

   (c) [4 points] The gambler will lose $1 on around ______ of the 75 plays, give or take ______ or so. (i.e., what percentage; fill in the blanks)

9. [6 points] A box contains 6 tickets numbered 1, 2, 3, 4, 5 and 6. Three tickets are drawn without replacement from the box. Find the chance that the three tickets left in the box are 4, 5 and 6.

10. [6 points] At a certain university, 20% of the students have GPAs of 3.5 or better. The registrar's office is going to take a simple random sample of 400 students. Shown below is the probablity histogram for the percentage of students in the sample with GPAs of 3.5 or higher. What does the shaded area represent?



11. [10 points] A certain town has 25,000 families. The average number of children per family is 2.6 with SD 0.80. The distribution is not normal, however; 25% of the families have no children at all. If we draw a random sample of 500 families, what are the chances that between 115 and 135 of the sample families will have no children?

12. [8 points] LA has about four times as many registered voters as San Diego. A simple random sample of registered voters is taken in each city, to estimate the percentage who will vote for school bonds. Other things being equal, a sample of 4,000 taken in LA will be about

    (i) four times as accurate

    (ii) twice as accurate

    (iii) as accurate

    as a sample of 1,000 taken in San Diego. Choose one option and justify your choice.

13. [8 points] One hospital has 218 live births during the month of January. Another has 536. Which is likelier to have 55% or more male births? Or is it equally likely? Explain. (There is about a 52% chance for a live-born infant to be male.)

14. [10 points] One study of grand juries in Alameda County, California, compared the demographic characteristics of jurors with the general population, to see if the jury panels were representative. Here are the results for age. (The first column gives the age range; the second the percentage of the county population within this range; and the third, the number of jurors within this range.)

Age	County-wide %	# of jurors
21 – 40	42	5
41 – 50	23	9
51 – 60	16	19
61 *and up*	19	33
Total	100	66

Were these 66 jurors selected at random from the population of Alameda County (age 21 and up)?

15. [8 points] A study on the relationship between birth order, family size and intelligence was conducted. The subjects consisted of all Dutch men who reached the age of 19 between 1963 and 1966. These men were required by law to take the Dutch army induction tests, includeing an intelligence test. The results showed that for each family size, measured intelligence decreased with birth order: first-borns did better than second-borns; second-borns did better than third-borns, and so on. And for any particular birth order, intelligence decreased with family size: for instance, first-borns in two-child families did better than first-borns in three-child families. These results remained true even after controlling for the social class of the parents. Taking, for instance, men from two-child families:

- the first-borns averaged 2.575
- the second-borns averaged 2.678

(The test scores range from 1 to 6, 1 being best.) The difference is small, but to show that it is real, a two-sample z-test was made. The SD for the test scores was around one point, both for the first-borns and the second-borns, and there were 30,000 of each. So,

$$\text{SE for average} \approx \frac{1}{\sqrt{30000}} \approx 0.006 \text{ points}$$
$$\text{SE for difference} \approx \sqrt{0.006^2 + 0.006^2} \approx 0.008.$$

Therefore, $z = \frac{2.575-2.678}{0.008} \approx -12.6$, and P is nearly 0%. Based on this, the researchers concluded that the difference was real and highly significant, especially considering the large number of cases.

Was it appropriate to make a two-sample z-test in this situation? Answer yes or no, and explain.

16. [8 points] The students in a high school physics class made 25 measurements of the weight of a piece of metal about the size of a nickel. On their first weighing, they got 5.29 grams. On their next one, they got 5.36 grams. On the third, they got 5.37 grams, and so on. All of the measurements are shown below, *in the order in which they were collected* (the number listed next to each measurement indicates the order of collection).

1	5.29	6	5.47	11	5.58	16	5.59	21	5.49
2	5.36	7	5.49	12	5.64	17	5.56	22	5.48
3	5.37	8	5.50	13	5.72	18	5.53	23	5.47
4	5.41	9	5.52	14	5.64	19	5.52	24	5.39
5	5.44	10	5.54	15	5.62	20	5.50	25	5.37

If appropriate, find an approximate 95% confidence interval for the weight of the piece of metal, using the fact that the average of the measurements is 5.50 grams and the SD is 0.10 grams. If it is not appropriate, explain why not.

17. Freshmen at public universities work 10.2 hours a week for pay, on average, with an SD of 7.6 hours; at private universities, the average is 9.2 hours, with an SD of 6.4 hours. Assume these data are based on two independent simple random samples, each of size 1,000. Determine whether the difference between the averages is due to chance by completing parts (a)–(c). Record your conclusion in part (d).

    (a) [4 points] Formulate the null hypothesis as a statement about a box model.

    (b) [4 points] Repeat part (a) for the alternative hypothesis.

    (c) [6 points] Calculate the appropriate test statistic.

    (d) [4 points] What do you conclude?

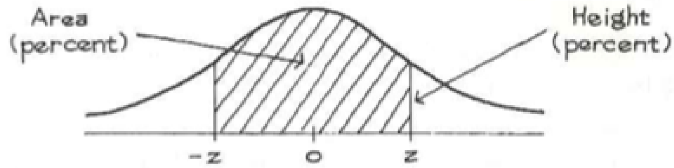
18. A laboratory makes 25 repeated measurements on the molecular weight of a protein. The average is 119 kilo-Daltons and the SD is 15. The lab now wants to estimate the likely size of certain chance errors. Fill in the blanks. Show your work below.

    (a) [5 points] The chance error in one measurement is about ______.

    (b) [5 points] The chance error in the average of the measurements is about ______.

19. [6 points] A die is rolled three times. Find the probability that an ace comes up on one or two of the rolls, but not all three.

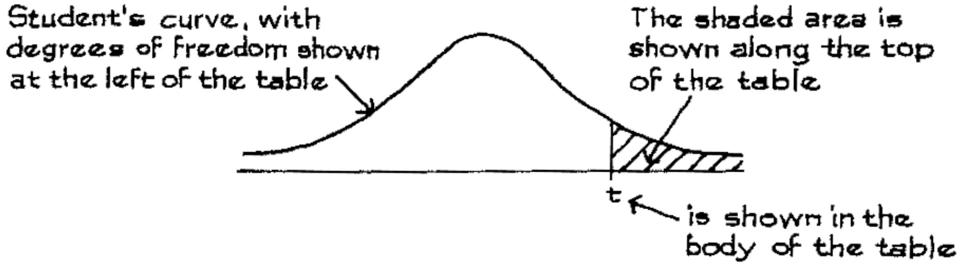
Tables



A NORMAL TABLE

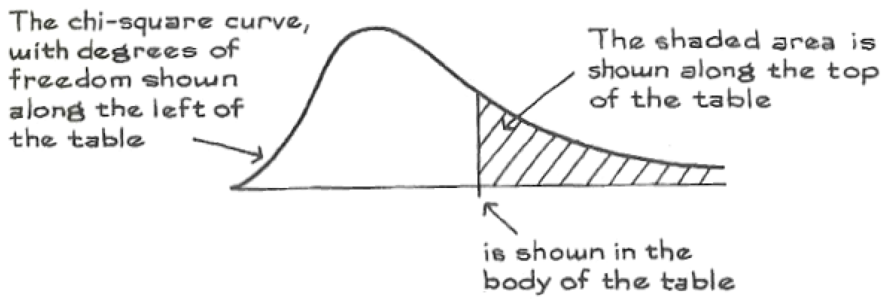
z	*Height*	*Area*	*z*	*Height*	*Area*	*z*	*Height*	*Area*
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	1.98	98.57	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973
1.25	18.26	78.87	2.75	0.91	99.40	4.25	0.005	99.9979
1.30	17.14	80.64	2.80	0.79	99.49	4.30	0.004	99.9983
1.35	16.04	82.30	2.85	9.69	99.56	4.35	0.003	99.9986
1.40	14.97	83.85	2.90	0.60	99.63	4.40	0.002	99.9989
1.45	13.94	85.29	2.95	0.51	99.68	4.45	0.002	99.9991

A t-TABLE



Degrees of freedom	*25%*	*10%*	*5%*	*2.5%*	*1%*	*0.5%*
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.96	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.60
5	0.73	1.48	2.02	2.57	3.36	4.03
6	0.72	1.44	1.94	2.45	3.14	3.71
7	0.71	1.41	1.89	2.36	3.00	3.50
8	0.71	1.40	1.86	2.31	2.90	3.36
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11
12	0.70	1.36	1.78	2.18	2.68	3.05
13	0.69	1.35	1.77	2.16	2.65	3.01
14	0.69	1.35	1.76	2.14	2.62	2.98
15	0.69	1.34	1.75	2.13	2.60	2.95
16	0.69	1.34	1.75	2.12	2.58	2.92
17	0.69	1.33	1.74	2.11	2.57	2.90
18	0.69	1.33	1.73	2.10	2.55	2.88
19	0.69	1.33	1.73	2.09	2.54	2.86
20	0.69	1.33	1.72	2.09	2.53	2.85
21	0.69	1.32	1.72	2.08	2.52	2.83
22	0.69	1.32	1.72	2.07	2.51	2.82
23	0.69	1.32	1.71	2.07	2.50	2.80
24	0.68	1.32	1.71	2.06	2.49	2.80
25	0.68	1.32	1.71	2.06	2.49	2.79

A CHI-SQUARE TABLE



Degrees of freedom	*99%*	*95%*	*90%*	*70%*	*50%*	*30%*	*10%*	*5%*	*1%*
1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	4.66	6.57	7.79	10.82	13.34	16.22	21.06	23.69	29.14
15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	5.81	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57

Source: Adapted from p. 112 of Sir R. A. Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver & Boyd. 1958).

Honor Statement:

I have neither given nor received help on this exam,
and all of the answers are my own.

Signature: ______________________________