Name:

Stat 1040, Fall 2004
Final Test, Wednesday December 15, 7:30–9:20 am

Show your work. The test is out of 100 points and you have 110 minutes.

1. Researchers think that eating "trans-fats" lowers the particle size of LDL molecules (so-called "bad cholesterol") in the body and hence increases the risk of heart disease. Stick margarine is high in "trans-fats" while butter is low in "trans-fats".

   In a recent randomized, controlled, double-blind study, subjects were put on special diets for a period of 3 months. The *treatment group* received a special diet in which 20% of the calories came from stick margarine, while the *control group* received a special diet in which 20% of the calories came from butter. The food was prepared by taking the same low-fat diet and mixing either stick margarine or butter into foods such as muffins, casseroles, and hot cereals. Participants were required to eat all of the food provided in the special diets, and nothing else. At the end of the study, researchers measured the LDL particle size and found that the average LDL particle size for the treatment group was smaller than the average LDL particle size for the control group, and that the difference was "statistically significant".

   (a) (2 points) Clearly explain what it means for the study to be *randomized.*

   (b) (2 points) Clearly explain what it means for the study to be *double-blind.*

   (c) (2 points) Clearly explain what it means for the result to be *statistically significant.*

   (d) (2 points) Why is it better to compare two groups like this instead of just putting all the people on the treatment diet (margarine) and comparing their LDL particle size at the beginning of the study to their LDL particle size at the end of the study?

   (e) (2 points) Why was the margarine or butter mixed into foods instead of being used as a spread?

2. Six children attend a party. There are 7 party favors: 3 pink favors and 4 blue favors. The children are each given a party favor at random.

   (a) (2 points) What is the chance that the first child gets a pink favor?

   (b) (2 points) What is the chance that the second child gets a pink favor?

   (c) (2 points) If I see that the first 2 children received pink favors, what is the chance that the third child also gets a pink favor?

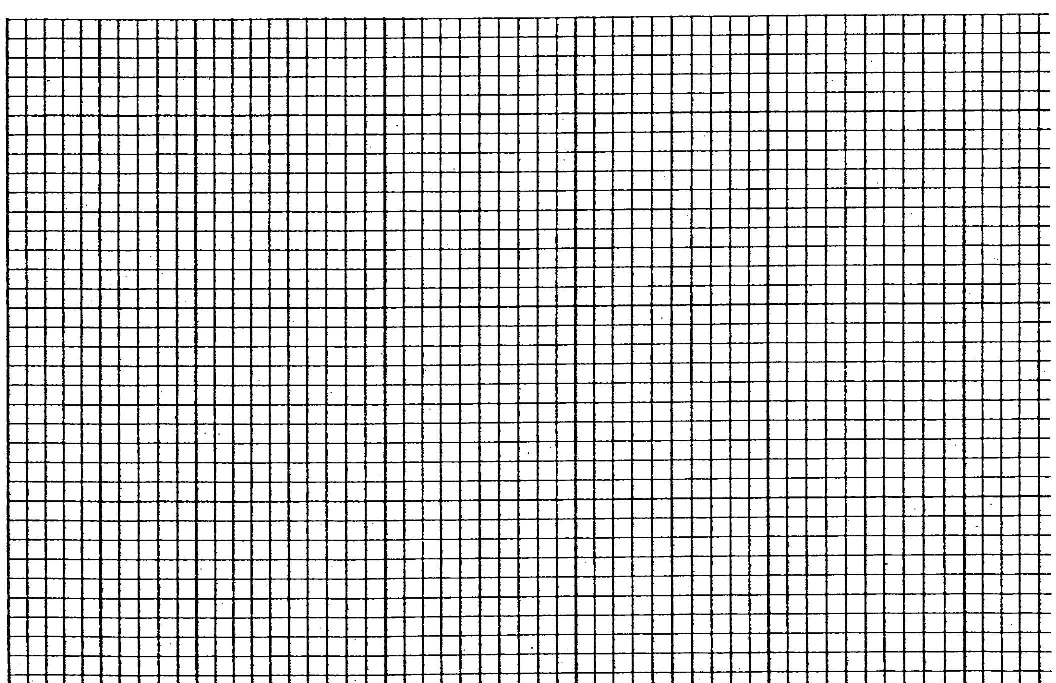
   (d) (2 points) What is the chance that the left-over favor is blue?

   (e) (2 points) What is the chance that the first two children get favors of different colors?

3. (6 points) The following table is for the gestational age of 1210 babies:

Gestational Age	Number
230–250	47
250–270	206
270–290	731
290–310	199
310–330	27

Draw a histogram for these data on the graph paper provided. Be sure to label the axes.

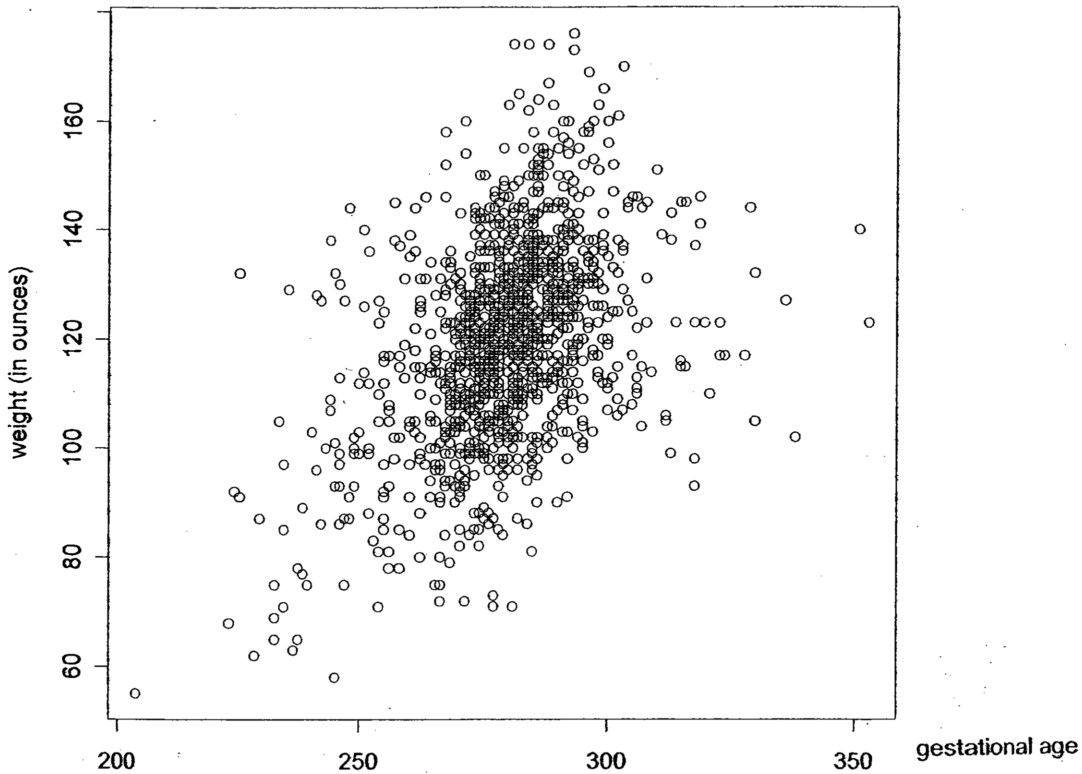


4. The following information was obtained from a group of 1221 babies:

Gestational age (in days): average = 280, SD = 15

Birth-weight (in ounces): average = 120, SD = 18

The correlation coefficient was .42.



(a) (5 points) Find the equation of the regression line for predicting birth-weight from gestational age.

(b) (3 points) Sketch the regression line on the scatter diagram.

(c) (2 points) Predict the birth-weight for a baby with gestational age of 260 days.

(d) (2 points) Find the rms error of your prediction in part (c).

(e) (2 points) Would you be surprised to find that the baby in part (c) weighed 170 ounces? Explain.

5. (4 points) In an article in the May 2004 issue of the *Journal of occupational and Environmental Medicine*, authors describe an observational study of over 23,000 Detroit auto workers. They found that physically active employees had, on average, \$250 a year lower average health-care costs than sedentary employees. Their conclusion was that **to decrease health-care costs, sedentary employees should be encouraged to exercise.** Suggest one possible confounding factor that could account for the difference in health-care costs and explain why your confounding factor could make their conclusion incorrect.

6. (6 points) The birth-weight of babies follows the normal curve with an average of 120 ounces and an SD of 18 ounces. If a baby is at the 10th percentile for birth-weight, how much does the baby weigh?

7. In a simple random sample of 4,700 U.S. adults, people were asked what foods and drinks they had consumed in the last 24 hours. It turned out that the average number of calories from sodas was 168 with an SD of 120.

   (a) (8 points) Find an approximate 95% confidence interval for the average number of calories from sodas for all U.S. adults.

   (b) (2 points) Is your confidence interval still valid if you find out that the histogram of calories from sodas does not follow the normal curve? Explain briefly.

   (c) (2 points) Is your confidence interval still valid if you find out that the sample is not a simple random sample? Explain briefly.

8. (12 points) Researchers think that eating margarine lowers the particle size of LDL molecules (so-called "bad cholesterol") compared to eating butter. In a randomized, controlled, double-blind experiment, 105 people in the treatment group (margarine diet) had an average LDL particle size of 252.6 Angstroms, with an SD of 4.6 Angstroms, while 110 people in the control group (butter diet) had an average LDL particle size of 254.8 Angstroms, with an SD of 4.1 Angstroms. Perform a test to determine whether the researchers' claim is correct. You must state a null and alternative hypotheses, compute a test statistic and a P-value, and clearly state your conclusions about the size of LDL molecules for people on margarine and butter diets such as those in this study.

9. (10 points) Researchers think anti-epileptic drugs accelerate bone loss in elderly women. To investigate, 12 women were randomly selected from all elderly women taking anti-epileptic drugs and they were monitored for a period of 5 years. At the end of the study, researchers measured their bone mineral density on a standardized scale. The average of the 12 measurements was -0.24 with an SD of 1.22. It is known that bone density measurements follow the normal curve. Test the hypothesis that the average for all such women is 0.0 against the alternative hypothesis that it is less than 0.0. State a null and an alternative hypothesis, find a test statistic and a P-value and clearly state your conclusions. (Note: negative values of bone mineral density correspond to accelerated bone loss.)

10. (12 points) The following table represents an exit poll of 794 Utah voters from the 2004 election.

Age	Bush	Kerry	total
18–29	157	37	194
30–44	171	52	223
45–59	154	86	240
60 and over	97	40	137
total	579	215	794

Treat this as though it is a simple random sample of Utah voters. Test to see whether reported voting behavior is independent of age for Utah voters for the 2004 Presidential election. You must clearly state a null and an alternative hypothesis, find a test statistic and a P-value, and clearly state your conclusions.

11. (4 points) A sociologist is interested in whether men and women have different reactions to stress in the workplace. He gives a survey with 50 questions. For each question he does a chi-square test for independence between the participant's gender and their response to the question. He obtains 50 P-values (one for each of the questions), and finds that 3 of them are statistically significant. Then he starts writing his paper and describes how men and women differ with respect to stress in the workplace. What mistake is he making?

Memory Aids

Please note that these are provided for your convenience, but it is your responsibility to know how and when to use them.

$$\text{rms error} = \sqrt{1-r^2} \times SD_Y$$

$$\text{slope} = r \times \frac{SD_Y}{SD_X}$$

$$\text{intercept} = \text{ave}_Y - \text{slope} \times \text{ave}_X$$

$$SD^+ = \sqrt{\frac{\text{number of draws}}{\text{number of draws} - 1}} \times SD$$

$$SD_{\text{box}} = \sqrt{\text{fraction of 0's} \times \text{fraction of 1's}}$$

$EV_{\text{sum}} =$ number of draws × ave_{box}

$SE_{\text{sum}} = \sqrt{\text{number of draws}} \times SD_{\text{box}}$

$EV_{\text{ave}} = \text{ave}_{\text{box}}$

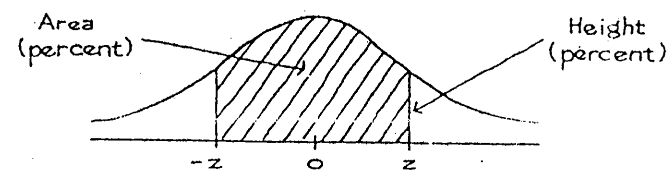
$SE_{\text{ave}} = \dfrac{SE_{\text{sum}}}{\text{number of draws}}$

$EV_{\%} =$ % of 1's in the box

$SE_{\%} = \left(\dfrac{SE_{\text{sum}}}{\text{number of draws}}\right) \times 100\%$

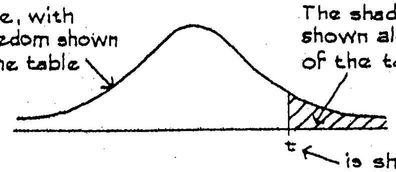
$SE_{\text{diff}} = \sqrt{a^2+b^2}$ where a is the SE for the first quantity, b is the SE for the second quantity, and the two quantities are independent

A NORMAL TABLE



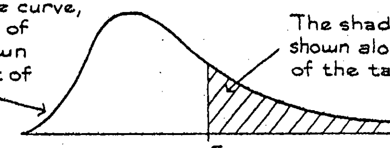
z	Area	z	Area	z	Area
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	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	2.81
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).