Stat 1040, Spring 2008 **Name:** _____

Final Test, May 1, 1:30pm–3:20pm

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

1. **Does Aspartame Cause Cancer?** Aspartame is an artificial sweetener found in thousands of products – sodas, chewing gum, dairy products and even many medicines. Some research has suggested that aspartame can cause lymphoma or leukemia in rats.

   A recent study by the National Cancer Institute involved 340,045 men and 226,945 women, ages 50 to 69. From surveys they filled out in 1995 and 1996 detailing food and beverage consumption, researchers calculated how much aspartame they consumed. Over the next five years, 2,106 developed cancers such as lymphoma or leukemia. No association was found between aspartame consumption and occurrence of these cancers.

   (a) (2 points) Was the study a controlled experiment or an observational study? Why?

   (b) (4 points) Suggest a possible confounding factor for this study and explain why your confounding factor might make you doubt their results.

   (c) (2 points) "It's very reassuring. It's a large study with a lot of power," said Richard Adamson, a senior science consultant to the American Beverage Association, the leading industry group. Does the large sample size prove that aspartame does not cause cancers such as lymphoma or leukemia? Explain.

2. A randomized, controlled, double-blind study published in March, 2008 shows the well-known "placebo effect" works even better if the placebo costs more. In the study, volunteers were given an electric shock and took a pill. Volunteers in the treatment group were told it was an expensive painkiller, while those in the control group were told it was a discounted painkiller. In fact, all the pills were placebos, but 85% of the volunteers who thought they were getting an expensive painkiller said they felt less pain after taking it, compared to 61% of those who thought they were getting a discounted painkiller.

   (a) (1 point) What is a placebo?

   (b) (3 points) Why is a placebo used in a controlled experiment?

   (c) (2 points) What sort of a test would you use if you wanted to test whether the difference between the two percentages could be due to chance error? (Circle the correct answer)

   - one-sample z-test
   - one-sample t-test
   - two-sample z-test
   - Chi-square test

3. (4 points) In a flyer by Horizon Textbook Publishing, a customized textbook manufacturer, they cite Dr. Blount, Gaston College, as follows:

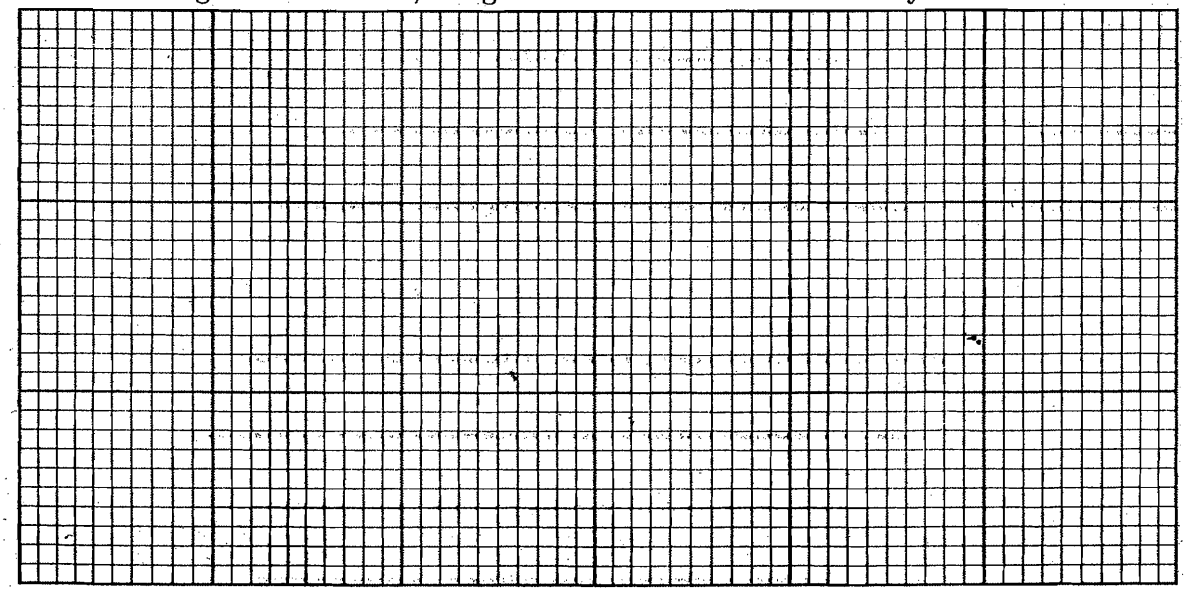
   > "After 4 years with my Horizon customized textbook, I've witnessed an increase in both grade point averages and instructor evaluation scores. Thanks, Horizon!"

   Assuming that his grade point averages and instructor evaluation scores really did increase, can we attribute the increase to the Horizon customized textbook? **Yes / No**? Circle your answer and explain, using the appropriate statistical terms. Provide two *different* reasons to justify your answer.

4. (8 points) The following table summarizes the lengths of 24 male painted turtles. Class intervals include the left endpoint but not the right.

Length (mm)	Number of turtles
90 to 100	3
100 to 110	6
110 to 120	7
120 to 130	6
130 to 140	2

Draw a histogram for the data, being careful to label the axes correctly.



5. (7 points) The length of female painted turtles follows the normal curve with an average of 136 mm and an SD of 21 mm. If the length of one of these turtles is at the 75th percentile, what is her length?

6. The length and width of 24 male painted turtles have the following summary statistics:

Length:	average = 113 mm	SD = 12 mm	r = 0.95
Width:	average = 88 mm	SD = 7 mm	

The scatter-diagram is football-shaped.

(a) (5 points) Predict the width of a turtle that is 130 mm in length.

(b) (1 point) What is the rms error for your prediction in part (a)?

7. A class of 26 fourth-graders has 14 boys and 12 girls. This class goes on a field trip. Two children are chosen at random to ride with the teacher.

(a) (1 point) What is the chance the first child is a boy?

(b) (2 points) What is the chance the second child is a boy?

(c) (2 points) What is the chance both children are boys?

(d) (2 points) What is the chance neither of the children are boys?

(e) (2 points) What is the chance one of the children is a boy and the other is a girl?

8. **German Internet Study** This question relates to a study published in April 2008 at `http://www.sevenoneinteractive.de/`. This was a telephone survey in which 1,009 Germans were asked questions about how they used the internet at home.

   (a) (12 points) One of the questions asked people how many Web sites they frequently revisited. For the 505 men in the study, the average was 9.4 with an SD of 8.3. For the 504 women in the study, the average was 6.4 with an SD of 6.0. Is this evidence that the average for all German men is higher than the average for all German women, or could the result just be due to chance error? (Assume these are two independent simple random samples from all German men and women.)

      i. Clearly state the null and alternative hypotheses.

      ii. Calculate the appropriate test statistic.

      iii. Find the P–value.

      iv. Do you reject the null hypothesis? Explain why or why not.

      v. State your conclusions.

(b) (10 points) According to an earlier study, German men visit an average of 20 new Web sites in a typical month. For the 505 men in the new study, the average number of new Web sites visited in a typical month was 20.8 with an SD of 21.6. Does the new study justify the following newspaper headline: "New study shows that German men visit an average of **more than 20** new Web sites in a typical month."? (Assume this is a simple random sample from all German men.)

i. Clearly state the null and alternative hypotheses.

ii. Calculate the appropriate test statistic.

iii. Find the P–value.

iv. Do you reject the null hypothesis? Explain why or why not.

v. State your conclusions.

(c) (8 points) Among the 350 people in this study aged 20 to 29 years, 12.6% visit more than 50 new Web sites in a typical month. Find an 85% confidence interval for the percentage of all Germans aged 20 to 29 years who visit more than 50 new Web sites in a typical month. (Assume this is a simple random sample of all Germans aged 20 to 29 years.)

(d) (2 points) Suppose we found out that the samples really came from an online questionnaire exclusively available to people who visited the German version of "myspace" (`myspace.de`). Which, if any, of the results from the previous three questions are still valid? Explain.

9. (8 points) For Utah men aged 50–80, the average number of hours of hard physical activity a week is 14 hours, with an SD of 15 hours. I plan to take a simple random sample of 225 Utah men aged 50-80. What is the chance that the average number of hours of hard physical activity a week for the men in the sample lies between 12.5 and 15.5?

10. (12 points) In one analysis of the data from the Utah Study of Nutrition and Bone Health they looked at the relationship between BSM1 vitamin D receptor genotype and whether or not a person has a hip fracture. The data for the women in the study are summarized in the table below. Assume this is a simple random sample from the population.

		Genotype		
		+/+	+/-	-/-
Hip Fracture?	**Yes**	183	281	95
	No	262	307	108

We are interested in whether or not genotype and hip fracture are independent in this population.

(a) Clearly state the null and alternative hypotheses.

(b) Calculate the appropriate test statistic.

(c) Find the P–value.

(d) Do you reject the null hypothesis? Explain why or why not.

(e) State your conclusions.

Memory Aids

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

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

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

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

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

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

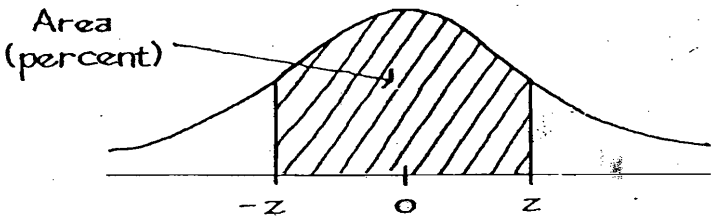
$$EV_{\text{sum}} = \text{number of draws} \times \text{ave}_{\text{box}}$$
$$SE_{\text{sum}} = \sqrt{\text{number of draws}} \times SD_{\text{box}}$$

$$EV_{\text{ave}} = \text{ave}_{\text{box}}$$
$$SE_{\text{ave}} = \frac{SE_{\text{sum}}}{\text{number of draws}}$$

$$EV_{\%} = \%\text{ of 1's in the box}$$
$$SE_{\%} = \left(\frac{SE_{\text{sum}}}{\text{number of draws}}\right) \times 100\%$$

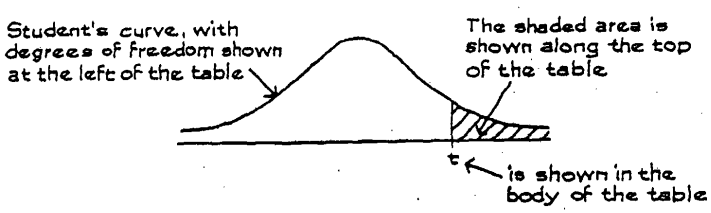
$$SE_{\text{diff}} = \sqrt{a^2 + b^2}$$ where a is the SE for the first quantity,
b is the SE for the second quantity, and the two quantities are independent

A NORMAL TABLE



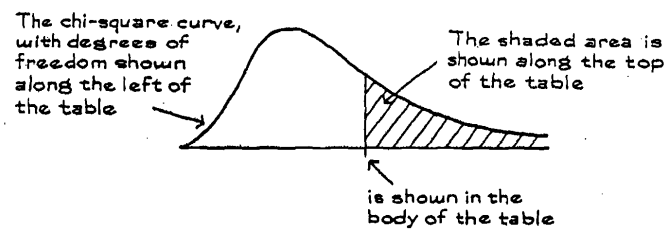
z	*Area*	*z*	*Area*	*z*	*Area*
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	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).