

Name:

STAT 1040, SPRING 2002  
Final Test, Tuesday April 30, 7:00–8:50 am

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

1. (5 points) To study the effectiveness of vitamin C in preventing colds, 200 volunteers were recruited. The researcher randomly assigned 100 of them to take vitamin C for 10 weeks and the remaining 100 to take nothing. The 200 participants recorded how many colds they had during the 10 weeks. The two groups were compared, and the researcher announced that taking vitamin C reduces the frequency of colds. Is there any problem with the design of the study? If your answer is yes, identify the problem and briefly describe how to properly design this particular study.
2. (6 points) A local politician who has to run for office sees the following article from The Utah Statesman (March 27 2002):

**Online poll**

Do you think the state's current liquor laws need to be changed? (Current results in bold.)

- Yes, the laws restrict the freedoms of the citizens of the state. **(42%)**
- No, the laws are not any tougher than those found in other states. **(38%)**
- I don't really care either way. **(11%)**
- I need more information on the subject. **(2%)**
- None of the above **(7%)**

Visit us on the Web at [www.utahstatesman.com](http://www.utahstatesman.com) to cast your vote and see results from past Utah Statesman online polls.

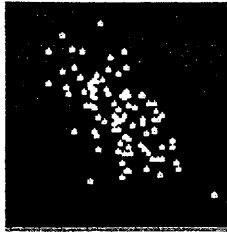
She wonders whether changing the state's liquor laws would be a good thing to have as part of her campaign platform, hoping to get more votes by supporting what the majority of people think. What do *you* think? Should she support a change in the current liquor laws to boost her election chances? Answer yes or no and give 3 different reasons to support your answer, based on information in the Statesman poll, not on your personal opinion.

3. (5 points) A large department store wants to know if consumers would be willing to pay slightly higher prices to have computers available throughout the store to help them locate items. The store posted an interviewer at the door and told her to collect a sample of 100 opinions by asking the next person who came in the door each time she had finished an interview. Is this a simple random sample, a cluster sample, or a sample of convenience? Why? Comment on the reliability of the information she would obtain.
4. A 10-sided die shows the numbers 1 through 10 with equal probability. A 6-sided die shows the numbers 1 through 6 with equal probability.
- (a) (4 points) Which is more likely, and why?
- Getting double-six when you roll two 6-sided dice
  - Getting double-ten when you roll two 10-sided dice
- (b) (4 points) If I roll one 6-sided die and one 10-sided die, what is the chance that the total number of spots is 10?
5. (10 points) There are approximately 20,000 students at USU. For a simple random sample of 500 USU students, we learn that 80% are satisfied with President Kermit Hall. If possible, construct a 95% confidence interval for the percentage of all USU students who are satisfied with President Hall. If you cannot do this, say why not.

6. (4 points) Match the four scatterplots with their correlations from the list:

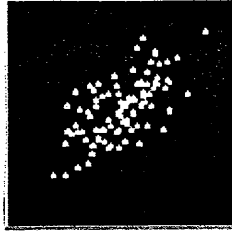
-1.03, -0.98, -0.62, 0.05, 0.65, 0.80, 0.98, 1.03

Plot A



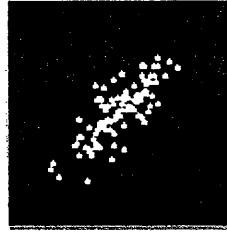
$r =$  \_\_\_\_\_

Plot B



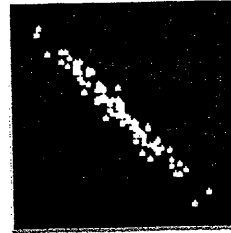
$r =$  \_\_\_\_\_

Plot C



$r =$  \_\_\_\_\_

Plot D



$r =$  \_\_\_\_\_

7. For a random sample of 20 car models, the average weight (in pounds) was 3236, with an SD of 523. The average gas mileage (in miles per gallon) was 21.4 with an SD of 4.2. The correlation between weight and gas mileage was -0.87. The scatter diagram was football shaped.

(a) (4 points) Assuming the histogram for gas mileage follows the normal curve, would you be surprised if someone told you that one of these cars got 27 miles per gallon? Explain your reasoning.

(b) (3 points) Find the equation of the regression line for predicting gas mileage from weight.

(c) (4 points) Predict the gas mileage of a car that weighs 3500 pounds.

(d) (4 points) Would you be surprised if someone told you that one of these cars weighing 3500 pounds got 27 miles per gallon? Explain your reasoning, using the rms error.

(e) (2 points) In one sentence, explain what the correlation coefficient tells you about the relationship between gas mileage and weight of cars like these.

8. For a simple random sample of 225 patients from Logan Regional Hospital, the average cholesterol count was 230 with an SD of 30. For a simple random sample of 400 patients from the University of Utah Hospital, the average cholesterol count was 240 with an SD of 40. You may assume that there are 5000 patients at each hospital.

The question we want to consider is whether or not the average cholesterol count of all 5000 patients at Logan Regional Hospital is the same as the average cholesterol count of all 5000 patients at the U of U Hospital.

- (a) (4 points) Formulate appropriate box models - indicate the number of tickets in each box, what the numbers on the tickets represent, and say whether you do or do not know the average and SD of each box.
- (b) (2 points) Clearly state the null and alternative hypotheses in terms of your box models.
- (c) (4 points) Compute an appropriate test statistic.
- (d) (2 points) Find the P-value.
- (e) (3 points) Using the 5% level, say whether or not you reject the null hypothesis and state your conclusion in terms of the original question.
- (f) (3 points) Did you assume that the histogram for the patients' cholesterol counts followed the normal curve? Why or why not?

9. (12 points) The table below shows the number of computers per household for a simple random sample of 600 households from Cache Valley and Salt Lake City.

Number of Computers	0	1	2	Total
Cache Valley	30	50	20	100
Salt Lake City	160	300	40	500
Total	190	350	60	600

For Cache Valley and Salt Lake City, is the number of computers per household independent of location? Answer with an appropriate statistical test. State the null and alternative hypotheses, compute a test statistic, estimate the P-value, and clearly state your conclusion.

10. In the United States, the average age of men at the time of their first marriage is 24.8 years. To determine if Cache Valley men marry for the first time at an earlier average age than men in the country as a whole, a sociologist took a simple random sample of 24 male Cache Valley residents who were or had been married and found that their average age at first marriage was 23.5 years with a standard deviation of 3.2 years.

- (a) (12 points) Did this survey show that Cache Valley men really marry for the first time at an earlier average age than the men in the country as a whole? Assume the histogram of mens' ages at first marriage closely follows the normal curve. Clearly state the null and the alternative hypothesis, conduct an appropriate test, and state your conclusions.

- (b) (3 points) Would your test in part (a) be valid even if the histogram of the mens' ages at first marriage had a very long right-hand tail? Why or why not? Explain briefly.

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

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

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

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

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

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

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

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

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

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

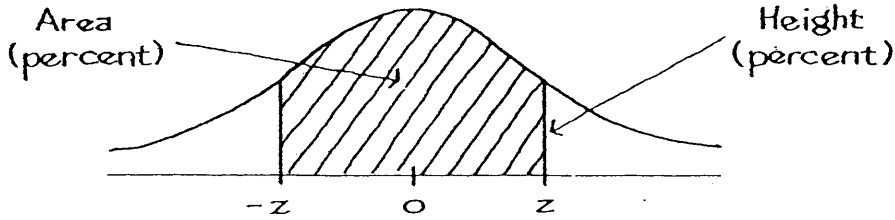
$$EV_{\%} = \text{\% of 1's in the box}$$

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

$$SE_{\text{diff}} = \sqrt{a^2 + b^2} \quad \text{where } a \text{ is the SE for the first quantity,}$$

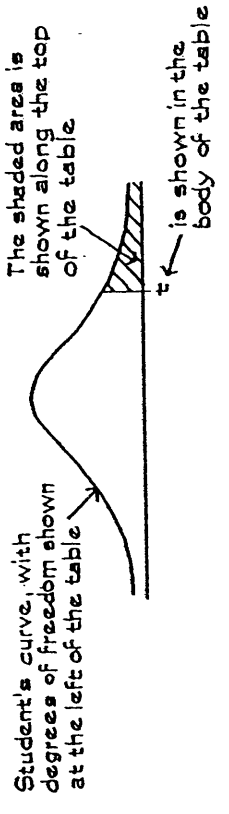
$b$  is the SE for the second quantity, and the two quantities are independent

# A NORMAL TABLE

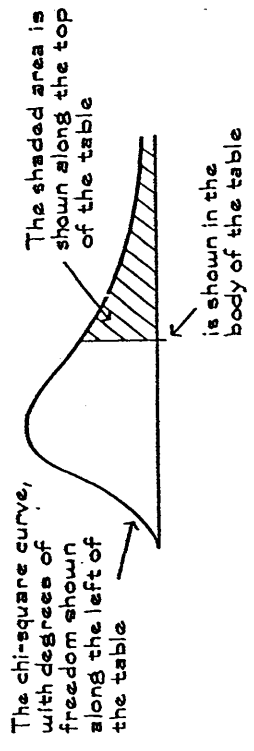


<i>z</i>	<i>Area</i>	<i>z</i>	<i>Area</i>	<i>z</i>	<i>Area</i>
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	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



A CHI-SQUARE TABLE



Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%	Degrees of freedom	99%	95%	90%	70%	50%	30%	10%	5%	1%
1	1.00	3.08	6.31	12.71	31.82	63.66	1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.82	1.89	2.92	4.30	6.96	9.92	2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.76	1.64	2.35	3.18	4.54	5.84	3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.74	1.53	2.13	2.78	3.75	4.60	4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.73	1.48	2.02	2.57	3.36	4.03	5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.72	1.44	1.94	2.45	3.14	3.71	6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	0.71	1.41	1.89	2.36	3.00	3.50	7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	0.71	1.40	1.86	2.31	2.90	3.36	8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	0.70	1.38	1.83	2.26	2.82	3.25	9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	0.70	1.37	1.81	2.23	2.76	3.17	10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	0.70	1.36	1.80	2.20	2.72	3.11	11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	0.70	1.36	1.78	2.18	2.68	3.05	12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	0.69	1.35	1.77	2.16	2.65	3.01	13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	0.69	1.35	1.76	2.14	2.62	2.98	14	4.66	6.57	7.79	10.82	13.34	16.22	21.06	23.69	29.14
15	0.69	1.34	1.75	2.13	2.60	2.95	15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	0.69	1.34	1.75	2.12	2.58	2.92	16	5.81	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	0.69	1.33	1.74	2.11	2.57	2.90	17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	0.69	1.33	1.73	2.10	2.55	2.88	18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	0.69	1.33	1.73	2.09	2.54	2.86	19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	0.69	1.33	1.72	2.09	2.53	2.85	20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57
21	0.69	1.32	1.72	2.08	2.52	2.83	21									
22	0.69	1.32	1.72	2.07	2.51	2.82	22									
23	0.69	1.32	1.71	2.07	2.50	2.81	23									
24	0.68	1.32	1.71	2.06	2.49	2.80	24									
25	0.68	1.32	1.71	2.06	2.49	2.79	25									

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