

Final Test, May 3, 1:30pm–3:20pm

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

Questions 1 to 5 below are based on survey results related to the Terri Schiavo case, as published in *Time*, April 4, 2005 (pp. 22–30). Terri Schiavo suffered severe brain damage during cardiac arrest in 1990. Over the next 15 years, Terri was kept alive through a feeding tube. For 10 years, Terri's husband, Michael Schiavo and Terri's parents, the Schindlers, battled in court over the issue of whether or not to remove Terri's feeding tube. After a final court order in March 2005, Terri's feeding tube was removed and Terri was allowed to die in early April 2005.

1. A simple random sample of 1010 Americans was polled by *Time* on the following question.

Question A: *“Do you agree with the decision to remove Schiavo's feeding tube?”*

Time reports that of the 1010 people,

595 (about 59%) agree

349 (about 35%) disagree

66 (about 7%) don't know.

- (a) (3 points) Assume that Terri's parents, who strongly opposed the removal of Terri's feeding tube, initiated an alternative poll with the following question:

Question B: *“Do you agree with the decision to remove Terri's feeding tube, thus letting her starve to a slow and painful death?”*

Other things being equal, the percentage of people that would have answered agree to question (B) would have been (a) higher than 59%, (b) about 59%, or (c) much less than 59%. Circle your answer and explain clearly.

- (b) (4 points) *Time* reports that “the margin of error for the percentage who agree is ± 3 percentage points.” Is this (a) our usual standard error, (b) about twice our usual standard error, or (c) about one half of our usual standard error. Circle your answer and show the necessary calculations.

(c) (3 points) *Time* further presents outcomes for different subgroups (such as for church goers/non church goers and for different political affiliations) and states that "the margin of error is higher for subgroups". Assume that we are looking at a subgroup that consists of about 253 people. For these 253 people, the standard error for the percentage who agree would be

- i. about four times as big
- ii. about twice as big
- iii. about $\sqrt{2}$ as big

as the standard error for the sample of 1010 people. Just circle your answer – no calculation is necessary.

2. A newspaper reporter wants to do some follow-up interviews with some of the participants in the survey and randomly selects 3 *different* people from the original 1010 participants. Recall that 595 answered agree, 349 answered disagree, and 66 answered don't know to the Question (A).

(a) (3 points) The chance that the 3 randomly selected participants for the follow-up interview all answered agree to question (A) is about _____ %.

(b) (3 points) The chance that the 3 randomly selected participants for the follow-up interview all answered disagree to question (A) is about _____ %.

(c) (3 points) The chance that *at least 1* of the 3 randomly selected participants for the follow-up interview answered agree to question (A) is about _____ %.

3. Use the fact that about 59% of the 1010 Americans in the *Time* sample answered agree to Question (A).

(a) (7 points) Find an approximate 99% confidence interval for the percentage of all Americans that would have answered agree to Question (A).

(b) (3 points) Is your confidence interval still valid if you find out that the 59% did not come from a simple random sample but from a Web page that belonged to Terri's husband (who strongly supported the removal of Terris' feeding tube)? Explain briefly.

4. (10 points) *Time* reports that 53% of Republicans answered agree to Question (A). The "religious right" in the Republican Party does not believe the percentage should be so high. To obtain a better idea about whether the majority of their members really agree, the Republican Party decides to do a simple random sample of Republicans. Suppose that the true (population) percentage of Republicans who agree is indeed 53%. What is the chance that in a simple random sample of 2600 Republicans, less than 1300 would answer agree? Indicate the box model, calculate the required EV's and SE's, and argue using the normal curve.

5. (12 points) Some political experts believe that Democrats agree with question (A) at a higher rate than Independents. In fact, *Time* reports that 65% of Democrats agree whereas only 59% of Independents agree. Suppose that 450 Democrats and 180 Independents were initially included in the survey. Perform a statistical test to determine whether the political experts' claim is correct. You must state a null and alternative hypothesis, compute a test statistic and P-value, and clearly state your conclusions about whether the political experts are correct or not.

6. In a recent news article, a researcher claims that being overweight in middle age causes dementia later in life, although the mechanism is unknown. The dementia rate in overweight people was about 8%, while in normal-weight "controls" it was only 7%.

(a) (2 points) The difference between dementia rates for overweight and normal-weight people was "highly statistically significant". What does this tell you about the P-value?

(b) (3 points) Consider the following statement:

The difference was "highly statistically significant" so if overweight people don't lose weight, they are extremely likely to get dementia.

Is this what statistical significance tells us? Explain.

(c) (2 points) Is this an observational study or a controlled experiment? Explain.

(d) (4 points) Suggest a confounding factor and clearly explain why this factor might be responsible for the association between weight and dementia.

7. Leonardo da Vinci (1452–1519) theorized that if you put your arms out to the side and measured from the fingertip of one hand to the fingertip of the other, this ‘wingspan’ distance would approximately equal your height. A group of fourth–grade students measured their height and wingspan and found

average height = 49.5 inches with an SD of 1.8 inches average wingspan = 48.9 inches with an SD of 2.1 inches

The scatter diagram was football-shaped and the correlation coefficient was 0.8.

- (a) (5 points) Find the equation of the regression line for predicting wingspan from height.

- (b) (2 points) Does the intercept make sense? Explain.

- (c) (3 points) Predict the wingspan of a fourth–grader who is 52 inches tall.

- (d) (2 points) Find the rms error for your answer in part (b).

- (e) (2 points) One of these fourth–graders is 52 inches tall. Is it likely that he or she would have a wingspan of 45 inches? Explain clearly using your answers to (c) and (d).

8. (12 points) The following table comes from a simple random sample of university students, each of whom was asked whether they agree or disagree with the statement "Cell phone use while driving should be prohibited".

	Male	Female	
Agree	13	16	29
Disagree	240	206	446
	253	222	475

Test to see whether male and female university students differ in their responses. You should clearly state the null and the alternative hypothesis, find a test statistic and an approximate P-value, and state your conclusions in everyday language.

9. (12 points) An English exam is taken by 2000 students. The exam scores are known to follow the normal curve. The teacher says that the average of all 2000 test scores is 75, but one of the students thinks the average is actually lower. She takes a simple random sample of 9 students and finds they got the following scores:

63, 53, 84, 82, 35, 50, 68, 73, 92

Test to determine whether the average really is 75, against the alternative that the student is correct. You should clearly state the null and the alternative hypothesis, find a test statistic and an approximate P-value, and state your conclusions in everyday language.

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

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

$$\text{slope} = r \times \frac{SDy}{SDx}$$

$$\text{intercept} = \text{avey} - \text{slope} \times \text{ave}_x$$

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

$$SD_{\text{box}} = \sqrt{\frac{\text{fraction of 0's} \times \text{fraction of 1's}}{\text{number of draws} - 1}} \times SD$$

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

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

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

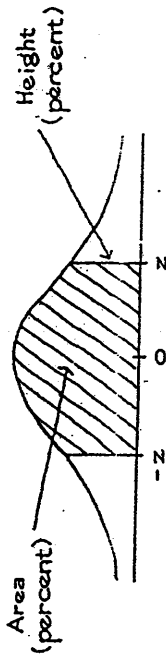
$$SE_{\text{ave}} = \frac{SE_{\text{sum}}}{\text{number of draws}}$$

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

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

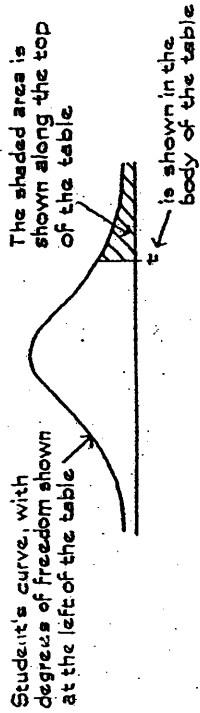
$$SE_{\text{diff}} = \sqrt{a^2 + b^2} \quad \text{where } a \text{ is the SE for the first quantity, } b \text{ is the SE for the second quantity, and the two quantities are independent}$$

A NORMAL TABLE

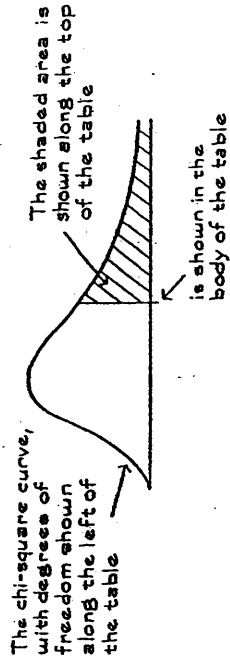


z	Area	z	Area	z	Area
0.00	0	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	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



A CHI-SQUARE 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

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, 1938).