Statistics 1040 Fall 2010 Exam 2

Do not open the exam until you are instructed to do so.

Directions: You have 75 minutes to complete the exam, which has 100 total possible points. Be sure to answer every question. Where numeric answers are required, you must show your work to receive credit. The formulas given below and the table on the last page may be useful for some of the questions. Where calculations are required, round to two decimal places.

Student Name: Solutions

Under your recitation leader's name, circle your recitation time:

Oleksandr Gromenko	Bryan Stephenson	Anthonie Nichols
10:30	8:30	8:30
11:30	12:30	10:30
12:30	1:30	11:30
		1:30

Possibly useful formulas for this exam

- calculator average: $\bar{x}$
- $SD = \sqrt{[\text{ave. of (value}^2)] - (\text{ave. of value})^2}$
- calculator SD: σ_n
- $r = \frac{(\text{ave. of x·y}) - (\text{ave. of x})(\text{ave. of y})}{(\text{SD for x})(\text{SD for y})}$
- $slope = r \cdot \frac{\text{SD for y}}{\text{SD for x}}$
- rms error $= (\sqrt{1 - r^2}) \cdot (\text{SD for y})$
- P(A and B) = P(A)×P(B|A)
- P(A or B) = P(A) + P(B) − P(A and B)
- $EV_{sum} = (\text{\# of draws}) \times ave_{box}$
- $SE_{sum} = \sqrt{\text{\# of draws}} \times SD_{box}$
- conf. int.: estimate $\pm z_\alpha \times (SE)$
- $SD_{box} = \sqrt{(\text{fraction of 0's}) \times (\text{fraction of 1's})}$
- $EV_{\%} = \%$ of 1's in box
- $SE_{\%} = \frac{SE_{sum}}{\text{\# of draws}} \times 100\% = \frac{SD_{box}}{\sqrt{\text{\# of draws}}} \times 100\%$
- $EV_{ave} = ave_{box}$
- $SE_{ave} = \frac{SE_{sum}}{\text{\# of draws}} = \frac{SD_{box}}{\sqrt{\text{\# of draws}}}$

1. Suppose that in a particular Stat 1040 recitation section, there are 37 students, of whom 9 are left-handed and 28 are not left-handed. The recitation leader calls 2 students (selected at random without replacement) to work an example at the board.

   (a) (6 points) What is the probability that both students are left-handed?

   $P(L_1 \text{ and } L_2) = P(L_1) \times P(L_2 | L_1)$ } +3 for mult. rule (of correct events)

   $= \frac{9}{37} \times \frac{8}{36}$ } +3 for correct numbers (−1 for each incorrect)

   (b) (6 points) What is the probability that that the first student is left-handed, but the second student is not?

   $P(L_1 \text{ and not } L_2) = P(L_1) \times P(\text{not } L_2 | L_1)$ } +3 for mult. rule (of correct events)

   $= \frac{9}{37} \times \frac{28}{36}$ } +3 for correct numbers (−1 for each incorrect)

   (c) (6 points) What is the probability that at least one of the two students is left-handed?

   $P(L_1 \text{ or } L_2) = P(L_1) + P(L_2) - P(L_1 \text{ and } L_2)$ } +3 for add. rule (correct events)

   $= \frac{9}{37} + \frac{9}{37} - \frac{9}{37} \times \frac{8}{36}$ } +3 for correct numbers (−1 for each incorrect)

   $1 - P(\text{never } L) = 1 - P(\text{not } L_1 \text{ and not } L_2) = 1 - P(\text{not } L_1) \times P(\text{not } L_2 | \text{not } L_1)$ } +3 for 1 − P(opposite) and mult. rule (of correct events)

   $= 1 - \frac{28}{37} \times \frac{27}{36}$ } +3 for correct numbers (−1 for each incorrect)

2. Suppose that in the same recitation section of the previous question (37 total students, 9 left-handed and 28 not left-handed), there are 12 male students and 25 female students, and all the left-handed students are male. Again 2 students are selected at random without replacement.

   (a) (2 points) Let event A be "the first student selected is female", and event B be "the first student selected is left-handed." Then events A and B are (circle one):

      i. independent → $P(B|A) = 0 \neq P(B)$
      +2 { (ii.) mutually exclusive → they can't both happen here
      iii. none of the above

   (b) (2 points) Let event A be "the first student selected is female" and event B be "the second student selected is left-handed." Then events A and B are (circle one):

      i. independent → $P(B|A) \neq P(B)$
      ii. mutually exclusive → they can both happen here
      +2 { (iii.) none of the above

3. (8 points) A fair coin is to be tossed many times – either 100 or 1000. Which number of tosses will make the following outcomes more likely? (Circle one of 100 or 1000 for each.)

+2 each

(a) Exactly 50% heads

(100) 1000

(b) More than 60% heads

(100) 1000

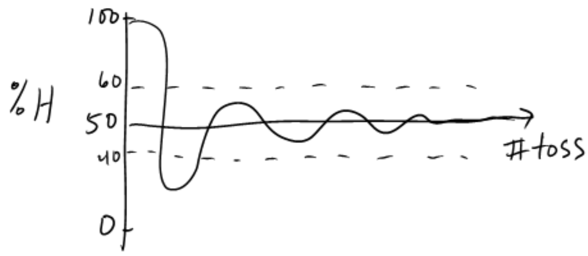
(c) Between 40% and 60% heads

100 (1000)

(d) More than 50% heads but less than 60% heads

100 (1000)

Visual aid:



4. (10 points) According to the 2000 U.S. Census, the average household size of a renter-occupied unit in Utah was 2.75. In the same year, a team of researchers took a SRS of 1200 rental units in Utah and found an average household size of 2.85. Clearly identify each of the following here:

+2 each

(a) population

all renter-occupied units in Utah

(b) parameter

2.75 (or the pop. average household size)

(c) sample

1200 selected rental units (-1 if only "1200")

(d) statistic

2.85 (or the sample average household size)

(e) absolute chance errror

2.85 − 2.75 = 0.10 (negative okay)

5. (6 points) Manhattan, Kansas has a population of about 53,000, while Manhattan, New York has a population of about 1.6 million. A SRS of 500 residents of Manhattan, Kansas is selected, and their average weight is calculated. Assuming the average (and SD) weight of residents in the two Manhattans is essentially the same, what sample size would be needed in a SRS of Manhattan, New York residents to achieve the same accuracy as the result of the Manhattan, Kansas sample? Give your answer and explain briefly.

+3 { 500

+3 { Accuracy (or SE) depends only on sample size

not necessary { – when pop. composition (ave & SD) is same and sample size is small relative to pop. size sample ↓ population ↓

6. (5 points) If you wanted to obtain a representative sample of 100 property owners in Salt Lake County, Utah, explain clearly what you would do to make your sample a SRS.

Proposed approach must include the following (or their equivalents):

+2 { • Assign all property owners in county a unique number

+3 { • Randomly draw (without replacement) 100 numbers from this list

(not necessary to state)

7. According to the U.S. Census Bureau, 85% of Utah residents have health insurance. Suppose you are going to take a SRS of 500 Utah residents.

   (a) (6 points) Fill in the blanks: The number of Utah residents in your sample that have health insurance will be around 425, give or take 8 or so.

+2 { Box: $\boxed{1} \times 85$ $\boxed{0} \times 15$ $ave_{box} = 0.85$ $SD_{box} = 0.36$

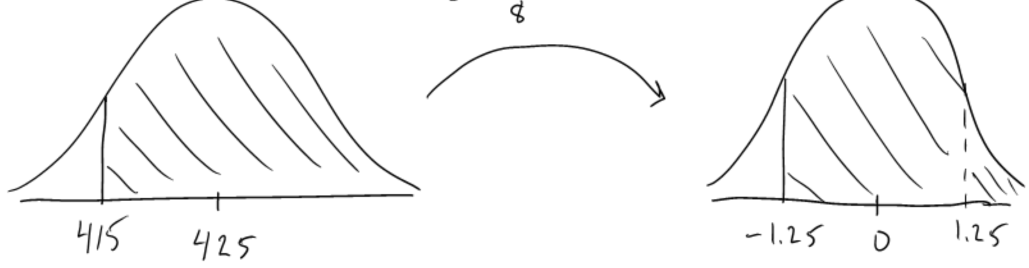
+1 { Sum of 500 draws

+1 { $EV_{sum} = 500 \times 0.85 = 425$

+2 { $SE_{sum} = \sqrt{500} \times 0.36 = 8$

   (b) (6 points) What is the chance that more than 415 Utah residents in your sample have health insurance?

$\frac{415-425}{8} = -1.25$



78.87% $\frac{100\% - 78.87\%}{2} = 10.57\%$

89.44%

Partial credit:

+2 for normal curve w/ EV & SE from (a)

+1 for conversion to standard units

+1 for % from normal table

+2 for appropriately using % from normal table

↗ the article, not the poll; this was not an online poll - it was a valid poll

8. (5 points) An October 2010 Associated Press article (posted on capitolhillblue.com) reported on a pre-election survey of about 1,000 likely voters, which found that 49 percent intended to vote Republican, compared to 43 percent Democrat (and 8 percent other). The margin of error was reported as plus or minus 3.5 percentage points. A reader posted the following comment on the website:

> *If the margin of error is 3.5 points, and the difference* [between percent Republican and percent Democrat] *is six points, that's the same as no difference. This poll is essentially saying they're neck and neck. When will reporters ever learn to interpret polls correctly?*

The reader is essentially calling this a statistical tie. Is it a statistical tie? Answer yes or no, and explain briefly.

+2 { No

+3 { the difference (6%) is greater than the margin of error (3.5%) [values not necessary]

9. In a SRS of 1000 U.S. adult males, the average height was 69.1 inches, with an SD of 3 inches.

   (a) (8 points) Construct a 95 percent confidence interval for the average height of U.S. adult males.

+2 for formula { estimate $\pm z_\alpha \times SE$

$69.1 \pm 1.95 \times .09$

for correct values → +1, +1, +3

$SE_{ave} = \frac{SD_{box}}{\sqrt{\# draws}} = \frac{3}{\sqrt{1000}} = .09$

69.1 ± 0.18

+1 for arithmetic { **68.92 to 69.28**

   (b) (6 points) An observer comments, "So approximately 95 percent of U.S. adult males have heights in this interval." Is the observer correct? If yes, explain why. If no, give a correct interpretation of the confidence interval.

+2 { No

+2 { We can be 95% confident

+2 { that the true pop. mean height is in this interval

-OR- +2 { if we repeated the SRS (and subsequent confidence interval calculation) many times, then we would expect 95% of the resulting intervals

+2 { to contain the true pop. mean height

10. In a poll of 400 likely voters in Utah's second congressional district just before the November 2010 elections, 48 percent responded they would vote for Jim Matheson (compared to 35 percent for Morgan Philpot, 7 percent for another candidate, and 10 percent undecided).

(a) (8 points) Based on the polling result (and treating these 400 likely voters like a SRS), construct a 90% confidence interval for the percentage of voters for Matheson in the actual election.

+2 for formula { estimate $\pm$ $z_\alpha \times SE$

$48\% \pm 1.65 \times 2.5\%$

for correct values → +1, +1, +3

box: $\underbrace{\boxed{1}^{\times 48} \quad \boxed{0}^{\times 52}}$ $\quad SD_{box} = .50$

$SE_{\%} = \frac{SD_{box}}{\sqrt{\#draws}} \times 100\%$

$= \frac{.50}{\sqrt{400}} \times 100\% = 2.5\%$

$48\% \pm 4.13\%$

+1 for arithmetic { **43.87% to 52.13%**

(b) (4 points) What is the correct interpretation of this confidence interval?

+2 { We can be 90% confident that

+2 { the true pop. percentage to vote for Matheson is in this interval.

-OR- +2 { if we were to repeat the SRS (and subsequent confidence interval calculation) many times, then we would expect 90% of the resulting intervals to

+2 { include the true pop. percentage to vote for Matheson

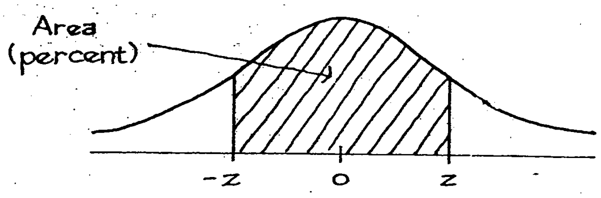
(c) (1 point) In the actual election, 51 percent voted for Matheson. Is this in your interval?

Yes (based on interval in (a))

+1

11. (5 points) Fill in the blank: Even when the distribution of box contents (or population values) is not approximately normal, we can use the normal approximation to address the chances of the sum (or percentage or average) of a large number of draws (or sample size) because of the ___Central___ limit theorem.

Tables



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