Statistics 1040, Sections 003 & 004, Quiz 1 (20 Points)

January 17, 2003

Your Name: ____________________

from: Stat 1040, Spring 2000, Final Test, Friday May 5, 2000, Question 1.

Question 1: Observational Studies and Experiments (**14 Points**)

In a recent study on SIDS (Sudden Infant Death Syndrome), one hospital collected data on 128 babies who died from SIDS in the last 12 months. They took a random sample of 500 babies (of similar ages) who did not die from SIDS (the "controls"), and they compared the two groups with respect to several variables of interest (e.g. whether the child slept on his or her stomach, birthweight, time of year, whether the mother smoked, whether she breast–fed, socio–economic status, etc.).

1

1. (**2 Points**) Is this a controlled experiment or an observational study? Circle your answer and explain.

   Workbook: 1 for correct explanation

   It is an observational study - there was no intervention.

2. (**6 Points**) One physician noticed that 63% of the SIDS babies had mothers who smoked during pregnancy, whereas only 26% of the control babies had mothers who smoked during pregnancy. Another physician claimed that low birthweight could be a "confounding factor". Explain what it means for low birthweight to be a "confounding factor". Be specific.

   2 - 6 for explanation

   Workbook: Perhaps smoking causes low birthweight and it is the low birthweight rather than smoking itself that is leading to higher rates of SIDS.

3. (**6 Points**) If you had access to the data, what would you do to "control for" birthweight?

   2 - 6 for explanation

   Workbook: Study babies with similar birthweights separately. eg. break up the comparison into groups of, say, babies 6-6.5 lb, 6.5-7 lb, 7-7.5 lb, etc.

Note: Other confounding factors could be:
- sleeping on stomach (yes/no)
- time of year
- mother breast-feeds the baby (yes/no)
- socio-economic status

etc.

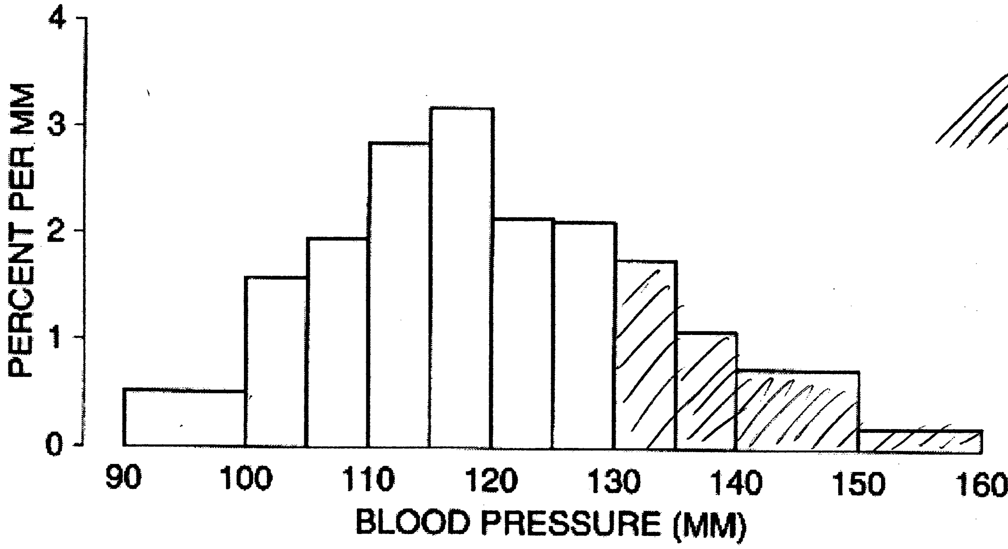
Please turn over!

The observed outcome is whether a child dies of SIDS (yes/no).
In a controlled experiment, one might put half of the children sleep on stomach and the other on the back, breast-feed half of the children, etc. - and see if a child dies of SIDS.

from: FPP, p. 51, Review Exercise 4., abc

Question 2: Histograms (**6 Points**)

The figure below is a histogram showing the distribution of blood pressure for all 14,148 women in a Drug Study (more details about this study can be found in your textbook, page 45).



//// area for part 1.

in 1., 2., 3.: ① if incorrect value, but some explanation
⓪ if incorrect value, no explanation

Use the histogram to answer the following questions:

1. (**2 Points**) Is the percentage of women with blood pressures above 130 mm around 25%, 50%, or 75%? ②

   130 − 135: 5 × 1.8% = 9.0%
   135 − 140: 5 × 1.1% = 5.5%
   140 − 150: 10 × 0.8% = 8.0%
   150 − 160: 10 × 0.2% = 2.0%

   9.0% + 5.5% + 8.0% + 2.0% = 24.5% ≈ 25%

   Also note that the shaded area (////) is far less than half of the entire area − so 25% is the only reasonable answer here.

2. (**2 Points**) Is the percentage of women with bloood pressures between 90 mm and 160 mm around 1%, 50%, or 99%? ②

   90 to 160 makes up (almost) the entire area of the histogram − so this should be 99% (or even 100% if all values for blood pressure have been displayed in this histogram).

3. (**2 Points**) In which interval are there more women: 135–140 mm or 140–150 mm? ②

   The interval 135−140 has a higher percent per mm than the interval 140−150, i.e., 1.1% per mm (135−140) and 0.8% per mm (140−150).

   However, from part 1. we see that 135−140 contains 5.5% of the women while 140−150 contains 8.0% of the women − so 140−150 contains more women.

Statistics 1040, Sections 003 & 004, Quiz 2 (20 Points)

January 24, 2003

Your Name: ____________________

Question 1: Measures of Center and Spread (**20 Points**)

−1 each calculation error

1. **(10 points)** Find the average and the standard deviation of the following two lists of numbers:

	Numbers	Average	Standard deviation
List 1:	100, 100, 100, 100, 100	100 (2)	0 (2)
List 2:	90, 90, 100, 110, 110	100 (2)	8.94 (4)

Show your work (or give a short explanation for your answer)! Use the formulas provided on the back.

List 1: Nothing to calculate! Since all numbers are identical (100), this must also be the average (and the median). Also, since the SD is some average departure from the average, but no value in the list departs from the average, the SD is 0.

List 2: 1) $\text{avg} = \frac{90+90+100+110+110}{5} = \frac{500}{5} = 100$

2)
$90-100=-10$
$90-100=-10$
$100-100=0$
$110-100=10$
$110-100=10$

3)
$(-10)^2=100$
$(-10)^2=100$
$0^2=0$
$10^2=100$
$10^2=100$

4) $\frac{100+100+0+100+100}{5} = \frac{400}{5} = 80$

5) $\sqrt{80} = 8.94 = SD$

Please turn over!

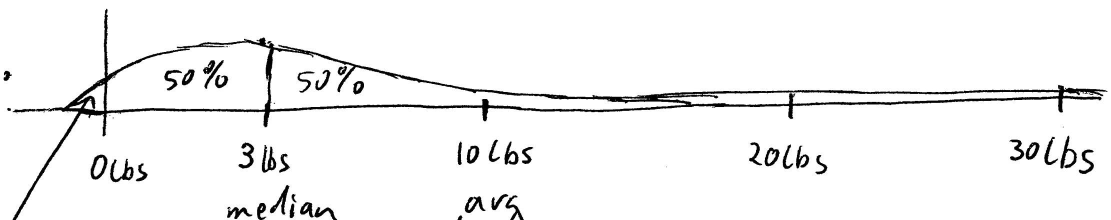
2. **(10 points)** Suppose an advertistment reported that the average weight loss after using a certain exercise machine for 2 months was 10 pounds. You investigate further and discover that the median weight loss was 3 pounds.

   (a) Explain whether it is most likely that the histogram of all weight losses has a long right tail, has a long left tail, or is symmetric. 3

   2

   long right tail: the average is greater than the median

   e.g.



   who says that nobody gains weight? weight loss

   (b) As a consumer trying to decide whether to buy this exercise machine, would it have been more useful for the *company* to give you the mean (average) or the median? Explain. 2

   Customers buying such an exercise machine hope to loose as much weight as possible. Obviously, 10 lbs sounds better than 3 lbs – so the company should provide the mean (average) 3 and not the median. As a fact, the median reveals that 50% of the customers have lost only 3 lbs or less – while the mean is heavily influenced by some customers that might have lost 100 or even 200 lbs (you know those images from advertisements...).

Formulas:

$$\text{avg} = \frac{\text{sum of all numbers}}{\text{how many numbers}}$$

$$\text{SD} = \sqrt{\text{average of } [(\text{deviations from avg})^2]}$$

Statistics 1040, Sections 003 & 004, Quiz 3 (20 Points)

January 31, 2003

Your Name: ____________________

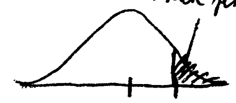
Question 1: Normal Approximation for Data (**20 Points**)

The Wall Street Journal (July 12, 1996) reported that a vacationing family of 4 spends a daily average of \$193 for lodging and food, with a standard deviation of \$38. Assuming that these expenditures approximately follow a normal curve, answer the questions below:

10

- The percentage of families who spend **at least \$250** on food and lodging is roughly 6.68 %.

which percentage?



① convert 250 into standard units: $\frac{250-193}{38} = 1.5$ s.u.

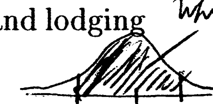
② area between −1.5 and 1.5 : 86.64 %

③ area above 1.5: $\frac{100\% - 86.64\%}{2} = \frac{13.36\%}{2} = \underline{6.68\%}$

10

- The percentage of families who spend **between \$155 and \$231** on food and lodging is about 68.27 %.

Which percentage?



2 possible ways to answer this:

i) Note that 155 is 1 SD below and 231 is 1 SD above the average. So the area (= percentage) is about 68%.

ii) Formal approach: ① convert 155 and 231 into standard units: $\frac{155-193}{38} = -1$ s.u., $\frac{231-193}{38} = 1$ s.u.

② area between −1.0 and 1.0 : $\underline{68.27\%}$

Show your work!

−3 for each incorrect (or missing) s.u.
−3 for each incorrect table value
−3 for each incorrect final result
−2 for each minor (computational) error
+3 for each correct graph (and nothing else)

Statistics 1040, Sections 003 & 004, Quiz 4 (20 Points)

February 7, 2003

Your Name: ____________________

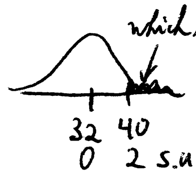
Question 1: Percentiles and the Normal Curve (**12 Points**)

The Trail Making Test is frequently used by clinical psychologists to test for brain damage. Patients are required to connect consecutively numbered circles on a sheet of paper. It has been determined that the average length of time required for a patient to perform this task is 32 seconds with a standard deviation of 4 seconds. Assume that the lengths of time required to connect the circles closely follow the normal curve.
Fill in the blanks and **show your work!**

6

1. The proportion of patients who need longer than 40 seconds to perform the task is about 2.3%. This also means that 40 seconds is the 97.7 th percentile.

which percentage?



(1) convert 40 into s.u.: $\frac{40-32}{4} = \frac{8}{4} = 2$ s.u. (1)

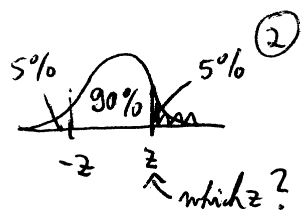
(2) area between −2 and 2: 95.45% (1)

(3) area above 2: $\frac{100\% - 95.45\%}{2} = \frac{4.55\%}{2} = 2.275\% \approx 2.3\%$ (2)

if 2.3% above 40 sec, then 97.7% below 40 sec; so 40 sec is the 97.7th percentile (2)

6

2. A psyhologist would like to retest those persons with completion times in the highest 5% of all required times. Thus, a person who exceeds a time of 38.6 seconds on the Trail Making Test will be considered for retesting.



(2)

(1) area between −1.65 and 1.65: 90.11% ≈ 90% (2)

(2) convert 1.65 s.u. into original units:

$1.65 \cdot 4 + 32 = 38.6$ (2)

Question 2: Correlation (**8 Points**) from: FPP, p. 135, Review Exercise 2

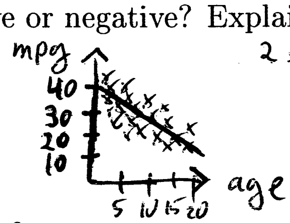
5

1. For a representative sample of cars, would the correlation between the age of the car and its gasoline economy (miles per gallon) be positive or negative? Explain!

2 for correct explanation

Worklook: (3)

e.g.



2. (a) **Negative:** older cars are less fuel-efficient.

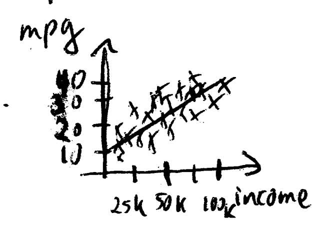
3

2. The correlation between gasoline economy and income of owner turns out to be positive. How do you account for this positive association?

3 for correct explanation

Worklook:

e.g.



(b) Richer people tend to own newer cars and maintain them better.

(and newer cars are more fuel-efficient!)

Statistics 1040, Sections 003 & 004, Quiz 5 (20 Points)

February 21, 2003

Your Name: ____________________

Question 1: The Regression Line (**20 Points**)

A study is made of Math and Verbal SAT scores for the entering class at a certain college. The summary statistics is:

		Average	SD
X:	M–SAT	560	120
Y:	V–SAT	520	110

The correlation coefficient r is 0.66 and we can assume that the scatterplot is football–shaped. **Show your work!**

1. (10 Points) Find the regression equation for predicting the V–SAT score from the M–SAT score.

slope $= r \cdot \frac{SD_Y}{SD_X} = 0.66 \cdot \frac{110}{120} = 0.605$

intercept $=$ avg$_Y$ $-$ slope $\cdot$ avg$_X$ $= 520 - 0.605 \cdot 560 = 520 - 338.8 = 181.2$

regression equation: $\boxed{\text{V-SAT} = 181.2 + 0.605 \cdot \text{M-SAT}}$

or $\boxed{Y = 181.2 + 0.605 \cdot X}$

2. (5 Points) If a student scores 680 on the M–SAT, the predicted V–SAT score is 592.6. (Use the regression equation from part 1!)

$\text{V-SAT} = 181.2 + 0.605 \cdot 680 = 181.2 + 411.4 = 592.6$

3. (5 Points) Find the r.m.s. error for predicting V–SAT scores from M–SAT scores.

r.m.s. error $= \sqrt{1-r^2} \cdot SD_Y = \sqrt{1-0.66^2} \cdot 110 = \sqrt{1-0.4356} \cdot 110$

$= \sqrt{0.5644} \cdot 110 = 0.751 \cdot 110 = 82.6$

Please turn over!

(i.e., on average, the predicted V-SAT score will be 82.6 points off the actual value)

Formulas:

$$\text{r.m.s. error} = \sqrt{1 - r^2} \times \text{SD}_y$$

$$\text{slope} = r \times \frac{\text{SD}_y}{\text{SD}_x}$$

$$\text{intercept} = \text{avg}_y - \text{slope} \times \text{avg}_x$$

Grading Criteria:

in 1., 2., & 3.:
- −2 for each calculation error
- −2 for each incorrect value used
- −2 for swapping x and y

in 1.:
- −3 for incorrect formula for slope
- −3 for incorrect formula for intercept
- −2 if no final equation stated
- −1 if only part of the equation stated (e.g., $181.2 + 0.605 \cdot x$)
- −1 if not specifying x & y (but using x & y in equation)

in 2.:
- −3 for incorrect formula for prediction
- −1 if correct result, but according to old method

in 3.:
- −3 for incorrect formula for r.m.s. error

Statistics 1040, Sections 003 & 004, Quiz 6 (20 Points)

February 28, 2003

Your Name: ____________________

Question 1: Chance/Probability **(20 Points)**

1. In a box of 15 chocolates, 5 are mint, 3 are orange, 4 are caramel, and 3 are cherry. I choose two chocolates at random (without replacement)!

   (a) **(3 Points)** What is the chance that the first is mint or orange? | mutually exclusive events

   $$\frac{5}{15} + \frac{3}{15} = \frac{8}{15} = 53.3\%$$

   Addition rule

   (b) **(4 Points)** What is the chance that the first two are both caramel? | dependent events

   $$\frac{4}{15} \cdot \frac{3}{14} = \frac{12}{210} = \frac{2}{35} = 5.7\%$$

   (c) **(4 Points)** What is the chance that the first is cherry and the second is caramel? | dependent events

   Multiplication rule

   $$\frac{3}{15} \cdot \frac{4}{14} = \frac{12}{210} = \frac{2}{35} = 5.7\%$$

   (d) **(4 Points)** If I like only mint, what is the chance that I like neither of the chocolates I choose? | dependent events

   $$\frac{10}{15} \cdot \frac{9}{14} = \frac{90}{210} = \frac{3}{7} = 42.9\%$$

   Multiplication rule

2. **(5 Points)** Two cards will be dealt off the top of a well–shuffled deck. You have a choice:

   (a) To win \$1 if the first is a ♡.

   (b) To win \$1 if the first is a ♡ and the second is a ◇.

   Which option is better? Or are they the same? Explain briefly.

Option (a) is better; for option (b), we have to fulfill two conditions and not just one as in (a).

optional: (a) probability of winning: $\frac{13}{52} = 25\%$

(b) probability of winning: $\underbrace{\frac{13}{52}}_{\text{first } \heartsuit} \cdot \underbrace{\frac{13}{51}}_{\text{second } \diamondsuit} = 6.4\%$

and

Grading Criteria:

1) general: −1 for calculation error

−1 for incorrect 1st chance

−2 for incorrect 2nd chance (e.g., if not conditional)

−2 for incorrect rule (e.g., addition ⟺ multiplication)

2) −3 for incorrect answer

−1 or −2 { if no (or incorrect) explanation

Statistics 1040, Sections 003 & 004, Quiz 7 (20 Points)

March 7, 2003

Your Name: ____________________

based on: FPP, p. 286, Review Exercise 7

Question 1: Box Models, EV, and SE (**14 Points**)

A quiz has 20 multiple choice questions. Each question has 4 possible answers, one of which is correct. A correct answer is worth 5 points, but a point is taken off for each incorrect answer. A student answers all the questions by guessing at random.

1. (**4 Points**) Find the box model.

−2 for minor mistake
−3 for major mistake
−1 if number of draws not stated

$\boxed{5}\ \boxed{-1}\ \boxed{-1}\ \boxed{-1}$

Number of draws: 20

2. (**5 Points**) Find the expected value, i.e., the number of points a student would get when answering all questions by guessing.

in 2. & 3.: −1 for each calculation error
−2 for minor mistake
−3 for major mistake (e.g., step missing)

$$\text{box average} = \frac{5+3\cdot(-1)}{4} = \frac{2}{4} = \frac{1}{2} = 0.5$$

$$EV_{sum} = 20 \cdot \frac{1}{2} = 10$$

3. (**5 Points**) Find the standard error.

$$\text{box } SD = \sqrt{\frac{(5-0.5)^2 + 3\cdot(-1-0.5)^2}{4}}$$

$$= \sqrt{\frac{(4.5)^2 + 3\cdot(-1.5)^2}{4}} = \sqrt{\frac{20.25+6.75}{4}} = \sqrt{\frac{27}{4}} = \sqrt{6.75}$$

$$= 2.598 \approx 2.6$$

$$SE_{sum} = \sqrt{20}\cdot 2.6 = 4.47 \cdot 2.6 = 11.6$$

Please turn over!

based on: FPP, p. 286, Review Exercise 9

Question 2: Law of Averages (**6 Points**)

A box contains red and green marbles; there are more green marbles than red ones. Marbles are drawn one at a time from the box, at random with replacement. You win a dollar if a red marble is drawn more often than a green one. There are two choices:

- A: 50 draws are made from the box.
- B: 500 draws are made from the box.

Choose one of the four options below. **Explain your answer.**

(4)

1. A gives a better chance of winning.
2. B gives a better chance of winning.
3. A and B give the same chance of winning.
4. Can't tell without more information.

(2) Option A is best. Say the percentage of reds in the box is 40%.
Then, to win, we want the percentage error to be big, so that the actual percentage will be greater than 50%.
This is more likely in the short run.

Formulas:

$$\text{box average} = \frac{\text{sum of all numbers in box}}{\text{how many numbers in box}}$$

$$\text{box SD} = \sqrt{\text{average of [(deviations from box average)}^2]}$$

$$\text{EV}_{sum} = \text{number of draws} \times \text{box average}$$

$$\text{SE}_{sum} = \sqrt{\text{number of draws}} \times \text{box SD}$$

Statistics 1040, Sections 003 & 004, Quiz 8 (20 Points)

March 21, 2003

Your Name: ____________________

Question 1: EV, SE, and Normal Curve (**20 Points**)

In a certain town, there are 40,000 registered voters, of whom 15,000 are Democrats. A survey organization is about to take a simple random sample of 1,000 registered voters.

1. **(4 Points)** Find the box model.

1= Democrat $\boxed{1}$
0= other $\boxed{0}$

$$\boxed{15{,}000 \times \boxed{1} \quad 25{,}000 \times \boxed{0}}$$

number of draws = 1,000

-1 if slightly incorrect number of $\boxed{0}$ / $\boxed{1}$'s in box
-2 if box given as $\boxed{\boxed{0}\ \boxed{1}}$ etc.
-3 if box something else than $\boxed{0}$/$\boxed{1}$'s
-1 if number of draws missing or incorrect

2. **(8 Points)** The expected number of Democrats in this sample of 1,000 is <u>375</u> with an SE of <u>15.3</u>.

$$\text{box avg} = \frac{15{,}000}{40{,}000} = 0.375 = 37.5\%$$

$$\text{box SD} = \sqrt{\frac{15{,}000}{40{,}000} \cdot \frac{25{,}000}{40{,}000}} = \sqrt{0.375 \cdot 0.625} = \sqrt{0.234} = 0.484$$

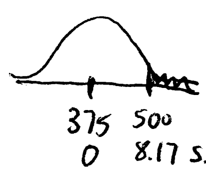
$$EV_{sum} = 1{,}000 \cdot 0.375 = 375$$

$$SE_{sum} = \sqrt{1{,}000} \cdot 0.484 = 31.6 \cdot 0.484 = 15.3$$

-1 for each calculation error
-1 for each minor mistake
-2 for each major mistake (or step missing)

3. **(8 Points)** The chance that **at least** 500 of the voters in the sample are Democrats is about <u>0</u> %.

$$\text{s.u.:} \ \frac{500-375}{15.3} = \frac{125}{15.3} = 8.17$$



area between -4.45 and 4.45: 99.9991%
area between -8.17 and 8.17: almost 100%
area above 8.17: about 0%
It is extremely unlikely that we end up with a sample that contains at least 500 Democrats.

-1 for each calculation error
-2 for incorrect curve parameters, i.e., anything else than EV and SE
-2 for incorrect s.u.
-2 for incorrect table value
-2 for incorrect area under the curve

Please turn over!

Statistics 1040, Sections 003 & 004, Quiz 9 (20 Points)

April 4, 2003

Your Name: ____________________

Question 1: EV%, SE%, and Normal Curve (**20 Points**) −2 each calculation error

A recently conducted survey at the USU has shown that 80% of the approximately 20,000 USU students are satisfied with President Kermit Hall. If we take a random sample of 135 USU students, the chance that at most 70% of them is satisfied with the President is around ___0.185___ %.

Show your work!

1: satisfied
0: not satisfied

box: $\boxed{20 \times \boxed{0} \quad 80 \times \boxed{1}}$ 3

draws = 135 1

box avg $= \frac{80}{100} = 0.8$

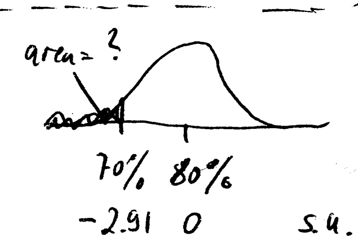
box SD $= \sqrt{\frac{80}{100} \cdot \frac{20}{100}} = \sqrt{0.8 \cdot 0.2} = \sqrt{0.16} = 0.4$ 2

$EV_{\%} = 80\%$ 2 [$EV_{sam} = 135 \cdot 0.8 = 108$ not needed]

$SE_{sam} = \sqrt{135} \cdot 0.4 = 11.6 \cdot 0.4 = 4.65$ 2

$SE_{\%} = \frac{4.65}{135} \cdot 100\% = 3.44\%$ 2

s.u.: $\frac{70\% - 80\%}{3.44\%} = -2.91$ 4



area from −2.90 to 2.90: 99.63% 2

area below −2.90: $\frac{100\% - 99.63\%}{2} = \underline{\underline{0.185\%}}$ 2

Please turn over!

The chance that at most 70% of the students in the sample of 135 students are satisfied with the President is about 0.185%.

Statistics 1040, Sections 003 & 004, Quiz 10 (20 Points)

April 11, 2003

Your Name: ____________________

based on: Stat 1040, Spring 1999, Final Test, May 3, 1999, Question 1.a,b,

Question 1: Confidence Intervals (**20 Points**)

In a school district with 1500 kindergarten children, the heights of 68 randomly chosen children are measured. The average height of these 68 children is 49.7 inches with an SD of 2.7 inches. Suppose that the heights of kindergarten children are known to follow the normal curve.

Show your work!

−1 for each calculation error

1. (**12 Points**) If possible, find a 89%-confidence interval for the average height of all 1500 kindergarten children in that school district. If this is not possible, explain why not.

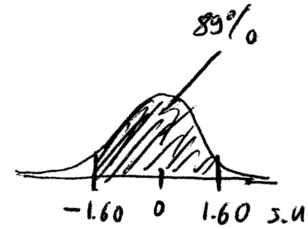
10 for "no" & explanation

10 for work with %

box: unknown

box avg = sample avg = 49.7 ①

box SD = sample SD = 2.7 ① } bootstrap



$SE_{sum} = \sqrt{68} \cdot 2.7 = 8.25 \cdot 2.7 = 22.26$ ②

$SE_{avg} = \frac{22.26}{68} = 0.327$ ②

It is possible to construct this CI since the data follows the normal curve and the average also will follow the normal curve.

89% CI: $49.7 \pm 1.60 \cdot 0.327 = 49.7 \pm 0.5232 =$ 49.18 inches to 50.22 inches ① ① ① ② ①

2. (**8 Points**) Approximately 89% of all the kindergarten children in that school district have heights in the interval 45.38 inches to 54.02 inches (which is centered around the average). If it is not possible to determine this interval, explain why not.

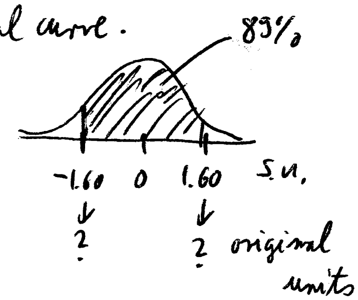
−6 for "no" & explanation

−5 for "yes" (but no work or result found)

−5 for result found.

It is possible to construct this interval since the data follows the normal curve.

s.u.	original units
−1.60	$-1.60 \cdot 2.7 + 49.7 = -4.32 + 49.7 = 45.38$ inches
1.60	$1.60 \cdot 2.7 + 49.7 = 4.32 + 49.7 = 54.02$ inches



or: 89% of all children in interval avg $\pm 1.60 \cdot SD = 49.7 \pm 1.60 \cdot 2.7 = 45.38$ inches to 54.02 inches

Please turn over!

Statistics 1040, Section 003 & 004, Quiz 11 (20 Points)

April 18, 2003

Your Name: ____________________

based on: Stat 1040, Fall 1998, Final Test, December 15, 1998, Question 10(a)

Question 1: Tests of Significance (**20 Points**)

A high school teacher working at an inner city high school is concerned about the time students spend working after school. She randomly selects 12 of her students and finds the average time spent working after school is 13.1 hours per week, with an SD of 10.5 hours per week.
If the national average is 10.7 hours per week, and assuming that the hours worked follow the normal curve, conduct an appropriate test to see whether this teacher's students work, on average, **longer** than those in the nation as a whole.

-1 for each calculation error

1. (3 points) State the null and the alternative hypothesis for this problem, in words and in terms of the box model. 1

   -2 if flipped

   null: avg for students at this high school is the same as the national avg, i.e. box avg = 10.7 1/2

   alternative: avg for these students is higher than the national avg, i.e., box avg > 10.7 1/2 1

2. (5 points) Calculate the appropriate test statistic.

   t-test: observed (avg): 13.1
   expected (avg): 10.7
   $SD^+ = 10.5 \cdot \sqrt{\frac{12}{11}} = 10.5 \cdot 1.044 = 10.97$ 2
   $SE_{sum} = \sqrt{12} \cdot 10.97 = 3.46 \cdot 10.97 = 37.96$ 1

   $SE_{avg} = \frac{37.96}{12} = 3.16$ 1
   $t = \frac{13.1 - 10.7}{3.16} = 0.76$ 1

3. (5 points) Obtain the (approximate) P–value (use the appropriate table!).

   -4 for normal table

   t-table with 11 degrees of freedom: 1 0.70 → 25% ; 0.76 ; 1.36 → 10% 2 } → P-value is between 10% and 25% 2

4. (5 points) State your conclusions in terms of rejecting (or not rejecting) the null hypothesis and in your own words.

   3 do not reject the null (P-value > 5%);
   2 there is not enough evidence to say that on avg students work longer at this high school than the national avg, i.e., these students work 10.7 hours on average

5. (2 points) Explain briefly why you chose this particular test to answer the question.

   -2 for z-test

   t-test since: • sample size = 12 < 30 1
   • box SD unknown (only sample SD known) 1/2
   • assumption that data follows normal curve 1/2

Please turn over!

Quiz 12 - Solutions

-15 for incorrect test

Question 1:

1/ null: use of hypertension drugs and developing cancer are independent, i.e. boxes are the same (2)

alternative: use of hypertension drugs and developing cancer are dependent, i.e., at least one box differs (2)

2/ χ^2 - test for independence:

obs ¦ exp	Beta		ACE		Calcium		Total
yes	28	34.5	6	10	27	16.5	61
No	396	389.5	118	114	175	185.5	689
Total	424		124		202		750

expected: $\frac{424 \cdot 61}{750} = 34.5$ $\quad \frac{124 \cdot 61}{750} = 10$ $\quad \frac{202 \cdot 61}{750} = 16.5$ (4)

389.5 114 185.5

$$\chi^2 = \frac{(28-34.5)^2}{34.5} + \frac{(6-10)^2}{10} + \frac{(27-16.5)^2}{16.5} + \frac{(396-389.5)^2}{389.5} + \frac{(118-114)^2}{114} + \frac{(175-185.5)^2}{185.5}$$

$= 10.35$ (4)

3/ $df = (2-1)\cdot(3-1) = 2$ (2)

χ^2-statistic is 10.35: above $\underset{\downarrow\ 1\%}{9.21}$

P-value < 1% (2)

4/ Conclusion:

- reject the null (P-value < 1%) (2)
- result is highly statistically significant (1)
- there is high evidence that use of hypertension drugs and developing cancer are dependent (1)

Question 2:

−8 for incorrect test

box A: urban

$$\text{sample }\%_A = \frac{63}{100} = 63\%$$

box B: suburban

$$\text{sample }\%_B = \frac{59}{110} = 53.6\%$$

1, null: no difference in percentage of residents who favor nuclear plant, i.e., 1

$$\text{box }A\% - \text{box }B\% = 0$$

alternative: difference in percentage of residents who favor nuclear plant, i.e., 1

$$\text{box }A\% - \text{box }B\% \neq 0$$

[two-tailed test!]

2, 2-sample z-test:

$$SD\ box_A = \sqrt{\frac{63}{100}\cdot\frac{37}{100}} = 0.48 \quad \tfrac{1}{2} \qquad SD\ box_B = \sqrt{\frac{59}{110}\cdot\frac{51}{110}} = 0.50 \quad \tfrac{1}{2}$$

$$SE\ sum_A = \sqrt{100}\cdot 0.48 = 4.8 \quad \tfrac{1}{2} \qquad SE\ sum_B = \sqrt{110}\cdot 0.50 = 5.2 \quad \tfrac{1}{2}$$

$$SE_{\%}\ A = \frac{4.8}{100}\cdot 100\% = 4.8\% \quad \tfrac{1}{2} \qquad SE_{\%}\ B = \frac{5.2}{110}\cdot 100\% = 4.7\% \quad \tfrac{1}{2}$$

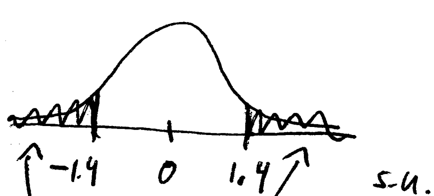
$$SE_{diff\ \%} = \sqrt{4.8\%^2 + 4.7\%^2} = 6.7\% \quad 1$$

observed % (difference): $63\% - 53.6\% = 9.4\%$

expected % (difference): 0%

$$z = \frac{9.4\% - 0\%}{6.7\%} = 1.4 \quad 1$$

3, P-value:



P-value: both "tails" together : $100\% - 83.85\% = 16.15\%$ 1

(do not divide by 2 !)

4, Conclusion:

- do not reject the null (P-value > 5%) 1
- there is no difference in percentage of residents who favor nuclear plant 1

Question 3:

-8 for incorrect test

box A: public

avg_A: 12.2

SD_A: 10.5

sample $size_A$: 1000

box B: private

avg_B: 9.2

SD_B: 9.9

sample $size_B$: 1000

1/ null: avg hours of work for pay are the same, i.e., $avg_A - avg_B = 0$ (1)

alternative: avg hours of work for pay are different, i.e., $avg_A - avg_B \neq 0$ (1)

[two-tailed test!]

2/ 2-sample z-test:

observed (difference): $12.2 - 9.2 = 3.0$

expected (difference): 0

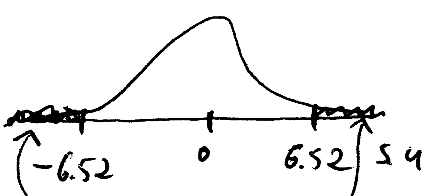
$SE_{sum_A} = \sqrt{1000} \cdot 10.5 = 332.0$ (1/2) $SE_{sum_B} = \sqrt{1000} \cdot 9.9 = 313.1$ (1/2)

$SE_{avg_A} = \frac{332.0}{1000} = 0.332$ (1/2) $SE_{avg_B} = \frac{313.1}{1000} = 0.313$ (1/2)

$$SE_{diff} = \sqrt{0.332^2 + 0.313^2} = 0.46 \quad (1)$$

$$z = \frac{3.0 - 0}{0.46} = 6.52 \quad (1)$$

3/ P-value:



P-value: both "tails" together: 100% - almost 100% = almost 0% (1)

(do not divide by 2!)

4/ Conclusion:

- reject the null (P-value < 1%) (1)
- result is highly statistically significant (1)
- there is high evidence that the avg hours of work for pay are different, (1/2) typically, "students in private universities come from wealthier families, (1/2) and have more support from home" (see FPP, p. A-96, Q7.)