Statistics 1040, Section 004, Quiz 1 (20 Points)

September 5, 2003

Your Name: _____

from: Stat 1040, Fall 1999, Final Test, Friday December 17, 1999, Question 1

Question 1: Observational Studies and Experiments (**10 Points**)

As a part of a study on exercise and health, a group of 800 people was followed for 6 years. At the beginning of the study, the researchers asked each person whether they exercised regularly or not. At the end of the study, the researchers measured several health–related variables, and in doing so, they noticed that the death rate for the exercise group was much lower than for the no–exercise group. Answer the following two questions:

1. (**5 Points**) Is this an observational study or a controlled experiment? Circle your answer and **explain**. (3) ⑤-② for explanation

   Worklook:

   They were asked what they did- they weren't told what to do.

2. (**5 Points**) Does this result necessarily imply that if people who do not exercise start to exercise regularly, they will live longer, on average, than if they do not? Explain clearly.

   Worklook: (3) ⑤-② for explanation

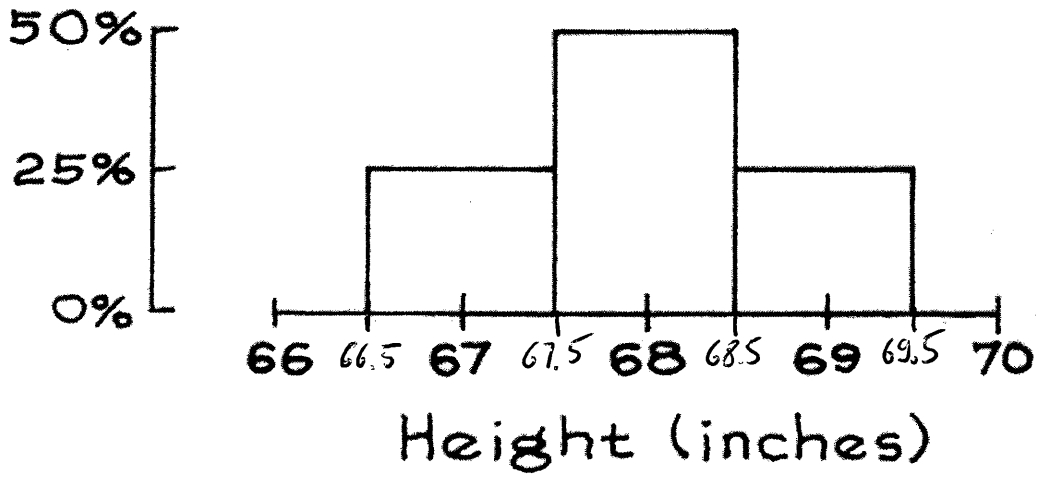
   No- this result merely shows that the type of people who choose to exercise differ from the type who do not. This is an observational study and there could be many confounding factors. For example, people who exercise might smoke less and it might be smoking that's harmful not exercise that's beneficial. Taking a smoker & making them exercise might have no impact.

Please turn over!

from: FPP, p.52, Review Exercise 6

Question 2: Histograms (**10 Points**)

(Hypothetical) In one study, 100 people had their heights measured to the nearest eighth of an inch. A histogram for the results is shown below. Two of the following lists have this histogram. Which ones, and why? (**Explain briefly!**)



1. 25 people, 67 inches tall; 50 people, 68 inches tall; 25 people, 69 inches tall.
2. 10 people, $66\frac{3}{4}$ inch tall; 15 people, $67\frac{1}{4}$ inches tall; 50 people, 68 inches tall; 25 people, 69 inches tall.
3. 30 people, 67 inches tall; 40 people, 68 inches tall; 30 people, 69 inches tall.

(2) for each correctly identified list

(0) – (4) for explanation

Worklook:

Lists 1. and 2. but not 3.

Explanation:

Histogram indicates: 25% of measurements in 66.5" to 67.5" class
50% of measurements in 67.5" to 68.5" class
25% of measurements in 68.5" to 69.5" class

List 1: $\frac{25}{100}$ = 25% are 67", i.e., in 66.5" to 67.5" class ✓
$\frac{50}{100}$ = 50% are 68", i.e., in 67.5" to 68.5" class ✓
$\frac{25}{100}$ = 25% are 69", i.e., in 68.5" to 69.5" class ✓
} histogram represents this data

List 2: $\frac{10}{100}$ = 10% are $66\frac{3}{4}$" and $\frac{15}{100}$ = 15% are $67\frac{1}{4}$";
Combined 10% + 15% = 25% are in 66.5" to 67.5" class ✓
other classes same as for List 1 ✓
} histogram represents this data

List 3: e.g. $\frac{30}{100}$ = 30% are 67", i.e., in 66.5" to 67.5" class –, but there should only be 25%

Statistics 1040, Section 004, Quiz 2 (20 Points)

September 12, 2003

Your Name: _____

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

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

-1 each calculation error

	Numbers	Average	Standard deviation
List 1:	17, 17, 17, 17, 17	17 (3)	0 (2)
List 2:	15, 16, 17, 18, 19	17 (2)	$\sqrt{2} \approx 1.4$ (4)

Show your work! Use formulas provided on the back where necessary.

List 1: Nothing to calculate! Since all numbers are identical (17), 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) $\text{avg} = \frac{15+16+17+18+19}{5} = \frac{85}{5} = 17$

2) $15 - 17 = -2$
$16 - 17 = -1$
$17 - 17 = 0$
$18 - 17 = 1$
$19 - 17 = 2$

3) $(-2)^2 = 4$
$(-1)^2 = 1$
$0^2 = 0$
$1^2 = 1$
$2^2 = 4$

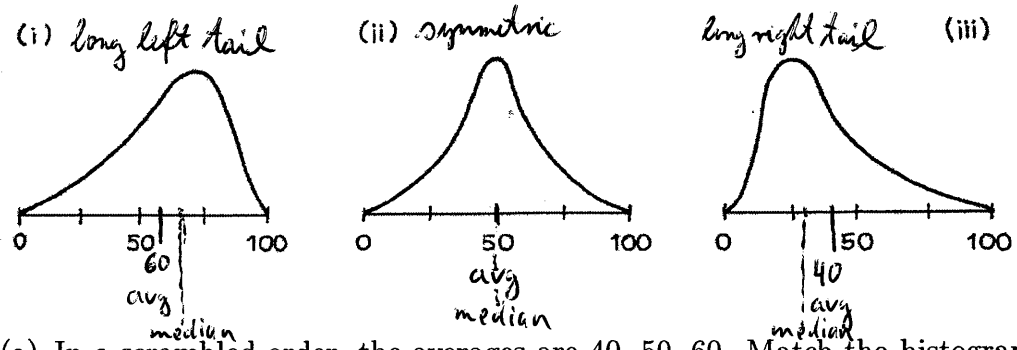
4) $\frac{4+1+0+1+4}{5} = \frac{10}{5} = 2.0$

5) $SD = \sqrt{2.0} = 1.414 \approx 1.4$

Please turn over!

from: FPP, p. 75, Review Exercise 6, a) b) d)

2. **(10 Points)** Below are sketches of histograms for three lists.



(a) In a scrambled order, the averages are 40, 50, 60. Match the histograms with averages:

Histogram (i): average = 60

Histogram (ii): average = 50

Histogram (iii): average = 40

(1) for each correctly identified average / histogram

(b) Match the histograms with the description (circle your answer):

- The median is less than the average. Histogram (i) (ii), or (iii) [long right tail]
- The median is about equal to the average. Histogram (i), (ii) or (iii). [symmetric]
- The median is bigger than the average. Histogram (i), (ii), or (iii) [long left tail]

(c) The SD for histogram (i) is a lot smaller than that for histogram (iii). **True** or **false**? Circle your answer and explain:

(2)

The two histograms are almost mirror images (2) and have about the same SD.

Formulas:

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

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

Statistics 1040, Section 004, Quiz 3 (20 Points)

September 19, 2003

-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

Your Name: _____

+3 for each correct graph (and nothing else)

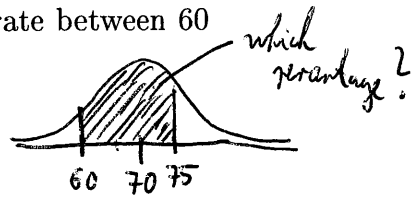
Question 1: Normal Approximation for Data (**20 Points**)

The histogram of resting pulse rates of men follows approximatelly the normal curve with an average of 70 beats per minute and a standard deviation of 8 beats per minute.

Answer the questions below.

Show all your work & use the Normal Table on the back!

1. (**10 Points**) About what percentage of men have a resting pulse rate between 60 and 75?



Answer: about 63.65 % [or 62.01% - based on rounding]

① convert 60 and 75 into standard units:

$\frac{60-70}{8} = -1.25$ s.u. $\quad \frac{75-70}{8} = 0.625 \approx 0.65$ s.u. [or ≈ 0.60]

② area between -1.25 and 1.25: 78.87%

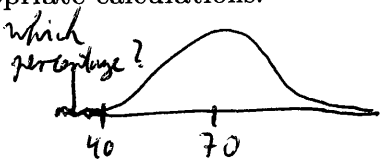
area between -0.65 and 0.65: 48.43% [-0.60 to 0.60: 45.15%]

③ area between -1.25 and 0.65: $\frac{78.87\%}{2} + \frac{48.43\%}{2} = 63.65\%$

[or $\frac{78.81\%}{2} + \frac{45.15\%}{2} = 62.01\%$]

2. (**10 Points**) Would it be unusual for a man to have a resting pulse rate less than 40? Yes or No? Explain and support your answer with appropriate calculations.

(3)



2 possible ways to answer this:

i) Note that 40 is more than 3.5D below the average.

So less than $\frac{0.3\%}{2} = 0.15\%$ of all men will have such a low resting pulse rate.

ii) $\frac{40-70}{8} = -3.75$ s.u.

area between -3.75 and 3.75: 99.982%

area below -3.75: $\frac{100\% - 99.982\%}{2} = 0.009\%$ - very unusual to have such a low pulse rate

Statistics 1040, Section 004, Quiz 4 (20 Points)

September 26, 2003

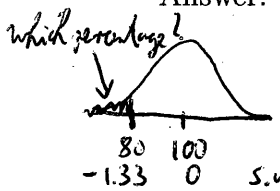
Your Name: _____

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

The scores of a reference population on a Wechsler Intelligence Scale for Children (WISC) approximately follow the normal curve with an average of 100 and a standard deviation of 15 points. **Show your work!**

1. (**6 Points**) What percent of this population have WISC scores below 80?

   Answer: about 8.85 %



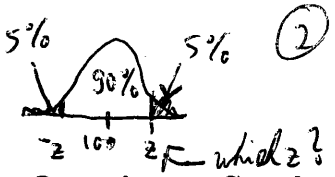
① convert 80 into s.u.: $\frac{80-100}{15} = -1.33$ (2)

② area between −1.35 and 1.35: 82.30% (2)

③ area below −1.35: $\frac{100\% - 82.30\%}{2} = \frac{17.70\%}{2} = 8.85\%$ (2)

2. (**6 Points**) What WISC score is needed to place a child in the top 5% of the population?

   Answer: a score of about 125



(2)

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

② convert 1.65 s.u. into original units:

$1.65 \cdot 15 + 100 = 124.75$ (2)

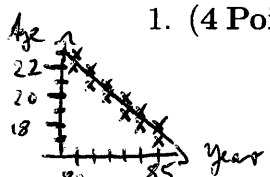
(1.65 · SD + avg)

Question 2: Correlation (**8 Points**)

Chapter 8, Exercise Set B, Question 7, page 130

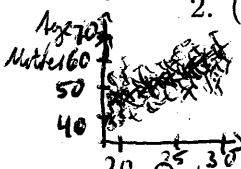
Investigators are studying registered students at a university. The students fill out questionnaires giving their year of birth, age (in years), age of mother, and so forth. Fill in the blanks, using the options given below, and **explain briefly**.

1. (**4 Points**) The correlation between student's age and year of birth is (b): nearly −1



Book: "The older you are, the earlier you were born. But there is some fuzz, depending on whether your birthday is before or after the day of the questionnaire."

2. (**4 Points**) The correlation between student's age and mother's age is (e): somewhat positive



The older the student, the older the mother. However, there is a lot of variation when to give birth (ages between 12 to 50 are possible).

Options: (a) -1 (b) nearly -1 (c) somewhat negative (d) 0
(e) somewhat positive (f) nearly 1 (g) 1

"slightly" wrong answers:
1: (a), (c), (f) 1
2: (d), (f)

0: "totally" wrong answer, no explanation
1: "totally" wrong answer, some explanation
2: "slightly" wrong answer, no explanation
3: "slightly" wrong answer, some explanation

Statistics 1040, Section 004, Quiz 5 (20 Points)

October 10, 2003

Your Name: _____

from: Stat 1040, Spring 2000, Midterm 1, February 17, 2000, Question 5

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

In a study of a large number of California boys it was found that:

x: Average height at age 6 = 46 inches, SD = 1.7 inches

y: Average height at age 18 = 70 inches, SD = 2.5 inches, r = 0.8

The scatter plot for the heights was football–shaped. **Show your work!**

1. (**10 Points**) Find the equation of the regression line for predicting height at age 18 from height at age 6.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.8 \cdot \frac{2.5}{1.7} = 1.176$$

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 70 - 1.176 \cdot 46 = 70 - 54.1 = 15.9$$

regression equation: $\boxed{\text{height at 18} = 15.9 + 1.176 \cdot \text{height at 6}}$

or $\boxed{y = 15.9 + 1.176 \cdot x}$

2. (**5 Points**) Use the regression equation from part 1. to predict the height at age 18 of a boy who is 44 inches tall at age 6. The predicted height is: 67.6 inches

$$\text{height at 18} = 15.9 + 1.176 \cdot 44 = 15.9 + 51.7 = 67.6 \text{ inches}$$

3. (**5 Points**) Find the r.m.s. error for predicting height of boys at age 18 from height at age 6. The r.m.s. error is: 1.5 inches

$$\text{r.m.s. error} = \sqrt{1-r^2} \cdot SD_y = \sqrt{1-0.8^2} \cdot 2.5 = \sqrt{1-0.64} \cdot 2.5$$

$$= \sqrt{0.36} \cdot 2.5 = 0.6 \cdot 2.5 = 1.5 \text{ inches}$$

(i.e., on average, the predicted height at 18 will be 1.5 inches off the actual value)

Please turn over!

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., 15.9+1.176·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, Section 004, Quiz 6 (20 Points)

October 17, 2003

Your Name: _____

from: Stat 1040, Spring 2003, Final Test, April 28, 2003, Question 5

Question 1: Chance/Probability (**16 Points**)

A drawer of socks contains 24 socks of which 5 are black, 10 are blue, and 9 are green. In the dark, a child chooses two socks at random to wear to school.

1. (**4 Points**) What is the chance that the first sock is green?

- 1 each calculation error or no final result

first green: $\frac{9}{24} = 0.375 = 37.5\%$

2. (**4 Points**) What is the chance that the first sock is green or blue?

first green: $\frac{9}{24}$
first blue: $\frac{10}{24}$
> mutually exclusive

first green or blue: $\frac{9}{24} + \frac{10}{24} = \frac{19}{24} = 0.792 = 79.2\%$

Addition Rule

3. (**4 Points**) What is the chance that the first two socks are both green?

first green: $\frac{9}{24}$
second green, given first green: $\frac{8}{23}$
> dependent

first and second green: $\frac{9}{24} \cdot \frac{8}{23} = \frac{3}{23} = 0.130 = 13.0\%$

Multiplication Rule

4. (**4 Points**) What is the chance that neither of the two socks is green?

first not green: $\frac{15}{24}$
second not green, given first not green: $\frac{14}{23}$
> dependent

first and second not green: $\frac{15}{24} \cdot \frac{14}{23} = \frac{210}{552} = 0.380 = 38.0\%$

Multiplication Rule

Question 2: Chance/Probability (**4 Points**)

There is something wrong in each of the following statements. Explain briefly what is wrong:

1. (**2 Points**) The probability that a randomly selected driver will be wearing a seat belt is .75, whereas the probability that he or she will not be wearing one is .30.

"wearing a seat belt" and "not be wearing one" are opposites. So, for example, prob of "not be wearing one" = 1 - prob of "wearing a seat belt" = 1 - .75 = .25.

2. (**2 Points**) The probability that a randomly selected car is red is 1.20.

But it is stated as 0.30. Incorrect!

Probabilities are numbers between 0 and 1.
A probability of 1.20 is incorrect.

Statistics 1040, Section 004, Quiz 7 (20 Points)

October 24, 2003

Your Name: _____

Question 1: Box Models, EV, and SE (**16 Points**)

You are participating in a new game that consists of tossing a 10–sided die, with sides numbered from 1 to 10. The die is fair, i.e., it has the same chance of landing on any side. Every time the die shows a number that is a multiple of 3 (i.e., 3, 6, or 9) you win \$3, otherwise you lose \$1, except when the die lands on 10, in which case you win (or lose) nothing (\$0). Assume you are tossing this die 200 times.

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

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

3	3	3	0	-1	-1	-1	-1	-1	-1

Number of draws: 200

or | 3 × 3 | 1 × 0 | 6 × -1 |

#draws: 200

2. (**6 Points**) Find the expected value of your gain/loss.

$$\text{box average} = \frac{3 \cdot 3 + 1 \cdot 0 + 6 \cdot (-1)}{10} = \frac{3}{10} = 0.3$$

$$EV_{sum} = 200 \cdot 0.3 = 60 \quad [\$]$$

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

3. (**6 Points**) Find the standard error of your gain/loss.

$$\text{box SD} = \sqrt{\frac{3 \cdot (3-0.3)^2 + 1 \cdot (0-0.3)^2 + 6 \cdot (-1-0.3)^2}{10}}$$

$$= \sqrt{\frac{3 \cdot 2.7^2 + 1 \cdot (-0.3)^2 + 6 \cdot (-1.3)^2}{10}}$$

$$= \sqrt{\frac{3 \cdot 7.29 + 0.09 + 6 \cdot 1.69}{10}}$$

$$= \sqrt{\frac{32.1}{10}} = \sqrt{3.21} = 1.79$$

Please turn over!

$$SE_{sum} = \sqrt{200} \cdot 1.79 = 14.14 \cdot 1.79 = 25.31 \quad [\$]$$

from: FPP, p. 285, Review Exercise 1

Question 2: Law of Averages (**4 Points**)

A box contains 10,000 tickets: 4,000 $\boxed{0}$'s and 6,000 $\boxed{1}$'s. And 10,000 draws will be made at random with replacement from this box. Which of the following best describes the situation, and why? **Explain briefly.**

1. The number of 1's will be 6,000 exactly.
2. The number of 1's is very likely to equal 6,000, but there is also some small chance that it will not be equal to 6,000.
3. The number of 1's is likely to be different from 6,000, but the difference is likely to be small compared to 10,000.

-3 for 1. or 2.

Wakelock: "Option 3. is the best because the chance error is not likely to be exactly zero, but it is likely to be small compared to the number of draws."

-1 for incorrect explanation

Formulas:

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

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

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

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

Statistics 1040, Section 004, Quiz 8 (20 Points)

October 31, 2003

part b) -1 if slightly incorrect number of 0/1's in box
-2 if box given as 0 1
-3 if box contains something other than 0/1's
-1 if # draws missing/incorrect

Your Name: _____

from: Stat 1040, Summer 2000, Test 2, July 17, 2000, Question 4.

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

Suppose it is known that 10% of all people in Utah have a specific blood type. Suppose I take a random sample of 500 Utah residents. **Show your work!**

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

1: specific blood type (we are interested in)
0: all other blood types

1	9 × 0

draws: 500

or

10 × 1	90 × 0

draws: 500

2. (**8 Points**) The expected number of Utah residents in this sample of 500 who have that specific blood type is 50. The corresponding SE is 6.7.

box average = fraction of 1's = $\frac{1}{10}$

box SD = $\sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{\frac{1}{10} \cdot \frac{9}{10}} = \sqrt{\frac{9}{100}} = \frac{3}{10} = .3$

$EV_{sum} = 500 \cdot \frac{1}{10} = 50$

$SE_{sum} = \sqrt{500} \cdot .3 = 22.36 \cdot .3 = 6.7$

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

3. (**8 Points**) Using the normal curve, the chance that **fewer than** 40 Utah residents in this sample have that blood type is 6.68%.

s.u.: $\frac{40-50}{6.7} = -1.49 \approx -1.5$

area between -1.5 and 1.5: 86.64%

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

-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, Section 004, Quiz 9 (20 Points)

November 14, 2003

Your Name: _____

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

There are approximately 20,000 students at USU. A random sample of 500 students had been asked whether or not they were satisfied with President Kermit Hall. The results of the survey showed that 400 of the 500 students in the sample were satisfied with President Kermit Hall.

-2 each calculation error

1. **(14 Points)** Find a 99.5% confidence interval for the percentage of all USU students who are satisfied with President Hall.

   -1 if % forgotten
   -1 if box indicated (box = population is unknown!)

   Show your work!

   (via Bootstrap procedure)

   sample % = $\frac{400}{500} = 0.8 = 80\%$ (2) = population %

   SD box = $\sqrt{0.8 \cdot 0.2} = \sqrt{0.16} = 0.4$ (2)

   $SE_{sum} = \sqrt{500} \cdot 0.4 = 22.36 \cdot 0.4 = 8.94$ (2)

   $SE_{\%} = \frac{8.94}{500} \cdot 100\% = 1.788\% \approx 1.8\%$ (2)

   99.5% CI: $80\% \pm \underset{\text{from table: related to 99.5\%}}{2.8} \cdot \underset{SE\%}{1.8\%} = 80\% \pm 5.04\%$ (2)

   $\approx 75\%$ to 85% (4)

2. **(6 Points)** Would it still be possible to construct a 99.5% confidence interval for the percentage of all USU students who are satisfied with President Hall if the survey showed that 498 (out of 500) students in the sample were satisfied with President Hall? **Yes** or **No**? (2) And explain why or why not?

   You do not have to construct the actual confidence interval, but you do have to show calculations necessary to support your answer.

   sample % = $\frac{498}{500} = 0.996 = 99.6\%$ (2)

   Please turn over!

   99.6% is too close to 100%. We cannot construct a 99.5% CI here (2)

Statistics 1040, Section 004, Quiz 10 (20 Points)

November 19, 2003

Your Name: _____

from: FPP, p. 426, Review Exercise 3

Question 1: The Accuracy of Averages (**20 Points**)

A real estate office wants to make a survey in a certain town, which has 50,000 households, to determine how far the head of household has to commute to work. A simple random sample of 1,000 households is chosen, the occupants are interviewed, and it is found that, on average, the heads of the sample households commuted 8.7 miles to work; the SD of the distances was 9.0 miles. (All distances are one–way; if someone isn't working, the commute distance is defined to be 0.)

-2 for each calculation error
-6 if working with $SE_{\%}$ or SE_{sum} in final stage

Answer the questions below and **show your work!**

1. (**8 Points**) The average commute distance of all 50,000 heads of households in the town is estimated as 8.7 miles, and this estimate is likely to be off by 0.28 miles or so.

Assume: sample avg = 8.7 miles = box avg (2)
sample SD = 9.0 miles = box SD (2)
$SE_{sum} = \sqrt{1{,}000} \cdot 9.0 = 31.6 \cdot 9.0 = 284.4$ miles (2)
$SE_{avg} = \frac{284.4}{1000} = 0.28$ miles (2)

[Note that the data itself do not follow the normal curve; just 1 SD below the avg are the negative numbers – and negative driving distances are impossible; but the average of 1000 draws will follow the normal curve!]

2. (**12 Points**) If possible, find a 95%–confidence interval for the average commute distance of all heads of households in the town. If this isn't possible, explain why not.

-10 for "it is not possible"

(4) It is possible to find a 95%-CI:

sample avg (2) ± multiplier (2) · SE_{avg} (2)
= 8.7 miles ± 2 · 0.28 miles
(2 ↑ related to 95%)
= 8.7 miles ± 0.56 miles
= 8.14 miles to 9.26 miles [1] (2)

Please turn over!

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

November 24, 2003

Your Name: _____

from: Stat 1040, Fall 1999, Final Test, December 17, 1999, Question 3

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

-2 for each calculation error

A journal article claims that 60% of North American adults are too sedentary. Suppose you think a lower percentage of Cache Valley adults are too sedentary. To test your belief, you take a simple random sample of 120 Cache Valley adults and find that only 68 of them are too sedentary. Test to see if your belief is correct. (You may assume that everybody is using the same definition of "too sedentary", although in practice this would be contentious). **Show your work!**

-2 if null, alt. mapped

Z-test:

1. (**5 Points**) State the null and the alternative hypothesis for this problem, in words and in terms of the box model.

(1) null: Cache Valley adults are also "too sedentary", i.e., box % = 60% (2) (1)

alternative: fewer Cache Valley adults are "too sedentary", i.e., box % < 60% (2)

2. (**5 Points**) Calculate the appropriate test statistic.

If the null is true, the sample % should be like the % of 1's in 120 draws from [60×1 40×0]:

(2) observed % = $\frac{68}{120} \cdot 100\% = 56.7\%$

expected % = 60% = EV %

$SD_{box} = \sqrt{0.6 \cdot 0.4} = \sqrt{0.24} = 0.49$ (1)

$SE_{sum} = \sqrt{120} \cdot 0.49 = 10.95 \cdot 0.49 = 5.37$ (1)

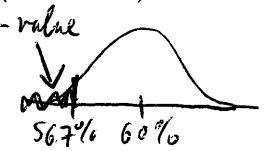
$SE_{\%} = \frac{5.37}{120} \cdot 100\% = 4.475\%$ (1)

$z = \frac{56.7\% - 60\%}{4.475\%} = -0.74$ (2)

(obs % ↓, EV % ↓, SE % ↑)

3. (**5 Points**) Obtain the P–value (use the normal table on the back).

(3)



area between −0.75 to 0.75: 54.67% (2) (1)

area below −0.75: $\frac{100\% - 54.67\%}{2} = 22.67\%$

P-value ≈ 23% (2)

4. (**5 Points**) State conclusions in terms of rejecting the null hypothesis and in your own words.

(3)

(4) ⇝ • do not reject null (P-value > 5%)

• Cache Valley adults are also "too sedentary" (i.e., no evidence that Cache Valley adults are less sedentary than the national figure) (2)

Please turn over!

Quiz 12 – Solutions

−15 for incorrect test

Question 1:

box A: Viagara box B: Placebo

$\text{sample \%}_A = \frac{117}{734} = 15.9\%$ $\text{sample \%}_B = \frac{29}{725} = 4.0\%$

1) null: no difference in percentage of men that experience headaches, i.e., (2)

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

alternative: higher percentage of men that take Viagara experience headaches, i.e., (2)

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

2) 2-sample z-test:

$\text{SD box}_A = \sqrt{\frac{117}{734} \cdot \frac{617}{734}} = 0.366$ (1) $\text{SD box}_B = \sqrt{\frac{29}{725} \cdot \frac{696}{725}} = 0.196$ (1)

$\text{SE sam}_A = \sqrt{734} \cdot 0.366 = 9.92$ (1) $\text{SE}_{\text{sam}_B} = \sqrt{725} \cdot 0.196 = 5.28$ (1)

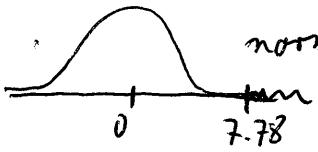
$\text{SE\%}_A = \frac{9.92}{734} \cdot 100\% = 1.35\%$ (1) $\text{SE\%}_B = \frac{5.28}{725} \cdot 100\% = 0.73\%$ (1)

$\text{SE diff\%} = \sqrt{1.35\%^2 + 0.73\%^2} = 1.53\%$ (2)

observed % (difference) = $15.9\% - 4\% = 11.9\%$ (1)

expected % (difference) = 0%

$z = \frac{11.9\% - 0\%}{1.53\%} = 7.78$ (2)

3) 

normal curve: 7.78 off the table, i.e., area above 7.78 is about 0%

⇝ p-value: about 0% (2)

4) Conclusion:

- reject the null (p-value < 1%) (1)
- result is highly statistically significant (1)
- there is high evidence that a higher percentage of men that take Viagara experience headaches than men not taking Viagara (1)

Question 2: -8 for incorrect test

1) null: mixe is Ok, i.e., peanuts : hazelnuts : cashews : pecans = 5 : 2 : 2 : 1 (1)

alternative: mixe is not Ok, i.e., peanuts : hazelnuts : cashews : pecans ≠ 5 : 2 : 2 : 1 (1)

2) χ^2-test:

Nut Type	obs	esep	$\frac{(obs-esp)^2}{esp}$
Peanuts	269	250	$\frac{(269-250)^2}{250} = 1.44$
Hazelnuts	112	100	$\frac{(112-100)^2}{100} = 1.44$
Cashews	74	100	$\frac{(74-100)^2}{100} = 6.76$
Pecans	45	50	$\frac{(45-50)^2}{50} = 0.50$
	500	500	

(2)

$\chi^2 = 1.44 + 1.44 + 6.76 + 0.50 = 10.14$ (1)

$df = 4 - 1 = 3$ (1)

3) $\chi^2 = 10.14$ is between 7.82 and 11.34
↓ 5% ↓ 1%

⇝ P-value is somewhere between 1% and 5% (2)

4) Conclusion:

- reject the null (P-value < 5%) (1/2)
- results is statistically significant (1/2)
- there is some evidence that the mixe is not Ok, i.e., the ratio of peanuts : hazelnuts : cashews : pecans is not 5:2:2:1 (1)

Question 3: -8 for incorrect test

1) null: high school seniors require usual amount of time, i.e., lose avg = 35 min (1)

alternative: high school seniors require different amount of time, i.e, lose avg ≠ 35 min (1)

2) t-test:
- sample size = 20 (< 30)
- SD of loss unknown (only SD of sample is known)
- times to complete quiz follow normal curve

two-tailed test since we ask for ≠ 35 min

observed (avg) = 33.1 min

expected (avg) = 35 min

$SD_+ = 4.3 \cdot \sqrt{\frac{20}{20-1}} = 4.4$ (1)

$SE_{sum} = \sqrt{20} \cdot 4.4 = 19.7$ (1)

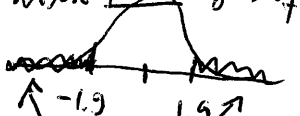
$SE_{avg} = \frac{19.7}{20} = 0.99 \approx 1.0$ (1)

$t = \frac{33.1 - 35}{1.0} = -1.9$ (1)

3) Note that the t-curve is symmetric around 0, so we have to look up 1.9 (instead of −1.9) in the table for the t-curve with 19 degrees of freedom:

t = 1.9 is between 1.73 and 2.09 (1)

↓ 5% ↓ 2.5%



P-value = both "tails" together

⇝ P-value: both "tails" together, i.e, between 5% and 10% (1)

4) Conclusion:
- do not reject the null (P-value > 5%) (1)
- there is not enough evidence to say that high school seniors require a time different from 35 min (1)

Note: A one-tailed test would reject the null!