

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

Friday, January 14, 2005

Your Name: \_\_\_\_\_

from: FPP, p. 25, Review Exercise 4

Question 1: Controlled Experiments/Observational Studies I (13 Points)

The Public Health Service studied the effects of smoking on health, in a large sample of representative households. For men and for women in each age group, those who never smoked were on average somewhat healthier than the current smokers, but the current smokers were on average much healthier than those who had recently stopped smoking.

- (6 Points) Why did they study men and women and the different age groups separately?

Workbook:

"They studied the groups separately to eliminate the effects of the confounding factors of age and gender." (4) (2)

- (7 Points) The lesson seems to be that you shouldn't start smoking, but once you've started, don't stop. Comment briefly.

Workbook:

"That is not an appropriate conclusion because there are confounding factors. For example, those who recently stopped smoking may have done so on doctor's orders, because they had severe health problems." (3) (4)

Please turn over!

from: FPP, p. 10-11, "Summary"

Question 2: Controlled Experiments/Observational Studies II (7 Points)

Fill the gaps in the following statements using the most appropriate words from the list below:

Statisticians want to know the effect of a treatment (or vaccine) (like the Salk vaccine) on a response (like getting polio). To find out, they compare the responses of a treatment group with a control group.

To make sure that the treatment group is like the control group, investigators put subjects into the treatment or the control group at random.

Whenever possible, the control group is given a placebo, which is neutral but resembles the treatment.

In a double-blind experiment, the subjects do not know whether they are in the treatment or in the control group; neither do those who evaluate the responses.

- ✓ placebo
- ✓ double-blind
- ✓ treatment group
- observational study
- ✓ random
- single-blind
- (✓) vaccine
- confounding factor
- objects
- ✓ control group
- controlled experiment
- ✓ subjects
- polio
- ✓ treatment

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

Friday, January 21, 2005

Your Name: \_\_\_\_\_

from: Stat 1040, Fall 2004, Final Test, December 15, 2004, Question 3

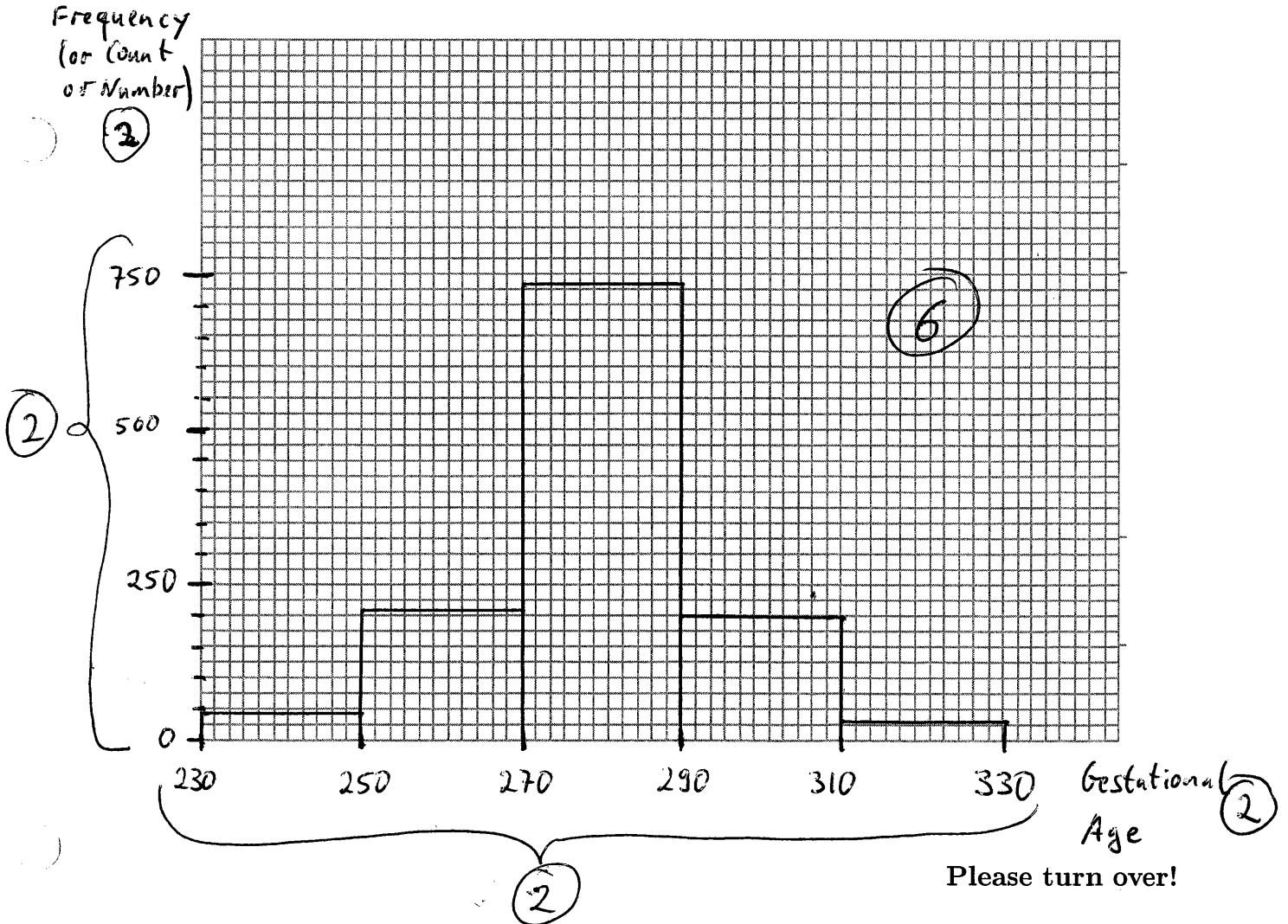
## Question 1: Histograms (14 Points)

The following table is for the gestational age of 1210 babies:

Note: class intervals are equally wide (20) - we could, but we don't have to use a density scale here.

Gestational Age	Number
230-250	47
250-270	206
270-290	731
290-310	199
310-330	27

Draw a histogram for these data on the graph paper provided. Make sure to label the axes.



**Question 2: Observational Studies / Controlled Experiments (6 Points)**

For each of the following studies, determine whether the study in question was a **randomized controlled experiment** or an **observational study** (circle the correct answer).

- Twenty male employees and twenty female employees participate in research designed to compare “attitudes towards the Social Security System” of men and women. Each individual responds to a series of questions on a survey. Mean scores are computed for men and for women.

randomized controlled experiment

observational study

2

- A researcher wants to learn whether regularly taking zinc supplements may reduce the risk of getting a cold. Volunteers in this study chose to (or chose not to) take a zinc supplement.

randomized controlled experiment

observational study

2

- A researcher wants to learn about whether computer simulations help students better understand statistical concepts. She puts the names of 20 volunteers into a box and randomly draws the names of 10 people who will use computer simulations to learn statistical concepts. The other 10 study participants will use a conventional approach, without computer simulations, to learn the same concepts.

2

randomized controlled experiment

observational study

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

Friday, January 28, 2005

Your Name: \_\_\_\_\_

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

1. (12 Points) Find the average, the median, and the standard deviation of the following list of numbers:

Numbers	Average	Median	Standard deviation
-8, 8, 8, -8	0 (3)	0 (3)	8

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

•  $avg = \frac{(-8) + 8 + 8 + (-8)}{4} = \frac{0}{4} = 0$

-1 for each calculation error  
-1 if no work shown/  
no explanation

• sorted list: -8      -8      8      8  
 $\frac{-8+8}{2} = \frac{0}{2} = 0 = \text{median}$

1) $avg = 0$ (1)	3) $(-8)^2 = 64$	4) $\frac{64+64+64+64}{4} = \frac{4 \cdot 64}{4} = 64$ (1)
2) $-8-0 = -8$	$8^2 = 64$	
$8-0 = 8$	$8^2 = 64$	5) $SD = \sqrt{64} = 8$ (2)
$8-0 = 8$	$(-8)^2 = 64$	
$-8-0 = -8$		

Please turn over!

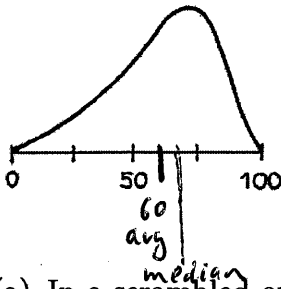
[Note: all numbers are equally far away from the average, but numbers are not equal; so the SD must be equal to this deviation from the average.]

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

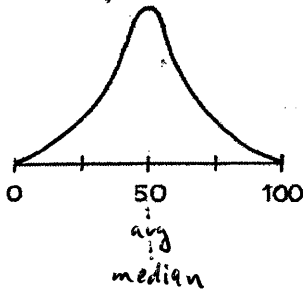
& Stat 1040, Fall 2003, Quiz 2, Sept. 12, 2003, Question 2

2. (8 Points) Below are sketches of histograms for three lists.

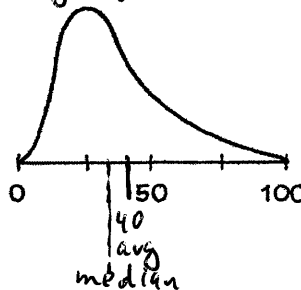
(i) long left tail



(ii) symmetric



long right tail (iii)



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

Histogram (i): average = 60 (1)

Histogram (ii): average = 50 (2)

Histogram (iii): average = 40 (1)

(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] (1)
- The median is about equal to the average. Histogram (i) (ii) or (iii). [symmetric] (2)
- The median is bigger than the average. Histogram (i) (ii), or (iii). [long left tail] (1)

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, Section 004, Quiz 4 (20 Points)

Friday, February 4, 2005

Your Name: \_\_\_\_\_

-2 for each calculation error  
+2 for correct graph (and nothing else)

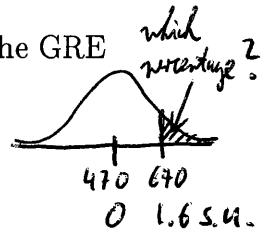
## Question 1: Normal Approximation for Data (20 Points)

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. For around 428,000 examinees who took the General GRE Test in 2001-02, the mean for the verbal ability portion of the exam was around 470 and the standard deviation was around 125 (<http://ftp.ets.org/pub/gre/994950.pdf>). Show your work!

- (7 Points) The percentage of examinees who scored more than 670 on the GRE test is roughly 5.48 %.

① convert 670 into standard units:

$$\frac{670 - 470}{125} = 1.6 \text{ s.u.} \quad \textcircled{3}$$



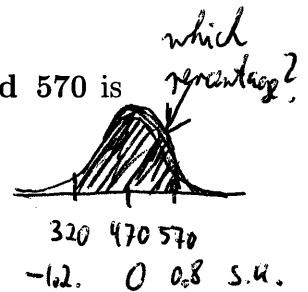
② area between -1.6 and 1.6: 89.04%  $\textcircled{2}$

③ area above 1.6:  $\frac{100\% - 89.04\%}{2} = \frac{10.96\%}{2} = \underline{5.48\%} \quad \textcircled{2}$

- (7 Points) The percentage of examinees who scored between 320 and 570 is about 67.31 %.

① Convert 320 and 570 into standard units:

$$\frac{320 - 470}{125} = -1.2 \text{ s.u.} \quad \textcircled{1} \quad \frac{570 - 470}{125} = 0.8 \text{ s.u.} \quad \textcircled{1}$$



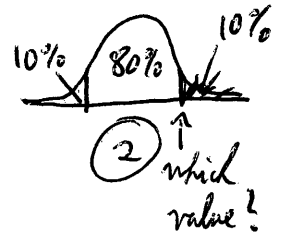
② area between -1.2 and 1.2: 76.99%  $\textcircled{1}$   
area between -0.8 and 0.8: 57.63%  $\textcircled{1}$

③ area between -1.2 and 0.8:  $\frac{76.99\%}{2} + \frac{57.63\%}{2} = \underline{67.31\%} \quad \textcircled{3}$

- (6 Points) In order to be among the top 10%, a student must have obtained a minimum GRE score of about 632.5.

① Find a value  $z$  such that the area between  $-z$  and  $z$  is about 80%.

$z = 1.30$  (gives 80.64%)  $\textcircled{2}$



② Transfer into original units:

$$1.30 \cdot 125 + 470 = \underline{632.5} \quad \textcircled{2}$$

Please turn over!

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

Friday, February 11, 2005

Grading Question 2:

"slightly" wrong answers:

1: (d)

2: (c), (e)

0: wrong answer, no explanation

1: wrong answer, some explanation

2: "slightly" wrong answer, no explanation

3: "slightly" wrong answer, some explanation OR

correct answer, no explanation

Your Name: \_\_\_\_\_

Question 1: Change Of Scale (12 Points)

Conversion of temperature from Celsius to Fahrenheit is another example of what statisticians call **change of scale**. The formula for conversion is

$$F^{\circ} = \frac{9}{5}C^{\circ} + 32^{\circ}$$

A group of people have an average body temperature of 37.0° Celsius, with a standard deviation of 0.2° Celsius.

1. (8 Points) If we translate these results into degrees Fahrenheit, the average temperature would be 98.6 degrees Fahrenheit, with a standard deviation of 0.36 degrees Fahrenheit.

$$\text{avg } F^{\circ} = \frac{9}{5} \cdot 37.0 + 32 = 66.6 + 32 = 98.6 \quad (4)$$

$$\text{SD } F^{\circ} = \frac{9}{5} \cdot 0.2 = 0.36 \quad (4)$$

2. (4 Points) Someone's temperature is 1.5 standard deviations above average on the Celsius scale. When converting this temperature to standard units for an investigator who is using the Fahrenheit scale, we have to report 1.5 standard units to this investigator.

Note: 1.5 SD above average on Celsius scale = 1.5 s.u. on Celsius scale  
= 1.5 s.u. on Fahrenheit scale

Question 2: Correlation (8 Points)

(grading: see top of page)

from: FPP, Chapter 8, p. 130, # 6 (→ Answer: p. A-56)  
& Stat 1040, Spring 2004, Quiz 5, Feb. 6, 2004, Question 2

1. (4 Points) If women always marry men who were five years older, the correlation between ages of husbands and wives would be (e) exactly 1. Choose one of the options below, and explain.

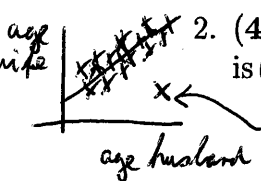
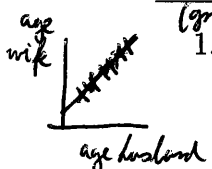
"All the points on the scatter diagram would lie on a line sloping up, so the correlation would be 1."

2. (4 Points) In reality, the correlation between ages of husbands and wives in the US is (d) close to 1. Choose one of the options below, and explain.

"Close to 1; this is like part (a), with some noise thrown into the data.  
Comment: In the March 1993 Current Population Survey, the correlation

- Options: (a) exactly -1 (b) close to -1 (c) close to 0  
(d) close to 1 (e) exactly 1

between the ages of the husbands and wives was about 0.95; the husbands were, on average, 2.7 years older than their wives."





# Statistics 1040, Section 004, Quiz 6 (20 Points)

Friday, February 25, 2005

Your Name: \_\_\_\_\_

*Based on: Stat 1040, Fall 1998, Final Test, December 15, 1998, Question 5 a)*

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

In a study, reading comprehension is tested for a large number of third grade students, once at the beginning of the school year and once at the end of the school year. During the school year, the students work on reading comprehension skills. The following results are obtained:

- x beginning-of-year: average score = 75; SD = 15;  
 y end-of-year: average score = 80; SD = 17;  $r = 0.6$ .

*in all parts:*

*-2 for each calculation error*

*-2 for each incorrect value used*

*-2 for swapping x and y*

The scatterplot of the data shows a football-shaped cloud. Show your work!

1. (10 Points) Find the equation of the regression line for predicting the end-of-year score from the beginning-of-year score.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.6 \cdot \frac{17}{15} = \underline{0.68} \quad (4)$$

*-2 if only part of the equation  
 i.e.  $29 + 0.68x$*

*-2 if not specifying x and y*

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 80 - 0.68 \cdot 75 = 80 - 51 = \underline{29} \quad (3)$$

$$\text{equation: } \boxed{\text{end-of-year score} = 29 + 0.68 \cdot \text{beginning-of-year score}} \quad \text{or} \quad \boxed{y = 29 + 0.68 \cdot x} \quad (3)$$

2. (5 Points) Use the regression equation from part 1. to predict the end-of-year score for a student who scored 85 on the beginning-of-year test.

The predicted end-of-year score is: 86.8

$$\text{end-of-year score} = 29 + 0.68 \cdot 85 = 29 + 57.8 = \underline{86.8}$$

(1) (1) (1) (2)

*-1 if correct result, but according to old method*

*-2 if result makes no sense at all!*

3. (5 Points) Find the r.m.s. error for predicting the end-of-year score from the beginning-of-year score.

The r.m.s. error is: 13.6

$$\begin{aligned} \text{r.m.s. error} &= \sqrt{1-r^2} \cdot SD_y \\ &= \sqrt{1-0.6^2} \cdot 17 \quad (3) \\ &= \sqrt{1-0.36} \cdot 17 \\ &= \sqrt{0.64} \cdot 17 \\ &= 0.8 \cdot 17 \\ &= \underline{13.6} \quad (2) \end{aligned}$$

Please turn over!

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

Friday, March 4, 2005

Your Name: \_\_\_\_\_

based on: Stat 1040, Spring 2003, February 28, 2003, Quiz 6, Question 1

## Question 1: Chance/Probability I (15 Points)

In a box of 15 chocolates, 5 are mint, 3 are orange, 5 are caramel, and 2 are cherry. I choose two chocolates at random (without replacement!).

Show your work!

-1 each calculation error  
(or no final result in %)  
-4 if % > 100% or % < 0%

1. (5 Points) What is the chance that the first is mint or orange?

The chance is 53.3 %.

first mint:  $\frac{5}{15}$

first orange:  $\frac{3}{15}$

> mutually exclusive  
addition rule (for mutually exclusive events)

first mint or orange:  $\frac{5}{15} + \frac{3}{15} = \frac{8}{15} = 0.533 = \underline{\underline{53.3\%}}$

2. (5 Points) What is the chance that the first two are both orange?

The chance is 2.86 %.

first orange:  $\frac{3}{15}$

second orange,  
given first orange:  $\frac{2}{14}$

> dependent  
multiplication rule

both orange:  $\frac{3}{15} \cdot \frac{2}{14} = \frac{6}{210} = 0.0286 = \underline{\underline{2.86\%}}$

3. (5 Points) What is the chance that the first is orange and the second is caramel?

The chance is 7.14 %.

first orange:  $\frac{3}{15}$

second caramel,  
given first orange:  $\frac{5}{14}$

> dependent

first orange and  
second caramel:

$\frac{3}{15} \cdot \frac{5}{14} = \frac{15}{210} = 0.0714 = \underline{\underline{7.14\%}}$

Please turn over!

from: FPP, Review Exercise 7, p. 235

Question 2: Chance/Probability II (5 Points)

A coin is tossed six times. Two possible sequences of results are

(i) H T T H T H

(ii) H H H H H H

(The coin must land on H or T in the order given; H = heads, T = tails).

Which of the following is correct?

Circle your answer and explain:

1. Sequence (i) is more likely.
2. Sequence (ii) is more likely.

3. Both sequences are equally likely.

Workbook answer:

"3. is correct, every possible string of H's and T's is equally likely." (2)

In fact, there are  $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^6 = 64$  possible sequences of H's & T's in six coin tosses. Thus, the chance for each of these sequences is  $\frac{1}{64} = 0.0156 = 1.56\%$ .

Note that this question did not ask whether getting 3H's is more or less likely than getting 6H's. In fact, when we write down all possible sequences of H's & T's in six coin tosses, we will see that there are far more (different) sequences with 3H's than there are sequences with 6 H's (just one!).

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

Friday, March 11, 2005

Your Name: \_\_\_\_\_

based on: Stat 1040, Fall 2003, October 24, 2003, Quiz 7, Question 1

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

A game consists of tossing an 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 an odd number (i.e., 1, 3, 5, 7, or 9) you lose \$2, otherwise you you win \$1, except when the die lands on 10, in which case you win (or lose) nothing (\$0). Assume you are playing the game 100 times.

1. (4 Points) Find the box model.

$$\boxed{5 \times [-2] \quad 1 \times [0] \quad 4 \times [1]}$$

# draws: 100

-2 for minor mistake

-3 for major mistake

-1 if number of draws not stated

2. (6 Points) Find the expected value of your gain/loss. It is -60 [\$]

$$\text{box avg} = \frac{5 \cdot (-2) + 1 \cdot 0 + 4 \cdot 1}{10} = \frac{-6}{10} = -0.6$$

$$EV_{\text{sum}} = 100 \cdot (-0.6) = \underline{\underline{-60}} \quad [\$]$$

in 2. & 3.:

-1 for each calculation error

-2 for each minor mistake

-3 for each major mistake

3. (6 Points) Find the standard error of your gain/loss. It is 14.28 [\$] (i.g. step missing)

$$\begin{aligned} \text{box SD} &= \sqrt{\frac{5 \cdot (-2 - (-0.6))^2 + 1 \cdot (0 - (-0.6))^2 + 4 \cdot (1 - (-0.6))^2}{10}} \\ &= \sqrt{\frac{5 \cdot (-1.4)^2 + 1 \cdot (0.6)^2 + 4 \cdot (1.6)^2}{10}} \end{aligned}$$

$$= \sqrt{\frac{5 \cdot 1.96 + 1 \cdot 0.36 + 4 \cdot 2.56}{10}}$$

$$= \sqrt{\frac{20.4}{10}} = \sqrt{2.04} = 1.428$$

Please turn over!

$$\begin{aligned} SE_{\text{sum}} &= \sqrt{100} \cdot 1.428 = 10 \cdot 1.428 \\ &= \underline{\underline{14.28}} \quad [\$] \end{aligned}$$

from: FPP, Review Exercise 1, p. 285

& Stat 1040, Fall 2003, October 24, 2003, Quiz 7, Question 2  
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? Circle your answer and 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.

Workbook: "Option 3. is the best because the chance error is not likely to be exactly 0, but it is likely to be small compared to the number of draws."

-1 for incorrect explanation

Formulas:

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

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

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

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

# Statistics 1040, Section 004, Quiz 9 (20 Points)

Friday, March 25, 2005

partly) -1 if slightly incorrect number of 0/1's in box  
 -2 if box given as  $\boxed{0 \quad 1}$

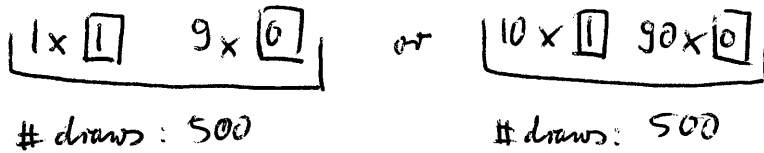
Your Name: \_\_\_\_\_

from: Stat 1040, Summer 2000, Test 2, July 17, 2000 Question 4.  
 & Stat 1040, Fall 2003, October 31, 2003 Quiz 8, Question 1  
 Question 1: EV, SE, and Normal Curve (16 Points)

-3 if box contains something other than 0/1's  
 -1 if # draws missing/incorrect

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

2. (6 Points) The expected number of Utah residents in this sample of 500 who have that specific blood type is 50 with an SE of 6.7.

box avg = fraction of 1's =  $\frac{1}{10} = 0.1$

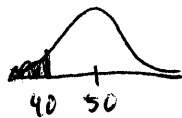
box SD =  $\sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{\frac{1}{10} \cdot \frac{9}{10}} = \sqrt{\frac{9}{100}} = \frac{3}{10} = 0.3$

EV<sub>sum</sub> =  $500 \cdot 0.1 = 50$

SE<sub>sum</sub> =  $\sqrt{500} \cdot 0.3 = 22.36 \cdot 0.3 = 6.7$

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

3. (6 Points) The chance that fewer than 40 Utah residents in this sample have that blood type is about 6.68 %.



-1.5 0 S.U.

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

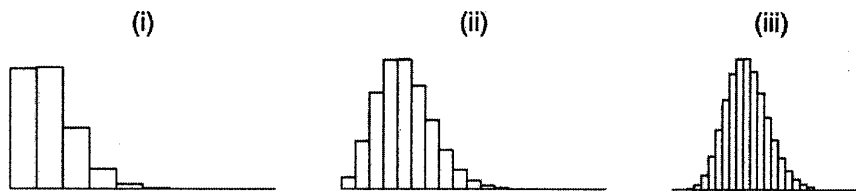
-1 for each calculation error  
 -2 for incorrect curve parameters, i.e., anything else than EV and SE  
 -2 for incorrect S.U.  
 -1 for incorrect table value  
 -1 for incorrect area under the curve

Please turn over!

from: Stat 1040, Fall 2003, November 7, 2003, Midterm 2, Question 5.2  
 & FPP, p. 324, Exercise Set C, Exercise 6

Question 2: Probability Histograms (4 Points)

Shown below are probability histograms for the sum of (a) 100, (b) 400, and (c) 900 draws from the box  $99 \times \boxed{0} \quad 1 \times \boxed{1}$ . Which histogram is which? Explain briefly.



(i) goes with sum (a) 100

(i)

(ii) goes with sum (b) 400

(ii)

(iii) goes with sum (c) 900

(iii)

Explanation: Textbook, p. A-78:

"The histograms get closer to the normal curve as the number of draws goes up."

(1)

Formulas:

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

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

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

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

Shortcut formulas for a box that contains only *two* different numbers:

$$\text{average} = \frac{(\text{smaller} \times \text{how many}) + (\text{bigger} \times \text{how many})}{\text{how many tickets in the box}}$$

$$\text{SD} = (\text{bigger} - \text{smaller}) \times \sqrt{\frac{\text{fraction bigger}}{\text{fraction smaller}}}$$

Shortcut formulas for a box that contains only  $\boxed{0}$ 's and  $\boxed{1}$ 's:

$$\text{average} = \frac{\text{number of } \boxed{1} \text{'s}}{\text{how many tickets in the box}}$$

$$\text{SD} = \sqrt{\frac{\text{fraction of } \boxed{1} \text{'s}}{\text{fraction of } \boxed{0} \text{'s}}}$$

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

Friday, April 8, 2005

Your Name: \_\_\_\_\_

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

-2 for each calculation error  
 -1 if % forgotten  
 -1 if box indicated  
 (box = population is unknown)

Are you sad to see USU President Kermit Hall leave?

This question has been inspired by the Utah Statesman's poll of 1/27/2005, but the numbers are hypothetical.

There are approximately 20,000 students at USU. A simple random sample of 300 USU students was asked the question: "Are you sad to see USU President Kermit Hall leave?". We learn that 100 students from this sample answered: "No, I didn't like what he did for the university."

1. (14 Points) If possible, construct a 95% confidence interval for the percentage of all USU students who were not sad to see President Hall leave, because they didn't like what he did for the University. If you cannot construct such a CI, explain why not.

Show your work.

1: not sad      box unknown:  $\frac{2 \times \boxed{1} \quad 2 \times \boxed{0}}{\# \text{ draws: } 300}$   
 0: sad

sample % =  $\frac{100}{300} = 0.333 = 33.3\%$  = population % (assumption) <sup>2</sup>  
 SD box =  $\sqrt{0.333 \cdot 0.667} = \sqrt{0.222} = 0.47$  (via calculator) <sup>2</sup>  
 $SE_{sam} = \sqrt{300} \cdot 0.47 = 17.32 \cdot 0.47 = 8.14$  <sup>2</sup>  
 $SE\% = \frac{8.14}{300} \cdot 100\% = 2.71\% \approx 2.7\%$  <sup>2</sup>

95% CI:  
 $sample\% \pm (mult. for 95\%) \cdot SE\%$   
 $= 33.3\% \pm 2 \cdot 2.7\%$   
 $= 33.3\% \pm 5.4\%$   
 $= 27.9\% \text{ to } 38.7\%$  <sup>2</sup>

2. (6 Points) Suppose that in the sample of 300 students, 298 students were "not sad to see President Hall leave, because they didn't like what he did for the University". Would it still be possible to construct a 95% confidence interval for the percentage of all USU students who were not sad to see President Hall leave?

Yes, possible or No, not possible? Why or why not? Explain!

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

sample % =  $\frac{298}{300} = 0.9933 = 99.33\%$  <sup>2</sup>

this % is too close to 100% (and we can't construct a 95% CI here) <sup>2</sup>  
 Please turn over!



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

Friday, April 15, 2005

Your Name: \_\_\_\_\_

From: Stat 1040, Fall 2003 Final Test, December 8, 2003, Question 4 a bc,  
Question 1: Confidence Intervals for Averages (20 Points)

A telephone answering service, at the end of each call, completes a report in which the length of the call is recorded. A simple random sample of 150 reports yields a mean length per call of 1.2 minutes with a standard deviation of 0.4 minutes.

1. (12 Points) Construct a 95% confidence interval for the average length of all the calls handled by the answering service.

$$SD = 0.4$$

$$SE_{\text{sam}} = \sqrt{150} \cdot 0.4 = 4.9 \quad (2)$$

$$SE_{\text{avg}} = \frac{4.9}{150} = 0.0327 \approx 0.03 \quad (2)$$

$$95\% \text{ CI: } 1.2 \text{ min} \pm 2 \cdot 0.03 \text{ min} = \underline{1.14 \text{ min}} \text{ to } \underline{1.26 \text{ min}} \quad (2)$$

-2 for each calculation error  
-1 if box indicated  
(box = population is unknown)

2. (5 Points) Because some of the calls are quite lengthy, call length does not follow the normal curve; it has a long right tail. Does this mean that your confidence interval calculated above is incorrect? Yes, incorrect or No, correct. Circle your answer and explain briefly. (3)

No, it is correct! We are looking at the average and this will follow the normal curve even if the original data does not. (2)

3. (3 Points) True or False (circle your choice - no further explanation needed): (3)  
95% of the calls received by the answering service have a length that falls between the lower limit and the upper limit of the confidence interval calculated in part 1. above.

We are 95% confident that the average falls between these two limits - but this does not mean that 95% of the calls received fall between these limits. [Please turn over!]

# Statistics 1040, Section 004, Quiz 12 (20 Points)

Friday, April 22, 2005

Your Name: \_\_\_\_\_

from: Stat 1040, Fall 2003, December 3-5, 2003, Quiz 12, Question 3  
Question 1: Tests of Significance (20 Points)

Past experience indicates that the time for high school seniors to complete a standardized test follows a normal distribution with a mean of 35 minutes. A simple random sample of 20 high school seniors was taken and it was found that on average, it took them 33.1 minutes to complete this test, with a standard deviation of 4.3 minutes.

Make an appropriate test to see whether or not these data suggest that the mean time needed by high school seniors to complete this test is **different from** 35 minutes.

Show your work!

- 15 for incorrect test  
(e.g. z-test instead of t-test)

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

1) null: high school seniors require the usual amount of time, (1)  
i.e., box avg = 35 min (1/2)

alternative: high school seniors require different amount of time, (1)  
i.e., box avg  $\neq$  35 min (1/2)

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

2) two-tailed t-test since we ask for  $\neq$  35 min

$$\text{observed (avg)} = 33.1 \text{ min}$$

$$\text{expected (avg)} = 35 \text{ min}$$

$$SD_t = 4.3 \cdot \sqrt{\frac{20}{20-1}} = 4.4 \quad (2)$$

$$SE_{\text{sam}} = \sqrt{20} \cdot 4.4 = 19.7 \quad (1)$$

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

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

Please turn over!

3. (4 Points) Obtain the (approximate) P-value (use the appropriate table!).

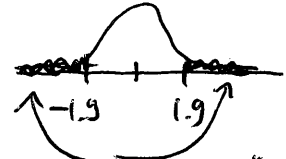
3)  $df = 20 - 1 = 19$  (1)

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 (2)

↓  
5%

↓  
2.5%



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

4. (6 Points) State your conclusions in terms of rejecting (or not rejecting) the null hypothesis and in your own words. (If appropriate, also speak of statistically significant or highly statistically significant.)

4) Conclusion:

• do not reject the null ( $P$ -value  $> 5\%$ ) (3)

• there is not enough evidence to say that high school seniors require a time different from 35 min (3)

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

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

$t$ -test: • sample size = 20 ( $< 30$ ) (1)

• SD of pop unknown (only SD of sample is known) (1/2)

• times to complete quiz follow normal curve (1/2)