

# Statistics 1040, Section 006, Quiz 1 (20 Points)

Friday, September 2, 2005

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

from: Quiz 1, Fall 2004

## Question 1: Controlled Experiments/Observational Studies I (14 Points)

(hypothetical) Does regularly taking vitamin C help protect people against flu?

A controlled experiment (2) was conducted to answer this question. The subjects (2) were 500 volunteering college students, assigned randomly (2) to two groups of 250 students. The students in the treatment group (2) took regularly a tablet of vitamin C, whereas those in the control group (2) took an identically looking and tasting pill, called placebo (2). Neither participating students nor personell administrating drugs to them knew who was taking which pill, in other words, it was a double-blind (2) experiment. After a couple of months, the numbers of flu cases in both groups were compared ...

Fill the gaps in the paragraph above using the most appropriate words from the following list:

- ✓ placebo
- ✓ double-blind
- haphazardly
- ✓ treatment group
- observational study
- ✓ randomly
- single-blind
- vaccine
- confounding factor
- objects
- ✓ control group
- ✓ controlled experiment
- ✓ subjects
- polio

Please turn over!

from: FPP, p. 24, Review Exercise 1 & Quiz 1, Fall 2004

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

In 1990, four passengers were killed by crashes on commuter airlines, compared to 39 killed on scheduled carriers (like United, TWA, and so forth). True or false? Circle your answer and explain: the data show that if you have to fly, it is safer to do so on a commuter airline.

4

Workbook:

"The statement is false - the data do not show that if you have to fly, it is safer to do so on a commuter airline. We cannot compare the numbers given - we need to compare rates.<sup>②</sup> To decide what the data DO show, we need to know how many people flew on commuter airlines versus scheduled carriers, and then we can calculate the rates and compare."

# Statistics 1040, Section 006, Quiz 2 (20 Points)

Friday, September 9, 2005

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

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

Question 1: Histograms (14 Points) & Quiz 2, Spring 2005

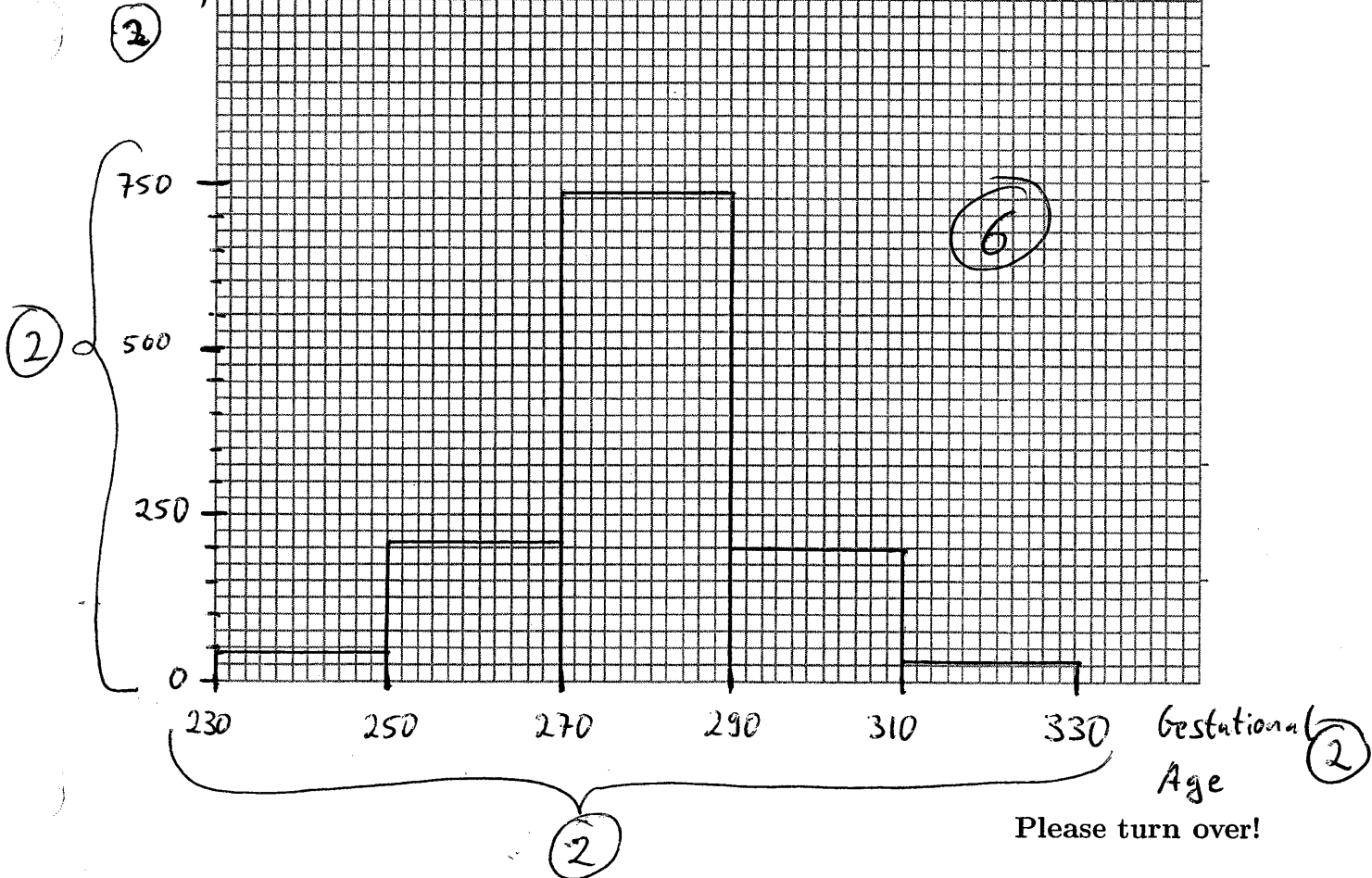
The following table is for the gestational age of 1210 babies:

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.

Frequency  
(or Count  
or Number)



from: Quiz 2, Spring 2005

**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 006, Quiz 3 (20 Points)

Friday, September 16, 2005

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

*from Quiz 2, Fall 2003*

## 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:

	Numbers	Average	Standard deviation
List 1:	17, 17, 17, 17, 17	<u>17</u> <sup>(2)</sup>	<u>0</u> <sup>(2)</sup>
List 2:	15, 16, 17, 18, 19	<u>17</u> <sup>(2)</sup>	<u><math>\sqrt{2} \approx 1.4</math></u> <sup>(4)</sup>

*- 1 each calculation error*

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,  $avg = \frac{15+16+17+18+19}{5} = \frac{85}{5} = \underline{\underline{17}}$

2,  $15 - 17 = -2$

$16 - 17 = -1$

$17 - 17 = 0$

$18 - 17 = 1$

$19 - 17 = 2$

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

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

3,  $(-2)^2 = 4$

$(-1)^2 = 1$

$0^2 = 0$

$1^2 = 1$

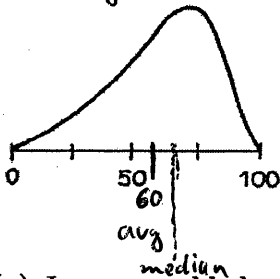
$2^2 = 4$

Please turn over!

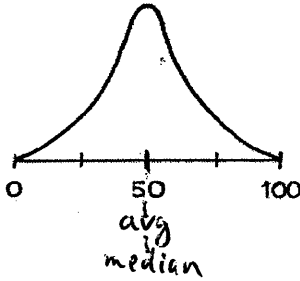
from: FPP, p. 75, Review Exercise 6, a) b) d) & Quiz 2, Fall 2003

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

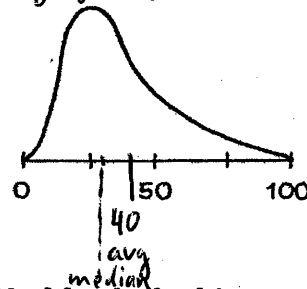
(i) long left tail



(ii) symmetric



(iii) long right tail



(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 and have about the same SD.

Formulas:

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

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

# Statistics 1040, Section 006, Quiz 4 (20 Points)

Friday, September 23, 2005

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

based on: Quiz 4, Spring 2005

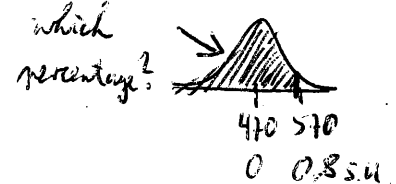
**Question 1: Normal Approximation for Data (20 Points)**

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

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 less than 570 on the GRE test is roughly 78.82 %.

① convert 570 into standard units:  
$$\frac{570 - 470}{125} = 0.8 \text{ s.u.} \quad \textcircled{3}$$

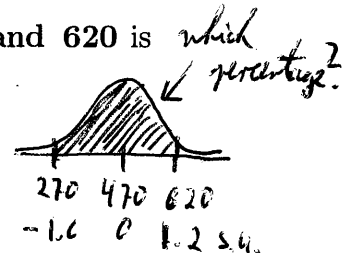


② area between -0.8 and 0.8: 57.63%  $\textcircled{2}$

③ area below 0.8:  $50\% + \frac{57.63\%}{2} = 50\% + 28.82\% = \underline{78.82\%} \quad \textcircled{2}$

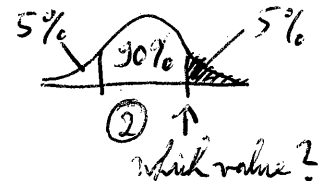
- (7 Points) The percentage of examinees who scored between 270 and 620 is about 83.02 %.

① convert 270 and 620 into standard units:  
$$\frac{270 - 470}{125} = -1.6 \text{ s.u.} \quad \textcircled{1} \quad \frac{620 - 470}{125} = 1.2 \text{ s.u.} \quad \textcircled{1}$$



② area between -1.6 and 1.6: 89.04%  $\textcircled{1}$  | ③ area between -1.6 and 1.2:  
area between -1.2 and 1.2: 76.99%  $\textcircled{1}$  | 
$$\frac{89.04\%}{2} + \frac{76.99\%}{2} = \underline{83.02\%} \quad \textcircled{3}$$

- (6 Points) In order to be among the top 5%, a student must have obtained a minimum GRE score of about 676.



① Find a value  $z$  such that the area between  $-z$  and  $z$  is about 90%.

$z = 1.65$  (gives about 90.11%)  $\textcircled{2}$

② Transfer into original units:

$$1.65 \cdot 125 + 470 = \underline{676.25} \quad \textcircled{2}$$

Please turn over!

# Statistics 1040, Section 006, Quiz 5 (20 Points)

Friday, September 30, 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)

based on: Quiz 5, Spring 2005, Question 1

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  $38.0^{\circ}$  Celsius, with a standard deviation of  $0.3^{\circ}$  Celsius.

1. (8 Points) If we translate these results into degrees Fahrenheit, the average temperature would be 100.4 degrees Fahrenheit, with a standard deviation of 0.54 degrees Fahrenheit.

$$\text{avg } F^{\circ} = \frac{9}{5} \cdot 38.0 + 32 = 68.4 + 32 = 100.4 \quad (4)$$

$$\text{SD } F^{\circ} = \frac{9}{5} \cdot 0.3 = 0.54 \quad (4)$$

2. (4 Points) Someone's temperature is 1.8 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 1.8 (4) standard units to this investigator.

*Note: 1.8 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) from: FPP Chapter 8, p. 130, # 6 (Answers: p. A-56)

(grading: see top of page)

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.

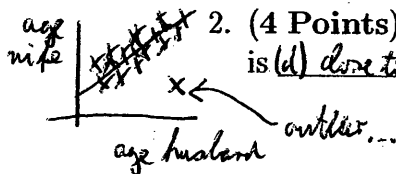
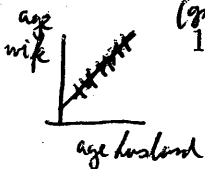
"All the points on the scatter diagram would lie on a line sloping up, so the correlation would be 1."

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 006, Quiz 6 (20 Points)

Friday, October 14, 2005

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

*from: Stat 1040, Spring 2000, Final Test, May 5, 2000, Question 10 a&b*

**Question 1: The Regression Line (20 Points) & Quiz 6, Fall 2004**

A researcher is interested in the extent to which lead particles emitted from automobiles are absorbed by competitive cyclists. For a large group of cyclists they found the following:

x	Hours of training:	average = 16.2,	SD = 5.9	
y	Blood lead ( $\mu\text{mol/L}$ ):	average = .42,	SD = .19,	$r=0.6$ .

*in all parts:*

*-2 for each calculation error*

*-2 for each incorrect value used*

*-2 for swapping x and y*

The scatter plot of the data is football-shaped.

**Show your work!**

1. (7 Points) Find the equation of the regression line for predicting blood lead from training time.

Equation:  $\text{blood lead} = 0.11 + 0.019 \cdot \text{hours}$  or  $y = 0.11 + 0.019 \cdot x$  (2)

$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.6 \cdot \frac{0.19}{5.9} = \underline{\underline{0.019}}$  (2)

$\text{intercept} = \text{avg } y - \text{slope} \cdot \text{avg } x = 0.42 - 0.019 \cdot 16.2 = 0.42 - 0.31 = \underline{\underline{0.11}}$  (3)

*-1 if only part of the equation (e.g.  $0.11 + 0.019 \cdot \text{hours}$ )*  
*-1 if not specifying x & y*

2. (3 Points) Use the regression equation from part 1. to predict the blood lead for a cyclist who trained for (21) hours.

Answer: about 0.51

$\text{blood lead} = 0.11 + 0.019 \cdot 21 = 0.11 + 0.399 = 0.509 \approx \underline{\underline{0.51}}$  (3)

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

Please turn over!

*-2 if result makes no sense at all*

3. (5 Points) Find the r.m.s. error for predicting blood lead from training time of cyclist.

Answer: 0.152

$$\begin{aligned}
 \text{r.m.s. error} &= \sqrt{1-r^2} \cdot SD_y \quad (3) \\
 &= \sqrt{1-0.6^2} \cdot 0.19 \\
 &= \sqrt{1-0.36} \cdot 0.19 \\
 &= \sqrt{0.64} \cdot 0.19 \\
 &= 0.8 \cdot 0.19 = \underline{\underline{0.152}} \quad (2)
 \end{aligned}$$

4. (5 Points) Would you be surprised to learn that a cyclist who trained for 3 hours had a blood lead of .8  $\mu\text{mol/L}$ ? Support your answer with a brief explanation and calculation.

Answer: (Yes, surprised) / No, not surprised

Explanation:

predicted after 3 hours:

$$\text{blood lead} = 0.11 + 0.019 \cdot 3 = 0.11 + 0.057 = 0.167 \quad (2)$$

observed after 3 hours: 0.8

$$\text{how unusual is this?} : \frac{0.8 - 0.167}{0.152} = \frac{0.633}{0.152} = \underline{\underline{4.16}}$$

or 0.8 is more than 4 r.m.s. errors away from the predicted  $\rightarrow$  very unusual!

Formulas:

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

$$\text{slope} = r \times \frac{SD_y}{SD_x}$$

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

# Statistics 1040, Section 006, Quiz 7 (20 Points)

Friday, October 21, 2005

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

from: Quiz 6, Fall 2002

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

1. A deck of 52 cards is shuffled and two cards are drawn without replacement.

(a) (3 Points) What is the chance that the first card is a ♡ or a ♦?

mutually exclusive events

$$\left\{ \begin{array}{l} \text{chance that 1st is } \heartsuit : \frac{13}{52} \\ \text{chance that 1st is } \spadesuit : \frac{13}{52} \end{array} \right.$$

$$\text{chance that 1st is } \heartsuit \text{ or } \spadesuit : \frac{13}{52} + \frac{13}{52} = \frac{26}{52} = \frac{1}{2} = \underline{\underline{50\%}}$$

Addition rule

(b) (4 Points) What is the chance that the first card is a ♡ and the second card is a ♦?

dependent events

$$\left\{ \begin{array}{l} \text{chance that 1st is } \heartsuit : \frac{13}{52} \\ \text{chance that 2nd is } \spadesuit, \\ \text{given that 1st is } \heartsuit : \frac{13}{51} \end{array} \right.$$

$$\text{chance that 1st is } \heartsuit \text{ and 2nd is } \spadesuit : \frac{13}{52} \cdot \frac{13}{51} = \frac{1}{4} \cdot \frac{13}{51} = \frac{13}{204} = 0.0637 = \underline{\underline{6.4\%}}$$

(c) (4 Points) What is the chance that both cards are ♡?

dependent events

$$\left\{ \begin{array}{l} \text{chance that 1st is } \heartsuit : \frac{13}{52} \\ \text{chance that 2nd is } \heartsuit, \\ \text{given that 1st is } \heartsuit : \frac{12}{51} \end{array} \right.$$

$$\text{chance that both cards are } \heartsuit : \frac{13}{52} \cdot \frac{12}{51} = \frac{1}{4} \cdot \frac{12}{51} = \frac{3}{51} = 0.0588 = \underline{\underline{5.9\%}}$$

Multiplication rule

(d) (4 Points) What is the chance that neither card is a ♡?

dependent events

$$\left\{ \begin{array}{l} \text{chance that 1st is not } \heartsuit : \frac{39}{52} \\ \text{chance that 2nd is not } \heartsuit, \\ \text{given that 1st is not } \heartsuit : \frac{38}{51} \end{array} \right.$$

$$\text{chance that neither card is } \heartsuit : \frac{39}{52} \cdot \frac{38}{51} = \frac{3}{4} \cdot \frac{38}{51} = \frac{114}{204} = 0.5588 = \underline{\underline{55.9\%}}$$

Multiplication rule

2. (5 Points) There are two options:

(i) (a) You toss a coin 100 times; on each toss, if it lands heads you win \$1, if it lands tails you lose \$1.

(ii) (b) You draw 100 times at random with replacement from the box 1 0.  
On each draw, you are paid (in dollars) the number on the ticket.

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

FPP, Chapter 13, Review question 11, page 236; & Quiz 6, Fall 2002

Workbook answer:

"Option (ii) is better. You have the same 50-50 chance of winning one dollar, but for option (i) you have no way to lose money."

## Grading Criteria:

- 1) general:
- 1 for calculation error
  - 1 for incorrect 1st chance
  - 2 for incorrect 2nd chance (e.g., if not conditional)
  - 1 for incorrect rule (e.g., addition  $\leftrightarrow$  multiplication)

- 2)
- 3 for answer (a)
  - 2 if no (or incorrect) explanation

# Statistics 1040, Section 006, Quiz 8 (20 Points)

Friday, October 28, 2005

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

From: Quiz 8, Spring 2003 & Quiz 9, Fall 2004

## Question 1: EV, SE, and Normal Curve (15 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. Show your work!

1 = Democrat     
0 = other   

1. (5 Points) Find the box model.

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

# draws = 1,000

- 1 if slightly incorrect number of  $\boxed{0}/\boxed{1}$ 's in box
- 2 if box given as  $\boxed{0} \boxed{1}$  etc.
- 3 if box contains something else than  $\boxed{0}/\boxed{1}$ 's
- 1 if # draws missing <sup>or</sup> 375 incorrect

2. (5 Points) The expected number of Democrats in this sample of 1,000 is \_\_\_\_\_ with an SE of 15.3.

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

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

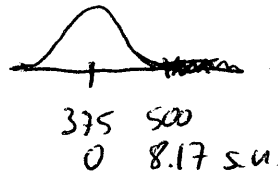
$$EV_{\text{sum}} = 1,000 \cdot 0.375 = \underline{375}$$

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

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

3. (5 Points) The chance that at least 500 of the voters in the sample are Democrats is about 0 %.

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



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

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.

Please turn over!

from Quiz 8, Fall 2004 (based on: FPP, p. 286, Review Exercise 9)

**Question 2: Law of Averages (5 Points)**

A box contains red and green marbles; there are more red marbles than green 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 (i.e., circle) one of the four options below. **Explain your answer.**

- ③
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.

Option B is best. Say the percentage of reds in the box is 60%. Then, to win, we want the percentage error to be small, so that the actual percentage will be greater than 50%. This is more likely in the long run.

②

**Formulas:**

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

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

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

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

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

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

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

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

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

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

# Statistics 1040, Section 006, Quiz 9 (20 Points)

Friday, November 4, 2005

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

from: FPP, p. 328, Review Exercise 6 (Solution: → Workbook!)

## Question 1: Normal Approximation for Probability Histograms I (12 Points)

A programmer is working on a new program, COIN, to simulate tossing a coin. As a preliminary test, he sets up the program to do one million tosses. The program returns with a count of 502,015 heads. The programmer looks at this and thinks: "Hmmm. Two thousand and fifteen off. That's a lot. No, wait. Compare it to the million. Two thousand - forget the fifteen - out of a million is two out of a thousand. That's one in five hundred. One fifth of a percent. Very small. Good. COIN passes."

Do you agree that COIN passes? Answer **yes** or **no**, and **explain**. You should use box model calculations to support your answer.

H: 1  
T: 0

Box Model:  $\boxed{1 \times \boxed{1} \quad 1 \times \boxed{0}}$  (1)

# draws = 1,000,000 (1)

$$\text{box avg} = \frac{\text{fraction of 1's}}{2} = \frac{1}{2} \quad (1)$$

$$\text{box SD} = \sqrt{\frac{\text{fraction of 1's}}{2} \cdot \frac{\text{fraction of 0's}}{2}} = \sqrt{\frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{2} \quad (1)$$

$$EV_{\text{sam}} = 1,000,000 \cdot \frac{1}{2} = 500,000 \quad (1)$$

$$SE_{\text{sam}} = \sqrt{1,000,000} \cdot \frac{1}{2} = 1,000 \cdot \frac{1}{2} = 500 \quad (1)$$

$$s.u. = \frac{502,015 - 500,000}{500} \approx 4.03 \quad (2)$$

This is more than 4 s.u. away from the  $EV_{\text{sam}}$  and thus very unlikely to happen just by chance!

area between -4.05 to 4.05: 99.9949%

area above 4.05:  $\frac{100\% - 99.9949\%}{2} \approx 0\%$

Something must be wrong with COIN!

(2) Please turn over!

based on: FPP, p. 328, Review Exercise 8  
 & Midterm 2, Spring 2005, Question 4

**Question 2: Normal Approximation for Probability Histograms II (8 Points)**

A coin is tossed 100 times. True or false? **Just circle your answer.** You don't have to give any explanation. **Answer each of the following questions separately!**

H: 1, T: 0 box:  0  1

1. The expected value for the number of heads is 50.  
 ② True / False [correct;  $EV_{sum} = 50$ ]
2. The expected value for the number of heads is 50, give or take 5 or so.  
 True / False ② [false, the  $EV_{sum}$  is exactly 50, no give or take here]
3. The number of heads will be 50.  
 True / False ② [false, the number of heads most likely will not be exactly 50, but it will be relatively close to 50]
4. The number of heads will be around 50, give or take 2 or so.  
 True / False ② [false,  $EV_{sum}$  indeed is 50, but  $SE_{sum}$ , the "give or take" part, is 5 and not 2]

# draws = 100  
 box avg =  $\frac{1}{2}$   
 $EV_{sum} = 100 \cdot \frac{1}{2} = 50$   
 $SE_{sum} = \sqrt{100} \cdot \frac{1}{2} = 5$

**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:

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

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

Shortcut formulas for a box that contains only  0 's and  1 's:

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

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



# Statistics 1040, Section 006, Quiz 10 (20 Points)

Friday, November 18, 2005

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

from: FPP, Chapter 21, p. 392, Question 4 [Solutions → Workbook]

## Question 1: The Accuracy of Percentages (20 Points)

The National Assessment of Educational Progress administers standardized achievement tests to nationwide samples of 17-year-olds in school. One year, the tests covered history and literature. You may assume that a simple random sample of size 6,000 was taken. Only 36.1% of the students in the sample knew that Chaucer wrote *The Canterbury Tales*, but 95.2% knew that Edison invented the light bulb.

*each part: -9 for "No"  
-1 for each calculation error*

1. (10 Points) Is it possible to find a 95% confidence interval for the percentage of all 17-year-olds in school who knew that Chaucer wrote *The Canterbury Tales*?
- ② **Yes** or No? – Circle your answer.

If yes, calculate this CI (and show your work). If no, clearly indicate why this is not possible.

$$\begin{aligned} \text{sample \%} &= 36.1\% = \text{population \% (assumption)} \\ \text{SD}_{\text{pop}} &= \sqrt{0.361 \cdot 0.639} = 0.48 \quad \text{① (via shortcut)} \\ \text{SE}_{\text{sam}} &= \sqrt{6,000} \cdot 0.48 = 37.2 \quad \text{①} \\ \text{SE\%} &= \frac{37.2}{6,000} \cdot 100\% = 0.62\% \quad \text{①} \end{aligned}$$

$$\begin{aligned} \text{95\% CI:} \\ \text{sample \%} \pm (\text{multiplier for 95\%}) \cdot \text{SE\%} \\ = 36.1\% \pm 2 \cdot 0.62\% \\ = 36.1\% \pm 1.24\% \\ = 34.86\% \text{ to } 37.34\% \quad \text{②} \end{aligned}$$

2. (10 Points) Is it possible to find a 95% confidence interval for the percentage of all 17-year-olds in school who knew that Edison invented the light bulb?
- ② **Yes** or No? – Circle your answer.

*-4 if SE% from 1. reused*

If yes, calculate this CI (and show your work). If no, clearly indicate why this is not possible.

$$\begin{aligned} \text{sample \%} &= 95.2\% = \text{population \% (assumption)} \\ \text{SD}_{\text{pop}} &= \sqrt{0.952 \cdot 0.048} = 0.21 \quad \text{① (via shortcut)} \\ \text{SE}_{\text{sam}} &= \sqrt{6,000} \cdot 0.21 = 16.3 \quad \text{①} \\ \text{SE\%} &= \frac{16.3}{6,000} \cdot 100\% = 0.27\% \quad \text{①} \end{aligned}$$

$$\begin{aligned} \text{95\% CI:} \\ \text{sample \%} \pm (\text{multiplier for 95\%}) \cdot \text{SE\%} \\ = 95.2\% \pm 2 \cdot 0.27\% \\ = 95.2\% \pm 0.54\% \\ = 94.66\% \text{ to } 95.74\% \quad \text{②} \end{aligned}$$

Note: Although the assumed  $\text{SD}_{\text{pop}}$  is lopsided here, the sample size of 6,000 justifies the use of the normal curve and therefore the calculation of this CI.  
Please turn over!  
[see FPP, Chapter 21, end of Section 2, p. 382-383 & FPP, Chapter 18, Section 5, p. 319-325]

# Statistics 1040, Section 006, Quiz 11 (20 Points)

Wednesday, November 30, 2005

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

from FPP, p. 427, Chapter 23, Review Exercise 8 [Solutions: → Workbook]

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

One year, there were about 3,000 institutions of higher learning in the U.S. (including junior colleges and community colleges). As part of a continuing study of higher education, the Carnegie Commission took a simple random sample of 400 of these institutions. The average enrollment in the 400 sample schools was 3,700, and the SD was 6,500. The Commission estimates the average enrollment at all 3,000 institutions to be around 3,700; they put a give-or-take number of 325 on this estimate.

Say whether each of the following statements is true or false, and explain. If you need more information to decide, say what you need and why.

1. (4 Points) An approximate 68%–confidence interval for the average enrollment of all 3,000 institutions runs from 3,375 to 4,025.

② Yes, No, or Need more information.

Circle your answer and explain.

$$\text{sample avg} = 3,700$$

$$\text{sample SD} = 6,500$$

$$SE_{\text{sam}} = \sqrt{400} \cdot 6,500 = 130,000$$

$$SE_{\text{avg}} = \frac{130,000}{400} = 325 \quad \text{①}$$

68% CI:

$$\text{sample avg} \pm (\text{multiplier for 68\%}) \cdot SE_{\text{avg}} \quad \text{①}$$

$$= 3,700 \pm 1 \cdot 325$$

$$= 3,375 \text{ to } 4,025 \text{ (so this is correct)}$$

2. (4 Points) If a statistician takes a simple random sample of 400 institutions out of 3,000, and goes one SE either way from the average enrollment of the 400 sample schools, there is about a 68% chance that this interval will cover the average enrollment of all 3,000 schools.

③ Yes, No, or Need more information.

Circle your answer and explain.

This is one of two possible interpretations of a CI. See p. 181 in our notes.

①

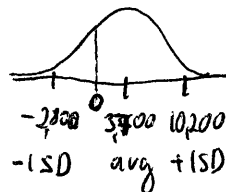
Please turn over!

3. (4 Points) About 68% of the schools in the sample had enrollments in the range  $3,700 \pm 6,500$ .

Yes, **No**, or Need more information.

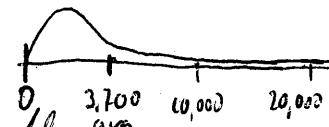
Circle your answer and explain.

Can the data look like this:



Obviously not - this would mean that a huge percentage of schools has a negative enrollment. More likely, the data will look like this:

This means, the data do not follow the normal curve and the statement is most likely false.



4. (4 Points) It is estimated that 68% of the 3,000 institutions of higher learning in the U.S. enrolled between  $3,700 - 325 = 3,375$  and  $3,700 + 325 = 4,025$  students.

Yes, **No**, or Need more information.

Circle your answer and explain.

The give-or-take is the SE and not the SD. This gives the 68% CI as calculated in part 1.). And even with the SD, this statement is still false, as seen in part 3.), due to the non-normal curve of the data.

5. (4 Points) The normal curve can't be used to figure confidence levels here at all, because the data don't follow the normal curve.

Yes, **No**, or Need more information.

Circle your answer and explain.

It is true that the data do not follow the normal curve, but we are dealing with CI calculations for the average. Therefore, calculations based on the normal curve are valid! Compare with the example on p. 189 in our notes (the data do not follow the normal curve, but the probability histogram for the sample average does).

# Statistics 1040, Section 006, Quiz 12 (20 Points)

Wednesday, December 7, 2005

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

from: Stat 1040, Final Test, April 23, 2004 (Spring 2004), Question 11

Question 1: Tests of Significance (20 Points) & Stat 1040, Handout, Fall 2005, Mo 12/5/05 !!!

Many companies are experimenting with "flex-time", which is supposed to reduce absenteeism. One company employees have averaged 6.3 days off work in the past. The company introduces "flex-time" and a year later a simple random sample of 100 employees is selected. They average 5.5 days off work with a standard deviation of 2.9. Test to determine if "flex-time" reduces absenteeism. Clearly state the null and alternative hypotheses, calculate the appropriate test statistic, find the P-value, and state your conclusion.

Show your work!

-1 for each calculation error  
-2 if null, alt swapped  
-15 for incorrect test

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

1) Null: flex-time has no effect on absenteeism, i.e.,  $\mu_{\text{avg}} = 6.3$  ①  
Alternative: flex-time reduces absenteeism, i.e.,  $\mu_{\text{avg}} < 6.3$  ①

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

observed avg: 5.5

expected avg: 6.3

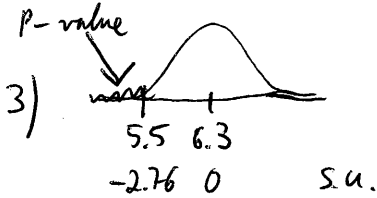
2)  $SE_{\text{sam}} = \sqrt{100} \cdot 2.9 = 10 \cdot 2.9 = 29$  ①

$SE_{\text{avg}} = \frac{29}{100} = 0.29$  ②

$z = \frac{5.5 - 6.3}{0.29} = -2.76$  ②

Please turn over!

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



area between  $-2.75$  and  $2.75$ : 99.40% (2)

$$\rightarrow \text{P-value} = \frac{100\% - 99.40\%}{2} = 0.3\% \quad (2)$$

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) • reject the null ( $P\text{-value} < 5\%$ ) (2)

• result is highly statistically significant ( $P\text{-value} < 1\%$ ) (2)

• flex-time reduces absenteeism (2)

5. (2 Points) The test that has to be used in this question is a

(1) z-test / t-test / 2-sample z-test.

Circle your answer and explain briefly why you chose this particular test to answer the question.

why z-test? - sample size  $\geq 30$  ! (1)