Statistics 1040, Sections 007 & 009, Quiz 1 (20 Points)

Friday, August 31, 2007

Your Name: _____

from: Quiz 1, Fall 2004 & Quiz 1, Fall 2005 (Solutions: → Course Web Page)

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: Quiz 1, Fall 2004 & Quiz 1, Fall 2005 (Solutions → Course Web Page)

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)

"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. (2) 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, Sections 007 & 009, Quiz 2 (20 Points)

Friday, September 7, 2007

Your Name: _____

from: Stat 1040, Fall 2004, Final Test, December 15, 2004, Question 3
& Quiz 2, Spring 2005
& Quiz 2, Fall 2005 (Solutions: → Course Web-Page)

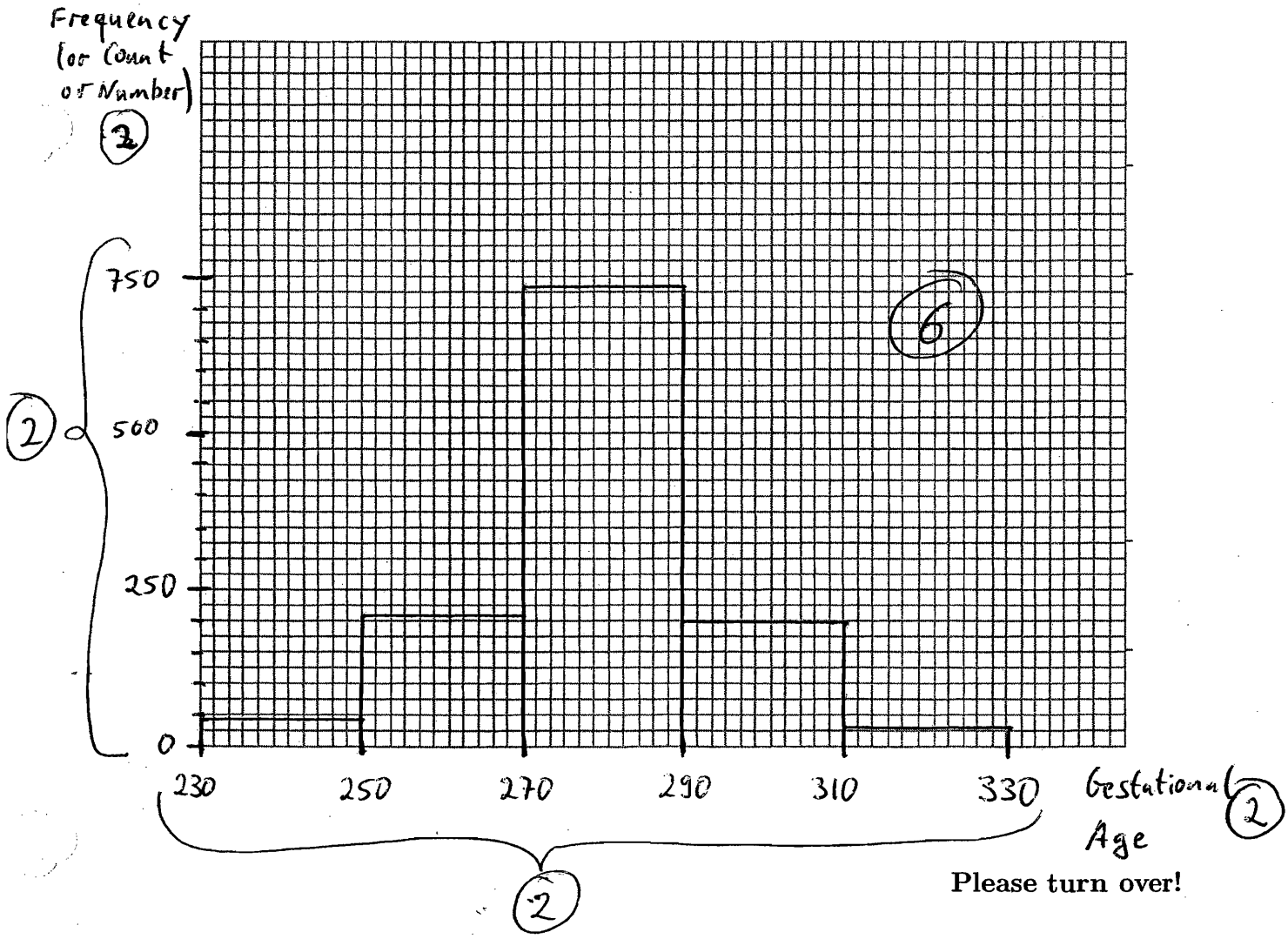
Question 1: Histograms **(14 Points)**

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.



Please turn over!

from: Quiz 2, Spring 2005 & Quiz 2, Fall 2005 (Solutions: → Course Web Page)

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, Sections 007 & 009, Quiz 3 (20 Points)

Friday, September 14, 2007

Your Name: _____

from: FPP, Chapter 4, Review Exercise 3 (Solutions: → Workbook)

Question 1: The Average and the SD I (**10 Points**)

Here is a list of numbers:

0.7 1.6 9.8 3.2 5.4 0.8 7.7 6.3 2.2 4.1
8.1 6.5 3.7 0.6 6.9 9.9 8.8 3.1 5.7 9.1

1. Without doing any arithmetic, guess whether the average is around (i) 1, (ii) 5, or (iii) 10. **Circle your answer and explain.** (3)

   Workbook: "The average should be in the middle of the distribution: only three of the numbers are smaller than 1, and none are bigger than 10." (2)

2. Without doing any arithmetic, guess whether the SD is around (i) 1, (ii) 3, or (iii) 6. **Circle your answer and explain.** (3)

   Workbook: "If the SD is 1, the entries 0.6 and 9.9 are much too far from average. The SD can't be 6, because none of the numbers are more than 6 away from the average." (2)

from: FPP, Chapter 4, Review Exercise 7(bc) (Solutions: → Workbook)

Question 2: The Average and the SD II (**10 Points**)

A study on college students found that the men had an average weight of about 66 kg and an SD of about 9 kg. The women had an average weight of about 55kg and an SD of about 9 kg (Note that 1 kg = 2.2 lb).

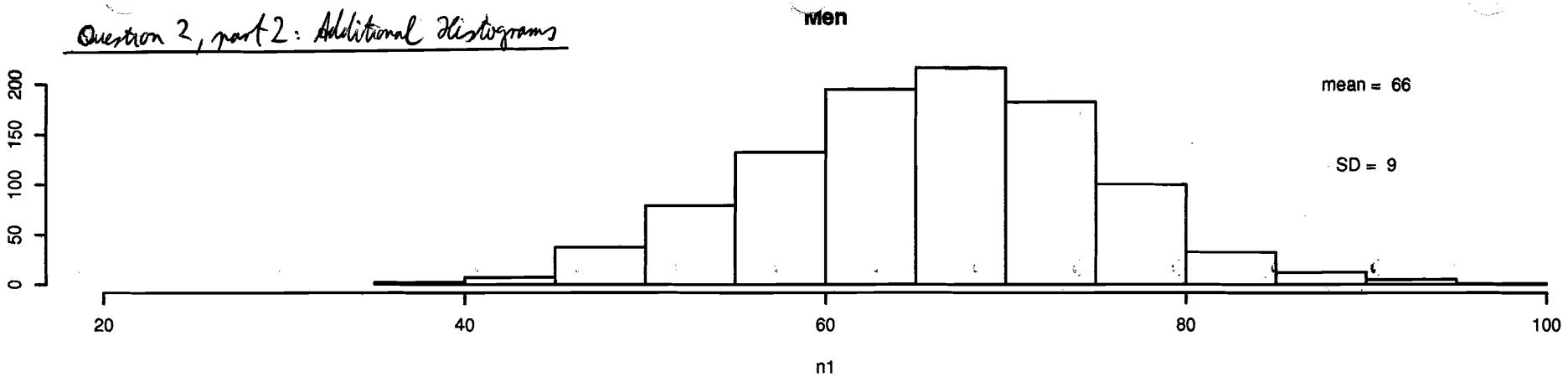
1. Just roughly, what percentage of the men weighted between 57 kg and 75 kg?
   Answer: 68 % (3)
   Fill in your answer and explain.

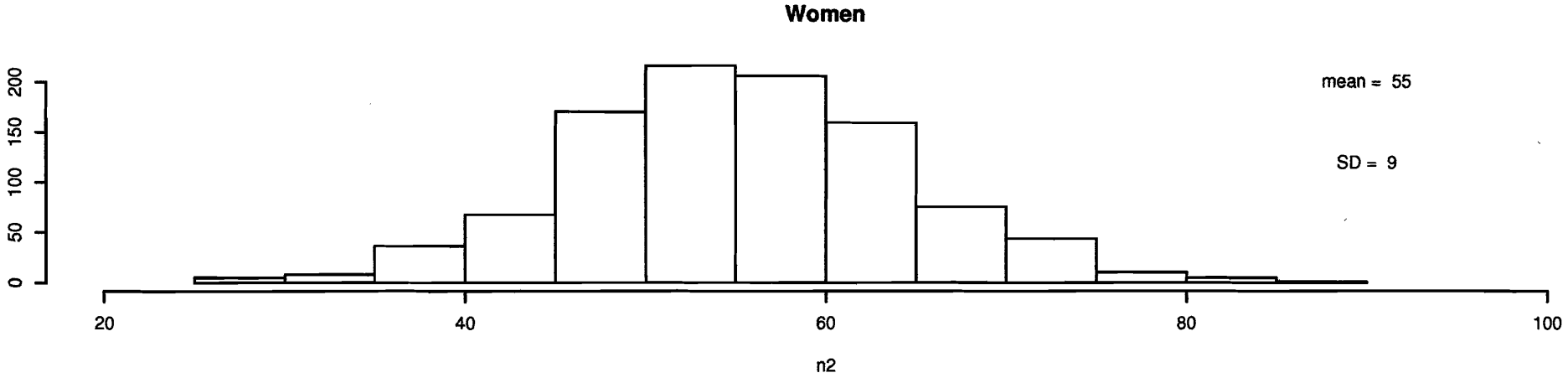
   Workbook: "The range is average ± 1 SD": 66 − 9 = 57, 66 + 9 = 75 (2)
   (our interval of interest)

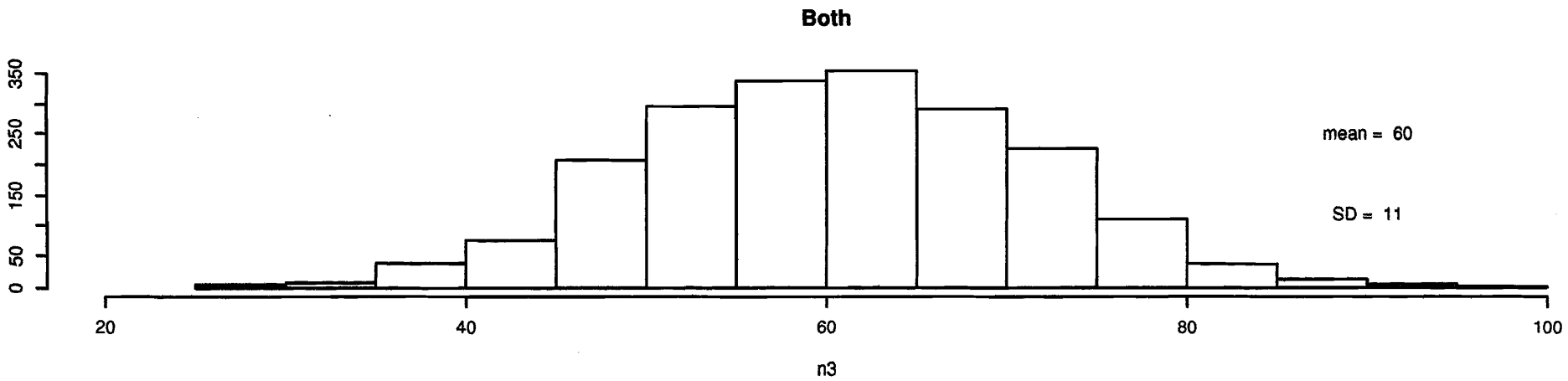
2. If you took the men and women together, would the SD of their weights be (i) smaller than 9 kg, (ii) just about 9 kg, or (iii) bigger than 9 kg? (3)
   Circle your answer and explain.

   Workbook: "If you take the men and women together, the spread in weights goes up." (2)

Question 2, part 2: Additional Histograms







Statistics 1040, Sections 007 & 009, Quiz 4 (20 Points)

Friday, September 21, 2007

Your Name: _____

based on: Quiz 4, Spring 2006 [new numbers!]

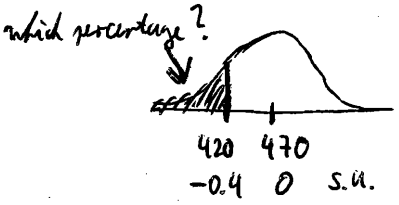
-2 for each calculation error

+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. We can assume that the histogram follows a normal curve. **Show your work!**

- (**7 Points**) The percentage of examinees who scored **less than 420** on the GRE test is roughly 34.46 %.



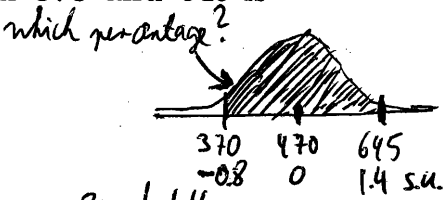
(1) convert 420 into standard units:

$\frac{420-470}{125} = -0.4$ s.u. (3)

(2) area between −0.4 and 0.4: 31.08% (2)

(3) area below −0.4: $\frac{100\% - 31.08\%}{2} = \frac{68.92\%}{2} = 34.46\%$ (2)

- (**7 Points**) The percentage of examinees who scored **between 370 and 645** is about 70.74 %.



(1) convert 370 and 645 into standard units:

$\frac{370-470}{125} = -0.8$ s.u. (1), $\frac{645-470}{125} = 1.4$ s.u. (1)

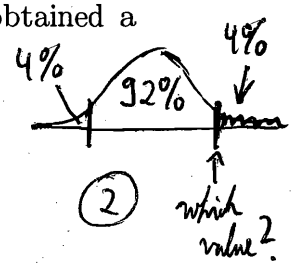
(2) area between −0.8 and 0.8: 57.63% (1)

area between −1.4 and 1.4: 83.85% (1)

(3) area between −0.8 and 1.4:

$\frac{57.63\%}{2} + \frac{83.85\%}{2} = 70.74\%$ (3)

- (**6 Points**) In order to be among the top 4%, a student must have obtained a minimum GRE score of about 689.



(1) Find a value z such that the area between −z and z is about 92%:

z = 1.75 (2) (gives 91.99%)

(2) Transfer z into original units:

1.75 • 125 + 470 = 688.75 ≈ 689 (2)

Please turn over!

Statistics 1040, Sections 007 & 009, Quiz 5 (20 Points)

Friday, September 28, 2007

Your Name: _____

from: FPP, Chapter 6, Review Exercise 4 (Solutions → Workbook)

Question 1: Measurement Error (**8 Points**)

You send a yardstick to a local laboratory for calibration, asking that the procedure be repeated three times. They report the following values:

35.96 inches 36.01 inches 36.03 inches

If you send the yardstick back for a fourth calibration, you would expect to get 36 inches, give or take

(a) .01 inches or so (b) .03 inches or so (c) .06 inches or so (4)

Circle your answer and explain.

$\text{avg} = \frac{35.96 + 36.01 + 36.03}{3} = 36$ (2)

$SD = \sqrt{\frac{(35.96-36)^2 + (36.01-36)^2 + (36.03-36)^2}{3}} \approx 0.03$ = "give or take" (2)

Question 2: Correlation (**6 Points**) from: FPP, Chapter 8, Review Exercise 11 (Solutions → Workbook)

A teaching assistant gives a quiz to his section. There are 10 questions on the quiz and no part credit is given. After grading the papers, the TA writes down for each student the number of questions the student got right and the number wrong. The average number of right answers is 6.4 with an SD of 2.0; the average number of wrong answers is 3.6 with the same SD of 2.0.

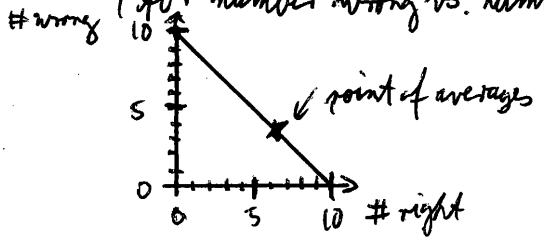
The correlation coefficient between the number of right answers and the number of wrong answers is

(a) exactly 0 (b) -0.50 (c) +0.50
(d) -1.0 (e) +1.0 (f) -2.0 (g) +2.0
(h) can't tell without the data

(4)

Circle your answer and explain.

Number wrong = 10 − Number right, so all the points on a scatter diagram (for number wrong vs. number right) lie on a straight line which slopes down. (2)



Please turn over!

from: FPP, Chapter 9, Review Exercise 8 (Solutions: → Worklurk)

Question 3: Correlation (**6 Points**)

For women age 25 and over in the U.S. in 2005, the relationship between age and educational level (years of schooling completed) can be summarized as follows:

average age $\approx$ 50 years, SD $\approx$ 16 years
average ed. level $\approx$ 13.2 years, SD $\approx$ 3.0 years, r $\approx$ -0.20

True or **false** and explain: As you get older, you become less educated. If this statement is false, what accounts for the negative correlation? (4)

The data are cross-sectional, not longitudinal. Younger people were born later and educational levels have been going up over time. (2)

Formulas:

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

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

Statistics 1040, Sections 007 & 009, Quiz 6 (20 Points)

Friday, October 12, 2007

Your Name: _____

from: Stat 1040, Spring 2005, Midterm 2, Question 3
& Stat 1040, Spring 2006, Midterm 2, Question 3

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

A selection of 65 varieties of cereal were tested for calories and sodium (in milligrams) for an one–cup serving. The results can be summarized as follows:

-2 each calculation error
-2 if x, y flipped
-2 if x, y not specified

y Average sodium = 240 mg; SD = 131 mg;
x Average calories = 149 calories; SD = 62 calories; $r = 0.53$.

Show your work!

1. (**7 Points**) Find the equation of the regression line for predicting number of mg sodium in an one–cup serving of cereals from calories.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.53 \cdot \frac{131}{62} = 1.12 \quad (3)$$

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 240 - 1.12 \cdot 149 = 240 - 166.88 = 73.12 \quad (2)$$

regression equation: $\boxed{\text{sodium} = 73.12 + 1.12 \cdot \text{calories}}$ (2)

or $\boxed{y = 73.12 + 1.12 \cdot x}$

2. (**4 Points**) Predict the number of mg sodium in an one–cup serving of cereals that has 200 calories per cup.

-3 for old method, incorrect result

for 200 calories:

$$\text{sodium} = 73.12 + 1.12 \cdot 200 = 73.12 + 224 = \underline{297.12 \text{ mg}} \quad (4)$$

-3 if result makes no sense at all

Please turn over!

3. (**5 Points**) Find the r.m.s. error for predicting mg sodium from calories.

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

$= \sqrt{1-0.53^2} \cdot 131$

$= 111.1$ mg (5)

−2 for each major mistake, e.g., SD_x instead of SD_y, $\sqrt{}$ of everything, r instead of r^2, etc.

4. (**4 Points**) Explain why it would not be a good idea to use the information in the question to estimate the amount of sodium for a cereal with 350 calories per cup.

350 calories: $\frac{350-149}{62} = 3.2$ s.a. (2)

350 is more than 3 s.a. above the average (of 149 calories); this is extrapolation (2) and the result will be meaningless [perhaps this cereal contains big chocolate chips that are high in calories, but may be low on sodium]

Formulas:

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

$$\text{slope} = r \times \frac{\text{SD}_y}{\text{SD}_x} \qquad \text{intercept} = \text{avg}_y - \text{slope} \times \text{avg}_x$$

Statistics 1040, Sections 007 & 009, Quiz 7 (20 Points)

Monday, October 22, 2007

Your Name: _____

from: FPP, Chapter 13, Review Exercise 6 (Solutions → Workbook)

Question 1: Chance/Probability I (**6 Points**)

A deck of cards is shuffled and the top two cards are placed face down on a table.
Circle your answer:

1. There is a 1 chance in 52 for the first card to be the ace of clubs.
(2) (True) / **False**

2. There is a 1 chance in 52 for the second card to be the ace of diamonds.
(2) (True) / **False** [We don't know anything about the first card!]

3. The chance of getting the ace of clubs and then the ace of diamonds is $1/52 \times 1/52$.
True / (False) (2) [The two events are dependent. The chance is $1/52 \times 1/51$.]

from: FPP, Chapter 13, Review Exercise 7 (Solutions → Workbook)

Question 2: Chance/Probability II (**3 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:

1. Sequence (i) is more likely.

2. Sequence (ii) is more likely.

(3) 3. Both sequences are equally likely. [Both sequences are equally likely, having chance $\frac{1}{2^6}$. Of course, there are a lot more sequences with 3 heads and 3 tails, but that's not the question.]

Please turn over!

from: FPP, Chapter 14, Review Exercise 4 (Solutions → Workbook)

Question 3: Chance/Probability III (**3 Points**)

Two cards will be dealt off the top of a well–shuffled deck. You have a choice:

(i) To win \$1 if at least one of the two cards is a queen.

(ii) To win \$1 if the first card is a queen.

Circle your answer:

(3) 1. Option (i) is better for winning.

2. Option (ii) is better for winning.

3. Both options are equally good for winning.

[Option (i) is better. Even if you miss the first time, you get a second try at the money.]

from: FPP, Chapter 14, Review Exercise 6 (Solutions → Workbook)

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

One event has chance 1/2, another has chance 1/3. Fill in the blanks, using one phrase from each pair below, to make **two true sentences**. Write out both sentences.

"If you want to find the chance that **at least one of the two events / both events** will happen, check to see if they are **independent / mutually exclusive**. If so, you can **add / multiply** the chances."

Correct Sentence 1: "If you want to find the chance that **at least one of the two events** will happen, check to see if they are mutually exclusive (1). If so, you can add (1) the chances."

Correct Sentence 2: "If you want to find the chance that **both events** will happen, check to see if they are independent (1). If so, you can multiply (1) the chances."

from: FPP, Chapter 14, Review Exercise 12 (no solutions published)

Question 5: Chance/Probability V (**4 Points**)

A coin is tossed 10 times. **Circle your answer:**

1. The chance of getting 10 heads in a row is 1/1,024.

(2) True / **False** [Getting ten heads in a row is very unlikely ($\frac{1}{2^{10}}$) - before you start tossing.]

2. Given that the first 9 tosses were heads, the chance of getting 10 heads in a row is 1/2.

(2) True / **False** [If you get nine heads in a row, however, the chance is 50-50 to get a tenth head: coin tosses are independent!]

Statistics 1040, Sections 007 & 009, Quiz 8 (20 Points)

Friday, October 26, 2007

Your Name: _____

from: FPP, Chapter 16, Review Exercise 1 (Solutions → Workbook)

Question 1: Law of Averages I (**2 Points**)

A box contains 10,000 tickets: 4,000 [0] 's and 6,000 [1] 's. And 10,000 draws will be made at random with replacement from this box. Which of the following best describes the situation? **Circle your answer.** You don't have to provide any explanations.

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.

(2) (2)

[The chance error is not likely to be 0 exactly, but should be small relative to the number of draws.]

from: FPP, Chapter 16, Review Exercise 4 (Solutions → Workbook)

Question 2: Law of Averages II (**8 Points**)

Circle your answer for each of the following four parts. You don't have to provide any explanations.

1. (**2 Points**) A die will be rolled some number of times, and you win \$1 if it shows an ace ([•]) more than 20% of the time.

   Which is better: **60 rolls** or **600 rolls**? (2)

2. (**2 Points**) As in 1.), but you win the dollar if the percentage of aces is more than 15%.

   Which is better: **60 rolls** or **600 rolls**? (2)

3. (**2 Points**) As in 1.), but you win the dollar if the percentage of aces is between 15% and 20%.

   Which is better: **60 rolls** or **600 rolls**? (2)

4. (**2 Points**) As in 1.), but you win the dollar if the percentage of aces is exactly $16\frac{2}{3}$%.

   Which is better: **60 rolls** or **600 rolls**? (2)

[additional details in Workbook]

from: FPP, Chapter 17, Review Exercise 7 (Solutions → Worklook)

Question 3: The Expected Value and Standard Error (**10 Points**)

One hundred draws are made at random with replacement from the box

1	2	3	4	5	6

.

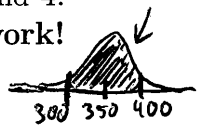
1. (**1 Point**) If the sum of the draws is 321, what is their average?

$$\text{avg} = \frac{321}{100} = 3.21 \quad (1)$$

2. (**1 Point**) If the average of the draws is 3.78, what is the sum?

$$\text{sum} = 3.78 \cdot 100 = 378 \quad (1)$$

3. (**8 Points**) Estimate the chance that the average of the draws is between 3 and 4. You have to make use of the normal curve to answer this part! **Show your work!**



- avg between 3 and 4 $\rightsquigarrow$ sum between 300 and 400
- box avg $= \frac{1+2+3+4+5+6}{6} = 3.5$ (1)

$$EV_{sum} = 100 \cdot 3.5 = 350 \quad (1)$$

$$\text{box SD} = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2}{6}}$$

$$= 1.7 \quad (1)$$

$$SE_{sum} = \sqrt{100} \cdot 1.7 = 17 \quad (1)$$

- s.u.: $\frac{300-350}{17} = -2.94$ (1)

$\frac{400-350}{17} = 2.94$ (1)

- area between -2.95 and 2.95:

99.68 % (2)

Formulas:

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

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

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

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

Statistics 1040, Sections 007 & 009, Quiz 9 (20 Points)

Friday, November 2, 2007

Your Name: _____

from: FPP, Chapter 18, Review Exercise 6 (Solutions → Workbook)
& Stat 1040, Quiz 9, Fall 2005, Question 1 (Solutions → Web)

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. (2)

H: 1
T: 0

Box Model: [1 × 1 | 1 × 0] (1)

draws = 1,000,000 (1)

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

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

$EV_{sum} = 1{,}000{,}000 \cdot \frac{1}{2} = 500{,}000$ (1)

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

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

This is more than 4 s.u. away from the EV_{sum} and thus very unlikely to happen just by chance! (2)

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!

Please turn over!

from: Stat 1040, Quiz 9, Fall 2005, Question 2 (Solutions → Web)
[based on: FPP, Chapter 18, Relevant Exercise 8 (some values changed)]

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

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]

Formulas:

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

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

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

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

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

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

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

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

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

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

Statistics 1040, Sections 007 & 009, Quiz 10 (20 Points)

Friday, November 16, 2007

Your Name: _____

from: FPP, Chapter 21, Review Exercise 5 (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.

sample % = 36.1% = population % (assumption)
SD box = $\sqrt{0.361 \cdot 0.639} = 0.48$ (1) (via bootstrap)
$SE_{sum} = \sqrt{6{,}000} \cdot 0.48 = 37.2$ (1)
$SE_{\%} = \frac{37.2}{6{,}000} \cdot 100\% = 0.62\%$ (1)

95% CI:
sample % ± (multiplier for 95%) · SE%
(1) (1) (1)
= 36.1% ± 2 · 0.62%
= 36.1% ± 1.24%
= 34.86% to 37.34% (2)

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

sample % = 95.2% = population % (assumption)
SD box = $\sqrt{0.952 \cdot 0.048} = 0.21$ (1) (via bootstrap)
$SE_{sum} = \sqrt{6{,}000} \cdot 0.21 = 16.3$ (1)
$SE_{\%} = \frac{16.3}{6{,}000} \cdot 100\% = 0.27\%$ (1)

95% CI:
sample % ± (multiplier for 95%) · SE%
(1) (1) (1)
= 95.2% ± 2 · 0.27%
= 95.2% ± 0.54%
= 94.66% to 95.74% (2)

Note: Although the assumed box is lopsided here, the sample size of 6,000 justifies the use of the normal curve and therefore the calculation of this CI.
[see FPP, Chapter 21, end of Section 2, p. 382-383 & FPP, Chapter 18, Section 5, p. 319-325]

Please turn over!

Statistics 1040, Sections 007 & 009, Quiz 11 (20 Points)

Wednesday, November 28, 2007

Your Name: _____

from: FPP, Chapter 23, Review Exercise 8 (Solutions → Worklook)

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.

sample avg = 3,700
sample SD = 6,500
$SE_{sum} = \sqrt{400} \cdot 6{,}500 = 130{,}000$
$SE_{avg} = \frac{130{,}000}{400} = 325$ (1)

68% CI:
sample avg ± (multiplier for 68%) · SE_{avg}
= 3,700 ± 1 · 325 (1)
= 3,375 to 4,025 (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)

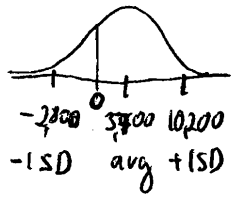
Please turn over!

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.** [either option is correct] 3

Circle your answer and explain.

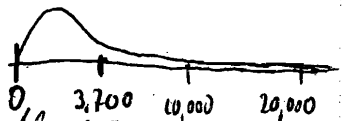
Can the data look like this:



2?

1

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 don't 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.** 3

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. 1

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.** 3

Circle your answer and explain.

It is true that the data do <u>not</u> follow the normal curve, but we are dealing with CI calculations for the <u>average</u>. Therefore, calculations based on the normal curve <u>are</u> 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 <u>average</u> does). 1

Quiz 12 - Solutions (from: Stat 1040, Quiz 12, Spring 2004, Question 1; Solutions → Web)

-15 for incorrect test

Question 1:

1) null: nose length as usual, i.e., true avg = 44 mm (1) (1)

alternative: longer noses at PS, i.e., true avg > 44 mm (1) (1)

2) t-test: (1)
- sample size = 18 (< 30)
- SD of true unknown (we can only calculate SD of sample)
- nose lengths follow normal curve

observed avg = $\frac{41+57+\dots+37+48}{18} = 44.78$ mm (2)

expected avg = 44 mm

$SD = \sqrt{\frac{(41-44.78)^2+(57-44.78)^2+\dots+(37-44.78)^2+(48-44.78)^2}{18}}$

$= \sqrt{\frac{837.11}{18}} = 6.82$ mm (2)

$SD_t = 6.82 \cdot \sqrt{\frac{18}{17}} = 7.02$ mm (2)

$SE_{sum} = \sqrt{18} \cdot 7.02 = 29.78$ mm (1)

$SE_{avg} = \frac{29.78}{18} = 1.65$ mm (1)

$t = \frac{44.78-44}{1.65} = 0.47$ (2), $df = 18-1 = 17$ (1)

3) P-value:

$t = 0.47$ is left of 0.69 (1)

↓

25%

⇝ P-value is > 25% (1)

4) Conclusion:
- do not reject the null (P-value > 5%) (1)
- there is not enough evidence to say that students at Penn State on average have noses that are longer than 44 mm (1)