Statistics 1040, Section 003, Quiz 1 (20 Points)

Friday, August 29, 2008

Your Name: _____

from: Quiz 1 { Fall 2004 / Fall 2005 / Fall 2007 }, Question 1 (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 (1) was conducted to answer this question. The subjects (5) 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 / Fall 2005 / Fall 2007 }, Question 2 (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, Section 003, Quiz 2 (20 Points)

Friday, September 5, 2008

from: Stat 1040, Fall 2004, Final Test, December 15, 2004, Question 3
& Quiz 2, { Spring 2005 / Fall 2005 / Fall 2007 }, Question 1

Your Name: _____

(Solutions: → Course Web Page)

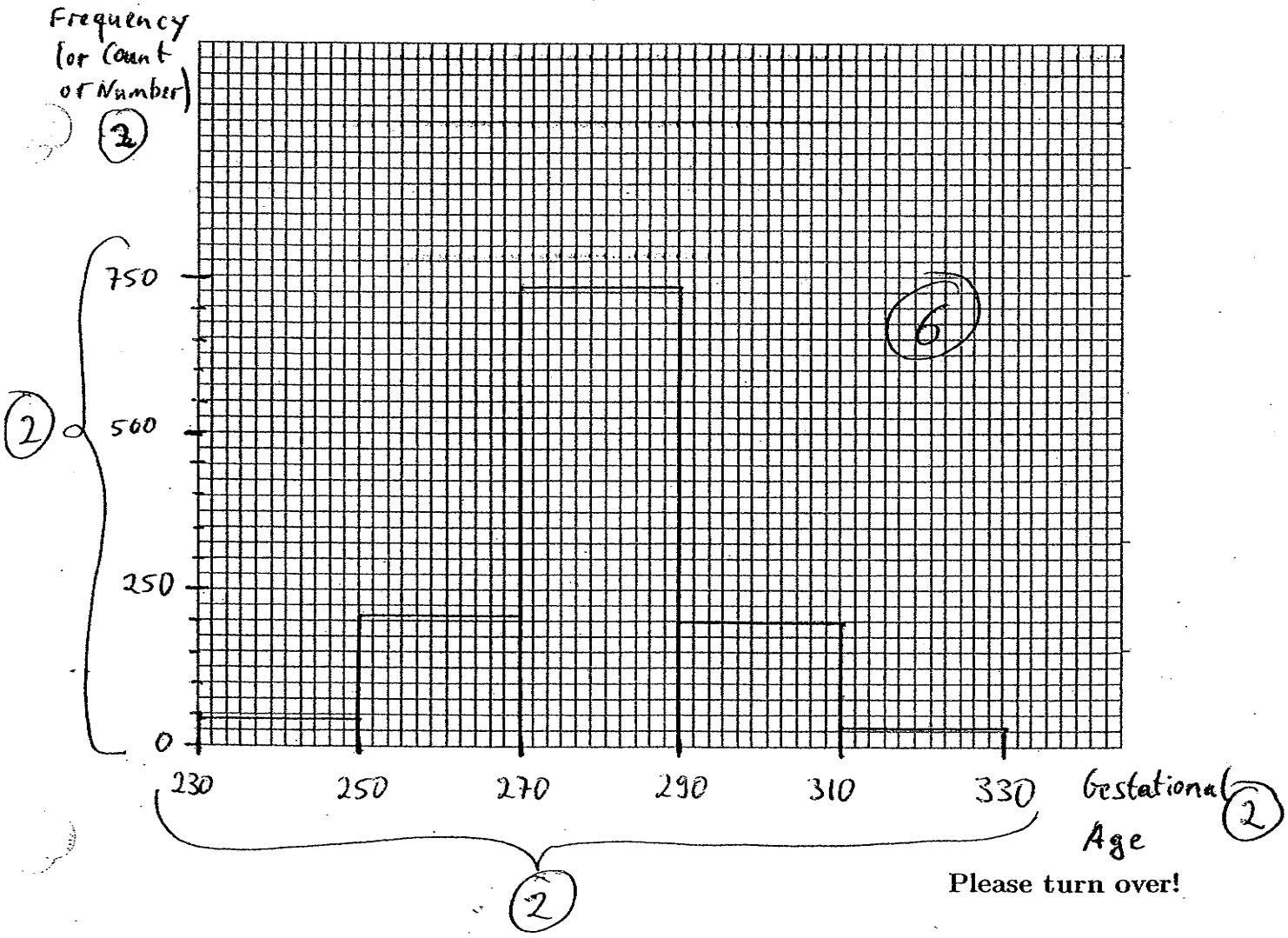
Question 1: Histograms (14 Points)

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

Gestational Age	Number	optional: Percent	optional: Percent per Day
230-250	47	3.9%	0.195%
250-270	206	17.0%	0.850%
270-290	731	60.4%	3.020%
290-310	199	16.4%	0.820%
310-330	27	2.2%	0.110%

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

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 / Fall 2005 / Fall 2007 }, Question 2 (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, Section 003, Quiz 3 (20 Points)

Friday, September 12, 2008

Your Name: _____

(Solutions: → Course Web Page)

from: Quiz 2, Fall 2003, Question 1
Quiz 3, Fall 2005, Question 1

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

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

-1 each calculation error

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

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

List 1: Nothing to calculate! Since all numbers are identical (17), this must also be the average (and the median). Also, since the SD is some average departure from the average, but no value in the list departs from the average, the SD is 0.

List 2: 1) $\text{avg} = \frac{15+16+17+18+19}{5} = \frac{85}{5} = 17$ } (2)

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

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

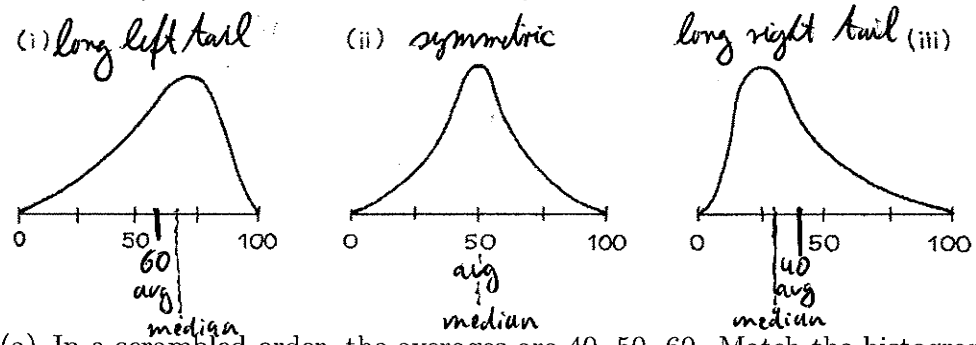
4) $\frac{4+1+0+1+4}{5} = \frac{10}{5} = 2.0$ } (1)

5) $SD = \sqrt{2.0} = 1.414 \approx 1.4$ } (1)

Please turn over!

from: FPP, p. 75, Review Exercise 6, a) b) d) (Solutions: → Worklook)
& Quiz 2, Fall 2003, Question 2
Quiz 3, Fall 2005, Question 2 (Solutions: → Course Web Page)

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



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

Histogram (i): average = 60

Histogram (ii): average = 50

Histogram (iii): average = 40

① for each correctly identified average & histogram

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

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

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

②

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

Formulas:

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

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

Statistics 1040, Section 003, Quiz 4 (20 Points)

Friday, September 19, 2008

Your Name: _____

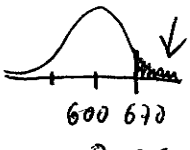
New Question (no solutions available)

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 112,000 male US citizens who took the General GRE Test in 2005–06, the mean for the quantitative ability portion of the exam was about 600 and the standard deviation was about 140 (`http://www.ets.org/Media/Tests/GRE/pdf/05-06_factors_final.%20pdf.pdf`). We can assume that the histogram follows a normal curve. **Show your work!**

-1 for each calculation error

- (**7 Points**) The percentage of male US citizens who scored **more than 670** on the GRE test is roughly 30.86 %.



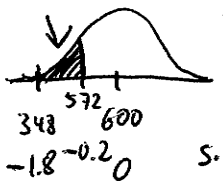
+2 for correct graph (but nothing else) (3)

(1) s.u.: $\frac{670-600}{140} = 0.5$ s.u.

(2) area from −0.5 to 0.5: 38.29% (2)

(3) area above 0.5: $50\% - \frac{38.29\%}{2} = 30.855\% \approx 30.86\%$ (2)

- (**7 Points**) The percentage of male US citizens who scored **between 348 and 572** is about 38.48 %.



(1) s.u.: $\frac{348-600}{140} = -1.8$ s.u. (1)

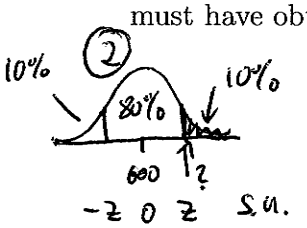
(2) area from −1.8 to 1.8: 92.81% (1)

s.u.: $\frac{572-600}{140} = -0.2$ s.u. (1)

area from −0.2 to 0.2: 15.85% (1)

(3) area from −1.8 to −0.2: $\frac{92.81\% - 15.85\%}{2} = \frac{76.96\%}{2} = 38.48\%$ (3)

- (**6 Points**) In order to be among the top 10% of all male US citizens, a student must have obtained a minimum GRE score of about 782.



(1) area from −1.30 to 1.30: 80.64% (closest to 80%) (2)

(2) original units:

$1.30 \cdot 140 + 600 = 182 + 600 = 782$ (2)

Please turn over!

Statistics 1040, Section 003, Quiz 5 (20 Points)

Friday, September 26, 2008

Your Name: _____

from: FPP, Chapter 6, Review Exercise 4, p. 104 (Solutions → Workbook)

Question 1: Measurement Error (**6 Points**) & Quiz 5, Spring 2008, Question 1 (→ Course Web Page)

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 (3)

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

Circle your answer and do the necessary calculations.

Calculations:

$$\text{avg} = \frac{35.96+36.01+36.03}{3} = 36 \quad (1)$$

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

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

from: FPP, Chapter 8, Exercise Set B, Question 7, p. 130 (Solutions: → Textbook, p. A-56)

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

1. The correlation coefficient between student's age and year of birth is (b) nearly -1.
2. The correlation coefficient between student's age and mother's age is (e) somewhat positive

(a) -1 (b) nearly -1 (c) somewhat negative
(d) 0 (e) somewhat positive (f) nearly 1 (g) 1 (h) -2.0 (i) 2.0
(j) can't tell without the data

Additional Explanation:

1.) The older you are, the earlier you were born; but there is some fuzz, depending on whether your birthday is before or after the day of the questionnaire.

2.) It is possible to give birth over a range of more than 30 years.

Grading:

1.) (b): (4)
(a) or (c): (2)

2.) (e): (4)
(d) or (f): (2)

Please turn over!

from: FPP, Chapter 5, Review Exercise 8, p. 95 (Solutions → Workbook)

Question 3: Change Of Scale (**6 Points**)

True or false? **Circle your answer!**

1. If you add 7 to each entry on a list, that adds 7 to the average.
(1) (True) / False [see pp. 92 ff in textbook]
2. If you add 7 to each entry on a list, that adds 7 to the SD.
True / (False) (1) [all the deviations from the average stay the same]
3. If you double each entry on a list, that doubles the average.
(1) (True) / False [see pp. 92 ff]
4. If you double each entry on a list, that doubles the SD.
(1) (True) / False [all the deviations from the average are doubled]
5. If you change the sign of each entry on a list, that changes the sign of the average.
(1) (True) / False [see pp. 92 ff]
6. If you change the sign of each entry on a list, that changes the sign of the SD.
True / (False) (1) [all the deviations from the average have their signs changed, but that goes away in the squaring; the SD has to be positive (or, exceptionally, zero)]

Formulas:

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

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

Statistics 1040, Section 003, Quiz 6 (20 Points)

Friday, October 10, 2008

Your Name: _____

from: FPP, Chapter 12, Review Exercise 2, p. 214 (Solutions: → Workbook)

Question 1: Regression I **(17 Points)** & additional parts

For women age 25–34 in the HANES5 sample, the relationship between height and income can be summarized as follows:

-1 for each calculation error (in any part)

X	height:	average = 64.0 inches	SD = 2.5 inches
Y	income:	average = $21,000	SD = $20,000
		r = 0.20	

-1 if x, y flipped
-1 if x, y not specified

Fill the blanks in the statements below and **show all the work needed to obtain the answers.**

(either mark what x & y are, or work with "height" & "income")

1. (5 Points) Find the regression equation for predicting income from height.

   The equation is: income = −81,400 + 1,600 · height
   or: $Y = -81{,}400 + 1{,}600 \cdot X$ } (1) for one of these

$$\text{slope} = r \cdot \frac{SD_Y}{SD_X} = 0.20 \cdot \frac{20{,}000}{2.5} = 1{,}600 \quad (2)$$

$$\text{intercept} = \text{avg}_Y - \text{slope} \cdot \text{avg}_X = 21{,}000 - 1{,}600 \cdot 64 = -81{,}400 \quad (2)$$

2. (3 Points) Using your regression equation from (1.), what is the regression estimate for the predicted income for a woman who is 60 inches tall?

   The answer is: $ 14,600

predicted income for a woman who is 60 inches tall =

$$-81{,}400 + 1{,}600 \cdot 60 = \$14{,}600 \quad (3)$$

Please turn over!

3. (3 Points) Find the r.m.s. error for your answer in the previous part.

The answer is $\approx$ $19,596

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

$= \sqrt{1-0.20^2} \cdot 20,000$ (3)

$= 19,595.92$

4. (3 Points) Can we use the regression equation from (1.) to predict the income for a woman who is 80 inches tall? **Yes** / (No.) Circle your answer. If **yes**, then calculate the predicted income. If **no**, then explain why not.

(2)

No! $\frac{80-64}{2.5} = 6.4$ s.u. (1) for reasonable explanation

An 80" tall woman is 6.4 s.u. above average; predicting the income would be extreme extrapolation! [Such a woman easily might obtain a top salary in the WNBA...]

5. (3 Points) The regression equation indicates: The taller a woman, the higher her income on average. Indicate a major confounding factor and explain which effect this factor has.

(2) for valid factor

(1) for valid explanation

Race! (also: Schooling, Education, ...)

According to http://en.wikipedia.org/wiki/Human_height

Avg female height	Sample population (Age range)
158.5cm (5', 2.5")	White Americans (20-39)
160.0cm (5', 3.6")	African-Americans (20-39)
156.1cm (5', 1.7")	Mexican-Americans (20-39)

[Note: Age is not a valid answer here as we only consider age 25-34; age might be relevant if we'd consider all age groups.]

& based on http://en.wikipedia.org/wiki/Personal_income_in_the_United_States

Race (both genders)	Overall median (all age 25+) Income
White	$ 33,030
African	$ 27,101
Hispanic or Latino	$ 23,613

Explanation: White & African American (women) are taller than Mexican/Hispanic/Latino American (women) and have higher (median) income.

Please turn over!

from: FPP, Chapter 12, Review Exercise 9, p. 215 (Solutions: → Workbook)

Question 2: Regression II (**3 Points**)

In a large study (hypothetical) of the relationship between parental income and the IQs of their children, the following results were obtained:

x	income:	average = $90,000	SD = $45,000
y	IQ:	average = 100	SD = 15
		r = 0.50	

For each income group ($0–$9999, $10,000–$19,999, $20,000–$29,999, etc.), the average IQ of children with parental income in that group was calculated and then plotted above the midpoint of the group ($5,000, $15,000, $25,000, etc.). It was found that the points on this graph followed a straight line very closely. The slope of this line (in IQ points per dollar) would be about:

6,0000 3,000 1,500 500 1/500 1/1,500 1/3,000 1/6,000 (3)

Circle your answer!

optional: $\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.50 \cdot \frac{15}{45{,}000} = \frac{1}{2} \cdot \frac{1}{3{,}000} = \frac{1}{6{,}000}$

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, Section 003, Quiz 7 (20 Points)

Wednesday, October 15, 2008

Your Name: _____

based on: Quiz 7, Fall 2004, Question 1 (new values; no solutions available)

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

A drawer of socks contains 22 socks of which 8 are black, 10 are green, and 4 are blue. In the dark, a child chooses two socks at random to wear to school. **Show your work!**

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

1. **(5 Points)** What is the chance that the first sock is blue?

   The chance is: 18.2 %

   1st blue: $\frac{4}{22}$ (5) $= 0.182 = 18.2\%$

2. **(5 Points)** What is the chance that both socks are blue?

   The chance is: 2.6 %

   1st blue: $\frac{4}{22}$ (1)

   2nd blue, given 1st blue: $\frac{3}{21}$ (2) → dependent

   both blue: $\frac{4}{22} \cdot \frac{3}{21} = \frac{12}{462} = 0.026 = 2.6\%$ (3) multiplication rule

3. **(5 Points)** What is the chance that one sock is blue and the other sock is green?

   The chance is: 17.3 %

   (BG) & (GB) mutually exclusive

   (BG) 1st blue: $\frac{4}{22}$

   2nd green, given 1st blue: $\frac{10}{21}$ (2) → dependent

   (GB) 1st green: $\frac{10}{22}$

   2nd blue, given 1st green: $\frac{4}{21}$ (2) → dependent

   one blue, one green: $\frac{4}{22} \cdot \frac{10}{21}$ (2) + (1) $\frac{10}{22} \cdot \frac{4}{21}$ (2) $= \frac{40}{462} + \frac{40}{462} = \frac{80}{462} = 0.173 = 17.3\%$ addition rule

4. **(5 Points)** What is the chance that both socks are the same color?

   The chance is: 34.2 %

   (BlBl) & (GG) & (BB) mutually exclusive

   (BlBl) both black: $\frac{8}{22} \cdot \frac{7}{21}$ (1) $= \frac{56}{462}$

   (GG) both green: $\frac{10}{22} \cdot \frac{9}{21}$ (1) $= \frac{90}{462}$

   (BB) both blue: $\frac{4}{22} \cdot \frac{3}{21}$ (1) $= \frac{12}{462}$

   same color: $\frac{56}{462}$ (1) + $\frac{90}{462}$ (1) + $\frac{12}{462} = \frac{158}{462} = 0.342 = 34.2\%$ addition rule

Statistics 1040, Section 003, Quiz 8 (20 Points)

Friday, October 24, 2008

Your Name: _____

based on: Quiz 8, Spring 2006, Question 1 (new values & extra part; no solutions available)

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

You are participating in a new game that consists of tossing a 10–sided die, with sides numbered from 1 to 10. The die is fair, i.e., it has the same chance of landing on any side. Every time the die shows a number that is a multiple of 3 (i.e., 3, 6, or 9) you win \$4, otherwise you lose \$1, except when the die lands on 10, in which case you win (or lose) nothing (\$0). Assume you are tossing this die 200 times. **Show your work!**

1. (**3 Points**) Find the box model.

   [3 × 4 | 1 × 0 | 6 × -1]

   # draws: 200

   -1 for minor mistake
   -2 for major mistake (e.g. 3, 6, 9 in base) [or base missing]
   -1 if # draws not stated

2. (**4 Points**) Find the expected value of your gain/loss.

   $\text{base avg} = \frac{3 \cdot 4 + 1 \cdot 0 + 6 \cdot (-1)}{10} \overset{②}{=} \frac{6}{10} = 0.6$

   $EV_{sum} = 200 \cdot 0.6 \overset{②}{=} 120$ [\$]

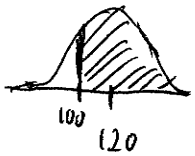
   in 2, 3, & 4:
   -1 for each calculation error
   -1 for each minor mistake
   -2 for each major mistake [or step missing]

3. (**4 Points**) Find the standard error of your gain/loss.

   $\text{base SD} = \sqrt{\frac{3 \cdot (4-0.6)^2 + 1 \cdot (0-0.6)^2 + 6 \cdot (-1-0.6)^2}{10}} \overset{②}{=} \sqrt{\frac{50.4}{10}} = 2.245$

   $SE_{sum} = \sqrt{200} \cdot 2.245 \overset{②}{=} 31.75$ [\$]

4. (**5 Points**) Assuming that many players participate in this game, find the percentage of players who win \$100 or more (after 200 tosses each).



s.u.: $\frac{100-120}{31.75} \overset{②}{=} -0.63$

area from −0.65 to 0.65: 48.43% ②

area above −0.65: $50\% + \frac{48.43\%}{2} \overset{①}{=} 74.22\%$

Please turn over!

from: FPP, p.285, Review Exercise 4 & Quiz 8, Spring 2006, Question 2

(Solutions → Workbook)
(Solutions → Course Web Page)

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

Circle your answer for each of the following four parts. You don't have to provide any explanations. [Explanations from Web Page]

1. (**1 Point**) A die will be rolled some number of times, and you win \$1 if it shows an ace (⚀) more than 20% of the time. [To win, you need a large percentage error, and that is more likely in 60 rolls.]

   Which is better: **60 rolls** or **600 rolls**? (1)

2. (**1 Point**) As in 1.), but you win the dollar if the percentage of aces is more than 15%. [Now you want a small percentage error.]

   Which is better: **60 rolls** or **600 rolls**? (1)

3. (**1 Point**) As in 1.), but you win the dollar if the percentage of aces is between 15% and 20%. [Again - you want a small percentage error.]

   Which is better: **60 rolls** or **600 rolls**? (1)

4. (**1 Point**) As in 1.), but you win the dollar if the percentage of aces is exactly $16\frac{2}{3}\%$. [Because to get exactly the expected value means getting exactly zero chance error, and that is more likely with fewer rolls.]

   Which is better: **60 rolls** or **600 rolls**? (1)

Formulas:

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

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

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

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

Statistics 1040, Section 003, Quiz 9 (20 Points)

Friday, October 31, 2008

in 1.22: −1 for each calculation error
−1 for each step missing

Your Name: _____

from: FPP, Chapter 18, Review Exercise 2, p. 327 (Solution Outline: → Worklook)

Question 1: Probability Histograms I (14 Points)

Four hundred draws will be made at random with replacement from the box [1] [3] [5] [7].
Show your work!

#draws = 400

6 points 1. Estimate the chance that the sum of the draws will be more than 1,500.
The chance is: 98.78 %

box avg: $\frac{1+3+5+7}{4} = 4$ (1)

box SD: $\sqrt{\frac{(1-4)^2+(3-4)^2+(5-4)^2+(7-4)^2}{4}}$

$= \sqrt{5} = 2.24$ (1)

$EV_{sum} = 400 \cdot 4 = 1,600$ (1)

$SE_{sum} = \sqrt{400} \cdot 2.24 = 44.8$ (1)

S.u.: $\frac{1,500-1,600}{44.8} = -2.23$ (1)

−2.23 0 s.u.

area between −2.25 and 2.25: 97.56%

area above −2.25: $\frac{97.56\%}{2} + 50\% = 98.78\%$ (1)

8 points 2. Estimate the chance that there will be fewer than 90 [3]'s.
The chance is: 12.51 %

in 2.: −5 for incorrect box, but correct steps
−7 for incorrect box & incorrect steps

new box with 1: [3]
0: [1], [5] or [7]

$1 \times [1] \quad 3 \times [0]$ (2)

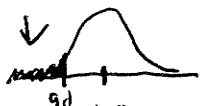
#draws: 400

box avg: $\frac{1}{4}$ (1)

box SD: $\sqrt{\frac{1}{4} \cdot \frac{3}{4}} = 0.433$ (1)

$EV_{sum} = 400 \cdot \frac{1}{4} = 100$ (1)

$SE_{sum} = \sqrt{400} \cdot 0.433 = 8.66$ (1)



S.u.: $\frac{90-100}{8.66} = -1.15$ (1)

−1.15 0 s.u.

area between −1.15 and 1.15: 74.99%

area below −1.15: $\frac{100\% - 74.99\%}{2} = 12.51\%$ (1)

Please turn over!

from: FPP, Chapter 18, Review Exercise 9, p. 328 (Solution Outline: → Worklook)

Question 2: Probability Histograms II (4 Points)

One hundred draws are made at random with replacement from a box with ninety–nine tickets marked "0" and one ticket marked "1". **Circle the correct answer.**

1. The sum will be around 1, give or take 1 or so.
(2) (True) / False

box avg = $\frac{1}{100}$

box SD = $\sqrt{\frac{1}{100} \cdot \frac{99}{100}} = 0.0995$

$EV_{sum} = 100 \cdot \frac{1}{100} = 1$

$SE_{sum} = \sqrt{100} \cdot 0.0995 = 0.995 \times 1$

2. There is about a 68% chance that the sum will be in the range 0 to 2.
True / (False) (2)

The probability histogram for this box is even more asymmetric than the one for the example in FPP, p. 320. So, just 100 draws does not justify the use of the normal approximation, even though we are dealing with a "sum" here.

Question 3: Sampling (2 Points)

from: FPP, Chapter 19, Review Exercise 5, p. 351 (Solution: → Worklook)

(Hypothetical) A survey is carried out by the finance department to determine the distribution of household size in a certain city. They draw a simple random sample of 1,000 households. After several visits, the interviewers find people at home in only 653 of the sample households. Rather than face such a high non–response rate, the department draws a second batch of households, and uses the first 347 completed interviews in the second batch to bring the sample up to its planned strength of 1,000 households. The department counts 3,087 people in these 1,000 households, and estimates the average household size in the city to be about 3.1 persons.

(2)

Is this estimate likely to be **too low**, (**too high**), or **about right**?

With smaller households, the interviewer is less likely to find someone at home.

Circle your answer! So, the survey procedure is, on average, replacing smaller households by larger ones.

Please turn over!

Statistics 1040, Section 003, Quiz 10 (20 Points)

Friday, November 14, 2008

Your Name: _____

from FPP, Chapter 21, Review Exercise 5, p. 392 (Solution Outline: → Worklook)

Question 1: The Accuracy of Percentages (20 Points)

& Quiz 10, Fall 2007, Question 1 (Solution: → Course Web Page)

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: -3 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)

$= 36.1\% \pm 2 \cdot 0.62\%$

$= 36.1\% \pm 1.24\%$

$= 34.86\%$ 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)

$= 95.2\% \pm 2 \cdot 0.27\%$

$= 95.2\% \pm 0.54\%$

$= 94.66\%$ to 95.74% (2)

Note: Although the assumed box is bimodal 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, Section 003, Quiz 11 (20 Points)

Wednesday, November 19, 2008

Your Name: _____

from: FPP, Chapter 23, Review Exercise 8, p. 427 (Solution Outline: → Worksheet)

Question 1: The Accuracy of Averages (20 Points) & Quiz 11, Fall 2007, Question 1 (Solution: → Course Web Page)

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)

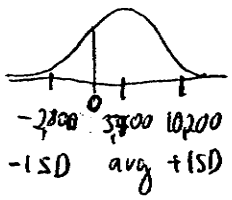
Please turn over!

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]

Circle your answer and explain.

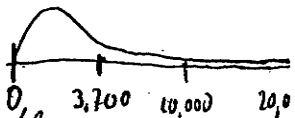
Can the data look like this:



??

(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. (3)

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)

Statistics 1040, Section 003, Quiz 12 (20 Points)

Monday, November 24, 2008

Your Name: _____

from: Quiz 12, Fall 2005, Question 1
Quiz 11, Spring 2008, Question 1
& Stat 1040, Final, Spring 2004, Question 11
(Solutions: → Course Web Page)

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

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.

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

Show your work!

1. (**2 Points**) The test that has to be used in this question is a
   (1) **z–test** / t–test.
   Circle your answer and explain briefly why you chose this particular test to answer the question.

   why z-test? - sample size ≥ 30 ! (1)

2. (**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., box avg = 6.3
   (1)
   Alternative: flex-time reduces absenteeism, i.e., box avg < 6.3 (1)

Please turn over!

3. **(5 Points)** Calculate the appropriate test statistic.

observed avg: 5.5
expected avg: 6.3

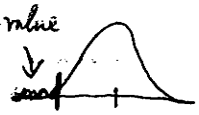
2| $SE_{sum} = \sqrt{100} \cdot 2.9 = 10 \cdot 2.9 = 29$ (1)

$SE_{avg} = \frac{29}{100} = 0.29$ (1)

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

4. **(4 Points)** Obtain the (approximate) P–value (use the appropriate table!).

3|



area between −2.75 and 2.75: 99.40% (2)

⇝ P-value: $\frac{100\% - 99.40\%}{2} = 0.3\%$ (2)

5. **(6 Points)** State your conclusions in terms of rejecting (or not rejecting) the null hypothesis and in your own words. (If approriate, also speak of statistically significant or highly statistically significant.)

4|
- reject the null (P-value < 5%) (2)
- result is highly statistically significant (P-value < 1%) (2)
- flex-time reduces absenteeism (2)

Please turn over!