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

Friday, September 3, 2004

Your Name: _____

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 objects (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, Spring 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. (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 006, Quiz 2 (20 Points)

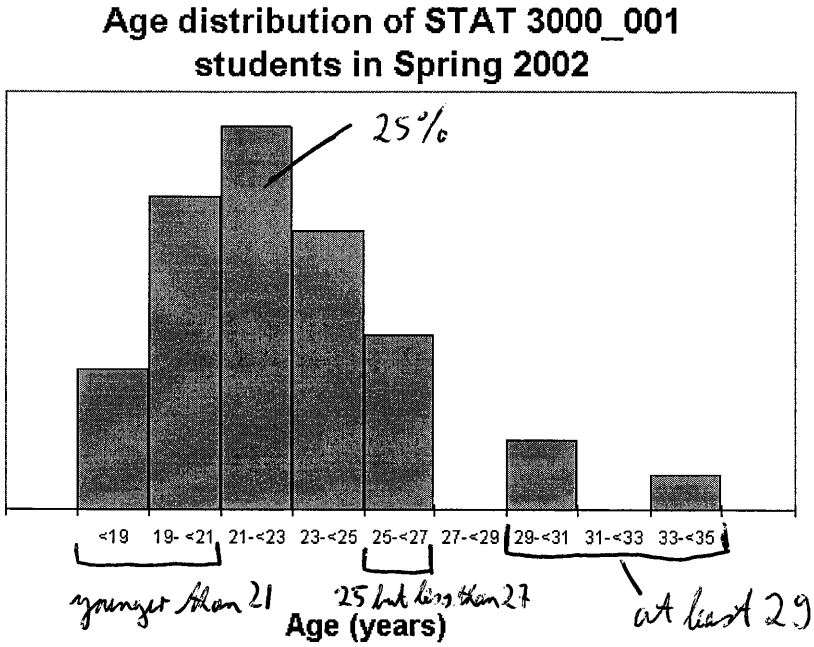
Friday, September 10, 2004

Your Name: _____

from Quiz 1, Spring 2002

Question 1: Histograms I (**12 Points**)

The histogram below shows the age distribution of Stat 3000, Section 001, students for the Spring 2002 semester. Unfortunately, the labels on the vertical axis have been deleted. However, the instructor recalls that there have been about 25% of students who were at least 21 but less than 23 years old. Try to help the instructor to fill in some of the missing percentages.



-3 if a percentage makes no sense, i.e., is outside the "acceptable answers"

(4) 1. What approximate percentage of students were at least 25 but less than 27 years old?

answer: about 12% acceptable answers: anything between 10% and 13%

explanation (not required): the height of the bar from 25 to 27 is slightly less than ½ of the height of the bar from 21 to 23, so slightly less than ½ · 25%

(4) 2. What approximate percentage of students were younger than 21 years of age?

answer: about 30% acceptable answers: anything between 26% and 40%

explanation (not required): stack the 17 to 19 bar on top of the 19 to 21 bar and this gets higher than the 21 to 23 bar; the extra height represents about 5%

(4) 3. What approximate percentage of students were at least 29 years old?

answer: about 8% acceptable answers: anything between 5% and 11%

Please turn over!

explanation (not required): stack the 33 to 35 bar on top of the 29 to 31 bar and this still is less than the 25 to 27 bar (which is about 12%)

From: FPP, p. 41, Exercise Set C, Exercise 3

Question 2: Histograms II (**8 Points**)

An investigator draws a histogram for some height data, using the metric system. She is working in centimeters (cm). The vertical axis shows density, and the top of the vertical axis is 10 percent per cm. Now she wants to convert to millimeters (mm). There are 10 millimeters to the centimeter. On the horizontal axis, she has to change 175 cm to 1,750 (2) mm, and 200 cm to 2,000 (2) mm. On the vertical axis, she has to change 10 percent per cm to 1 (2) percent per mm, and 5 percent per cm to 0.5 (2) percent per mm.

Textbook (page A-46):

"The idea on density: If you spread 10 percent evenly over 1 cm = 10 mm, there is 1 percent in each mm, that is, 1 percent per mm."

Statistics 1040, Section 006, Quiz 3 (20 Points)

Friday, September 17, 2004

Your Name: _____

based on: Quiz 2, Spring 2003

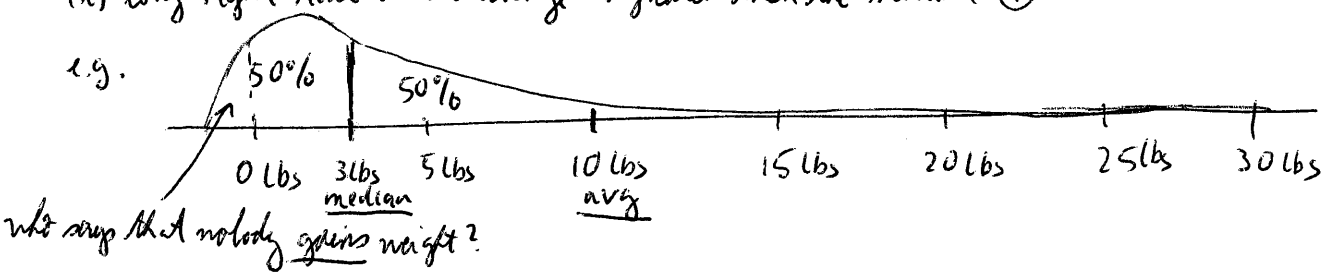
Question 1: Measures of Center and Spread I (**10 Points**)

Suppose an advertisement reported that the average weight loss after using a certain exercise machine for 2 months was 10 pounds. You investigate further and discover that the median weight loss was 3 pounds. (3)

1. Is it most likely that the histogram of all weight losses (i) has a long right tail, (ii) has a long left tail, or (iii) is symmetric? Circle your answer and explain!

(i) long right tail: the average is greater than the median (1)

e.g.



who says that nobody gains weight?

2. As a **consumer** trying to decide whether to buy this exercise machine, would it be more helpful for you to know the mean or the median? Circle your answer and explain! (2)

the median reveals that 50% of the customers have lost only 3 lbs or less - the exercise machine may not be worth the money (1)

3. As a **sales manager of the company** producing these exercise machines, would it be more commercially profitable for you (and your company) to publish the mean or the median? Circle your answer and explain! (2)

customers buying such an exercise machine hope to loose as much weight as possible; obviously, 10 lbs sounds better than 3 lbs (1)

Please turn over!

Question 2: Measures of Center and Spread II (**10 Points**)

Find the average and the standard deviation of the following two lists of numbers:

	Numbers	Average	Standard deviation
List 1:	7, 7, 7, 7	7 (2)	0 (2)
List 2:	7, –7, 7, –7	0 (2)	7 (2)

Show your work and/or give a short explanation for your answer.
Note: A calculation is NOT always necessary!
Use the formulas provided below (if needed).

Explanation:

List 1: average = 7 (because all numbers are equal, the average must be the same)
SD = 0 (because all numbers are equal, there is no spread) (1)

List 2: average = 0 (the sum is 0, so the average is 0)
SD = 7 (because all numbers are equally far away from the average [but numbers are not equal], the SD must be the same) (1)

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 006, Quiz 4 (20 Points)

Friday, September 24, 2004

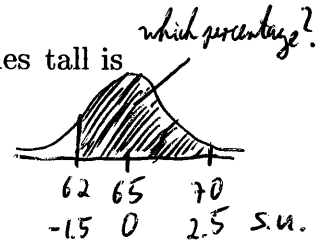
Your Name: _____

-2 for each calculation error
+3 for correct graph and nothing else

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

The heights of women approximately follow the normal curve with an average of **65 inches** and a standard deviation of **2.0 inches**. Answer the questions below:

- **(12 Points)** The percentage of women who are between 62 and 70 inches tall is roughly 92.70 %.



(1) convert 62 and 70 into standard units: (3)

$\frac{62-65}{2.0} = -1.5$ s.u., $\frac{70-65}{2.0} = 2.5$ s.u. (3)

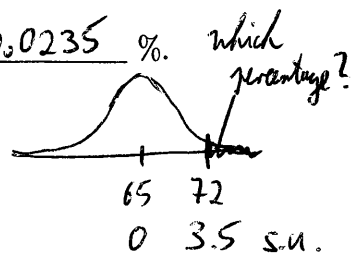
(2) area between -1.5 and 1.5: 86.64% (2)

area between -2.5 and 2.5: 98.76% (2)

(3) area between -1.5 and 2.5: $\frac{86.64\%}{2} + \frac{98.76\%}{2} = 92.70\%$ (2)

- **(8 Points)**

The percentage of of women who are taller than 72 inches is about 0.0235 %.



(1) convert 72 into standard units:

$\frac{72-65}{2.0} = 3.5$ s.u. (3)

(2) area between -3.5 and 3.5: 99.953% (2)

(3) area above 3.5: $\frac{100\% - 99.953\%}{2} = \frac{0.047\%}{2} = 0.0235\%$ (3)

Show your work!

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

Friday, October 1, 2004

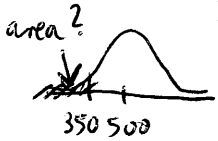
Your Name: _____

from: FPP, p. 95, Review Exercise 7 & Quiz 4, Spring 2002

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

Among freshmen at a certain university, scores on the Math SAT followed the normal curve, with an average of 500 and an SD of 100. Fill in the blanks. **Show your work!**

1. (**6 Points**) A student who scored 350 on the Math SAT was at the ___7___th percentile of the score distribution.



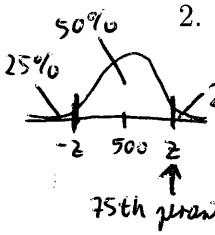
transfer 350 in s.u.: $\frac{350-500}{100} = -1.5$

area from -1.5 to 1.5 (from Table): 86.64%

area below -1.5: $\frac{100\% - 86.64\%}{2} = \frac{13.36\%}{2} = 6.68\%$, i.e., about 7% of the area lies below 350

-2 for incorrect s.u.
-2 for incorrect table value
-2 for incorrect area

2. (**6 Points**) To be at the 75th percentile of the distribution, a student needed a score of about ___565 (or 570)___ points on the Math SAT.



Find a value z such that the area between -z and z is about 50%: z = 0.65 (48.43%) or z = 0.70 (51.61%)

Transfer into original units: for z = 0.65: 0.65 · 100 + 500 = 565
for z = 0.70: 0.70 · 100 + 500 = 570

Note: either 0.65 or 0.70 are valid answers.

A score of 565 (or 570) was needed to be at the 75th percentile.

-2 for incorrect area
-2 for incorrect z (not 0.65 or 0.70)
-2 for incorrect original units

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

from: FPP, p. 130, Exercise Set B, Question 8 (Answers: p. A-56)

Investigators take a sample of DINKS (dual–income families, where husband and wife both work and have no kids). The investigators have data on the husband's income and the wife's income. By definition,

family income = husband's income + wife's income.

The average family income was around $50,000, and 10% of the couples had family income in the range $45,000–$55,000. Fill in the blanks, using the options given below, and **explain briefly**:

1. (**4 Points**) The correlation between wife's income and family income is ___(e) somewhat positive___.

Textbook: "Although wife's income must be less than family income, the two are positively associated."

2. (**4 Points**) Among couples whose family income is in the range $45,000–$55,000, the correlation between wife's income and husband's income is ___(b) nearly -1___.

Textbook: "If family income is practically constant, the more the wife makes, the less the husband can make."

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

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

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

Statistics 1040, Section 006, Quiz 6 (20 Points)

Friday, October 15, 2004

Your Name: _____

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

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

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 (μmol/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: blood lead = 0.11 + 0.019 · hours or y = 0.11 + 0.019 · x (2)

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

   $\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 0.42 - 0.019 \cdot 16.2 = 0.42 - 0.31 = 0.11$ (3)

   -1 if only part of the equation (e.g. 0.11 + 0.019 · 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

   blood lead = 0.11 + 0.019 · 21 = 0.11 + 0.399 = 0.509 ≈ 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

r.m.s. error $= \sqrt{1 - r^2} \cdot SD_y$ (3)

$= \sqrt{1 - 0.6^2} \cdot 0.19$

$= \sqrt{1 - 0.36} \cdot 0.19$

$= \sqrt{0.64} \cdot 0.19$

$= 0.8 \cdot 0.19 = 0.152$ (2)

4. (**5 Points**) Would you be surprised to learn that a cyclist who trained for 3 hours had a blood lead of .8 μmol/L? Support your answer with a brief explanation and calculation. (1)

Answer: **Yes, surprised** / **No, not surprised**

Explanation:

predicted after 3 hours:

blood lead $= 0.11 + 0.019 \cdot 3 = 0.11 + 0.057 = 0.167$ (2)

observed after 3 hours: 0.8

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

so 0.8 is more than 4 r.m.s. errors away from the predicted → very unusual! (2)

Formulas:

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

$$\text{slope} = r \times \frac{\text{SD}_y}{\text{SD}_x}$$

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

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

Friday, October 22, 2004

Your Name: _____

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

A drawer of socks contains 20 socks of which 6 are black, 10 are green, and 4 are blue. In the dark, a child chooses two socks at random to wear to school.

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

   The chance is: 20.0 %

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

   first blue: $\frac{4}{20} = 0.20 = 20.0\%$ (2) (1) (2)

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

   The chance is: 3.2 %

   first blue: $\frac{4}{20}$ > dependent

   second blue, given first blue: $\frac{3}{19}$ multiplication rule

   both blue: (1) (2) $\frac{4}{20} \cdot \frac{3}{19} = \frac{12}{380} = 0.032 = 3.2\%$

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

   The chance is: 21.1 %

   (BG) & (GB) mutually exclusive

   (BG) first blue: $\frac{4}{20}$
   second green, given first blue: $\frac{10}{19}$ > dependent

   (GB) first green: $\frac{10}{20}$
   second blue, given first green: $\frac{4}{19}$ > dependent

   one blue, one green: (1) (2) $\frac{4}{20} \cdot \frac{10}{19} + \frac{10}{20} \cdot \frac{4}{19} = \frac{40+40}{380} = \frac{80}{380} = 0.211 = 21.1\%$

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

   The chance is: 34.7 %

   (BLBL) & (GG) & (BB) mutually exclusive

   (BLBL) both black: $\frac{6}{20} \cdot \frac{5}{19}$

   (GG) both green: $\frac{10}{20} \cdot \frac{9}{19}$

   (BB) both blue: $\frac{4}{20} \cdot \frac{3}{19}$

   same color: (1) (1) (1) (1) (1) $\frac{6}{20} \cdot \frac{5}{19} + \frac{10}{20} \cdot \frac{9}{19} + \frac{4}{20} \cdot \frac{3}{19} = \frac{30+90+12}{380} = \frac{132}{380} = 0.347 = 34.7\%$

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

Friday, October 29, 2004

Your Name: _____

from: Quiz 7, Fall 2003

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

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

-2 for minor mistake
-3 for major mistake (e.g. % Symbol)
-1 if # draws not stated

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

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

Number of draws: 200

or

| 3 × 3 | 1 × 0 | 6 × -1 |

draws: 200

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

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

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

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

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

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

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

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

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

Please turn over!

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

based on: FPP, p. 286, Review Exercise 9

Question 2: Law of Averages (**4 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.**

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.

(3) (2.) (1)

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.

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, Section 006, Quiz 9 (20 Points)

Friday, November 5, 2004

Your Name: _____

from: Quiz 8, Spring 2003

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.

15,000 × [1] 25,000 × [0]

draws = 1,000

- 1 if slightly incorrect number of 0/1's in box
- 2 if box given as [0] [1] etc.
- 3 if box contains something else than 0/1's
- 1 if # draws missing or 375 incorrect

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

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

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

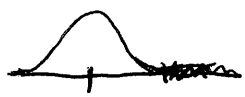
$EV_{sum} = 1,000 \cdot 0.375 = 375$

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

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



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.

- ½ 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

Please turn over!

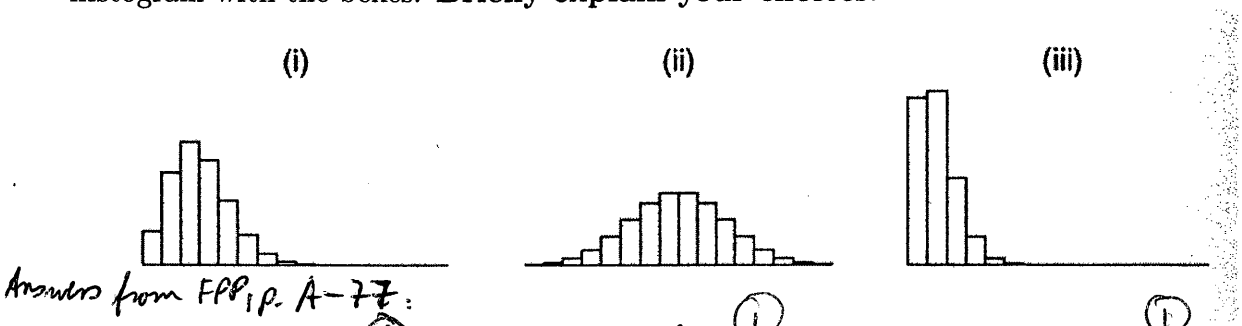
From: FPP, p. 324, Exercise Set C, Question 5
& Stat 1040, Fall 2003, Midterm 2, Question 5.1

Question 2: Probability Histograms (**5 Points**)

Twenty–five draws are made at random with replacement from the each of the boxes below:

A) $\boxed{\boxed{0}\ \boxed{1}}$ B) $\boxed{9\ \boxed{0}\text{'s}\ \boxed{1}}$ C) $\boxed{24\ \boxed{0}\text{'s}\ \boxed{1}}$

The probability histograms for the sums are shown below, in scrambled order. Match the histogram with the boxes. **Briefly explain your choices.**



Answers from FPP, p. A-77:

(i) goes with B) (2); (ii) goes with A) (1); (iii) goes with C) (1)

Explanation: "The more lopsided the box, the more skewed the histogram." (1)

[see back for additional comment]

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 $\boxed{0}$'s and $\boxed{1}$'s:

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

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

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

Friday, November 19, 2004

Your Name: _____

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

Recent political events are currently in focus of many surveys and polls nationwide. With four members of the Bush Cabinet resigning over the last few days, a natural concern for every U.S. citizen is: *Will the Bush Cabinet resignations have a positive or negative impact on U.S. policy?*

This question was asked to a sample of 787 U.S. citizens: 299 of them answered "Positive".

1. (**14 Points**) Construct a 87% confidence interval for the percentage of all U.S. citizens who think that the Bush Cabinet resignations will have a positive impact on U.S. policy.

   Show your work.

box unknown: | 2 × 1 2 × 0 |

1: Positive
0: Negative

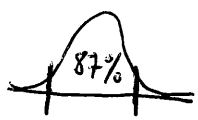
draws: 787

sample % = $\frac{299}{787}$ = 0.38 = 38% (2) = population % (assumption)

SD box = $\sqrt{0.38 \cdot 0.62} = \sqrt{0.2356} = 0.485$ (2) (via bootstrap)

$SE_{sam} = \sqrt{787} \cdot 0.485 = 28.05 \cdot 0.485 = 13.6$ (2)

SE % = $\frac{13.6}{787} \cdot 100\% = 1.73\%$ (2)



87% CI: sample % ± (multiplier for 87%) · SE %

= 38% ± 1.50 (2) · 1.73%

= 38% ± 2.6%

= 35.4% to 40.6% (4)

-2 each calculation error
-1 if % forgotten
-1 if box indicated

Please turn over!

(box = population is unknown!)

2. (**6 Points**) For each of the following situations, explain **why** or **why not** it would be possible to constuct a 87% confidence interval for the percentage of all U.S. citizens who think that the Bush Cabinet resignations will have positive impact on U.S. policy. **Please do not construct the actual confidence interval – just answer each question with Yes or No and provide a very brief explanation.**

- The sample of 787 U.S. citizens was obtained by using a computer to randomly generate a sufficent number of valid telephone numbers (including area code) and calling these numbers until 787 valid answers were collected.

  Is it possible to construct a 87% CI here? – Yes or (No?) (1)

  Explanation:

  - this is not a SRS, but biased in favor of people with more than 1 phone line (e.g., residential & cell phone) & biased against people with caller ID (that often do not pick up calls when they can't identify the caller) => this clearly does not result in a CI for all U.S. citizens (1)

- The sample of 787 U.S. citizens was obtained as a SRS from all U. S. citizens, but 780 of the responders said "Positive" (i.e., thought that the Bush Cabinet resignations will have positive impact on U.S. policy).

  Is it possible to construct a 87% CI here? – Yes or (No?) (1)

  Explanation:

  - although this is a SRS, it is

    sample % = $\frac{780}{787}$ = 0.991 = 99.1% (1)

    which is too close to 100%

- The 787 answers come from the Quick Poll at the CNN Web page (http://www.cnn.com).

  Is it possible to construct a 87% CI here? – Yes or (No?) (1)

  Explanation:

  - this is not a SRS, but biased in favor of people that have internet access, read the CNN Web page, and may have some strong opinion (1)

    => this clearly does not result in a CI for all U.S. citizens

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

Monday, November 29, 2004

Your Name: _____

From: Stat 1040, Summer 2000, Test 2, July 17, 2000, Question 5

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

When the latest Harry Potter book was released, a local bookstore sold 500 copies in one day. They randomly sampled 300 of these people and found that the average age for these 300 people was 13.3 years with a standard deviation of 8.2 years.

1. (**14 Points**) Find a 95% confidence interval for the average age of all 500 customers.

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

Assume: sample avg = 13.3 years = base avg (2)
sample SD = 8.2 years = base SD (1)

$SE_{sum} = \sqrt{300} \cdot 8.2 = 17.3 \cdot 8.2 = 142.0$ years (2)

$SE_{avg} = \frac{142.0 \text{ years}}{300} = 0.473$ years (2)

95% CI: sample avg ± multiplier · SE_{avg} (1)
(1) (2) (1)

$= 13.3 \pm 2 \cdot 0.473$
(2 ↖ related to 95%)

$= 13.3 \pm 0.946$

$\approx$ 12.35 years to 14.25 years (2)

2. (**6 Points**) True or false? — Answer these three questions (**no** explanation needed):

   (a) 95% of the 300 sampled customers had ages in the interval from part 1. **True** or **False**? Just circle your answer. (2)

   [A 95% CI for the average does not mean that 95% of the data falls in this interval.]

   (b) We cannot approximate the percentage of customers who had ages in the interval from part 1. because the ages do not follow the normal curve. (2) **True** or **False**? Just circle your answer.

   [Less than 2 SD below the avg are the negative numbers - and negative ages are impossible.]

   (c) The interval from part 1. is invalid because ages do not follow the normal curve. (2) **True** or **False**? Just circle your answer.

   [Even if the data do not follow the normal curve, the avg of 300 draws most likely will follow the the normal curve - so the interval is valid.]

Please turn over!

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

Friday, December 3, 2004

Your Name: _____

from: Stat 1040, Fall 2002, Quiz 11
& FPP, p. 499, Review Exercise 8

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

Bookstores like education, because national data show that 71% of college graduates have read a book in the past year, compared to 54% of the general population age 18 and over. The data also show the nationwide average educational level to be 13 years of schooling completed, with an SD of about 3 years, for persons age 18 and over.
A bookstore is doing a market survey in a certain county, and takes a sample of 1,000 people age 18 and over. They find the average educational level to be 14 years, and the SD is 5 years. Can the difference in average educational level between the sample and the nation be explained by chance variation? If not, what other explanations can you give? Please **follow the steps below** in answering these questions.

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

Null: County's avg education level matches nationwide level, i.e., box avg = 13 years

Alternative: County's avg education level is different from nation, i.e., box avg ≠ 13 years

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

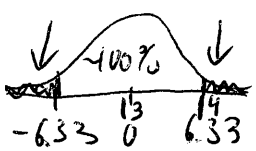
observed avg = 14
expected avg = 13
sample SD = 5 (use this!)

$SE_{sum} = \sqrt{1{,}000} \cdot 5 = 158.1$

$SE_{avg} = \frac{158.1}{1{,}000} = 0.158$

$\rightsquigarrow z = \frac{obs - exp}{SE} = \frac{14-13}{0.158} = 6.33$

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



P-value is area on both sides, but this is almost 0%

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

Reject null hypothesis; the result is highly statistically significant (P-value ≈ 0% < 1%).
The avg education level for this county is different from the national avg;
more specifically, it is above the national avg; this could be a rich suburban county.

Please turn over!

Grading Criteria:

1) Wrong null and alternative hypotheses, e.g.:

- mixed null and alternative	-3
- "14" in hypothesis instead of "13"	-3
Hypotheses stated in words only or in numbers only	-1 each

2) incorrect z, e.g. $\frac{exp-obs}{SE}$ or $\frac{obs-exp}{SD}$ -2 [correct: $\frac{obs-exp}{SE}$]

incorrect SE_{avg}	-2
calculation error	-1 each

3)

incorrect area	-2
incorrect table value	-2
calculation error	-1

4)

reject null if p-value > 5%	-4
(or do not reject null if p-value < 5%	-4)
no explanation (e.g., reject, but no conclusion)	-2
Correctly rejecting, but explanation missed up	-2
if not speaking of rejecting / not rejecting	-2
if not speaking of (highly) statistically significant	-1