Statistics 1040, Section 009, Quiz 1 (20 Points)

Friday, January 13, 2006

Your Name: _____

from: FPP, p. 25, Review Exercise 4 & Quiz 1, Spring 2005

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

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

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

Worklock:

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

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

Worklock: (3)

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

Please turn over!

from: FPP, p. 10-11, "Summary" & Quiz 1, Spring 2005

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

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

Statisticians want to know the effect of a treatment (or vaccine) (1) (like the Salk vaccine) on a response (like getting polio). To find out, they compare the responses of a treatment group (1) with a control group (1).

To make sure that the treatment group is like the control group, investigators put subjects (1) into the treatment or the control group at random (1).

Whenever possible, the control group is given a placebo (1), which is neutral but resembles the treatment.

In a double-blind (1) experiment, the subjects do not know whether they are in the treatment or in the control group; neither do those who evaluate the responses.

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

Statistics 1040, Section 009, Quiz 2 (20 Points)

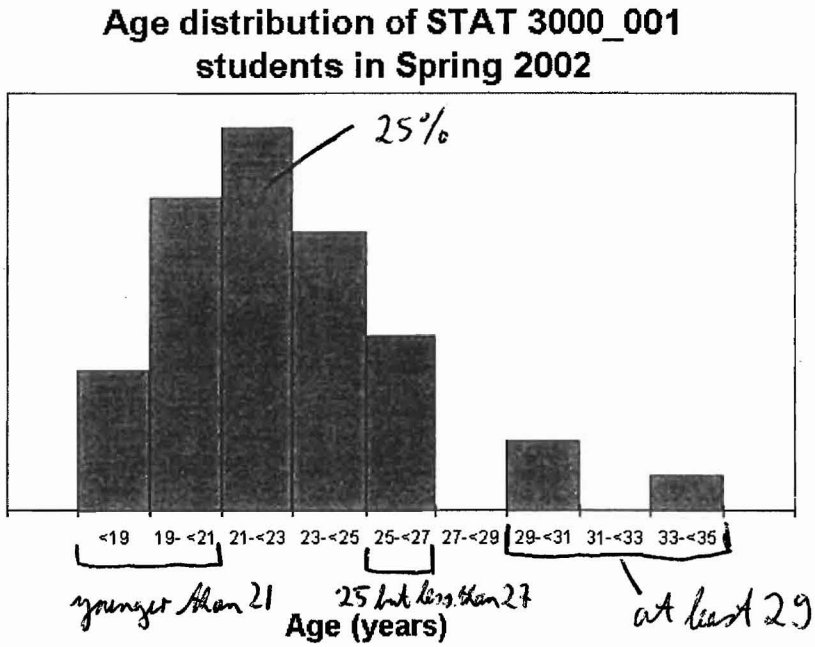
Friday, January 20, 2006

Your Name: _____

from: Quiz 1, Spring 2002 & Quiz 2, Fall 2004

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%

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%)

Please turn over!

From: FPP, p. 41, Exercise Set C, Exercise 3 & Quiz 2, Fall 2004

Question 2: Histograms II (**8 Points**) [Answers: → Textbook !]

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 009, Quiz 3 (20 Points)

Friday, January 27, 2006

Your Name: _____

Question 1: Measures of Center and Spread I (**14 Points**)

Below are the temperatures (in degrees Celsius) for five locations in Utah on Tuesday, January 20, 2004, at 9pm SMT, as found on `www.wunderground.com`:

City	Temperature
Bryce Canyon	-15
Logan	-14
Ogden	-12
St. George	5
Salt Lake City	-4

-1 for each calculation error

Show your work!

1. (**5 Points**) Find the **average temperature** in degrees Celsius for these locations in Utah.

$$\text{avg} = \frac{(-15)+(-14)+(-12)+5+(-4)}{5} = \frac{-40}{5} = -8^\circ C$$

(3) (2)

2. (**3 Points**) Find the **median temperature** in degrees Celsius for these locations in Utah.

sorted list: $-15 \quad -14 \quad -12 \quad -4 \quad 5$ (3)

median $= -12^\circ C$

3. (**6 Points**) Find the **standard deviation** of the temperatures for these locations in Utah.

1) avg $= -8$ (1)

2) $-15-(-8) = -7$
$-14-(-8) = -6$
$-12-(-8) = -4$
$5-(-8) = 13$
$-4-(-8) = 4$ (1)

3) $(-7)^2 = 49$
$(-6)^2 = 36$
$(-4)^2 = 16$
$13^2 = 169$
$4^2 = 16$ (1)

4) $\frac{49+36+16+169+16}{5} = \frac{286}{5} = 57.2$ (1)

5) $SD = \sqrt{57.2} = 7.56^\circ C$ (2)

1

Please turn over!

Question 2: Measures of Center and Spread II (**6 Points**)

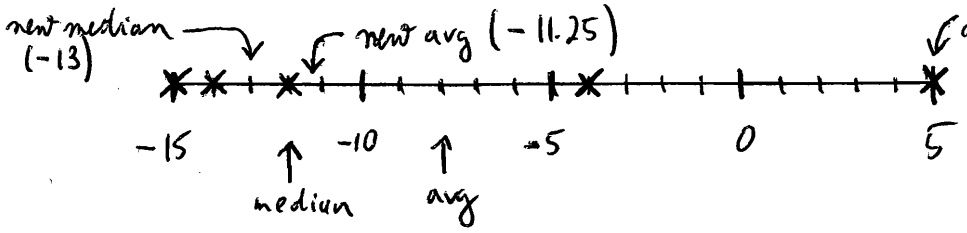
To answer the questions below, you need to apply your knowledge about average, median, and standard deviation. **No calculation is needed!**

1. (**3 Points**) If the St. George temperature (the only positive value) is removed from the list, what will happen to the average and median? Choose the most appropriate answer and **explain** briefly:

(2) (a) The average will change more than the median;

(b) The median will change more than the average;

(c) Both average and median will stay exactly the same.



Explanation: (1)
+5 is a very large value. We have seen in class how such a large value pulls the average towards it. If such a large value is removed, the average will change considerably (-8 to -11.25) while the median only changes a bit (-12 to -13).

2. (**3 Points**) If the St. George temperature (the only positive value) is removed from the list, what will happen to the standard deviation? Choose the most appropriate answer and **explain** briefly:

(a) The SD will become bigger;

(2) (b) The SD will become smaller;

(c) The SD will become negative;

(d) The SD will not change at all.

Explanation: (1)
The SD describes the spread of the data. If the largest value is removed, the spread can only become smaller (from 7.56 to 4.32).
The SD is <u>never</u> negative (and the SD is 0 only if all numbers are exactly the same – meaning there is no spread).

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 009, Quiz 4 (20 Points)

Friday, February 3, 2006

Your Name: _____

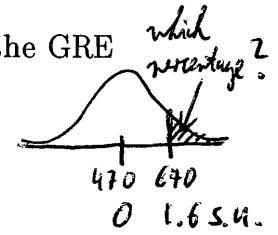
from: Quiz 4, Spring 2005

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

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

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

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



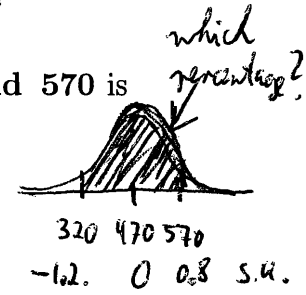
① convert 670 into standard units:

$\frac{670-470}{125} = 1.6$ s.u. ③

② area between −1.6 and 1.6: 89.04% ②

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

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



① Convert 320 and 570 into standard units:

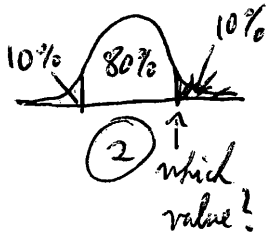
$\frac{320-470}{125} = -1.2$ s.u. ① $\frac{570-470}{125} = 0.8$ s.u. ①

② area between −1.2 and 1.2: 76.99% ①
area between −0.8 and 0.8: 57.63% ①

③ area between −1.2 and 0.8:

$\frac{76.99\%}{2} + \frac{57.63\%}{2} = 67.31\%$ ③

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



① Find a value z such that the area between −z and z is about 80%:

z = 1.30 (gives 80.64%) ②

② Transfer into original units:

$1.30 \cdot 125 + 470 = 632.5$ ②

Please turn over!

Statistics 1040, Section 009, Quiz 5 (20 Points)

Friday, February 10, 2006

Your Name: _____

Question 1: Change Of Scale (**10 Points**)

In a class experiment last week, we measured the length of a pencil (including the eraser) nine times. After adjustment of one outlier, all values looked reasonable. The average length of our nine measurements was 7.5 inches, with an SD of 0.14 inches. Recall that 1 inch = 2.54 cm.

If we translate these results into cm, the average length will be 19.05 (4) cm, with a standard deviation of 0.3556 (4) cm.

Be precise and report **all** digits from your calculator this time (e.g., if your calculator shows 27.8835, then report this number and do not report 28 instead).

Show your work! avg (in cm): $7.5 \cdot 2.54$ (1) $= 19.05$ cm

SD (in cm): $0.14 \cdot 2.54$ (1) $= 0.3556$ cm

from: FPP, p. 138, Review Exercise 11 [Solution → Workbook!]

Question 2: Correlation (**10 Points**)

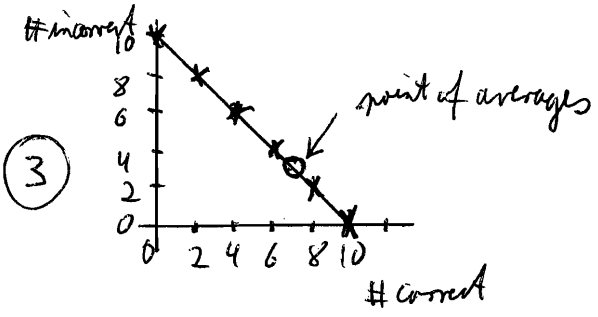
A teaching assistant gives a quiz to his section. There are 10 questions on the quiz and no part credit is given. After grading the papers, the TA writes down for each student the number of questions the student got right and the number wrong. The average number of right answers is 6.4 with an SD of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

The correlation coefficient is

(7) (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

Circle your answer and explain.

The number of correct answers and the number of incorrect answers add up to 10! (3)



Statistics 1040, Section 009, Quiz 6 (20 Points)

Friday, February 24, 2006

Your Name: _____

from: Quiz 6, Spring 2005

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

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

in all parts:
-2 for each calculation error
-2 for each incorrect value used
-2 for mixing x and y

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

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

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

-2 if only part of the equation (e.g. $29 + 0.68x$)
-2 if not specifying x & y

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

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

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

2. (**5 Points**) Use the regression equation from part 1. to predict the end–of–year score for a student who scored 85 on the beginning–of–year test.
The predicted end–of–year score is: 86.8

-1 if correct result, but according to old method
-2 if result makes no sense at all!

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

(1) (1) (1) (1) (1)

3. (**5 Points**) Find the r.m.s. error for predicting the end–of–year score from the beginning–of–year score.
The r.m.s. error is: 13.6

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

$= \sqrt{1 - 0.6^2} \cdot 17$

$= \sqrt{1 - 0.36} \cdot 17$

$= \sqrt{0.64} \cdot 17$

$= 0.8 \cdot 17$

$= 13.6$ (2)

Please turn over!

Statistics 1040, Section 009, Quiz 7 (20 Points)

Friday, March 3, 2006

Your Name: _____

based on: Quiz 6, Question 1, Spring 2003
& Quiz 7, Question 1, Spring 2005

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

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

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

1. (**5 Points**) What is the chance that the first is **not** mint?

   The chance is 66.7 %.

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

   first not mint: $1 - \frac{5}{15} = \frac{15}{15} - \frac{5}{15} = \frac{10}{15} = 0.667 = 66.7\%$

   ① ② ②

   ↖ opposite rule

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

   The chance is 2.86 %.

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

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

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

   ① ② ②

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

   The chance is 4.76 %.

   first cherry: $\frac{2}{15}$

   second caramel, given first cherry: $\frac{5}{14}$ > dependent

   first cherry and second caramel: $\frac{2}{15} \cdot \frac{5}{14} = \frac{10}{210} = 0.0476 = 4.76\%$ — multiplication rule

   ① ② ②

Please turn over!

from: FPP, Review Exercise 7, p. 235 & Quiz 7, Question 2, Spring 2005

Question 2: Chance/Probability II **(5 Points)**

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

(i) H T T H T H (ii) H H H H H H

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

Which of the following is correct?

Circle your answer and explain:

1. Sequence (i) is more likely.
2. Sequence (ii) is more likely.
3. Both sequences are equally likely.

(3)

Workbook answer:

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

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

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

Statistics 1040, Section 009, Quiz 8 (20 Points)

Friday, March 10, 2006

Your Name: _____

from: Quiz 7, Fall 2003
& Quiz 8, Fall 2004

Question 1: Box Models, EV, and SE (**12 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.

Show your work!

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

   3 × [3] 1 × [0] 6 × [-1]

   # draws: 200

   - 1 for minor mistake
   - 2 for major mistake (e.g., 3, 6, 9 in box)
   - 1 if # draws not stated

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

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

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

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

3. (**5 Points**) Find the standard error of your gain/loss.

   $$\text{box SD} = \sqrt{\frac{3 \cdot (3-0.3)^2 + 1 \cdot (0-0.3)^2 + 6 \cdot (-1-0.3)^2}{10}}$$
   $$= \sqrt{\frac{3 \cdot 2.7^2 + 1 \cdot (-0.3)^2 + 6 \cdot (-1.3)^2}{10}}$$
   $$= \sqrt{\frac{3 \cdot 7.29 + 0.09 + 6 \cdot 1.69}{10}}$$
   $$= \sqrt{\frac{32.1}{10}} = \sqrt{3.21} = 1.79$$

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

Please turn over!

from: FPP, p 285, Review Exercise 4 [Answers → Worklook !]

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

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

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) [To win, you need a large percentage error, and that is more likely in 60 rolls.]
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) [Now you want a small percentage error.]
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) [Again - you want a small percentage error.]
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) [Because to get exactly the expected value means getting exactly zero chance error, and that is more likely with fewer rolls.]

Formulas:

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

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

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

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

Statistics 1040, Section 009, Quiz 9 (20 Points)

Friday, March 24, 2006

part 1.):
-1 if slightly incorrect number of 0/1's in box
-1 if box given as [0 1]
-2 if box contains something other than 0/1's
-1 if # draws missing/incorrect

Your Name: _____

from Quiz 9, Spring 2004

Question 1: EV, SE, and Normal Curve **(14 Points)**

According to the U.S. Census Bureau's "QuickFacts" Web sited (`http://quickfacts.census.gov/qfd/states/49000.html`), about 26% of Utah residents age 25 and older have a bachelor degree or higher. Suppose that 500 Utah residents age 25 and older have been randomly chosen to participate in a survey.

1. **(2 Points)** Find the box model.

1: bachelor degree or higher
0: no such degree

26 × [1]	74 × [0]

draws: 500

2. **(6 Points)** Find the expected number of Utah residents in this sample of 500 who have a bachelor degree or higher. What is the corresponding SE?

box avg = fraction of 1's = $\frac{26}{100} = 0.26$

box SD = $\sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{\frac{26}{100} \cdot \frac{74}{100}} = \sqrt{0.1924} = 0.4386 \approx 0.44$

$EV_{sum} = 500 \cdot 0.26 = 130$

$SE_{sum} = \sqrt{500} \cdot 0.44 = 22.36 \cdot 0.44 = 9.84$

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

3. **(6 Points)** Using the normal curve, find the chance that **at most** 120 of the Utah residents in the sample have a bachelor degree or higher.

120 130
-1 0 s.u.

s.u.: $\frac{120-130}{9.84} = -1.02 \approx -1.0$

area between −1.0 and 1.0: 68.27%

area below −1.0: $\frac{100\% - 68.27\%}{2} = \frac{31.73\%}{2} = 15.87\%$

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

Please turn over!

based on FPP: p. 328, Chapter 18, Review Exercise 8 & Quiz 9, Spring 2004

Question 2: Normal Approximation for Probability Histograms (**6 Points**)

A coin is tossed 100 times. True or false? Circle your answer. Answer each of the following questions separately. **No explanation is needed.**

1. (**1 Point**) True (1) or false: The expected value for the number of heads is 50.

#H: 1
T: 0
box: [0] [1]
draws: 100

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

so $EV_{sum} = 50$? yes (see left)
⇒ True

2. (**1 Point**) True or false: (1) The expected value for the number of heads is 50, give or take 5 or so.

The EV_{sum} is exactly 50, no give or take
⇒ False

3. (**2 Points**) True or false: (2) The number of heads will be 50.

The number of heads most likely will not be exactly 50, but it will be relatively close to 50
⇒ False (2)

4. (**2 Points**) True or false: (2) The number of heads will be around 50, give or take 3 or so.

As calculated in 1,
$EV_{sum} = 50$
and
$SE_{sum} = 5$ = "give or take" part ($\neq 3$).
⇒ False (overall since the SE_{sum} is incorrect)

Please turn over!

Statistics 1040, Section 009, Quiz 10 (20 Points)

Wednesday, April 12, 2006

Your Name: _____

from: Quiz 10, Fall 2004

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

Political events in the Fall of 2004 were in focus of many surveys and polls nationwide. With four members of the Bush Cabinet resigning within a few days in Fall 2004, a natural concern for every U.S. citizen at that time was: *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.

1: Positive
0: Negative

box unknown: | ? × 1 ? × 0 |

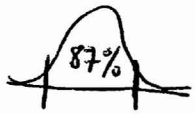
draws: 787

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 box strategy)

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

    $$\text{sample \%} = \frac{780}{787} = 0.991 = 99.1\%$$ (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 009, Quiz 11 (20 Points)

Wednesday, April 19, 2006

Your Name: _____

from: Stat 1040, Fall 1999, Final Test, December 17, 1999, Question 3

Question 1: Tests of Significance **(20 Points)** & Quiz 11, Fall 2003

-2 for each calculation error

A journal article claims that 60% of North American adults are too sedentary. Suppose you think a lower percentage of Cache Valley adults are too sedentary. To test your belief, you take a simple random sample of 120 Cache Valley adults and find that only 68 of them are too sedentary. Test to see if your belief is correct. (You may assume that everybody is using the same definition of "too sedentary", although in practice this would be contentious). **Show your work!**

-2 if null, alt. mapped

z-test:

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

(2) (1)

null: Cache Valley adults are also "too sedentary", i.e., box % = 60% (1)

(2)

alternative: fewer Cache Valley adults are "too sedentary", i.e., box % < 60%

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

If the null is true, the sample % should be like the % of 1's in 120 draws from [60×1, 40×0]:

(2) observed % = $\frac{68}{120} \cdot 100\% = 56.7\%$

expected % = 60% = EV %

$SD_{box} = \sqrt{0.6 \cdot 0.4} = \sqrt{.24} = 0.49$ (1)

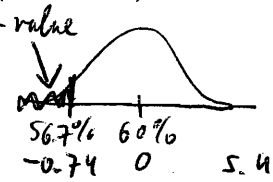
$SE_{sum} = \sqrt{120} \cdot 0.49 = 10.95 \cdot 0.49 = 5.37$ (1)

$SE_{\%} = \frac{5.37}{120} \cdot 100\% = 4.475\%$ (1)

$z = \frac{\text{obs \%} - \text{EV \%}}{SE_{\%}} = \frac{56.7\% - 60\%}{4.475\%} = -0.74$ (2)

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

(3)



(2) area between −0.75 & 0.75: 54.67% (1)

area below −0.75: $\frac{100\% - 54.67\%}{2} = 22.67\%$

P-value ≈ 23% (2)

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

(3)

(4) ⇝ do not reject null (P-value > 5%)

• Cache Valley adults are also "too sedentary" (i.e., no evidence that Cache Valley adults are less sedentary than the national figure)

(2)

Please turn over!

Statistics 1040, Section 009, Quiz 12 (20 Points)

Wednesday, April 26, 2006

Your Name: _____

from: Stat 1040, Fall 2004, Final Test, December 15, 2004, Question 9

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

Researchers think anti–epileptic drugs accelerate bone loss in elderly women. To investigate, 12 women were randomly selected from all elderly women taking anti–epileptic drugs and they were monitored for a period of 5 years. At the end of the study, researchers measured their bone mineral density on a standardized scale. The average of the 12 measurements was -0.24 with an SD of 1.22. It is known that bone density measurements follow the normal curve. (Note that negative values of bone mineral density correspond to accelerated bone loss.)

Test the hypothesis that the average for all such women is 0.0 against the alternative hypothesis that it is less than 0.0. State a null and an alternative hypothesis, find a test statistic and a P–value, and clearly state your conclusions.

-15 for incorrect test

Indicate whether this is a **z–test**, **t–test** or **2–sample z–test**. Circle your answer and explain why you haven chosen that test.

- sample size < 30 ✓
- SD for bone unknown (only sample SD known) ✓ } (1)
- data follow normal curve

Show your work!

-2 for each calculation error
-2 if null, alt swapped

1/ null: drugs have no effect on bone loss (1), i.e., bone avg = 0 (½)
alternative: drugs accelerate bone loss (1), i.e., bone avg < 0 (½)

2/ observed (avg) = -0.24
expected (avg) = 0.0
$SD^t = \sqrt{\frac{12}{11}} \cdot 1.22 = 1.044 \cdot 1.22 = 1.27$ (2)
$SE_{sum} = \sqrt{12} \cdot 1.27 = 3.46 \cdot 1.27 = 4.41$ (2)
$SE_{avg} = \frac{4.41}{12} = 0.368$ (2)
$t = \frac{-0.24 - 0}{0.368} = -0.65$ (2)
$df = 12 - 1 = 11$ (2)

3/ $t = -0.65$ above -0.70
↓
25%
⇝ P-value greater than 25% (2)

4/ • do not reject the null (P-value > 5%) (2)
• drugs have no effect on bone loss (2)