

Statistics 1040, Section 004, Quiz 1 (20 Points)

Monday, January 12, 2004

Your Name: _____

from: Stat 1040, Spring 2000, Final Test, Friday May 5, 2000, Question 1

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

In a recent study on SIDS (Sudden Infant Death Syndrome), one hospital collected data on 128 babies who died from SIDS in the last 12 months. They took a random sample of 500 babies (of similar ages) who did not die from SIDS (the "controls"), and they compared the two groups with respect to several variables of interest (e.g. whether the child slept on his or her stomach, birthweight, time of year, whether the mother smoked, whether she breast-fed, socio-economic status, etc.).

1. (5 Points) Is this a controlled experiment or an observational study? Circle your answer and explain.

Workbook:

"It is an observational study - there was no intervention." (2)

(3)

Note: Dying (not dying from SIDS) is outcome of the study - and also would be outcome of an ^{experiment!}

2. (5 Points) One physician noticed that 63% of the SIDS babies had mothers who smoked during pregnancy, whereas only 26% of the control babies had mothers who smoked during pregnancy. Another physician claimed that low birthweight could be a "confounding factor". Explain what it means for low birthweight to be a "confounding factor". Be specific.

(0)...(5)

Workbook:

"Perhaps smoking causes low birthweight and it is the low birthweight rather than smoking itself that is leading to higher rates of SIDS."

Note: Prematurely born (underweight) babies also have a higher SIDS rate...

3. (4 Points) If you had access to the data, what would you do to "control for" birthweight?

(0)...(4)

Workbook:

"Study babies with similar birthweights separately, e.g., break up the comparison into groups of, say, babies 6-6.5 lb, 6.5-7 lb, 7-7.5 lb, etc."

Please turn over!

from FPP: p. 24, Review Exercise 1

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 (such as United, TWA, and so forth). True or false 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 many people flew on commuter airlines versus scheduled carriers, and then we can calculate the rates and compare."

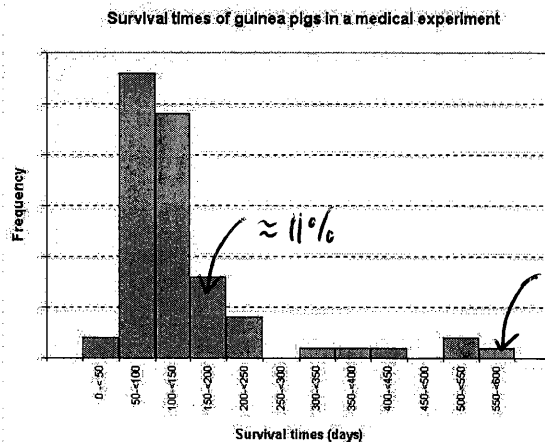
Statistics 1040, Section 004, Quiz 2 (20 Points)

Friday, January 16, 2004

Your Name: _____

Question 1: Histograms I (12 Points)

The histogram below shows the distribution of survival times in days of 72 guinea pigs after they were injected with tubercle bacilli in a medical experiment. A lab assistant is working on a report on the results of the experiment that he has to present in 15 minutes to the Company's Board of Directors. Unfortunately, the lab assistant accidentally deleted the numbers on the vertical axis! There is no way to go back to the original file in such a short time – the only thing he recalls is that 8 guinea pigs out of 72 (i.e., about 11%) survived more than 150 but less than 200 days and only one guinea pig (i.e., about 1.4%) survived at least 550 days. Try to help the lab assistant to fill in the missing percentages below so he can still give a decent presentation.



-2 if outside $x\% \pm 5\%$
 -4 if order not maintained
 -4 if sum of % $\gg 100\%$
 (e.g. 110% etc.)

1. (4 Points) What approximate percentage of guinea pigs survived less than 100 days?

0-50 & 50-100: together, about 4 times as high as 150-200:

$$\leadsto 4 \cdot 11\% \approx 44\%$$

2. (4 Points) What approximate percentage of guinea pigs survived more than 100 but less than 150 days?

100-150: about 3 times as high as 150-200:

$$\leadsto 3 \cdot 11\% \approx 33\%$$

3. (4 Points) What approximate percentage of guinea pigs survived more than 300 days?

300-350: 1.4%
 350-400: 1.4%
 400-450: 1.4%
 450-500: 0%
 500-550: 2.8%
 550-600: 1.4%
 $\approx 8.4\%$

check result: 44% [0-100]
 + 33% [100-150]

Please turn over!

+ 11% [150-200]

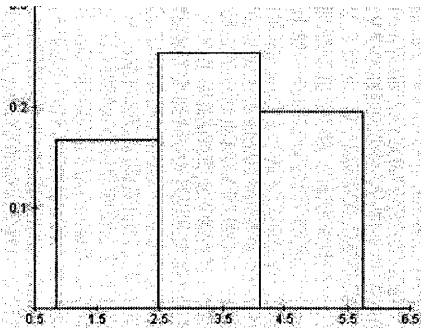
+ 5.5% [200-300]

+ 8.4% [300-600]

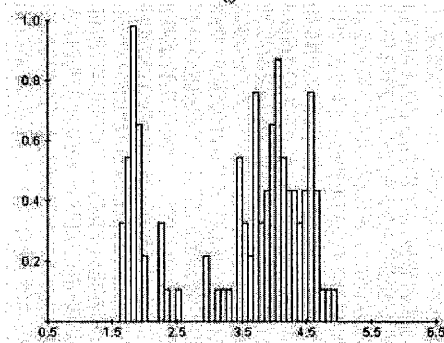
$\approx 101.9\%$ [reasonable approximation]

Question 2: Histograms II (8 Points)

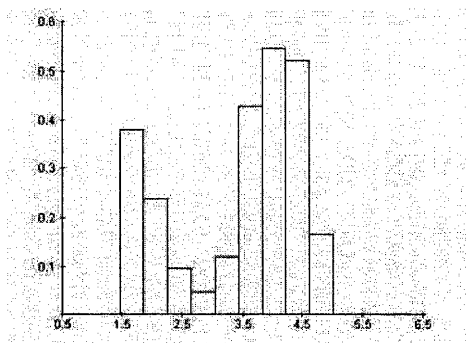
The three histograms below represent the same Old Faithful data set. The observations are the duration (in minutes) for eruptions of the Old Faithful geyser in Yellowstone National Park. Which of the three histograms you think is the best? Explain clearly why you prefer this particular histogram to represent the Old Faithful data and not the other two.



(i)



(ii)



(iii)

The best histogram is: (iii) (4)

Explanation: (i) has too few classes (hides information) (2)
(ii) has too many classes (too spiky) (2)

Note that there seem to be shorter eruptions (centered around 2 min) and longer eruptions (centered around 4 min). This is best visible in (iii).

Statistics 1040, Section 004, Quiz 3 (20 Points)

Friday, January 23, 2004

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.

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

③
②

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

sorted list: -15 -14 -12 -4 5 ③

↑

median = -12°C

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

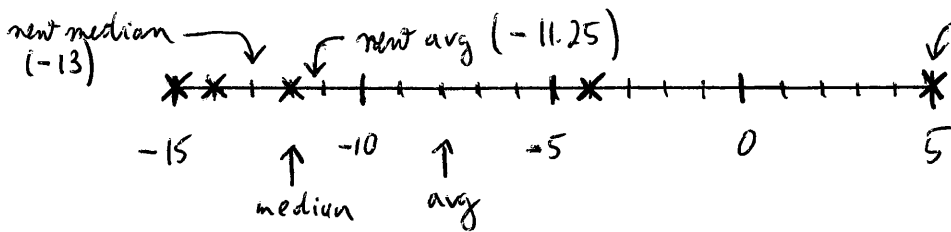
1) avg = -8 ① 2) -15 - (-8) = -7 -14 - (-8) = -6 -12 - (-8) = -4 5 - (-8) = 13 -4 - (-8) = 4 ①	3) (-7) ² = 49 (-6) ² = 36 (-4) ² = 16 13 ² = 169 4 ² = 16 ①	4) $\frac{49 + 36 + 16 + 169 + 16}{5} = \frac{286}{5} = 57.2$ ① 5) SD = $\sqrt{57.2} = 7.56^{\circ}\text{C}$ ② <p style="text-align: right;">Please turn over!</p>
---	--	--

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:

- ② (a) The SD will become bigger;
 (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 never negative (and the SD is 0 only if all numbers are exactly the same - meaning there is no spread).

Formulas:

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

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

Statistics 1040, Section 004, Quiz 4 (20 Points)

Friday, January 30, 2004

Your Name: _____

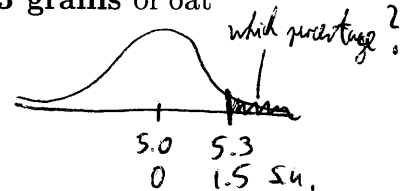
-2 for each calculation error

Question 1: Normal Approximation for Data (20 Points) *+3 for correct graph (and nothing else)*

With Americans becoming more health conscious, many food companies are emphasizing the benefits of oat bran in a wide variety of products — from muffins to potato chips. One such company makes oat bran donuts that contain on average **5 grams** of oat bran. Assuming that the amount of oat bran per donut approximately follows the normal curve, with a standard deviation of **0.2 gram**, answer the questions below:

- (8 Points) The percentage of donuts that contain **more than 5.3 grams** of oat bran is roughly 6.68 %.

① Convert 5.3 into standard units: $\frac{5.3 - 5.0}{0.2} = 1.5 \text{ s.u.}$ (3)

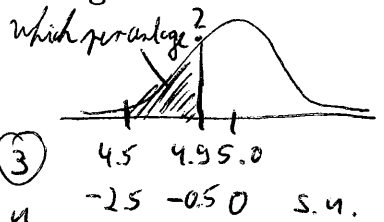


② area between -1.5 and 1.5: 86.64% (2)

③ area above 1.5: $\frac{100\% - 86.64\%}{2} = \frac{13.36\%}{2} = \underline{\underline{6.68\%}}$ (3)

- (12 Points) The percentage of donuts that contain **between 4.5 grams and 4.9 grams** of oat bran is about 30.24 %.

① Convert 4.5 and 4.9 into standard units:
 $\frac{4.5 - 5.0}{0.2} = -2.5 \text{ s.u.}$ (3) $\frac{4.9 - 5.0}{0.2} = -0.5 \text{ s.u.}$ (3)



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

area between -0.5 and 0.5: 38.29% (2)

③ area between -2.5 and -0.5: $\frac{98.76\% - 38.29\%}{2} = \frac{60.47\%}{2} = \underline{\underline{30.24\%}}$ (2)

Show your work!

Statistics 1040, Section 004, Quiz 5 (20 Points)

Grading question 2:

Friday, February 6, 2004

"slightly" wrong answers:

- 1: (d)
- 2: (c), (e)

0: wrong answer, no explanation

1: wrong answer, some explanation

Your Name: _____

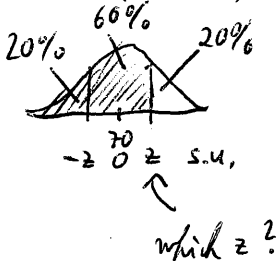
2: "slightly" wrong answer, no explanation

3: "slightly" wrong answer, some explanation OR

Question 1: Percentiles and the Normal Curve (12 Points) correct answer, no explanation

The police department in a large city gives an entrance exam to all applicants for positions in the department. The distribution of scores follows approximately a normal curve with a mean of 70 and a standard deviation of 8. Fill in the blanks and show your work!

1. (8 Points) Suppose that applicants are considered for a job only if their score on the entrance exam is above the 80th percentile. Thus, an entrance score of ≈ 77 (or more) is required for job consideration.



③

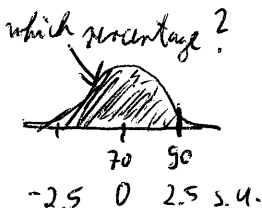
① area between -0.85 and 0.85 : $60.47\% \approx 60\%$

② convert 0.85 s.u. into original units:

$$0.85 \cdot 8 + 70 = 76.8 (\approx 77) \quad \text{③}$$

SD avg

2. (4 Points) An applicant who scored 90 on the entrance exam was at the 99.38th percentile of the score distribution.



① convert 90 into s.u.: $\frac{90-70}{8} = 2.5$ ①

② area between -2.5 and 2.5 : 98.76% ①

③ area below 2.5: $50\% + \frac{98.76\%}{2} = 99.38\% (\approx 99\%)$ ②

Question 2: Correlation (8 Points) from: FPP, Chapter 8, p. 130, #6 (\rightarrow Answers: p. A-56)

(grading: see top of page)

age wife

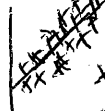


1. (4 Points) If women always marry men who were five years older, the correlation between ages of husbands and wives would be (e) exactly 1. Choose one of the options below, and explain. "All the points on the scatter diagram would be on a line sloping up, so the correlation would be 1."

age husband

2. (4 Points) In reality, the correlation between ages of husbands and wives in the U.S. is (d) close to 1. Choose one of the options below, and explain.

age wife



X ← outlier ...

- Options: (a) exactly -1 (b) close to -1 (c) close to 0
(d) close to 1 (e) exactly 1

age husband

"Close to 1; this is like part (a), with some noise thrown into the data.

Comment: In the March 1993 Current Population Survey, the correlation between the ages of the husbands and wives was about 0.95; the husbands were, on average, 2.7 years older than their wives."

Statistics 1040, Section 004, Quiz 6 (20 Points)

Friday, February 20, 2004

Your Name: _____

data from: FPP, p. 570, Special Review Exercises, Question 13

Question 1: The Regression Line (20 Points)

For women age 25–54 working full time in the U.S. in 1993, the relationship between income and education (years of schooling completed) can be summarized as follows:

- X Average education ≈ 13.7 years, $SD \approx 2.4$ years
 Y Average income $\approx \$24,300$, $SD \approx \$15,900$, $r \approx 0.42$

The scatterplot of the data is football-shaped. **Show your work!**

1. (10 Points) Find the equation of the regression line for predicting income from education for these women.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.42 \cdot \frac{15,900}{2.4} = 2782.5$$

$$\begin{aligned} \text{intercept} &= \text{avg } y - \text{slope} \cdot \text{avg } x = 24,300 - 2782.5 \cdot 13.7 \\ &= 24,300 - 38120.25 = -13,820.25 \end{aligned}$$

regression equation: $\text{income} = -13,820.25 + 2782.5 \cdot \text{education}$

or $y = -13,820.25 + 2782.5 \cdot x$

2. (5 Points) Use the regression equation from part 1. to predict the income of a woman who has 8 years of education. Her predicted income is: \$ 8,439.75 \approx \$8,440

$$\begin{aligned} \text{income} &= -13,820.25 + 2782.5 \cdot 8 = -13,820.25 + 22,260 \\ &= 8439.75 \end{aligned}$$

3. (5 Points) Find the r.m.s. error for predicting income from education for women age 25–54 working full time in the U.S. in 1993. The r.m.s. error is: \$ 14,430

$$\begin{aligned} \text{r.m.s. error} &= \sqrt{1-r^2} \cdot SD_y \\ &= \sqrt{1-0.42^2} \cdot 15,900 \\ &= \sqrt{1-0.1764} \cdot 15,900 \\ &= \sqrt{0.8236} \cdot 15,900 \end{aligned}$$

$$= 0.9075 \cdot 15,900$$

$$= 14,429.25$$

$$\approx 14,430$$

Please turn over!

(i.e., on average the predicted income will be off about \$14,430 from the actual income)

Formulas:

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

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

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

Grading Criteria:

- in 1., 2.1 & 3.1:
- 2 for each calculation error
 - 2 for each incorrect value used
 - 2 for swapping x and y
- in 1.1:
- 3 for incorrect formula for slope
 - 3 for incorrect formula for intercept
 - 2 if no final equation stated
 - 1 if only part of the equation stated (e.g., $-13,820.25 + 2782.5x$)
 - 1 if not specifying x & y (but using x & y in equation)
- in 2.1:
- 3 for incorrect formula for prediction
 - 1 if correct result, but according to old method
- in 3.1:
- 3 for incorrect formula for r.m.s. error

Statistics 1040, Section 004, Quiz 7 (20 Points)

Friday, February 27, 2004

Your Name: _____

from: Stat 1040, Fall 2001, Final Test, Thursday, December 13, 2001, Question 4

Question 1: Chance/Probability I (13 Points)

An elementary school in Logan employs 15 teachers: 11 are women and 4 are men. Two teachers are selected at random to meet the governor and attend a reception in SLC.

1. (4 Points) What is the chance that both are women?

The chance is: 52.4 %

-1 each calculation error
(or no final result)

first woman: $\frac{11}{15}$ $\left\{ \begin{array}{l} \text{dependent} \\ \text{multiplication rule} \end{array} \right.$
 second woman, given first woman: $\frac{10}{14}$
 both women: $\frac{11}{15} \cdot \frac{10}{14} = \frac{110}{210} = 0.524 = \underline{\underline{52.4\%}}$

-4 if % > 100% or % < 0%

2. (4 Points) What is the chance that at least one is a woman?

The chance is: 94.3 %

both men: $\frac{4}{15} \cdot \frac{3}{14} = \frac{12}{210}$ ②

opposite:

at least one woman: $1 - \frac{12}{210} = \frac{210-12}{210} = \frac{198}{210} = 0.943 = \underline{\underline{94.3\%}}$
 ↖ opposite rule

3. (5 Points) What is the chance that both are the same gender?

The chance is: 58.1 %

both women: $\frac{110}{210}$ ① (from 1.)
 both men: $\frac{12}{210}$ ① (from 2.) $\left\{ \begin{array}{l} \text{mutually exclusive} \end{array} \right.$

same gender (i.e., both women or both men):

Please turn over!

$\frac{110}{210} + \frac{12}{210} = \frac{122}{210} = 0.581 = \underline{\underline{58.1\%}}$

↖ addition rule

from: FPP, Chapter 13, p. 225-226, #1 (→ Answers: p. A-66)

Question 2: Chance/Probability II (7 Points)

A computer is programmed to compute various chances. Match the numerical answers with one of the following verbal descriptions (which can be used more than once):

Verbal description

- (i) This is as likely to happen as not.
- (ii) This is very likely to happen, but it's not certain.
- (iii) This won't happen.
- (iv) This may happen, but it's not likely.
- (v) This will happen, for sure.
- (vi) There is a bug in the program.

Numerical answer is matched by:

- (a) -50% (vi) ① each correct answer
- (b) 0% (iii)
- (c) 10% (iv)
- (d) 50% (i)
- (e) 90% (ii)
- (f) 100% (v)
- (g) 200% (vi)

Statistics 1040, Section 004, Quiz 8 (20 Points)

Friday, March 5, 2004

Your Name: _____

based on: FPP, p. 286, Review Exercise 7

Question 1: Box Models, EV, and SE (14 Points)

A quiz has 20 multiple choice questions. Each question has 5 possible answers: only one answer is completely correct, two are partially correct, and two are completely incorrect. The correct answer is worth 5 points, a partially correct answer is worth one point, and a point is taken off for a completely incorrect answer.

A student (who did not study at all) answers all the questions by guessing at random.

1. (4 Points) Find the box model.



Number of draws: 20

2. (5 Points) Find the expected value, i.e., the number of points a student would get when answering all questions by guessing. $EV_{sum} = \underline{20}$ (points)

$$\text{box average} = \frac{5 + 1 + 1 + (-1) + (-1)}{5} = \frac{5}{5} = 1$$

$$EV_{sum} = 20 \cdot 1 = 20 //$$

3. (5 Points) Find the standard error. $SE_{sum} = \underline{9.8}$ (points) (e.g., step missing)

$$\begin{aligned} \text{box SD} &= \sqrt{\frac{(5-1)^2 + 2 \cdot (1-1)^2 + 2 \cdot (-1-1)^2}{5}} \\ &= \sqrt{\frac{16 + 2 \cdot 0 + 2 \cdot 4}{5}} = \sqrt{\frac{24}{5}} = \sqrt{4.8} = 2.19 \end{aligned}$$

$$SE_{sum} = \sqrt{20} \cdot 2.19 = 4.47 \cdot 2.19 = 9.79 \approx \underline{9.8} //$$

Please turn over!

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

Question 2: Law of Averages (6 Points)

A box contains red and green marbles; there are more green marbles than red 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 one of the four options below. **Briefly explain your answer.**

- ④
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.

② option 1. is best. Say the percentage of reds in the box is 40%. Then, to win, we want the percentage error to be big, so that the actual percentage of red marbles will be greater than 50%. This is more likely in the short run.

Have a good Spring Break!

Formulas:

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

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

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

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

Statistics 1040, Section 004, Quiz 9 (20 Points)

Friday, March 19, 2004 *part 1,)*

Your Name: _____

- 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

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

According to the U.S. Census Bureau's "QuickFacts" Web site (<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.

$$\boxed{26 \times [1] \quad 74 \times [0]}$$

draws: 500

1: bachelor degree or higher
0: no such degree

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?

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

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

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

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

- 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%

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

- 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

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 or false: The expected value for the number of heads is 50.

#: 1
T: 0
box:

0	1
---	---

draws: 100

$\text{box avg} = \frac{1}{2}$ $\text{box SD} = \sqrt{\frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{2}$ $EV_{\text{sum}} = 100 \cdot \frac{1}{2} = 50$ $SE_{\text{sum}} = \sqrt{100} \cdot \frac{1}{2} = 5$	$\Rightarrow EV_{\text{sum}} = 50? \text{ yes (see left)}$ $\Rightarrow \text{True}$
---	--

2. (1 Point) True or false: 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
 \Rightarrow False

3. (2 Points) True or false: 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
 \Rightarrow False

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

As calculated in 1,

$$EV_{\text{sum}} = 50$$

and

$$SE_{\text{sum}} = 5 = \text{"give or take" part } (\neq 3).$$

\Rightarrow False (overall since the SE_{sum} is incorrect)

Please turn over!

Statistics 1040, Section 004, Quiz 10 (20 Points)

Friday, April 2, 2004

Your Name: _____

Question 1: Confidence Intervals (20 Points)

There are approximately 20,000 students at USU. For a simple random sample of 200 USU students, we learn that 140 students are satisfied with President Kermit Hall.

1. (14 Points) If possible, construct a 95% confidence interval for the percentage of all USU students who are satisfied with President Hall. If you cannot do this, explain why not.

box unknown: $\frac{2 \times \boxed{1} \quad 2 \times \boxed{0}}{\# \text{ draws} = 200}$ -2 each calculation error

Show your work.

sample % = $\frac{140}{200} = 0.7 = 70\%$ (2) *population % (assumption) -1 if % forgotten*

SD box = $\sqrt{0.7 \cdot 0.3} = \sqrt{0.21} = 0.46$ (2) *(via book/straps) -1 if box indicated (box = population is unknown!)*

SE_{sum} = $\sqrt{200} \cdot 0.46 = 14.14 \cdot 0.46 = 6.5$ (2)

SE% = $\frac{6.5}{200} \cdot 100\% = 3.25\%$ (2)

95% CI: $70\% \pm 2 \cdot 3.25\% = 70\% \pm 6.5\%$

sample % ↑ related to 95% ↑ SE%

$= 63.5\% \text{ to } 76.5\%$ (4)

2. (6 Points) (Hypothetical) Suppose that in the sample of 200 students, only 1 student is satisfied with President Hall. Would it still be possible to construct a 95% confidence interval for the percentage of all USU students who are satisfied with President Hall? Yes or No? (2) And explain why or why not.

You do not have to actually construct this confidence interval, but you do have to show calculations necessary to support your answer.

sample % = $\frac{1}{200} = 0.005 = 0.5\%$ ($= \frac{1}{2}\%$) (2)

$\frac{1}{2}\%$ is too close to 0%. We cannot construct a 95% CI here. (2)

Please turn over!

Statistics 1040, Section 004, Quiz 11 (20 Points)

Friday, April 9, 2004

Your Name: _____

from: FPP, p. 426, Review Exercise 3

Question 1: Confidence Intervals for Averages (14 Points)

A real estate office wants to make a survey in a certain town, which has 50,000 households, to determine how far the head of household has to commute to work. A simple random sample of 1,000 households is chosen, the occupants are interviewed, and it is found that on average, the heads of the sample households commuted 8.7 miles to work; the SD of the distances was 9.0 miles. (All distances are one-way; if someone isn't working, or is working at home, the commute distance is defined to be 0.)

-2 for each calculation error

Fill the blanks in the statements below and **show your work!**

-6 if working with SE% at SE_{sum} in final stage

1. (7 Points) The average commute distance of all 50,000 heads of households in the town is estimated as 8.7 miles, and this estimate is likely to be off by 0.28 miles or so.

Assume: sample avg = 8.7 miles = low avg (2)

sample SD = 9.0 miles = low SD (1)

$$SE_{sum} = \sqrt{1,000} \cdot 9.0 = 31.6 \cdot 9.0 = 284.4 \text{ miles} \quad (2)$$

$$SE_{avg} = \frac{284.4}{1,000} = 0.28 \text{ miles} \quad (2)$$

[Note that the data itself do not follow the normal curve; just 1 SD below the avg are the negative numbers - and negative driving distances are impossible; but the average of 1,000 draws will follow the normal curve!]

2. (7 Points) If possible, find a 85%-confidence interval for the average commute distance of all heads of households in the town. If this isn't possible, explain why not.

-5 for "it is not possible"

It is possible to find a 85% - CI:

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

$$= 8.7 \pm 1.45 \cdot 0.28 \quad (2)$$

↑ related to 85%

$$= 8.7 \pm 0.406$$

$$\approx \underline{8.3 \text{ miles}} \text{ to } \underline{9.1 \text{ miles}} \quad (2)$$

Please turn over!

From: FPP, Chapter 26, Exercise Sets

Question 2: Tests of Significance (6 Points)

Circle the correct answer. Set B, p. 479, question 1.

1. (2 Points) In order to test a null hypothesis, you need

- (a) data
- (b) a box model for the data
- (c) both of the above
- (d) none of the above.

2

Set D, p. 484, question 1(a).

2. (2 Points) A "highly significant" result cannot possibly be due to chance.

True or False.

Even if the null hypothesis is true, 1% of the time the experiment will give a result which is "highly significant".

2

Set C, p. 483, question 3(b).

3. (2 Points) If the observed significance level is 5%, there are 95 chances in 100 for the alternative hypothesis to be right.

True or False.

see FPP, p. 482, for more details

2

Please turn over!

Quiz 12 - Solutions

Question 1:

-15 for incorrect test

- 1) null: nose length as usual, i.e., $\mu = 44$ mm (1)
alternative: longest noses at PSU, i.e., $\mu > 44$ mm (1)

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

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

$$\text{expected } \underline{\text{avg}} = 44 \text{ mm}$$

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

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

$$\text{SD}_+ = 6.82 \cdot \sqrt{\frac{18}{17}} = 7.02 \text{ mm} \quad (2)$$

$$\text{SE}_{\text{sum}} = \sqrt{18} \cdot 7.02 = 29.78 \text{ mm} \quad (2)$$

$$\text{SE}_{\text{avg}} = \frac{29.78}{18} = 1.65 \text{ mm} \quad (1)$$

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

3) P-value:

$$t = 0.47 \text{ is left of } 0.69$$

↓
25%

$$\leadsto \text{P-value is } > 25\% \quad (2)$$

4) Conclusion:

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

Question 2:

-8 if any test conducted

We cannot conduct a 2-sample z-test (or any other test) here:

we have the whole populations, i.e., all student cars and all faculty cars registered with Parking Services - student cars are older on average!

Question 3:

-8 for incorrect test

1) null: "A" students are evenly distributed, i.e., $\frac{1}{3}$ in front, $\frac{1}{3}$ in middle, $\frac{1}{3}$ in back (1)

alternative: "A" students are not evenly distributed, i.e., front: middle: back $\neq \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$ (1)

2) χ^2 -test:

Location	obs	exp	$\frac{(obs - exp)^2}{exp}$
Front	19	$\frac{1}{3} \cdot 33 = 11$	$\frac{(19-11)^2}{11} = 5.82$
Middle	9	$\frac{1}{3} \cdot 33 = 11$	$\frac{(9-11)^2}{11} = 0.36$
Back	5	$\frac{1}{3} \cdot 33 = 11$	$\frac{(5-11)^2}{11} = 3.27$
Total	33	33	9.45

(2)

$$\chi^2 = 5.82 + 0.36 + 3.27 = 9.45$$

(1)

$$df = 3 - 1 = 2$$

(1)

3) P-value:

$$\chi^2 = 9.45 \text{ is right of } 9.21$$

↓
1%

\leadsto P-value is $< 1\%$

(2)

4) Conclusion:

- reject the null (P-value $< 5\%$) (1)

- result is highly statistically significant (P-value even $< 1\%$) (1/2)

- there is strong evidence that "A" students are not evenly distributed in the classroom. (1/2)