Statistics 1040, Section 008, Quiz 1 (20 Points)

Friday, January 11, 2008

Your Name: _____

from: FPP, Chapter 2, Review Exercise 4 & Quiz 1, Spring 2005 & Spring 2006, Question 1

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

(Solutions: → Course Web Page)

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?

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.

That is not (3) 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 & Spring 2006, Question 2

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

(Solutions: → Course Web Page)

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 008, Quiz 2 (20 Points)

Friday, January 18, 2008

Your Name: _____

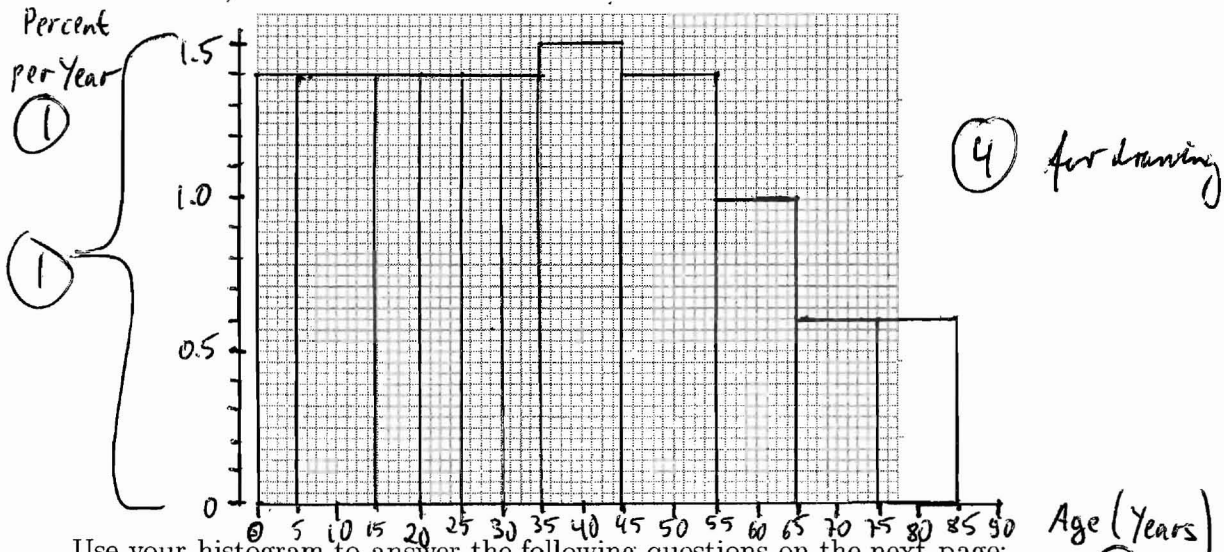
from: FPP, Chapter 3, Review Exercise 2 (Solutions: → Worklook)

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

The age distribution of people in the U.S. in 2004 is shown below. (2) (2)

Age	Percent of population	Width	Height
0-5	7	5	7/5 = 1.4
5-15	14	10	14/10 = 1.4
15-20	7	5	7/5 = 1.4
20-25	7	5	7/5 = 1.4
25-30	7	5	7/5 = 1.4
30-35	7	5	7/5 = 1.4
35-45	15	10	15/10 = 1.5
45-55	14	10	14/10 = 1.4
55-65	10	10	10/10 = 1.0
65-75 ~~85~~	6	10	6/10 = 0.6
75 and over	6	10	6/10 = 0.6

1. (**12 Points**) Draw a histogram for these data on the graph paper provided. (The class intervals include the left endpoint, not the right; for instance, on the second line of the table, 14% of the people were age 5 years or more but had not yet turned 15. The interval "75 and over" can be ended at 85. Men and women are combined in the data.) Make sure to label the axes.



(1) (1) (4) for drawing (1) (1)

Use your histogram to answer the following questions on the next page:

Please turn over!

2. (**2 Points**) Are there more children age 1 or elders age 71?
Circle your answer.

→ age 1: 1.4%

age 71: 0.6%

(2)

3. (**2 Points**) Are there more 21–year–olds or 61–year–olds?
Circle your answer.

→ age 21: 1.4%

age 61: 1.0%

(2)

4. (**2 Points**) Are there more people age 0–4 or 65–69?
Circle your answer.

→ age 0-4: $4 \cdot 1.4\% = 5.6\%$

age 65-69: $4 \cdot 0.6\% = 2.4\%$

(2)

5. (**2 Points**) The percentage of people age 35 and over is around 25% 50% or 75%?
Circle your answer.

(2)

sum up percentages up to age 35:

$7\% + 14\% + 7\% + 7\% + 7\% + 7\% = 49\%$

therefore, "age 35 and over" $= 100\% - 49\% = 51\%$ (closest to 50%)

[Note: explanations weren't required!]

Statistics 1040, Section 008, Quiz 3 (20 Points)

Friday, January 25, 2008

Your Name: _____

from: Quiz 3, Spring 2006, Question 1 (Solutions → Course Web Page)

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!

from: Quiz 3, Spring 2006, Question 2 (Solutions → Course Web Page)

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

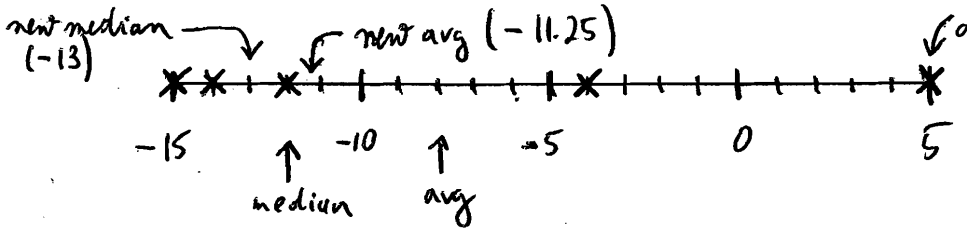
To answer the questions below, you need to apply your knowledge about average, median, and standard deviation. **No calculation is needed!**

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 never negative (and the SD is 0 only if all numbers are exactly the same – meaning there is no spread).

Formulas:

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

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

Statistics 1040, Section 008, Quiz 4 (20 Points)

Friday, ~~January 31,~~ 2008
February 1,

Your Name: _____

from: Midterm 1, Spring 2005, Question 1 (Solutions: → Course Web Page)

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 146,000 non–US citizens who took the General GRE Test in 2001–02, the mean for the quantitative ability portion of the exam was 700 and the standard deviation was 120. We can assume that the histogram follows a normal curve. **Show your work!**

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

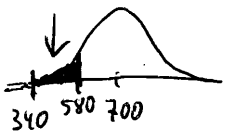
- (**7 Points**) The percentage of non–US citizens who scored **more than 670** on the GRE test is roughly 59.87 %.

① s.u.: $\frac{670-700}{120} = -0.25$ s.u. (3)

② area from −0.25 to 0.25: 19.74% (2)

③ area above −0.25: $\frac{19.74\%}{2} + 50\% = 9.87\% + 50\% = 59.87\%$ (2)

- (**7 Points**) The percentage of non–US citizens who scored **between 340 and 580** is about 15.73 %.

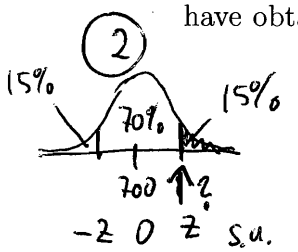


① s.u.: $\frac{340-700}{120} = -3.00$ s.u. (1) | s.u.: $\frac{580-700}{120} = -1.00$ s.u. (1)

② area from −3.00 to 3.00: 99.73% (1) | area from −1.00 to 1.00: 68.27% (1)

③ area from −3.00 to −1.00: $\frac{99.73\% - 68.27\%}{2} = \frac{31.46\%}{2} = 15.73\%$ (3)

- (**6 Points**) In order to be among the top 15% of all non–US citizens, a student must have obtained a minimum GRE score of about 826.



① area from −1.05 to 1.05: 70.63% (closest to 70%) (2)

② original units:

$1.05 \cdot 120 + 700 = 126 + 700 = 826$ (2)

Please turn over!

Statistics 1040, Section 008, Quiz 5 (20 Points)

Friday, February 8, 2008

Your Name: _____

from: Quiz 5, Fall 2007, Question 1 (Solutions → Course Web Page)

Question 1: Measurement Error (**7 Points**)

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

35.96 inches 36.01 inches 36.03 inches

If you send the yardstick back for a fourth calibration, you would expect to get 36 inches, give or take (4)

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

Circle your answer and explain.

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

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

Question 2: Correlation (**7 Points**) from: Quiz 5, Fall 2007, Question 2 (Solutions → Course Web Page)

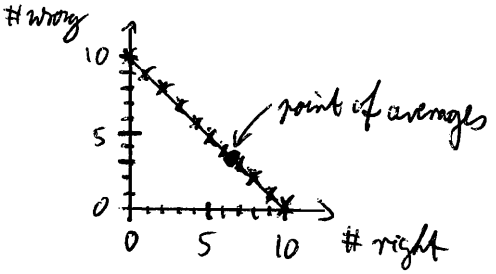
A teaching assistant gives a quiz to his section. There are 10 questions on the quiz and no part credit is given. After grading the papers, the TA writes down for each student the number of questions the student got right and the number wrong. The average number of right answers is 6.4 with an SD of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

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

(4)

(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.



Number wrong = 10 − Number right!

So, all the points on a scatter diagram (for # wrong vs. # right) lie on a straight line which slopes down. (3)

Please turn over!

from: Quiz 5, Spring 2006, Question 1 (Solutions: → Course Web Page)

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

In a class experiment last week, we measured the length of a pencil (including the eraser) 13 times. The average length of our 13 measurements was 7.5 inches, with an SD of 0.07 inches. Recall that 1 inch = 2.54 cm.

If we translate these results into cm, the average length will be 19.05 (2) cm, with a standard deviation of 0.1778 (2) 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): $0.07 \cdot 2.54 =$ (1) 0.1778 cm

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 008, Quiz 6 (20 Points)

Friday, February 22, 2008

Your Name: _____

from: Midterm 1, Fall 2007, Question 4 (Solutions → Web Page)

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

For 167 college students, the relationship between height and handspan size is summarized as follows:

x height:	average = 68.0 inches	SD = 4.0 inches	(point of averages; about 95% of data within 68 ± 2·4.0 = 60 to 76)
y handspan size:	average = 20.9 inches	SD = 1.9 inches	(about 95% of data within 20.9 ± 2·1.9 = 17.1 to 24.7)
	r = 0.75		

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

1. (4 Points) Six scatter diagrams are printed on the next page. Which of the scatter diagrams is the correct one for these data? **Circle the correct letter below (No explanation is needed for this part!)**:

   A (B) C D E F (4)

   [see figure for explanation – but not needed]

2. (6 Points) Using the summary statistics above, what is the regression estimate for handspan for a student who is 60 inches tall?

   The answer is: 18.05 inches

   (-1 for each calculation error in 2. & 3.)

   $su_x = \frac{60-68}{4} = -2$ (2)

   $su_y = -2 \cdot 0.75 = -1.5$ (2)

   $y = -1.5 \cdot 1.9 + 20.9 = 18.05$ inches (2)

3. (6 Points) Find the r.m.s. error for your answer in the previous part.

   The answer is: 1.26 inches

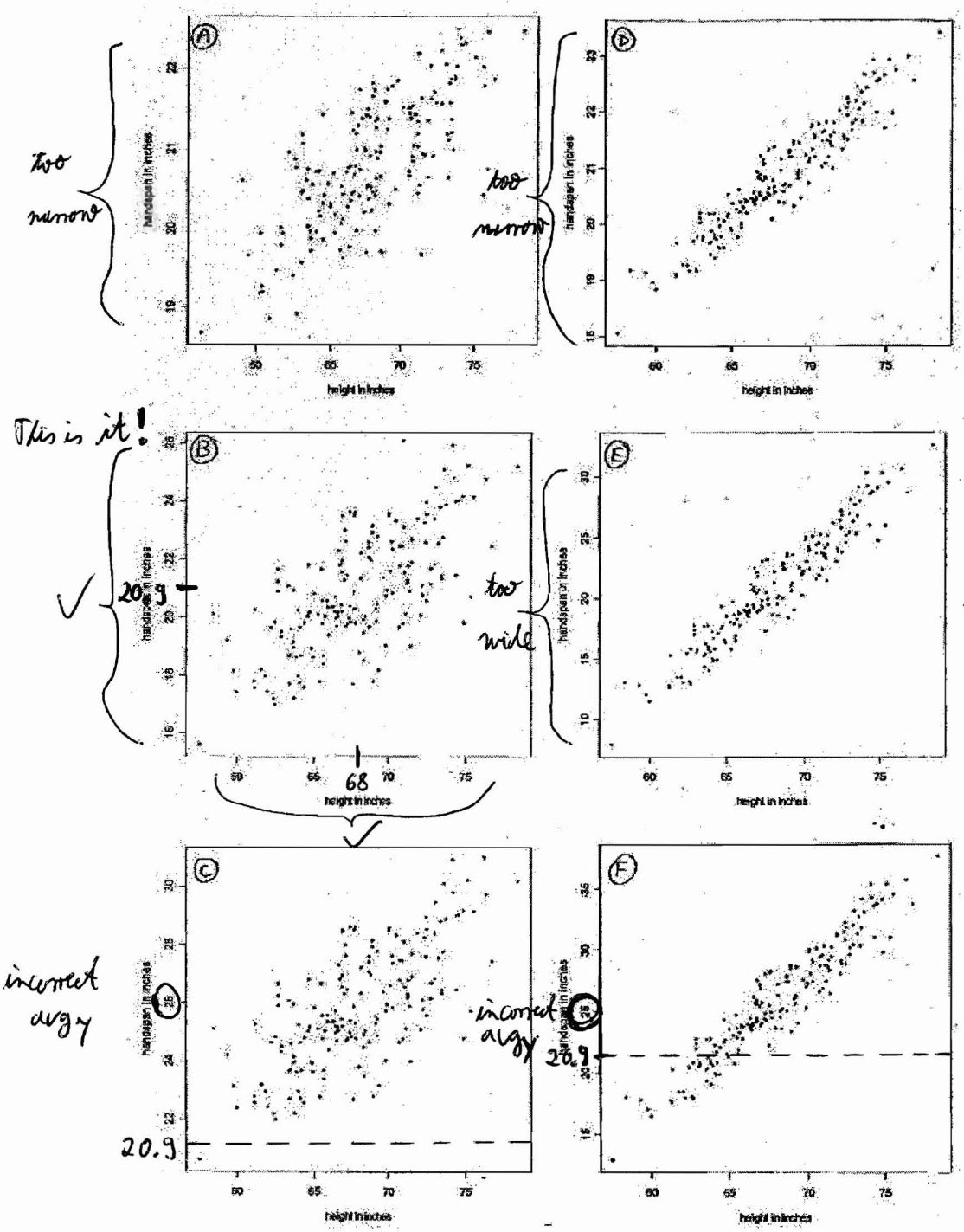
   (-2 for each mistake (e.g., no square or incorrect SD))

   r.m.s. error $= \sqrt{1-0.75^2} \cdot 1.9 = 1.26$ inches (6)

4. (4 Points) What would the correlation coefficient be if we changed all the handspan measurements to centimeters? (There are 2.54 centimeters in an inch).

   The answer is: $r = 0.75$ (it won't change!) (4)

Exclude impossible plots.



This is it!

Formulas:

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

Statistics 1040, Section 008, Quiz 7 (20 Points)

Friday, February 29, 2008

Your Name: _____

from: Quiz 7, Spring 2005, Question 1 (Solutions: → Course Web Page)

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 for each calculation error (or no final result in %)

1. (**5 Points**) What is the chance that the first is mint or orange?

   The chance is 53.3 %.

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

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

   first orange: $\frac{3}{15}$ → mutually exclusive

   addition rule (for mutually exclusive events)

   first mint or orange: $\frac{5}{15}_{①} + \frac{3}{15}_{②②} = \frac{8}{15} = 0.533 = 53.3\%$

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

   multiplication rule

   both orange: $\frac{3}{15}_{①} \cdot \frac{2}{14}_{②②} = \frac{6}{210} = 0.0286 = 2.86\%$

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

   The chance is 7.14 %.

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

   second caramel, given first orange: $\frac{5}{14}$ → dependent

   multiplication rule

   first orange and second caramel: $\frac{3}{15} \cdot \frac{5}{14} = \frac{15}{210} = 0.0714 = 7.14\%$

   ① ② ②

Please turn over!

from: Quiz 7, Spring 2005, Question 2 & FPP, Chapter 13, Review Exercise 7, p. 235

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

(Solutions: → Course Web Page & Workbook)

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)

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 008, Quiz 8 (20 Points)

Friday, March 7, 2008

Your Name: _____

from: Quiz 7, Fall 2003 & Quiz 8, Fall 2004, Quiz 8, Spring 2006 (Solutions: → Course Webpage)

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.

   $3 \times [3] \quad 1 \times [0] \quad 6 \times [-1]$

   # 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.

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

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

   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.

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

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

Please turn over!

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

Question 2: Law of Averages (**8 Points**) & Quiz 8, Spring 2006 [→ Course Web Page]

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:

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

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

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

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

Statistics 1040, Section 008, Quiz 9 (20 Points)

Friday, March 21, 2008

part 1.1 -1 if slightly incorrect number of 0/1's in box

Your Name: _____ -2 if box given as 10 1

from: Quiz 9, Spring 2005, Question 1

(Solutions: → Course Web Page)

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

-3 if box contains something other than 0/1's

Suppose it is known that 10% of all people in Utah have a specific blood type. Suppose I take a random sample of 500 Utah residents ... **Show your work!** -1 if # draws missing

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

1: specific blood type (we are interested in)

0: all other blood types

1 × 1 9 × 0 or 10 × 1 90 × 0

draws: 500 # draws: 500

2. (**6 Points**) The expected number of Utah residents in this sample of 500 who have that specific blood type is 50 with an SE of 6.7.

-1 for each calculation error

-1 for each minor mistake

-2 for each major mistake or step missing

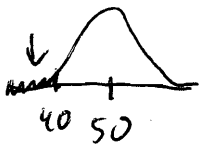
box avg: fraction of 1's $= \frac{1}{10} = 0.1$

box SD $= \sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{\frac{1}{10} \cdot \frac{9}{10}} = \frac{3}{10} = 0.3$

$EV_{sum} = 500 \cdot 0.1 = 50$

$SE_{sum} = \sqrt{500} \cdot 0.3 = 6.7$

3. (**6 Points**) The chance that **fewer than** 40 Utah residents in this sample have that blood type is about 6.68 %.



s.u.: $\frac{40-50}{6.7} = -1.49 \approx -1.5$

area between -1.5 and 1.5: 86.64%

area below -1.5: $\frac{100\% - 86.64\%}{2} = 6.68\%$

-1 for each calculation error

-2 for incorrect curve parameters, i.e., anything else than EV and SE

-2 for incorrect s.u.

-1 for incorrect table value

-1 for incorrect area under the curve

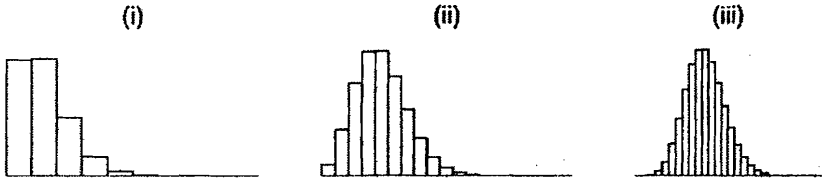
Please turn over!

from: Quiz 9, Spring 2005, Question 2

(Solutions: → Course Web Page)

Question 2: Probability Histograms (**4 Points**)

Shown below are probability histograms for the sum of (a) 100, (b) 400, and (c) 900 draws from the box $\boxed{99 \times \boxed{0} \quad 1 \times \boxed{1}}$. Which histogram is which? **Explain briefly.**



(i) goes with sum (a) 100 (1)

(ii) goes with sum (b) 400 (1)

(iii) goes with sum (c) 900 (1)

Explanation: The histograms get closer to the normal curve as the number of draws goes up. (1)

Formulas:

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

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

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

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

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 008, Quiz 10 (20 Points)

Wednesday, April 9, 2008

Your Name: _____

from: Quiz 10, Fall 2004 & Quiz 10, Spring 2006 (Solutions -> Course Web Page)

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 |

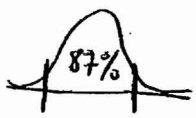
draws: 787

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 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

    $$\text{sample \%} = \frac{780}{787} = 0.991 = 99.1\%$$ (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 008, Quiz 11 (20 Points)

Wednesday, April 16, 2008

Your Name: _____

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

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

Many companies are experimenting with "flex–time", which is supposed to reduce absenteeism. One company employees have averaged 6.3 days off work in the past. The company introduces "flex–time" and a year later a simple random sample of 100 employees is selected. They average 5.5 days off work with a standard deviation of 2.9. Test to determine if "flex–time" reduces absenteeism. Clearly state the null and alternative hypotheses, calculate the appropriate test statistic, find the P–value, and state your conclusion.

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

Show your work!

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

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

2. (**3 Points**) State the null and the alternative hypotheses for this problem, in words and in terms of the box model.

   (1)
   Null: flex-time has no effect on absenteeism, i.e., box avg = 6.3
   (1) (1)
   Alternative: flex-time reduces absenteeism, i.e., box avg < 6.3

Please turn over!

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

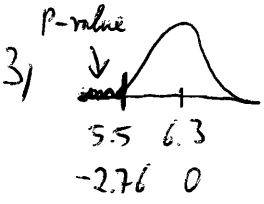
observed avg: 5.5
expected avg: 6.3

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

$SE_{avg} = \frac{29}{100} = 0.29$ (2)

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

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

3| 

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

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

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

4| • reject the null (P-value < 5%) (2)

• result is highly statistically significant (P-value < 1%) (2)

• flex-time reduces absenteeism (2)

Please turn over!

Statistics 1040, Section 008, Quiz 12 (20 Points)

Wednesday, April 23, 2008

Your Name: _____

from: Stat 1040, Final, Fall 2007, Question 10 (Solutions → Course Web Page)

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

In a randomized, controlled, double–blind study published in *The Journal of the American Medical Association* in October 2007, researchers followed 371 heavy drinkers for 14 weeks to try to determine whether the migraine drug Topamax could help them to quit drinking. By the end of the study, 27 of the 183 people in the Topamax group had quit drinking completely, while only 6 of the 188 people in the placebo group had quit drinking completely. Is this evidence that Topamax helps, or could the result just be due to chance error? Clearly state the null and alternative hypotheses, calculate the appropriate test statistic, find the P–value, and state your conclusion.

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

Show your work!

2-sample z-test:

T: Topamax group

C: Control group (placebo)

1/ null: T and C help quit drinking at the same rate, (1)

i.e., $\text{boze}_{T\%} - \text{boze}_{C\%} = 0\%$ (1/2)

alternative: T helps quit drinking at a higher rate, (1)

i.e., $\text{boze}_{T\%} - \text{boze}_{C\%} > 0\%$ (1/2)

2/

T	C
sample size T = 183	sample size C = 188
sample T% = $\frac{27}{183} = 14.75\%$ (1)	sample C% = $\frac{6}{188} = 3.19\%$ (1)
$SD_T = \sqrt{0.1475 \cdot 0.8525} = 0.355$ (1)	$SD_C = \sqrt{0.0319 \cdot 0.9681} = 0.176$ (1)
$SE_{sum_T} = \sqrt{183} \cdot 0.355 = 4.80$ (1)	$SE_{sum_C} = \sqrt{188} \cdot 0.176 = 2.41$ (1)
$SE_{\%_T} = \frac{4.80}{183} \cdot 100\% = 2.62\%$ (1)	$SE_{\%_C} = \frac{2.41}{188} \cdot 100\% = 1.28\%$ (1)

$SE_{diff\%} = \sqrt{(2.62\%)^2 + (1.28\%)^2} = 2.92\%$ (1 1/2)

$z = \frac{14.75\% - 3.19\%}{2.92\%} = 3.96$ (1 1/2)

3/ area between −3.95 and 3.95: 99.992% (1 1/2)

P-value = area above 3.95 = $\frac{100\% - 99.992\%}{2} = 0.004\%$ (1 1/2)

Please turn over!

4/ • reject the null (1)

• result is highly stat. significant (1)

• T helps quit drinking at a higher rate (1)