Your Name: _____

Instructions: Carefully check whether you have to provide an explanation or not. In case you have to provide an explanation, keep it short. Just 1 sentence (or 2 sentences at most) or a short calculation will be fine. If you do not have to provide an explanation, do not waste your time giving an unneeded explanation.

from: Stat 1040, Quiz 8, Fall 2005, Question 1   *(Solutions → Web)*

Question 1: EV, SE, and Normal Curve (**40 Points**)

In a certain town, there are 40,000 registered voters, of whom 15,000 are Democrats. A survey organization is about to take a simple random sample of 1,000 registered voters. **Show your work!**

1. (**8 Points**) Find the box model. $1 = $ *Democrat* $0 = $ *other*

$$\boxed{15,000 \times \boxed{1} \quad 25,000 \times \boxed{0}} \quad \text{⑤}$$
$$\text{# draws} = 1,000 \quad \text{③}$$

−2 if slightly incorrect number of $\boxed{0}/\boxed{1}$ *'s in box*

−5 if box given as $\boxed{\boxed{0}\boxed{1}}$ *etc.*

−5 if box contains something else than $\boxed{0}/\boxed{1}$ *'s*

−3 if # draws missing or incorrect

2. (**20 Points**) The expected number of Democrats in this sample of 1,000 is ___375___ with an SE of ___15.3___ .

$$\text{box avg} = \frac{15,000}{40,000} = 0.375 \quad \text{⑤}$$
$$\text{box SD} = \sqrt{\frac{15,000}{40,000} \cdot \frac{25,000}{40,000}} = \sqrt{0.375 \cdot 0.625} = \sqrt{0.234} = 0.484 \quad \text{⑤}$$
$$EV_{sum} = 1,000 \cdot 0.375 = \underline{\underline{375}} \quad \text{⑤}$$
$$SE_{sum} = \sqrt{1,000} \cdot 0.484 = 31.6 \cdot 0.484 = \underline{\underline{15.3}} \quad \text{⑤}$$
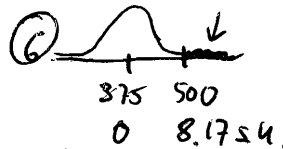
−2 for each calculation error

−2 for each minor mistake

−5 for each major mistake (or step missing)

3. (**12 Points**) The chance that **at least** 500 of the voters in the sample are Democrats is about ___0___ %.

$$S.u.: \quad \frac{500 - 375}{15.3} = \frac{125}{15.3} = 8.17 \quad \text{⑥}$$

375 500
0   8.17 s.u.

area between −4.45 and 4.45: 99.9991%   ③
area between −8.17 and 8.17: almost 100%
area above 8.17: about ___0%___   ③

−2 for each calculation error

−2 for each incorrect curve parameter

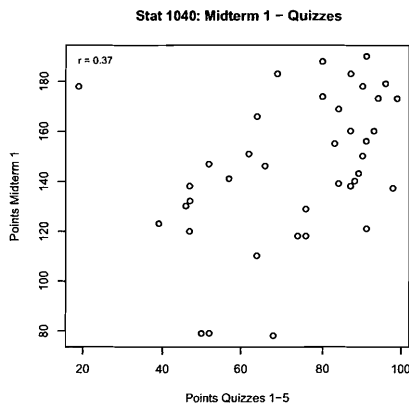It is extremely unlikely that we end up with a sample that contains at least 500 Democrats.

New Question!

Question 2: Regression (50 Points)

In a recent section of Stat 1040, the following scores for the sum of the first five quizzes and the first midterm were observed:

X   Quiz 1-5 score: avg = 73 points;   SD = 19 points;

Y   Midterm 1 score: avg = 145 points;   SD = 29 points; $r = 0.37$.

The scatterplot that shows the data is displayed below and can be assumed to be football–shaped.

Stat 1040: Midterm 1 − Quizzes

(scatterplot with "r = 0.37", x-axis "Points Quizzes 1-5" from 20 to 100, y-axis "Points Midterm 1" from 80 to 180)

Show your work!

−2 each calculation error

−2 if x, y flipped

−2 if x, y not specified

1. **(15 Points)** Find the regression equation for predicting the Midterm 1 score from the Quiz 1-5 score.

$$slope = r \cdot \frac{SD_y}{SD_x} = 0.37 \cdot \frac{29}{19} = 0.56 \qquad \textcircled{6}$$

$$intercept = avg_y - slope \cdot avg_x = 145 - 0.56 \cdot 73 = 104.1 \qquad \textcircled{6}$$

regression equation: $\boxed{Midterm\ 1\ score = 104.1 + 0.56 \cdot Quiz\ 1\text{-}5\ score} \qquad \textcircled{3}$

or: $\boxed{y = 104.1 + 0.56 \cdot x}$

2. **(8 Points)** Using your regression equation, estimate the Midterm 1 score for a student who had a Quiz 1-5 score of 60 points.

predicted Midterm 1 score for someone with 60 points in the quizzes =

$$104.1 + 0.56 \cdot 60 = \underline{137.7\ points} \qquad \textcircled{8}$$

−2 for old method, correct result

−7 for old method, incorrect result

−5 if result makes no sense at all

3. (**7 Points**) Find the r.m.s. error for predicting the Midterm 1 score from the Quiz 1-5 score.

$$\text{r.m.s. error} = \sqrt{1-r^2} \cdot SD_Y$$
$$= \sqrt{1-0.37^2} \cdot 29$$
$$= \underline{\underline{26.9 \text{ points}}} \quad ⑦$$

−3 for each major mistake, e.g. SD_x instead of SD_y, √ of everything, r instead of r^2 etc.

4. (**10 Points**) Can we use the regression equation to predict the Midterm 1 score for a student who had a Quiz 1-5 score of 19 points? ⟨YES⟩ or NO? Circle your answer and provide a short explanation. ④

$$S.u. = \frac{19-73}{19} = -2.84$$

This is within 3 SDs of the average where 99.7% of the data are located; this is not a problem and also no extrapolation (and actually we have an observation with a Quiz 1-5 score of 19 points : see the scatterplot) ⑥

5. (**10 Points**) Independently from your previous answer, let us assume that we can use the regression equation to predict the Midterm 1 score for a student who had a Quiz 1-5 score of 19 points. Would you be *surprised* that a student with 19 points in the quizzes got a score of 178 points in Midterm 1? **YES** or ⟨**NO?**⟩ **Circle your answer and provide a short explanation.** ④

predicted Midterm 1 score for someone with 19 points in the quizzes :
$$104.1 + 0.56 \cdot 19 = 114.7 \text{ points} \quad ②$$

observed Midterm 1 score = 178 points

$$S.u. = \frac{\text{observed} - \text{predicted}}{\text{r.m.s. error}} = \frac{178 - 114.7}{26.9} = 2.35$$

This is within the 3 r.m.s. error band of the regression line where 99.7% of the data are located; even though this point looks like an outlier in the scatterplot, the observed score of 178 is still not too surprising in the framework of the overall data (note the large r.m.s. error of almost 27 points!). By the way, a score of 196 points (or better) would have been really surprising for someone with 19 points in the quizzes. ④

Question 3: Chance Errors in Sampling (**40 Points**)

A group of 50,000 tax forms has an average gross income of $37,000, with an SD of $20,000. Furthermore, 20% of the forms have a gross income over $50,000. A group of 900 forms is chosen at random for audit. To estimate the chance that between 19% and 21% of the forms chosen for audit have gross incomes over $50,000, a box model is needed.

1. (**5 Points**) Should the number of tickets in the box be **900** or **(50,000?)** ⑤

   Circle your answer.

   Note: 900 is the sample size (i.e., # draws)

2. (**5 Points**) Each ticket in the box shows

   ⑤ **(a zero or a one)**   or   **a gross income**

   Circle your answer.

   Note: 0: gross income less than or equal to $50,000
        1: gross income over $50,000 ---

3. (**5 Points**) True or **false** the SD of the box is $20,000. ⑤

   Circle your answer.

   Note: box SD = $\sqrt{0.20 \cdot 0.80} = 0.4$

4. (**5 Points**) **True** or false: the number of draws is 900. ⑤

   Circle your answer.

5. (**12 Points**) Find the chance (approximately) that between 19% and 21% of the forms chosen for audit have gross incomes over $50,000. **Show your work!**
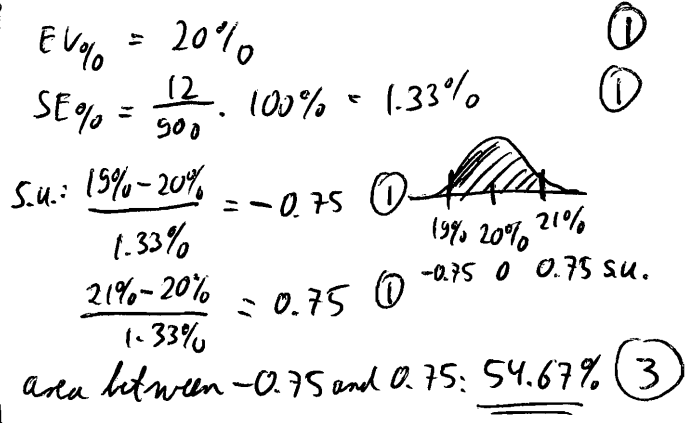
   The chance is: ___ 54.67 ___ %

   box: $\lfloor 10{,}000 \times \boxed{1} \quad 40{,}000 \times \boxed{0} \rfloor$ ①
        # draws = 900 ①

   box avg = $\frac{10{,}000}{50{,}000} = 0.2$ ①

   box SD = $\sqrt{0.2 \cdot 0.8} = 0.4$ ①

   $EV_{sum} = 900 \cdot 0.2 = 180$ (not required)

   $SE_{sum} = \sqrt{900} \cdot 0.4 = 12$ ①

   $EV_{\%} = 20\%$ ①

   $SE_{\%} = \frac{12}{900} \cdot 100\% = 1.33\%$ ①

   S.u.: $\frac{19\% - 20\%}{1.33\%} = -0.75$ ①

       19% 20% 21%

   $\frac{21\% - 20\%}{1.33\%} = 0.75$ ①

      -0.75 0 0.75 s.u.

   area between −0.75 and 0.75: 54.67% ③

6. (**8 Points**) With the information given, can you find the chance (approximately) that between 9% and 11% of the forms chosen for audit have gross incomes over $75,000? **Either find the chance, or explain why you need more information.**

   The chance is: _____ %

   We have no way to calculate this chance! We need to know the percentage of forms that have gross income → if attempt to calculate a % over $75,000 in order to find an EV and SE. ⑧

4

Question 4: Probability and Chance (**40 Points**)

−2 each calculation error

A bookshelf contains 8 novels, 7 books of poems, 1 dictionary, and 2 copies of Freedman, Pisani, and Purves's "Statistics" textbook. I pick two books at random without replacement. **Answer each of the following questions separately. Show your work!**

$8 + 7 + 1 + 2 = 18$ books.

1. (**5 Points**) What is the chance that the first book is a novel or a dictionary?

   The chance is ____50.0___ %

   chance first is novel: $\frac{8}{18}$ ①

   chance first is dictionary: $\frac{1}{18}$ ① ⟩ mutually exclusive ③ correct rule

   chance first is novel **or** dictionary: $\frac{8}{18} + \frac{1}{18} = \frac{9}{18} = \frac{1}{2} = 0.5 = \underline{\underline{50\%}}$

2. (**13 Points**) If I want to study statistics, what is the chance that I pick at least one copy of the Stats textbook?

   The chance is ____21.6___ %

   chance first is no stats book: $\frac{16}{18}$ ③

   chance second is no stats book, given that first is no stats book: $\frac{15}{17}$ ③ ③ correct rule

   chance both are no stats books: $\frac{16}{18} \cdot \frac{15}{17} = \frac{240}{306}$ ④ correct rule

   chance at least one stats book: $1 - \frac{240}{306} = \frac{306}{306} - \frac{240}{306} = \frac{66}{306} = 0.216 = \underline{\underline{21.6\%}}$

3. (**10 Points**) What is the chance that the first two books are both novels?

   The chance is ____18.3___ %

   chance first is novel: $\frac{8}{18}$ ①

   chance second is novel, given that first is novel: $\frac{7}{17}$ ④ ⑤ correct rule

   chance both are novels: $\frac{8}{18} \cdot \frac{7}{17} = \frac{56}{306} = 0.183 = \underline{\underline{18.3\%}}$

4. (**12 Points**) What is the chance that I pick one book of poems and a dictionary (in any order)?

   The chance is ____4.6___ %

   chance first is book of poems: $\frac{7}{18}$ ①

   chance second is dictionary, given that first is book of poems: $\frac{1}{17}$ ②

   chance first is book of poems & second dictionary: ② correct rule $\frac{7}{18} \cdot \frac{1}{17} = \frac{7}{306}$

   chance first is dictionary: $\frac{1}{18}$ ①

   chance second is book of poems, given that first is dictionary: $\frac{7}{17}$ ②

   chance first is dictionary & second book of poems: ② correct rule $\frac{1}{18} \cdot \frac{7}{17} = \frac{7}{306}$

   chance one book of poems and a dictionary (in any order) = $\frac{7}{306} + \frac{7}{306} = \frac{14}{306} = 0.046 = \underline{\underline{4.6\%}}$ ② correct rule

Question 5: Sampling (30 Points)

Part 1: (20 Points)

For each of the following, decide whether this describes a simple random sample (SRS). **Just circle your answer.**

- (5 Points) A student newspaper asked readers to respond to the question "Do you think that there should be more student activities on the weekends?" An overwhelming 95 percent said "yes". The article reporting the results concluded that 95 percent of all students feels this way.

  This is a SRS: yes / (no) (5)

  Note: this is a voluntary response survey

- (5 Points) A researcher selects a sample from a list of all patients at one of five large hospitals in the following manner. A patient is chosen from the first 25 on the list, then every 25th patient from that point forward is selected.

  This is a SRS: yes / (no) (5)

  Note: this is a systematic sample

- (5 Points) Fifteen state parks are to be selected from 1000 state parks in such a way that each has an equal chance of being selected. A random number generator on a computer is used to select 15 integers between 1 and 1000. Based upon those integers, the state parks are selected from a numbered list.

  This is a SRS: (yes) / no (5)

- (5 Points) A researcher chooses a random sample of households, then interviews every member of the selected households.

  This is a SRS: yes / (no) (5)

  Note: this is a cluster sample

Part 2: (10 Points)

A population consists of 100 individuals who have been numbered from 1 to 100 for the purpose of taking a simple random sample of ten individuals. Which of the following sets of ten is most likely to be chosen as the sample?

Just circle the correct answer:

- (a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- (b) 5, 10, 15, 20, 25, 30, 35, 40, 45, 50
- (c) 3, 17, 24, 39, 41, 47, 66, 73, 87, 96
- (d) These are all equally likely.

Note: In a SRS, each possible set of 10 (different) individuals has the same chance of being selected. This includes extreme cases such as (a) and (b) on the left.

(10)

6