(b) The 8,000 children were part of a "longitudinal study", that is, data is recorded about the same children over time. We are not told how the children were initially selected. Would it affect the outcome of the study if the sample were biased, for example if it contained too many children from well-off families or too many girls? Explain your answer.

(c) What is the causal link between handedness and hyperactivity discussed in the article?

(d) Does the research prove this causal link? Explain briefly.

(e) The article says that

"Mixed handed children aged seven and eight were twice as likely as their right handed peers to have difficulties with language".

i. Which two conditional probabilities are being discussed here?

\[
P(\text{language difficulty} \mid \text{mixed handed}),
\]

\[
P(\text{language difficulty} \mid \text{right handed})
\]

ii. What is the relationship given in the article between these two conditional probabilities?

\[
P(\text{language difficulty} \mid \text{mixed handed}) = 2 \times P(\text{language difficulty} \mid \text{right handed})
\]

(f) [BONUS] (2 marks) If you are told that a randomly chosen seven year old has language difficulties, are they more or less likely to be right handed or ambidextrous? How much more/less likely? [Hint: either manipulate conditional probabilities (hard) or consider 1000 kids and count (easier)]
Randomly chosen
7 year old.
- has language difficulties.

Is this child more likely to be right handed or mixed handed?

six 10x 20x 50x

twice as likely to be RH.

\[ P(\text{right handed} \mid \text{language difficulty}) \]
\[ P(\text{mixed handed} \mid \text{language difficulty}) \]
1 in 100 is ambidextrous.

10 mixed-handed.

\[ 990 \text{ right handed} \]

\[ P(\text{language difficulty} \mid \text{mixed handed}) = \frac{1}{11} \]

\[ P(\text{language difficulty} \mid \text{right handed}) = \frac{1}{2} \]

1 mixed handed kid has language difficulties.

990 \times \frac{1}{20} = 49.5 \text{ right handed kids have language difficulties.}

\[ P(\text{right handed} \mid \text{language difficulty}) = \frac{49.5}{49.5 + 1} = 0.98. \]

\[ P(\text{mixed handed} \mid \text{language difficulty}) = \frac{1}{49.5 + 1} = 0.02. \]
A - person is female

B - person is pregnant.

\[ P(B | A) = P(\text{pregnant} | \text{female}) = \frac{15}{80} = 1.875 \approx 2\% \]

\[ P(A | B) = P(\text{female} | \text{pregnant}) = 1 \]
July 1, 2009

President Mark G. Yudof  
Office of the President  
University of California  
1111 Franklin Street, 12th Floor  
Oakland, CA 94607-5200  

Via e-mail: president@ucop.edu

Dear Mark:

Re: Comments on Proposed Furlough/Salary Reduction Plan Options

Thank you for the opportunity to comment on the proposed Furlough/Salary Reduction Plan Options. The Santa Cruz campus solicited comments from Senate faculty, non-represented academic and staff employees, and managers and supervisors, and received over 650 comments in the brief turnaround time provided. While we received a wide variety of comments and suggestions, this letter articulates the most prevalent and key points expressed by the campus community. In addition, I have enclosed the comments received for your review and consideration.

As you will see in the following summary of comments, there were five specific areas in which both academic and staff employees articulated common viewpoints:

1) Option II was the preferred selection by an overwhelming margin;
2) Retirement, service credit, and leave accruals should not be negatively impacted;
3) Salary reductions should be graduated or progressive;
4) Specific sunset clause needed for whatever plan is chosen; and
5) Extramurally funded employees should not be included in any plan.

Summary of Employee Comments

We received over 300 comments from the academic members of the Santa Cruz campus, including comments from the Chair of the Santa Cruz Division of the Academic Senate and the Senate Committee on Planning and Budget. In addition, some 350 comments were received from our managers, supervisors and non-represented staff employees.

Of those who responded with a specific option selection, 84% of academic respondents and 88% of staff respondents chose Option II: 21 Unpaid Days Plan. The main reason conveyed by academics for this
(e) Is the mean likely to be less than, greater than or close to the median? Explain your answer.

(f) Does the histogram appear to follow the normal curve? Explain briefly.

(g) The SD of age is approximately 24 years. What proportion of the data lies within 2 SD of the mean?

5. (2 marks) At the end of this exam is the first page of a letter from UCSC Chancellor George R. Blumenthal to UC President Mark Yudof. This letter was part of a discussion about reducing the UC’s budget shortfall by reducing expenditure on staff and faculty.

Read the letter, paying particular attention to the section under the heading “Summary of Employee Comments”.

(a) How many comments were received overall?

(b) How many respondents chose Option II? Explain your answer.
once every 50 years
stopped flights for 1 week.

\[ \frac{1}{80} \times \frac{7}{365} \]
Inference.

\[ \text{Population} \rightarrow \text{Sample} \]

- everything I'm interested in.
- parameter of the population.
- what I've collected data about.
- what does the sample say about the population parameter?

Size of the chance error in the sample.
- it depends mostly on the size of the sample.
- much less on the size of the population.

(sampling with replacement, as sample size much less than population size).

Start by looking at a known population
+ studying size of chance error
- in 0% in the sample.
<table>
<thead>
<tr>
<th>Fresh</th>
<th>11</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh</td>
<td>82</td>
<td>39%</td>
</tr>
<tr>
<td>Junior</td>
<td>88</td>
<td>42%</td>
</tr>
<tr>
<td>Senior</td>
<td>29</td>
<td>14%</td>
</tr>
<tr>
<td>Grad</td>
<td>1</td>
<td>0% (approx.)</td>
</tr>
</tbody>
</table>

Spring 2010
AMS 5 students.

Sample (4th row) - 3 sophomores out of 10. (30%).
No a SRS as friends tend to sit together.

In my SRS there were 6 out of 11 = 55%.
Is this within the range of fluctuation we would expect?

Sample Size 11.

\[
EV = 11 \times \frac{82 + 0}{211} = 4.3.
\]

as a % of the sample.

\[
\frac{4.3}{11} \times 100 = 39%. 
\]
Size of fluctuations.

$$SE_{sum} = \sqrt{\text{#draws} \times \text{SD box}}$$
$$= \sqrt{11 \times (1 - 0) \times \frac{82}{211} \times \frac{129}{211}}$$
$$= 1.6$$

$$SE_\% = \frac{SE_{sum}}{\text{#draws}} \times 100 = \frac{1.6 \times 100}{11} = 15.0\%.$$

In a sample of $11$ AMS 5 students, expect $39\% \pm 15\%$ to be sophomores.

(5 so our value of $55\%$ is not unreasonable)

Imagine we have a $SE_{5}$ of size $44$

$$SE_{sum} = \sqrt{44} \times \sqrt{\frac{82}{211} \times \frac{129}{211}}$$
$$= 3.2.$$

$$SE_\% = \frac{3.2}{44} \times 100 = 7.3\%.$$

Sample increased in size by factor of 4
Error in % reduced by a factor of 2.
As sample size increases, chance error in \( \% \) decreases.

Irrespective of size of the population provided that the sample is small enough that we can consider the sample to be approximately drawn with replacement.

(SE for sample is exactly SE from box model)

What happens when the sample is a significant fraction of the population?

\[
\text{SE when drawing without replacement} = \text{SE when drawing with replacement} \times \text{correction factor}.
\]

\[
\text{correction factor} = \sqrt{\frac{\text{tickets in box} - \text{# draws}}{\text{# tickets} - 1}}
\]
Sample size 211 out of a population of 211

\[ \text{cf} = \sqrt{\frac{211 - 11}{211 - 1}} = 0.976. \]

Sample size 44

\[ \text{cf} = \sqrt{\frac{211 - 44}{211 - 1}} = 0.9. \]

Effect on chance error.

Sample size 11. Corr.\( \text{SE}_\% \) = \( 15 \times 0.976 \) = 14.6%.

44 Corr.\( \text{SE}_\% \) = \( 7.3 \times 0.9 \) = 6.6%.

Corrected \( \text{SE}_\% \) is reduced from the uncorrected values.

Sample of size 44, the difference starts to be important.
Alternatively.

If I fix the sample size, how large must the population be for the correction to be negligible?

Sample size 2500
Population 50% [ ]

<table>
<thead>
<tr>
<th>Population size</th>
<th>Correction factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>0.707</td>
</tr>
<tr>
<td>10,000</td>
<td>0.866</td>
</tr>
<tr>
<td>100,000</td>
<td>0.987</td>
</tr>
<tr>
<td>500,000</td>
<td>0.997</td>
</tr>
</tbody>
</table>

---

Population

Assumed known composition

Sample

Characteristics of sample %.
What can we say about the population?

From the size and composition of the sample, how accurate will the estimate of the population parameter be?

What range of values for the population parameter would we consider reasonable, based on the data we have?
Population

All UCSC student.

What is the % freshman on campus?

Sample size 250.
70 freshman.
(\text{\# sum of draws}).

\[
\frac{70}{250} \times 100 = 28\%.
\]

What's the SE on the % of freshman in my sample?

\[\text{SE}_{\text{sum}} = \sqrt{\text{\# draws}} \times \text{SD box}\]

\[\text{SD box} = (1 - \theta) \sqrt{\frac{\text{fraction of tickets with } 11's}{\text{tickets with } 11's} \times \frac{\text{fraction of tickets with } 00's}{\text{tickets with } 00's}}\]

UNKNOWN!

This is the parameter of interest.

So use the fraction in the sample in place of the fraction in the population.
\[ SD_{box} = (1 - 0) \sqrt{\frac{70}{250} \times \frac{150}{250}} = 0.449 \]

\[ SE = \sqrt{250 \times 0.449} = 7.1 \]

\[ SE_{\%\text{fresh}} = \frac{SE_{\text{fresh}}}{\text{sample size}} \times 100 = \frac{7.1}{250} \times 100 = 2.8\% \]

So from our sample of 250 students, we estimate:

\[ \% \text{ of fresh on campus} = 28\% \pm 2.8\% \]

Be careful in interpreting this.