

### 36-201: Introduction to Statistical Reasoning

1. [8 points; 2 each part].

- (a) small  $p$ -value bad for Null Hypothesis,  $H_0$ .
- (b) Correcting 1180 lbs to 118 lbs will most likely affect the mean and standard deviation.
- (c) A 95% percent confidence interval will be wider than a 90% confidence interval for this proportion.
- (d) A 95% confidence interval based on 400 students would probably be narrower than a 95% confidence interval based on 100 students.

2. [27 points]

(a) Weight-loss programs...

(i) [4 points]

$H_0$ : Both groups show same weight change

$H_A$ : The Support Group will show greater weight loss than the Did Nothing group

(ii) [3 points] Either of the first two, and the third, for full credit:

- 3.53 is a very large effect according to the 68-95-99.7 rule
- $p$ -value = 0.00042 is bad for  $H_0$  at almost any reasonable cutoff (0.05, 0.01, ...).
- There is strong evidence to reject  $H_0$ .

(iii) [2 points] How should the 50 overweight people have been selected to begin with?

Anything to the effect of select 50 who are representative of the population of overweight people. Here are some sample answers (any one would do):

- Randomly select from the population of overweight people
- do matching, stratified sampling, etc, to get a good mix of characteristics of overweight people in both the treatment and the control groups.

(b) The “Breathalyzer” crackdown...

(i) [4 points]

$H_0$ : No difference in highway casualties after the crackdown starts

$H_A$ : Highway fatalities lower after the crackdown starts

(ii) [2 points] You would use a A time series plot to study this data.

(iii) [3 points] The 30% drop in highway casualties is very dramatic, but we don’t know enough to say that the Beathalyzer crackdown was the only, or even a major, reason for the drop. Weather conditions, amount of traffic, changes in posted speed limits in Britain, etc., are all possible confounding variables that could have caused the drop even if no crackdown had taken place.

(c) IV specialist teams...

(i) [4 points]

$H_0$ : Complications due to IV therapy is independent of whether a special IV team is doing the work or not.

$H_A$ : IV complications and the presence of an IV team are dependent.

*Note:* Alternatively this could be done in terms of the proportion of IV complications (under  $H_A$  expect lower proportion of complications with the IV team).

(ii) [2 points] We would use a Chi-squared test to analyze the data.

If the first answer is in terms of means or proportions give lots of partial credit and accept confidence intervals here.

(iii) [3 points] It was a good idea, for this experiment, that new internal medicine patients were randomly assigned to each group, because this guarantees that the patients in each group have about the same mix of features that could matter; each feature is a potential lurking variable and randomization breaks any possible tie with lurkers.

3. [14 points] The Honolulu Heart Program...

(a) [3 points] This study is a/an Observational Study.

(b) [4 points] *Row Percents:*

Milk Consumption	Stroke?		Total
	Yes	No	
≥ 2 glasses per day	3.7	96.3	100%
< 2 glasses per day	7.9	92.1	100%

(c) [3 points] There is a negative association between milk consumption and suffering a stroke: men who drink at least 2 glasses per day have less than half the risk of getting a stroke.

(d) [4 points] Some sample answers:

**Confounding:** exercise (can't "cause" you to drink milk, but might be correlated with milk consumption and cause lower risk of stroke).

**Common response:** diet (could both "cause" you to drink milk and lower the risk of stroke).

4. [14 points] “cabbage” and “onion” in wine...

(a) [3 points] This study is a/an Experiment.

(b) [3 points] In this study,

concentration of ethanethiol is the *explanatory variable* and percent correct is the *response variable*.

(c) [4 points]

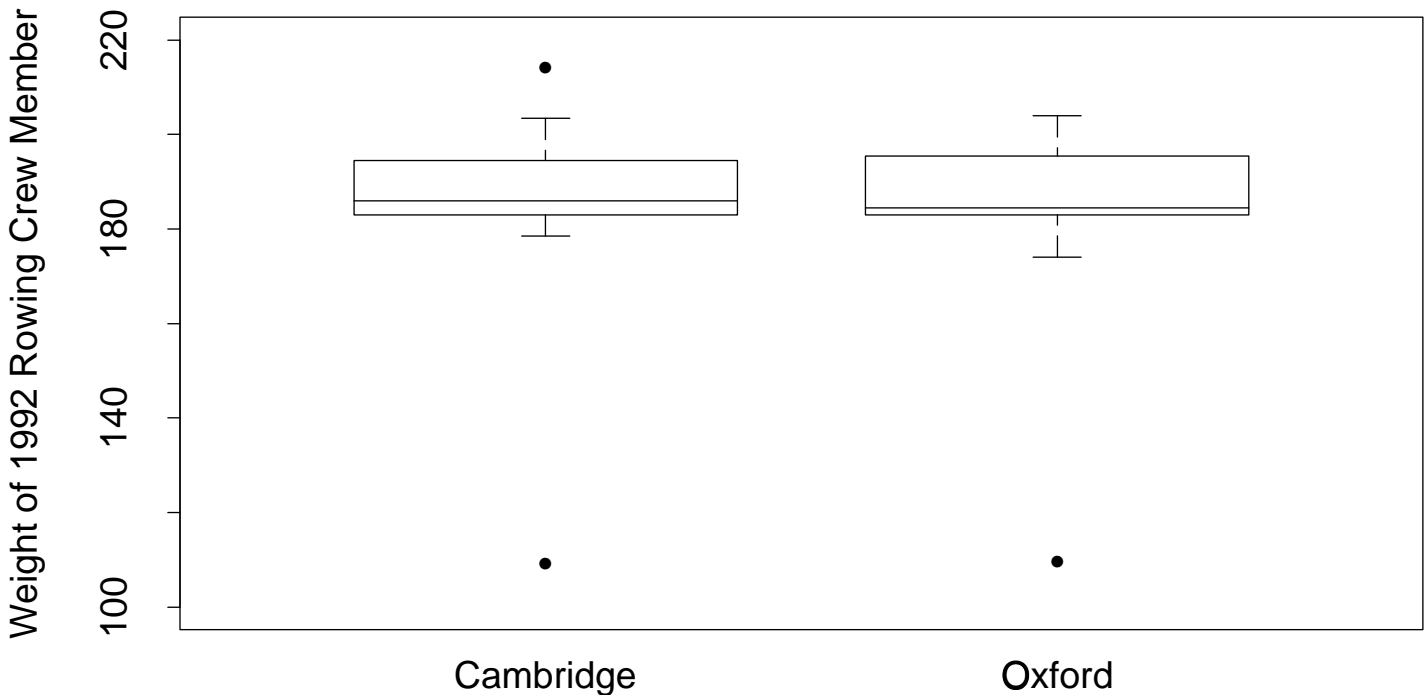
Straight line fits very well, except for two points well off the line.

Correlation of 0.93 is very high, indicating almost perfect positive linear association.

(d) [4 points] A horizontal line through 75% intersects the regression line at a log-concentration just below 0.7

5. [14 points] Rowing crews...

(a) [4 points] Here are boxplots, in S-PLUS. Only the two lower outliers are required for full credit (S-PLUS has a different definition of quartiles, hence IQR, hence outliers by  $1.5 \times \text{IQR}$  rule, than *Data Desk*).



(b) [3 points] List outliers...

The first outlier here is not needed for full credit (would be found only if student recomputes quartiles and IQR using the method of Handout #2 in class).

Value	School	Possible Reason
214	Cambridge	No obvious reason; just a big guy.
109	Cambridge	Coxswain
109.5	Oxford	Coxswain

(c) [4 points]

- If the outlier(s) for Cambridge were left out, the mean weight of the remaining Cambridge crew members would be somewhat higher than the current mean of 182.4 pounds.
- If the outlier(s) for Oxford were left out, the mean weight of the remaining Oxford crew members would be somewhat higher than the current mean of 180.4 pounds.

(d) [3 points] Can you use the confidence interval and hypothesis test to draw conclusions about the populations of all rowing crew members at each school? Explain.

- 1 pt for just comparing the CI and signif. test with the boxplots and saying “they all reach the same conclusion”.
- 3 pts for observing that the 1992 rowing crews are probably not a simple random sample from all row crew members.

6. [9 points] “Ask Marylin”...

(a) [5 points]

Observed Counts	White Collar Job			Blue Collar Job			Combined		
	Hired?		Total	Hired?		Total	Hired?		Total
Applicant:	Yes	No		Yes	No		Yes	No	
Male	30	170	200	300	100	400	330	270	600
Female	40	160	200	85	15	100	125	445	300
<b>Total</b>	70	330	400	385	115	500	455	445	900

Row	White Collar Job			Blue Collar Job			Combined		
	Hired?			Hired?			Hired?		
Percents	<u>Yes</u>	<u>No</u>	<u>Total</u>	<u>Yes</u>	<u>No</u>	<u>Total</u>	<u>Yes</u>	<u>No</u>	<u>Total</u>
Applicant:									
Male	15%	85%	100%	75%	25%	100%	55.00%	45.00%	100%
Female	20%	80%	100%	85%	15%	100%	41.67%	58.33%	100%

(b) [4 points] This is an example of Simpson's Paradox. Write a response from Marilyn...

By looking at the row percents we can see that the company actually hires a greater proportion of female job applicants than it does of male applicants, in both white and blue collar jobs. That is what the company president meant. The reason the overall percentages are reversed is that, although even numbers of men and women applied to the white collar jobs, many more men than women applied to the blue collar jobs. If a more equal number of men and women had applied to the blue collar jobs, then the overall percentages would have favored women. Combining the data sets means you can potentially lose important information about the problem.

7. Dole vs. Clinton...

(a) [7 points, one per blank] Fill in the blanks (if you can't tell, just write "can't tell"):

- What was the population? All registered voters.
- How was the sample selected? stratified or multistage rand. sample.
- How were the sample subjects contacted? Telephone.
- What was the sample size? 1035.
- What was the overall margin of error? ±3%.
- When was the survey conducted? March 21 to April 2, 1996.
- Can you find out the exact survey questions asked? I think Figure 4 gives them.

(b) [4 points] Each CI is the estimated percent from Figure 4, plus or minus 3%, which is the overall margin of error for the survey:

- For the question “*He would make the changes I would like to see in the welfare system*”,  
The proportion of voters who think this about Clinton  
is between 42% and 48%, with 95% confidence.  
The proportion of voters who think this about Dole  
is between 39% and 45%, with 95% confidence.
  - For the question “*I have confidence in his ability to deal with an international crisis*”,  
The proportion of voters who think this about Clinton  
is between 38% and 44%, with 95% confidence.  
The proportion of voters who think this about Dole  
is between 44% and 50%, with 95% confidence.
- (c) [3 points] The check (“√”) is made when the 95% CI’s don’t overlap or at worst just kiss. The check mark goes with the higher sample proportion; since there is no overlap in the CI’s we are confident that the population proportion is higher for that person too. This is what happens with the “*international crisis*” question (and we are confident Dole’s proportion is higher).
- With the “*welfare*” question, the two confidence intervals do overlap, so we cannot be confident about whose population proportion is higher.
- Note [not part of answer]:** The same explanation can be used for all the checkmarks in the graphic.