MEASURES, WORDS, RATES, RATIOS, AND PROPORTIONS

I. **Problems Involving Choice of a Measure**.

The purpose of these problems is to give you more examples of thinking about the different measures that might be used to answer a question, and the consequences of choosing one measure over another.

1. In order to consider whether progress is being made in treating cancer, you might consider the following as possible ways to measure progress:
     a. Total deaths from cancer in the country in one year.
     b. The percent of all people in the country who die from cancer in one year.
     c. The percent of all cancer patients who survive cancer for five years from the time the disease was discovered.

Discuss the usefulness of each of these in measuring the effectiveness of cancer treatment. First think about whether having the measure large or small is good or bad. Could the measure go in the "bad" direction even if cancer treatment is becoming more effective? Can you think of any better measures? If so, why do you think they are better?

2. The Center for Disease and Control Healthy Weight website (http://www.cdc.gov/healthyweight/assessing/bmi/) says,
     "Body Mass Index (BMI) is a number calculated from a person's weight and height. BMI provides a reliable indicator of body fatness for most people and is used to screen for weight categories that may lead to health problems."
Here is the classification in weight category by BMI:

BMI	Classification
Below 18.5	Underweight
18.5 – 24.9	Normal
25.0 – 29.9	Overweight
30.0 and Above	Obese

If height is measured in inches and weight in pounds, BMI is calculated by

$$\text{BMI} = 703\{\text{weight}/[(\text{height})^2]\}$$

(The factor 703 is a conversion factor – the original formula used height in meters and weight in kilograms).

Think about this formula and how good you might expect it to be as an indicator of healthy weight. What factors might it not take into account? Why do you think the website said, "for most people"?
 [Hint: Tom Cruise is 5'7" and weighs 201 lbs, giving him a body mass index of 31.5].

3. In studying the effect of sex differences on traffic fatalities, which measure below should we look at, and why? Or are there slightly different variants of the general question that will be better answered by one measure than the other? Or by still another measure? (If so, what?)
In the following, "fatalities" refers to fatalities while traveling in a car; "number of miles traveled" means number of miles traveled by car; etc.

      A. total number of male fatalities/total number of female fatalities

      B. (#male fatalities/number of males in population)/(#female fatalities/number of females in population)

      C. (#male fatalities/total number of miles traveled by males)/(#female fatalities/total number of miles traveled by females)

      D. (#male fatalities/total number of minutes traveled by males)/(#female fatalities/total number of minutes traveled by females)

4. In considering how may people are unemployed, we can sort people into the following categories:

      A. Employed full time (includes people "with a job but not at work," e.g., on temporary family leave, not currently working because of a strike)

      B. Employed part-time, want part-time work. (Includes full-time students 16 or older.)

      C. Employed part time, want full-time work. ("working part time for economic reasons." Includes full-time students 16 or older.)

      D. Work 15 hours or more in a week without pay in a family-operated enterprise. ("unpaid family workers")

      E. Not employed, actively looked for work in the last month, not on temporary layoff (Includes full-time students 16 years or older.)

      F. Not employed, on temporary layoff.

      G. Not employed, want a job now, looked for work in the last year, available for work. ("Marginally attached" to the labor force.") This includes "discouraged workers", who are not looking for work for one of the following reasons: 1) they believe no job is available to them in their line of work or area, 2) they had previously been unable to find work, 3) they lack the necessary schooling, training, skills or experience, or 4) employers think they are too young or too old, or they face some other type of discrimination.)

      I. Not employed, don't want a job now.

      J. Not employed, would like a job, but unable to hold a job because of family duties, disabilities etc., therefore not currently looking for work.

Note: Only people 16 or older not on active duty in the Armed forces and not in an institution (e.g., prison, mental institution) are included.

In the U.S., the official *unemployment rate* is defined as the number of people unemployed divided by the number or people in the labor force. The *number of people unemployed* is defined as those in categories E plus F. The *number of people in the labor force* is defined as those in categories A through F. Do you think that this definition gives a good measure? If not, what other measure would you propose?

1. Source: U.S. Department of Labor, Bureau of Labor Statistics, at
http://www.bls.gov/cps/cps_htgm.htm
2. Different countries use different means to calculate their unemployment rate.
3. Since about 1995, the U.S. Bureau of Labor Statistics has publishing some "alternate
measures of labor underutilization" in addition to the official unemployment rate. You
can find recent figures at http://www.bls.gov/news.release/empsit.t12.htm

5. Politician A said that SAT verbal scores had not increased from 1981 to 2002,
supporting his assertion by the fact that the average score for both years was 504.
Politician B disagreed, producing the following figures on average SAT verbal scores
from those years, broken down by ethnic group:

	1981	2002	Gain
White	519	527	+8
Black	412	431	+19
Asian	474	501	+27
Mexican	438	446	+8
Puerto Rican	437	455	+18
American Indian	471	479	+8

a. Must one politician be lying? Or could their figures both be correct? Hint: the
following information on the number and percentage of test-takers in each ethnic group
may be relevant:

	1981		2002	
	#	%	#	%
White	719,383	85	698,659	65
Black	75,434	9	122,684	11
Asian	29,753	3	103,242	10
Mexican	14,405	2	48,255	4
Puerto Rican	7,038	1	14,273	1
American Indian	4,655	0	7,506	1
Total		100		92

(2002 percentages do not sum to 100% because of 8 percent responding "Latin
American" or "Other," which were not given as response categories in 1981).

b. Compare Politician A and Politician B's measures of progress. Is one better than the
other?

II. **Words and Operations**

> " … in response to a challenging quantitative homework problem, [my students in environmental science classes] assert that they do not know the mathematics needed to solve the problem. In fact, the only mathematics needed may just be simple multiplication. The students know that branch of mathematics perfectly well! They just don't know *when* they need to multiply and *what* they need to multiply by what."
>
> John Harte
> Professor of Ecosystem Sciences
> University of California, Berkeley

The next set of problems has three purposes:
1. *To heed Prof. Harte's experience by doing some problems just involving arithmetic operations – but where the focus is on figuring out from context what operation to do to what numbers where.*
2. *To practice using possibly unfamiliar definitions.*
3. *To preview some concepts that will come up later in more complex contexts.*

1. (Warm-up with percents)
a. *Purchasing power* means the amount of goods you can purchase for a fixed amount of money. If prices go down by 15 percent, by what percent does your purchasing power increase?

b. In 2000, two law professors and a sociologist published a study of errors in death penalty cases in the U.S. from 1973 – 1995. They found that the following percentages of such convictions reversed in each of the following successive stages of the appeal process:

   41% were reversed on direct appeal to a state courts.
   10% of the remaining cases were reversed post-conviction state proceedings.
   40% of cases still remaining were reversed by federal habeas corpus.
What is the percentage of cases that were reversed overall?

Note: The word "proportion" has two different meanings in mathematics and applications.
   Meaning 1: "An equality between two ratios". This definition is used in middle grades math classes, but rarely elsewhere.
   Meaning 2: "A ratio of part to whole". This is the meaning used in this class and in many applications of mathematics and statistics to health, social sciences, etc.

2. In May 2000, the Missouri Department of Health received reports of eight people who had worked at the same microwave popcorn production plant and who had developed severe bronchiolitis obliterans (fixed obstructive lung disease) while working there. There were no obvious causes of the condition. Consequently, researchers studied workers in the plant, looking at how many had airway obstruction and their exposure to diacetyl, a chemical predominant in artificial butter flavoring and in the air at the plant. A

total of 116 workers were studied. Of them, 21 had airway obstruction. The researchers divided the workers into 58 with low exposure to diacetyl and 58 with high exposure. Of the 21 people with airway obstruction, 15 were from the high exposure group.

a. Use the above information to fill out the following *contingency table* (also known as a *two-way table*).

	Low exposure	High exposure	Total
Airway obstructed			
Airway not obstructed			
Total	58	58	116

b. i. What proportion of those with low exposure had airway obstructed?
   ii. What proportion of those with airway obstructed had low exposure?

c. What would be a good way to graph these data? Why?

d. The *risk* of a disease or other condition in a group is defined as the proportion (often expressed as a percentage) of the group that has the condition. The *relative risk* of two groups is the ratio of the risks of the two groups. Find
       i. The risk of having airway obstruction for the low exposure group
       ii. The risk of having airway obstruction for the high exposure group
       iii. The relative risk of having airway obstruction for the high exposure group compared to the low exposure group.

e. The *odds* of having a certain condition for a certain group are the ratio of the proportion of the group having the condition to the proportion of the group not having the condition. The *odds ratio* for two different groups is the ratio of the odds for the two different groups. (The standard convention is to put the group with the lower odds in the denominator, so that the odds ratio is greater than 1.)
       i. Find another, equivalent definition of odds in terms of numbers rather than proportions.
       ii. Find the odds of having airway obstruction for the low exposure group.
       iii. Find the odds of having airway obstruction for the high exposure group.
       iv. Find the odds of having airway obstruction for the total group.
       v. Find the odds ratio of having airway obstruction in the high exposure group compared to the low exposure group.

Comment: Here we are talking just about the group of 116 people studied. We aren't justified in extrapolating to others.

f. If the entries in the main part of the contingency table are as shown at right, find a simple formula (in terms of a, b, c, and d) for the odds ratio. (Actually, to follow the standard convention given above, you will need two formulas plus the phrase "whichever is larger")

a	b
d	c

3. a. The relative risk of Disease A for people who take Drug 1 is 50% (compared with people who do not take the drug). The relative risk of Disease B for people who take Drug 2 is also 50% (compared with people who do not take the drug). The risk of Disease A for people who do *not* take Drug 1 is 20%. The risk of Disease B for people who do *not* take Drug 2 is .0002%. What is the risk of Disease A for people who take Drug 1? What is the risk of Disease B for people who take Drug 2? Do Drugs A and B seem equally worthwhile?

   b. Sometimes drug ads, journal articles, and even government guidelines for physicians and consumers describe benefits in terms of relative risks and harms in terms of absolute risks. For example, the 2002 *Guide to Clinical Preventive Services* of the U.S. Preventive Services Task Force described the *benefits* of sigmoidoscopy screening by saying it "reduced the risk of death by 59% for cancers within reach of the sigmoidoscope" but described *harms* by saying, "Perforations are reported to occur in approximately 1 of 1,000-10,000 rigid sigmoidoscopic examinations." What do you think of this practice? Think about part a, and also the following quote from a drug ad: "Lipitor cuts the risk by nearly half. In patients with type 2 diabetes and at least one other risk factor for heart disease, Lipitor reduced the risk of stroke by 48%." The actual results of the clinical trial on which this claim was based were: After 4 years, 2.8% of patients taking the placebo (sugar pill) had a stroke, compared to 1.5% taking Lipitor.

      c. Another way of describing drug benefits is "number needed to treat": The (average) number who would need to take the drug to prevent one additional instance of the disease (or death, or …). In the Lipitor ad in part (b), what would be the number needed to treat (in the population studied) for 4 years to prevent one additional stroke?

4. Eight hundred patients with a disease took part in a clinical trial. Half were given an experimental drug, and half were given a placebo (a pill that looked like the drug). The experiment was well designed, so that neither the patients nor their doctors knew who got the drug and who got the placebo. At the end of the trial, patients were classified by their doctors as recovered or not. Here is a contingency table giving the results:

	Recovered	Not recovered	Total
Received drug	200	200	400
Received placebo	160	240	400
Total	360	440	800

a. Calculate the recovery rates of the two groups of patients (received drug, received placebo.) Which had the higher recovery rate?

b. Half the patients in the total study were male and half female (although the drug/placebo groups were not evenly distributed by sex). Here are the results for the males only:

Male	Recovered	Not recovered	Total
Received drug	180	120	300
Received placebo	70	30	100
Total	250	150	400

Calculate the recovery rates for males receiving the drug and for males receiving the placebo. Which had the higher recovery rate?

c. i. *Without doing any more calculations*, which group of females (received drug or received placebo) do you think had the higher recovery rate?

ii. Now make a contingency table for females in the study to check whether or not your prediction in part (i) was correct or not.