A recent report from a private research organization pointed out that at current rates of human consumption, the earth's drinkable water supply will be totally consumed by the year 2025. This report went on to note that "The rich get richer, the poor have children" and then blamed impoverished nations for the depletion of natural resources. Here is some information from the 1997 CIA World Factbook for 10 counties.

Country	Total Fertility Rate (i.e. Average number of Children born per woman)	Per capita Gross Domestic Product (GDP) in dollars (a measure of wealth)	Female Literacy Rate (i.e. women over age 15 who can read and write)
Afghanistan	6.07	$800	15%
Cambodia	5.81	$710	22%
Costa Rica	2.85	$5,500	95%
Indonesia	2.66	$3,770	78%
Italy	1.16	$19,600	96%
Jordan	4.94	$5,000	49%
Nigeria	6.17	$1,380	47%
Russia	1.35	$5,200	97%
United Arab Emirates	3.62	$23,800	80%
United States	2.06	$28,600	97%
AVERAGE	3.669	9436	67.6%
STAND. DEV.	1.8486	9888	30.27%

1. What is the correlation between gross domestic product and fertility (you can use the output on the next page to help you if you know what to do)?

The average of the products of GDP and Fertility = 24272.69
The correlation is:

$$\frac{24272.69 - (9436 * 3.669)}{9888 * 1.8486} = -.5661 \text{ or about } -.57$$

2. Using the data above or using the output shown on the next page, give the regression equation for predicting fertility from per capita Gross Domestic Product.

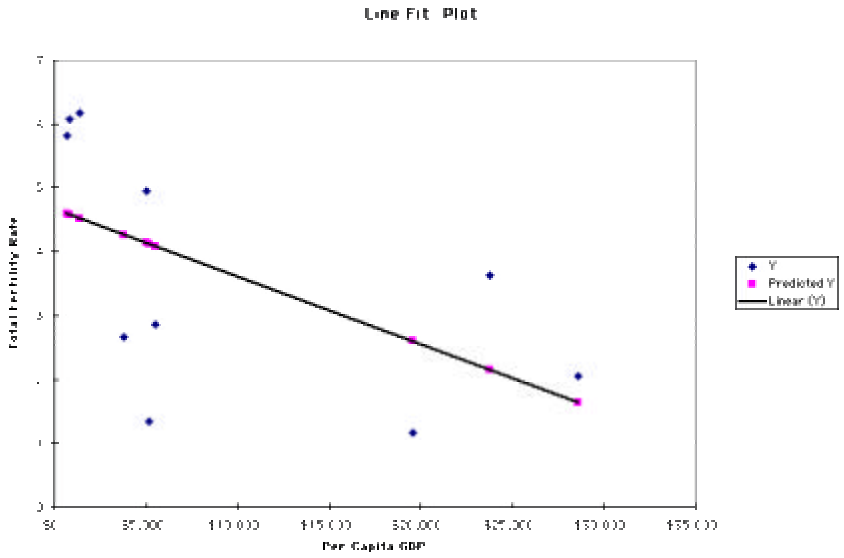
fertility = -.0001058(GDP) + 4.677

3. Cuba's GDP is $1,480. What is the predicted fertility rate for Cuba?

4.511 = -.0001058(1480) + 4.667

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.566123039					
R Square	0.320495295					
Adjusted R Square	0.235557207					
Standard Error	1.703676023					
Observations	10					
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	4.667676807	0.74469567	6.267898415	0.000240995	2.950404403	6.384949211
GDP	-0.000105837	5.44851E-05	-1.942493565	0.088006757	-0.00023148	1.9806E-05
RESIDUAL OUTPUT						
Observation	Predicted Y	Residuals	Standard Residuals			
1	4.583007302	1.486992698	0.925759305			
2	4.592532622	1.217467378	0.757960517			
3	4.085573963	-1.235573963	-0.76923316			
4	4.268671766	-1.608671766	-1.001513227			
5	2.593273944	-1.433273944	-0.892315538			
6	4.138492403	0.801507597	0.498995803			
7	4.521621911	1.648378089	1.026233253			
8	4.117325027	-2.767325027	-1.722857749			
9	2.148759045	1.471240955	0.915952718			

10	1.640742017	0.419257983	0.26101808			



4. Suppose I told you that Cuba's fertility rate is actually 1.54. Is the result you got in (3) different from 1.54? If it is different, give us some reasons why your predicted result might different from 1.54. If there is no difference, give us some reasons why your prediction is exactly on target.

The line is just the best fitting line, it's not expected to go through the points exactly. It is however, expected to go through the mean value of the Y variable for each level of X. What is implied is that there is a distribution of Y for each X and Cuba happens to be on the low end of the distribution.

Another thing to consider is the literacy information that was given, it actually has a higher correlation with fertility than GDP. In other words, there may be an unmeasured variable out there that better explains fertility.

5. Examining the computer output again, do you think GDP is a "good" predictor of fertility (hint, there might be a better one out there)? Which statistics from the output might you use to determine whether GDP is a "good" predictor of fertility?

GDP is not a good predictor. The t-test for the slope suggests that we cannot reject the null hypothesis that the true slope, b0 = 0. In other words, there is no compelling evidence to suggest that a nation's wealth is related to its fertility. Also, for the slope, the 95% confidence interval crosses zero which suggests that the slope for the population regression line could be zero or even positive (i.e. as nations get wealthier they have more children…). You might also say something about r-squared -- that this model explains about 32% of the variation in the y-variable (fertility) -- not exactly great.

The Public Health Service studied the effects of wine drinking on cholesterol in a large sample of representative households in the United States and in France. For men and for women in each age group in both counties, those who had drank moderate amounts of wine had lower cholesterol levels than those who drank no wine. But in the U.S. those who drank no wine had lower cholesterol levels than those who drank large amounts of wine. In France, those who drank no wine had ***higher*** cholesterol levels than those who drank large amounts of wine.

6. Why did they study men and women and the different age groups separately?

The researchers are interested in the relationship between wine consumption and cholesterol levels. Age and gender affect both wine consumption and cholesterol. Wine consumption is different for different age groups and men tend to drink more than women. Cholesterol levels are different for different age groups and for men and women. The researchers studied these groups separately to control for these confounding factors. [NOTE: identifying that confounding is the problem is key to answering this question correctly]

7. The lessons one learns from this study seems to be, if you drink lots of wine and are concerned about your cholesterol levels, you should live in France (if you do not already) and if you do not drink wine (and are concerned about your cholesterol levels), you should live in the United States. Is this correct? Explain. Be brief.

This is not correct. This is an observational study and we cannot conclude that wine drinkers who move to or live in France can reduce their cholesterol nor can we conclude that people who don't drink wine should move to or live in the US if they want to have low cholesterol. There could be unobserved factors (such as a stressful lifestyle) that could explain why wine in large amounts doesn't seem to help Americans like it seems to help the French.

[NOTE: identifying that this is an observational study, being able to discuss the problems associated with them and then providing an illustration to communicate your point is key here.]

Suppose it is known that the hourly wage for American workers is normally distributed with an average of $11.98 with a standard deviation of $6.21.

8. What percentage of American workers earn more than $20 per hour?

Should get a Z = 1.29 the resulting percentage to answer the question should be about 9.85% (or 1 - .9015)

9. A simple random sample of 81 American workers is drawn from the population. What is the chance that the sample average will fall between $11.50 per hour and $12.50 per hour?

Z for $12.50 is .75 and Z for 11.50 is about -.70, the area (chance) between them is about 53% (53.14 exactly perhaps)

10. Just this morning, the Department of Labor drew a sample of 1000 American workers from the population and found that their sample had a mean hourly wage of $12.49 and standard deviation of $8.42.

Test the hypothesis that there has been an increase in the hourly wage of American workers. State a null and an alternative hypothesis. Perform a test, state the resulting p-value and give a conclusion -- is there evidence to suggest that there has been an increase or not? Use a 5% level of significance to make your decision.

The null would be H0: mu = $11.98, the alternative Ha: mu > $11.98, the appropriate test is Z and it would (12.49 - 11.98) / (6.21 / root 1000) = 2.60 has a p-value of .0047. I would reject the null at the 5% level of significance, wages appear to be rising.

There are 20,000 restaurants in the County of Los Angeles, 60% of them received a letter grade of "A" during inspections, 30% received either a B or a C grade and 10% failed their inspections.

The University has hired you as a temporary worker. Your job is to schedule the next 10 of the Chancellor's dinners (he eats out all of the time). Unfortunately, you don't know about the rating system and you never eat out because you don't have the money. So you listen to your best friend and pick 10 restaurants at random from an internet database of restaurants in Los Angeles.

The Chancellor will give you +2 points if you choose "A" restaurants, no points if you choose "B" or "C" restaurants, and -10 points if you choose a restaurant with a failing grade.

11. What is the standard deviation for the average of samples of 10 restaurants?

The expected value is .2, the standard deviation 3.5157 and the standard deviation for samples of size 10 is 1.1118

12. To convert your temporary job into a permanent job, you must have an average of at least +2 points from the Chancellor after picking 10 restaurants for him. What's your chance of getting an average of at least +2 points after picking 10 restaurants?

1.619 = Z or about 1.62, should be about 5.26%

13. You wound up with an average of 1 point and did not get the job. Your best friend applied for the same job the next day, got it, and began picking restaurants for the Chancellor. After picking 10 restaurants, your friend had an average of -1.8 points and got fired. Is there evidence to suggest that your friend was deliberately trying to give the Chancellor food poisoning or was s/he using the random selection method too? State the null hypothesis and alternative hypothesis, perform a test, use a 1% level of significance as your rule, state the p-value and give us your conclusions.

Null is H0: mu = +.2 Alternative is H0: mu < +.2 The Z = -1.8 with a p-value of about 3.6%. You would not reject the null. There is not enough evidence (at the 1% level of significance) to suggest that your friend is trying to poison the Chancellor deliberately.

14. A poll on women's issues interviewed 1,025 women and 472 men randomly selected from the United States. The poll found that 47% of the women said they do not get enough time for themselves.

(a) Construct a 90% confidence interval for the percentage of women who say they do not get enough time for themselves.

(b) Your friend is taking Statistic 11 next quarter and unfortunately they are enrolled at 8am with Professor Lew and they ask you for help all of the time. Explain to your friend why we can't just say that 47% of all adult women in the U.S. do not get enough time for themselves.

(a) The 90% confidence interval is 47% + or - (1.645 * 1.56%) or something like 44.44% to 49.56%. The standard deviation is root ((47*53)/1025) = 1.56 why 1,025? We are only interested in the women's responses. (b) You might simply say that 47% is a sample percentage but what you would really like is the population percentage, but you don't have or know it. As long as we're only working with samples, we're going to be off of the true percentage by random chance. We do have tools, like the confidence interval, which allow us to make statements like "I am 90% confident that the true percentage is covered by the interval from 44.44% to 49.56%".

The following comes from a recent article in *The Wall Street Journal* :

Generation X-ers Aren't Relying On the Survival of Social Security

BY JOHN SIMONS

According to the most recent Wall Street Journal/NBC News poll, only 39% ofX-ers believe that Social Security will still be able to provide benefits when they retire. That compares to recent surveys of all Americans which show that 45% think so.

Assume that 45% figure is a stable, long-run, historical fact about American beliefs about Social Security benefits. Also assume that the survey of Generation X-ers had 121 respondents.

15. Test the hypothesis that the belief that Social Security will still be able to provide retirement benefits has decreased over time. State the null and the alternative, perform a test, and state a p-value. Please use a 5% level of significance as your decision rule. On the basis of your test results, do you think that Generation X-ers are like other Americans in their beliefs about social security or are they different?

Null is H0: p = .45 Alternative is Ha: p < .45 The Z = (39 - 45)/ root ((45*55)/121) = -1.33 with a p-value of about .0918 or 9.2%. You would not reject the null at the 5% level. This sample of Gen X-ers does not appear to be different from the population as a whole.

16. Suppose in a few years the Wall Street Journal decided to replicate this study (i.e.draw a new sample) on Generation Y-ers (that's you all, I think...). Let's assume that the 39% figure is now the stable, long run fact about belief in Social Security benefits by Americans.

What is the chance that a sample of 64 will have at least 30% of the surveyed Generation Y-ers believing in Social Security?

Z = (30 - 39)/ root ((39*61)/64) = -1.48. The chance that at
17. least 30% believe in social security will be about 93.1%

17. A simple random sample of 100 stocks was drawn from the entire market. The average return was 13%, and the SD was 6%; furthermore, the distribution of percentage returns in the sample was close to normally distributed.

Based on these data, is possible to construct a 95% confidence interval for the percentage of stocks in the market as a whole that had percentage returns greater than 20%?

Please answer yes or no. If you answer yes, please construct a confidence interval and please explain why this is possible to do. If you answer no, please explain why it is inappropriate to construct a confidence interval given this information.

Yes. Because the distribution is close to normal, we might use the normal approximation to figure out what fraction of stocks have percentage returns greater than 20%.

20% - 13%
---------- = 7/6 = 1.17
6%

and the area to the RIGHT of +1.17 from your table is:

.1210

so about 12.1% of the stocks had returns greater than 20%

to construct a confidence interval:

$$12.1\% \pm 1.96 * (\sqrt{\frac{.121 * .879}{100}}) * 100$$

18. 1996 was a particularly good year for the stock market. For 1996 as a whole, the mean return on all common stocks on the NYSE (New York Stock Exchange) was $\mu = 16.4\%$. The standard deviation was about $\sigma = 36\%$. Assume the distribution of annual returns is roughly normal.

(a) Suppose you selected 9 stocks at random from the NYSE stocks in 1996. What is the expected value (mean) and standard error of the returns of randomly chosen portfolios of 9 stocks?

That should read "standard deviation of the returns of randomly chosen portfolios of 9 stocks"

The expected mean is 16.4%
The standard deviation for the portfolios is 36/root(9) = 12

(b) What percentage of such portfolios of 9 stocks will lose money (i.e. have returns of zero or less)?

$$Z = \frac{0 - 16.4}{12} = -1.37 \text{ or } 8.53\%$$