Political Science 100a/200a
Fall 2001
Problem Set 2

1. Using **lifeexp.dta** from the course website, in Stata:

   (a) Produce a nice, professional looking histogram of the variable **lifeexp** (i.e., explore the **graph** command so that you know how to adjust the number of blocks, the labelling of the axes, the title, and so on).

   (b) What are the mean, median, standard deviation, and interquartile range of **lifeexp**? Interpret the meaning of the difference between the median and the mean in this case. How is this consistent with the skewness value? Does the distribution look "normal"? (Note: you will need to ask the command **summarize** to give you details to answer some parts of this question, that is, **sum lifeexp ,d**.)

   (c) What percentage of the observations lie within one s.d. of the mean? (Hint: an easy way to do this is to use **sum lifeexp if lifeexp > .. & lifeexp < ..** , filling in the ..'s as appropriate.)

   (d) **generate** a new variable that is the log of **wbgdp**, and repeat questions b and c.

   (e) Generate a new variable that represents life expectancy converted to standard units. If you wanted to convey information about a particular country's life expectancy, why might such a measure be useful?

   (f) Produce a histogram of **lifeexp** for the countries in each region. Do these distributions "look like" the distribution for the world as a whole? Does this suggest a possible explanation for the funny (i.e., not normal) look of the histogram for lifeexp for the whole world?

   (g) Produce a professional looking boxplot of **lifeexp**, by region, and use it to identify outlier cases in each region.

2. Use the identity established in class, $var(X) = \frac{1}{n}(\sum x_i^2) - \mu^2$ to show the following: Let X represent a binary variable that takes the value 1 for k cases and 0 $n-k$ cases. Then the mean of X is $p = k/n$ and the variance of X is $p(1-p)$.

3. A researcher decides to test the "democratic peace hypothesis" by collecting data on 65 bilateral interstate wars since 1815. ("Bilateral" means that the war involved only two states, one on each side.) She finds that only one of these dyadic wars involved a democracy fighting against another democracy, for a rate of .015. On this basis, she concludes that democracies are much less likely to fight against each other than are other sorts of interstate dyads.

   Is the researcher's reasoning correct? If not, show why by providing an example that is consistent with the data but does not support the conclusion.

4. For the countries with data, use Stata to calculate the correlation between the year the first McDonalds opened and the number of McDonalds operating circa 2000.

   (a) What does this statistic tell you?

   (b) Assuming that the underlying relationship between start year and number of currently operating restaurants continues into the future, would you expect the correlation to increase, decrease, or stay the same if we looked again in 2010 with 10 more years of data?

   (c) A friend tells you that the correlation coefficient is meaningless here because changing the scale of the "first year" variable (e.g., using months from 1955) would change its variance and thus affect the correlation coefficient. Is this logic correct? If not, show why.

5. Do problems 8 - 10 on pages 262-63 of FPP.

6. There are two al Qaeda members on a list of 100 names held by the F.B.I. The F.B.I. chooses 20 names on the list at random for interviews. What is the probability that at least one of the members is in the group of 20 chosen?

7. *Bayesian updating.* You have a coin that you believe is either fair or is biased in favor of heads in such a way that each flip has .75 chance of landing heads. You think these two possibilities ($Pr(\text{heads}) = .5$ and $Pr(\text{heads}) = .75$) are equally likely. You intend to perform an experiment to get data about the coin by flipping it.

   (a) What is your prior belief that the coin is biased?

   (b) What is your prior belief that the coin will land heads if tossed?

   (c) Suppose you flip the coin once and it lands heads. What is your updated belief that the coin is biased? What would it be if the coin had landed tails?

   (d) Suppose you flip the coin a second time and it lands tails (after landing heads on the first toss). What is your updated belief that the coin is biased now?

   (e) Suppose you flip the coin n times and observe k heads. Find the expression for the probability that the coin is biased in terms of n and k. (Hint: you will need the binomial distribution to give you the probabilities that you would observe this data under the two different hypotheses, fair and biased.)

8. You are interested in studying the extent to which governments cooperate in imposing international economic sanctions. You acquire the following data:

Outcome	x	$P(x)$
Significant cooperation	4	.11
Modest cooperation	3	.16
Minor cooperation	2	.21
No cooperation	1	.52

   (a) Is this probability distribution discrete or continuous? Draw a graph of the probability distribution and of its c.d.f.

   (b) Find $E(X)$ and $V(X)$. Explain what these values mean. Also calculate $E(7X)$, $E(2 + 3X)$, and $V(2 + 3X)$.

9. As discussed in class, a "uniform distribution on [0, 1]" is the probability distribution for a number drawn at random from the [0, 1] interval. Represented $U[0, 1]$, this uniform distribution has $\mu = .5$ and $\sigma^2 = 1/12$. You can draw a uniformly distributed variable in Stata by the following commands: **set obs 1000** (this creates space for a data set with 1000 observations), **gen x = uniform()**. Try this, listing x and looking at its histogram to check that you have done this right.

   (a) Use the *theoretical* results discussed in class to answer the following: If random variables X and Y are independent and have uniform distributions on [0, 1], then what is $E(X + Y)$? $E(X - Y)$ $var(X + Y)$? $var(X - Y)$? $var(5X)$?

   (b) Draw two such variables in Stata and confirm that these theoretical results are hold for the random variables you draw.

   (c) Use one of your variables to illustrate that in general it need not be true that $E(X^2) = E(X)^2$ or that $E(1/X) = 1/E(X)$. What is the general rule at work here?

10. *Extra credit.* A fair coin will be flipped until it comes up tails, or until it is flipped five times, whichever comes first. If you take the gamble, you will be paid 2^k dollars, where k is the number of coin flips.

    (a) What is the expected value of the gamble (in dollars)? (Hint: Think of the amount you get paid as a random variable, and figure out the probabilities associated with the different amounts to compute the expectation).

    (b) How much would you in fact be willing to pay to play this game? (obviously, this question has no universally "correct" answer)

    (c) Now imagine the coin will be flipped until it comes up tails. What is the expected value of the gamble? How much would in fact be willing to pay?