14.31, Fall 2000 – Problem Set One (Review)
Due: Tuesday, September 19
Clearly indicate your answer. No credit given for hard to find answers.

1. Suppose that a student is preparing to take the SAT exam. Why is her eventual SAT score properly viewed as a random variable?

2. Suppose that the joint probability distribution, p(x,y), of variables X and Y is

	X		
Y	1	2	3
1	0.03	0.06	0.06
2	0.02	0.04	0.04
3	0.09	0.18	0.18
4	0.06	0.12	0.12

Determine
a) the marginal unconditional probability distribution of X and Y.
b) the conditional probability distribution $p(X|Y_i)$ and $p(Y|X_i)$.
c) the conditional expectation of X and Y, $E(X|Y_i)$ and $E(Y|X_i)$.

3. The joint probability p(x,y) of random variables X and Y where X is the first year rate of return (%) expected from project A and Y is the first year rate of return (%) expected from project B is given by

	X			
Y	-10	0	20	30
20	0.27	0.08	0.16	0.00
50	0.00	0.04	0.10	0.35

a) Calculate the expected rate of return on project A, E(X).
b) Calculate the expected rate of return on project B, E(Y).
c) Are the rates of return on the two projects independent?

4. Suppose that the height of men in a population is normally distributed with mean= μ inches and standard deviation $\sigma = 2.5$ inches. A sample of 100 men drawn randomly from this population had an average height of 67 inches.
a) Establish a 95% confidence interval for the mean height in the population as a whole.
b) Could this sample with a mean of 67 have come from a population with a mean value of 69?

5. Let X denote the prison sentence in years, for people convicted of auto theft in a particular state in the United States. Suppose that the pdf of X is given by

$$f(x) = (1/9)x^2\,, 0 < x < 3$$

Use integration to find the expected prison sentence. What is the variance of the random variable X with this density function?

6. Suppose that the rainfall in a certain locality is normally distributed with $E(X) = 30.0$ inches. (This value has been established from a long history of weather data). In recent years, certain climatological changes seem to be affecting, among other things, the annual precipitation. It is hypothesized that in fact the annual rainfall has increased. The variance is assumed to be unknown since the suggested climatological changes may affect the variability of the rainfall, and hence past records on the variance are not meaningful. Assume that the past eight years have yielded the following annual precipitation (inches):
34.1, 33.7, 27.4, 31.1, 30.9, 35.2, 28.4, 32.1
Use a t-test to test $H_0 : \mu = 30.0$ and $H_i : \mu > 30.0$.

7. You have data on a random sample of 300 college students. The students were classified with respect to the size of the high school from which they graduated and with respect to their freshman year grade point average.

	High School		
Record	Small	Medium	Large
Above 2.5 average	18	51	46
Below 2.5 average	42	79	64

Use a Chi-square test to determine if there is any relationship between high school size and freshman grade point average.

8. Suppose that at a large university, college grade point average, GPA, and SAT score, SAT, are related by the conditional expectation E(GPA—SAT)=0.70+ 0.002SAT.
i) Find the expected GPA when SAT=800. Find E(GPA—SAT=1,400). Comment on the difference.
ii) If the average SAT in the university is 1,100, what is the average GPA? (Hint: recall that E[E(Y—X)]=E(Y).)

9. In some years some mutual funds will outperform the market (the return from holding their shares is higher than the return from holding a portfolio such as the S&P 500). Suppose that you examine the performance of 4,170 funds over a ten year period. If performance relative to the market is random then each fund has a 50-50 change of outperforming the market in any year and its performance is independent from year to year.

i) If performance relative to the market is random, what is the probability that any particular fund outperforms the market in all 10 years?
ii) Find the probability that at least one fund out of 4,170 outperforms the market in all 10 years. What do you make of your answer?

10. For a randomly selected county in the United States, let X represent the proportion of adults over age 65 who are employed, or the elderly employment rate. Then, X is restricted a value between 0 and 1. Suppose that the cumulative distribution function for X is given by $F(x) = 3x^2 - 2x^3$ for $0 \leq x \leq 1$. find the probability that the elderly employment rate is at least 0.6 (60%).

11. Suppose that the salary new graduates measured in *thousands* of dollars is 32.3, with a standard deviation of 14.6. What are the mean and standard deviation when salary is measured in dollars.

12. Let Y_1, Y_2, Y_3, and Y_4 be independent, identically distributed random variables from a population with mean μ and variance σ^2. Let $\bar{Y} = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$ denote the average of these four random variables.
i) What are the expected value and variance of $\bar{Y}$ in terms of μ and σ^2?
ii) Now, consider a different estimator of μ:

$$W = \frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4 \,. \tag{1}$$

This is a weighted average. Show that W is also an unbiased estimator of μ. Find the variance of W.
iii) Based on your answers to i) and ii) which estimator of μ do you prefer, $\bar{Y}$ or W?
iv) Now consider a more general estimator of μ, defined by

$$W_a = a_1 Y_1 + a_2 Y_2 + a_3 Y_3 + a_4 Y_4 \,, \tag{2}$$

where the a_i are constants. What condition is needed on the a_i for W_a to be an unbiased estimator of μ.
v) Compute the variance of the estimator W_a from iv).

13. You are hired by the governor to study whether a tax on liquor has decreased average liquor consumption in your state. You are able to obtain, for a sample of individuals selected at random, the difference in liquor consumption (in ounces) for the years before and after the tax. For person i who is sampled randomly from the population, Y_i denotes the change in liquor consumption. Treat these as a random variable from a Normal(μ,σ^2) distribution.
i) The null hypothesis is that there is no change in average liquor consumption. State this

formally in terms of μ.
ii) The alternative is that there was a decline in liquor consumption; state the alternative in terms of μ.
iii) Now, suppose that your sample size is n=900 and you obtain the estimates $\bar{y} = -32.8$ and $s = 466.4$. Calculate the t statistic for testing H_0 against H_1. Do you reject H_O at the 5% level? at the 1% level? (Note: sample size is large so you can use the standard normal distribution.)
iv) Would you say that the estimated fall in consumption is large in magnitude? comment on the practical vs statistical significance of this estimate.
v) What has been implicitly assumed in your analysis about other determinants of liquor consumption over the 2 year period in order to infer causality from the tax change to liquor consumption?

14. The new management at a bakery claims that workers are now more productive than they were under old management, which is why wages have "generally increased." Let W_i^b be Worker i's wage under the old management and let W_i^a be Worker i's wage after the change. The difference is $D_i = W_i^a - W_i^b$. Assume that the D_i are a random sample from a Normal(μ,σ^2) distribution.
i) Using the following data on 15 workers, construct an exact 95% confidence interval for μ.
ii) Formally state the hypothesis that there has been no change in average wages. In particular, what is $E(D_i)$ under H_0? If you are hired to examine the validity of the new management's claim, what is the relevant alternative hypothesis in terms of $\mu = E(D_i)$?
iii) Test the null hypothesis from ii) against the stated alternative at the 5% and 1% levels.

Worker	Wages Before	Wages After
1	8.30	9.25
2	9.40	9.00
3	9.00	9.25
4	10.50	10.00
5	11.40	12.00
6	8.75	9.50
7	10.00	10.25
8	9.50	9.50
9	10.80	11.50
10	12.55	13.10
11	12.00	11.50
12	8.65	9.00
13	7.75	7.75
14	11.25	11.50
15	12.65	13.00