Statistics 23, Section 1, Final Exam
Tuesday, December 14, 1999

Name: ____

Pledge: I have neither given nor received aid on this examination.

Signature: ____

Instructions: Show all work, but do not do hard arithmetic (an answer like $8.3+\frac{7}{\sqrt{5.1}}$ is fine).

1. An experiment results in one of three mutually exclusive outcomes, A, B or C. It is known that $P(A)=0.2$, $P(B)=0.4$ and $P(C)=0.3$.

   a. Find $P(A\ or\ B)$

   b. Find $P(B\ and\ C)$

   c. Find $P(A|C)$

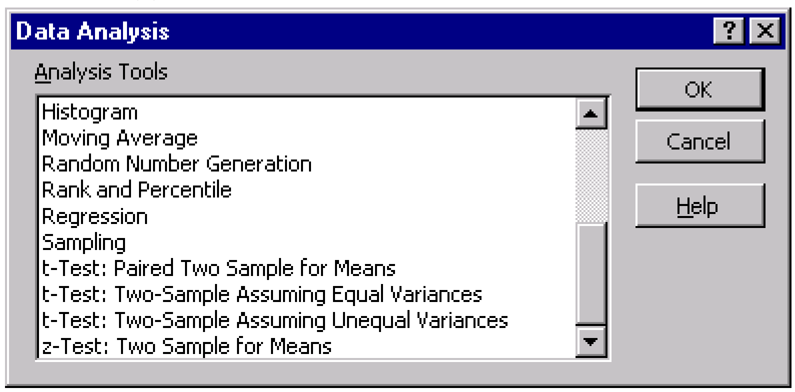
   d. Are A and B independent? Why or why not?

2. To study the effects of competition on cable television rates, 4 counties were selected at random, and their rates before and after the introduction of competition were put in an Excel spreadsheet as:

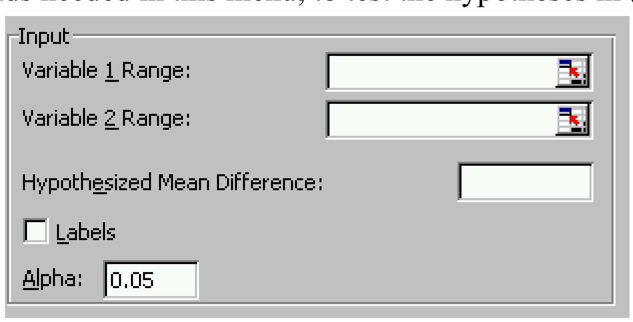
	A	B	C
1	County	Rate Before:	Rate After:
2	A	$21.35	$21.56
3	B	$25.73	$25.91
4	C	$18.92	$19.32
5	D	$22.07	$22.35

a. Formulate hypotheses to test whether there is a significant difference between rates before and after the introduction of competition.

b. Indicate, on the following menu, which Excel Data Analysis Tool you would use to test the hypotheses in (a)



c. Fill out the fields needed in this menu, to test the hypotheses in (a):



d. What assumptions are needed to use the above methods?

e. Here are two possible outputs from the appropriate Excel Data Analysis tool (one is right, the other is wrong):

t-Test: Paired Two Sample for Means

	Variable 1	Variable 2
Mean	22.0175	22.285
Variance	7.941825	7.487233
Observations	4	4
df	3	
t Stat	-5.47220925	
P(T<=t) one-tail	0.00599872	
t Critical one-tail	2.35336302	
P(T<=t) two-tail	0.01199744	
t Critical two-tail	3.18244929	

t-Test: Two-Sample Assuming Equal V

	Variable 1	Variable 2
Mean	22.0175	22.285
Variance	7.941825	7.487233
Observations	4	4
df	6	
t Stat	-0.1362	
P(T<=t) one-tail	0.448058	
t Critical one-tail	1.943181	
P(T<=t) two-tail	0.896116	
t Critical two-tail	2.446914	

Choose the one that you think is right, and use it to give a p-value, with both yes-no $(\alpha = 0.01)$ and gray level interpretations, to test the hypotheses in (a).

f. Now focus only on the rates after the introduction of competition. Write an Excel formula for how large a sample size would be needed to estimate the mean of this population with an accuracy of 0.2, 90% of the time. (Hint: you can get needed information out of the tables in (e) above).

g. If Excel gives the numerical answer of 537.173407 to (f), what would you round it to?

3. 50% of all schools subscribe to CCN. Of these subscribers, 10% actually never use CCN, while 20% use CCN at least 5 times per week.

a. Find the probability that a randomly selected school subscribes to CCN, but never uses it.

b. If a school is selected randomly from the subscribers, find the probability that it uses CCN less than 5 times per week.

c. Find the probability that a randomly selected school never uses CCN (hint: this includes those that don't subscribe, and also those that subscribe, but don't use it).

d. If a randomly selected school doesn't use CCN, what is the probability they have subscribed to it?

4. For the probability distribution:

x	0	1	2	3	4
$f(x)$	0.1	0.2	0.4	0.2	0.1

a. Find $P\{X \geq 1 \mid X < 3\}$.

b. Why is $EX = 2$?

c. Write down a calculation which shows that $\operatorname{var}(X) = 1.2$.

d. What is the standard deviation of X ?

e. For $g(x) = (x-2)^2$, what is $Eg(X)$?

5. To evaluate the accuracy of a Metlar scale, an item whose weight is known to be 14.01 ounces is weighed five times. The weights are entered in an Excel spreadsheet as shown here:

	H
37	14.04
38	14.01
39	13.99
40	14.03
41	14.02

a. Write an Excel formula to calculate a p-value to test whether there is a statistically significant difference between the average value and 14.01.

b. Interpret the result, if the answer to (a) is $p - val = 0.405$.

c. What assumptions are needed to work parts (a) and (b)?

d. Write an Excel formula to give the endpoints of a 98% confidence interval for the mean.

6. a. A list of 101 numbers has $\bar{x} = 2$ and $s = 2$. Find $\sum_{i=1}^{n} x_i$ and $\sum_{i=1}^{n} x_i^2$.

b. Can a list of 10 numbers have $\sum_{i=1}^{n} x_i = 20$ and $\sum_{i=1}^{n} x_i^2 = 30$? Why or why not?

7. In a random sample of 500 small business operators, 23% were motivated by a desire to be their own boss.

a. Write down the steps you would use to check whether the sample size is large enough to use the Normal approximation for confidence intervals and hypothesis tests (but don't actually check).

b. Suppose that the steps in (a) revealed that the Normal Approximation *is not* satisfactory. Write an Excel formula to calculate a p-value to determine whether the population percentage of all small business operators, who are motivated by a desire to be their own boss, is significantly more than 20%.

c. Suppose that the steps in (a) revealed that the Normal Approximation *is* satisfactory. Repeat part (b), using a Normal approximation (with continuity correction).

d. Suppose that the steps in (a) revealed that the Normal Approximation *is* satisfactory. Write an Excel formula to give the endpoints of a level 90% conservative confidence interval for the population proportion of all small business operators, who are motivated by a desire to be their own boss.

e. Suppose that the steps in (a) revealed that the Normal Approximation *is* satisfactory. Write an Excel formula to calculate how large (use the “best guess” method) a sample size should be used to estimate the population proportion so that the accuracy is within 0.005, with probability 0.98.