Midterm Exam
Psychology 406
Fall 2007

Name: ______________________________

Whenever possible, use fractions throughout.

I. ELISA tests are used to screen donated blood for the presence of the AIDS (HIV) virus. The test actually detects antibodies, substances that the body produces when the virus is present. When antibodies are present, ELISA is positive with probability 0.997 and negative with probability 0.003. When the blood tested is not contaminated with AIDS antibodies, ELISA gives a positive result with probability 0.015 and a negative result with probability 0.985. Suppose that 1% of a large population carries the AIDS antibody in their blood.

The following results may be useful in answering the questions below:

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

(1) What is the probability that the ELISA test for AIDS is positive for a randomly chosen person from this population?

(2) What is the probability that a person has the antibody given that the ELISA test is positive (for a randomly chosen person from this population)?

(3) What is the probability that a person doesn't have the antibody given that the ELISA test is positive (for a randomly chosen person from this population)?

(4) What is the joint probability that a person doesn't have the antibody and the ELISA test is positive (for a randomly chosen person from this population)?

II. Suppose I have a set of observations, $X_1, \ldots, X_N$, and I turn them into Z-scores, $Z_1, \ldots, Z_N$. If I carry out a further transformation to $T_i = 10Z_i + 50$, what is the resulting mean and variance of $T_1, \ldots, T_N$?

Besides just giving the answers, demonstrate how you got them using our rules for summations (and how I might have done this in class).

III. Suppose X is a random variable that takes on the following values with the probabilities listed:

P(X = -1) = 10/32; P(X = 0) = 20/32; P(X = 1) = 1/32; P(X = 2) = 1/32

a) E(X):

b) Var(X):

c) Mode:

d) Median:

e) E(X^2):

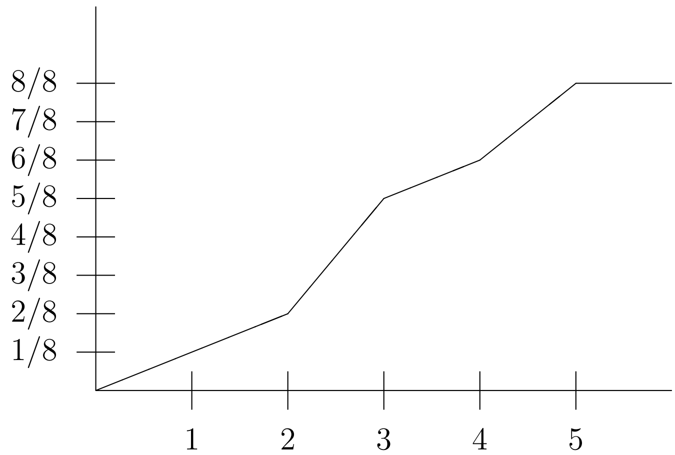
f) Standard deviation of X:

g) P($X \geq 0$):

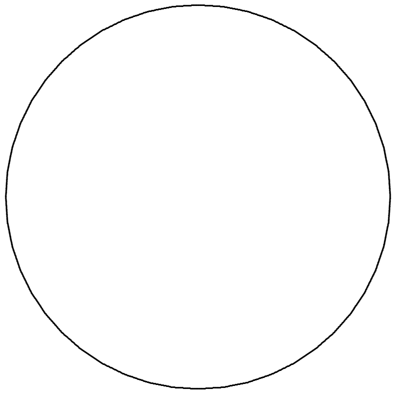
h) E(-3X - 7):

i) Var(-3X - 7):

j) X represents a "fair" game (yes or no):



IV. Let X be a continuous random variable with the cumulative distribution function given above.



a) Calibrate the pointer above to correspond to the cumulative distribution function.

b) P$(X = 2)$:

c) P$(3 \leq X \leq 5)$:

V. As part of an article that appeared in the *NY Times* (10/9/2003) (which is attached), and after Arnie won the California recall, a section is given on "How the Poll was Conducted" beginning on the second page. Explain and show what the section means, in terms of formulas you know, about the margin of error in estimating population proportions. Does the plus or minus two points require any rounding up to two decimal places?

VI. On a particular standardized reading test, the mean score for 50 low-achieving students was 61.3 with a standard deviation of 10.2. According to the publisher's manual, the mean score in the general population should be 70. For now, assume that the population from which the low-achieving students were drawn can be represented by a random variable X that is N(μ,σ^2). If necessary, use interpolation in the tables to obtain the appropriate t-value for use (and show how you did this).

(a) To assess whether the population mean for the low-achieving group might differ from that of the general population used to standardize the reading test, carry out a test of $H_o : \mu = 70$ versus $H_1 : \mu < 70$ at a fixed alpha level of .01.

(b) Construct a 99% confidence interval for the unknown mean μ for the low-achieving population.

(c) Assuming that the standard deviation in the low-achieving population is a known value of 10.2, what sample size would be required to have a 99% confidence interval be 4 points in total length? Show your work.

VII. In a taste test of two versions of cola (e.g., pepsi versus coke), I asked 6 people independently which they prefer (blindfolded and randomly deciding on each trial whether pepsi or coke is given first to the person to taste). Suppose the preference for pepsi is represented by a random variable Y taking on a value of 1 and the preference for coke represented by the random variable Y taking on a value of 0.

We let $\hat{p}$ denote the observed proportion of the 6 people who prefer pepsi to coke, and assume in the population from which these people are drawn that $\mathrm{P}(Y = 1) = 1/2$.

(a) Find the probability distribution for $\hat{p}$:

Values of $\hat{p}$ Probability

(b) $\mathrm{E}(\hat{p}) =$

(c) $\mathrm{Var}(\hat{p}) =$

VIII. I have sampled 10 hyperactive children and have blood pressure scores before and after the administration of a tranquilizer. The data are as follows:

Subject	Before	After
1	155	110
2	160	120
3	140	145
4	145	125
5	155	145
6	160	115
7	150	120
8	175	155
9	170	145
10	165	140

Suppose μ_B and μ_A refer to the population means for the before and after scores, respectively.

(a) Using the usual t-test procedure for dependent samples, carry out a test of $H_o : \mu_B - \mu_A = 0$ versus $H_1 : \mu_B - \mu_A > 0$ at a fixed alpha level of .01.

(b) Construct a 95% confidence interval on $\mu_B - \mu_A$.

(c) Using the sign test, state the hypothesis you would now test that would be analogous to what as carried out in (a). Provide the exact p-value.

IX. I am interested in the effects of a certain chemical that is typically sprayed on fruit to preserve it during transit, and decide to carry out an experiment on maze learning behavior in rats to see if the chemical might have any noticeable influence. Given the 36 rats I have available, 18 are randomly assigned to a "chemical additive" diet and 18 to a regular diet. Some of the data follow, where the dependent variable is the number of trials it takes a rat to make five consecutive successful runs (no rat was given more than 20 trials). Use the summary statistics to answer the two questions, (a) and (b).

Chemical Additive Diet		Regular Diet	
7	20	8	9
8	9	9	13
12	14	10	20
20	16	14	18
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Sum:	270	Sum:	222
Sum of Squares:	4430	Sum of Squares:	2950

The assumptions are made that the data in the (independent) groups I and II came from a $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively.

(a) Give the 95% confidence interval on $\mu_1 - \mu_2$ using the approximate degrees-of-freedom calculated from the data. If necessary, use interpolation in the tables to obtain the appropriate t-value for use (and show how you did this).

(b) Carry out a test of $H_o : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$, using the "pooled" version of the t-test at a fixed alpha level of .05. Again, if necessary find the appropriate t-value for use by interpolation (and show how you did this).

X. Completion

a) The sampling distribution for the number of successes in n independent observations on a population characterized by a dichotomous outcome of success/failure, is called the ____________________.

b) If I have 6 individuals and need to form a committee of 3 from this group, how many different ways could this be done?

____

c) The sample proportion $\hat{p}$, since $\mathrm{E}(\hat{p}) = p$, is said to be a(n) ____________________ estimator of p.

d) An estimator $\hat{\theta}$ is said to be ____________________ if $\hat{\theta}$ converges to θ asympototically.

e) The robustness of several hypothesis testing procedures is based on the fact that sample means are approximately normal when simple random samples are taken and the sample size gets large. This result is called the ____________________.

f) Generally, and other things being equal, power decreases both as the sample size ____________________, and as the alternative hypothesis value being considered gets ____________________ the value specified under H_o.

g) One-tailed hypothesis tests are also referred to as ____________________ hypothesis tests.

h) In calculating power for a specific value within the alternative hypothesis, 1.0 minus the power is called the ____________________.

i) The estimated degrees-of-freedom for a two-independent samples t-test with 10 observations in one group and 15 in the second must fall between the values of ____________________ and ____________________.

j) The assumption in the "pooled" two-independent sample t-test that $\sigma_1^2 = \sigma_2^2$ appears to not be very important (i.e., we have robustness) as long as ____________________.

k) 5 "factorial" is equal to ____________________.

l) If $X_1 \sim N(0, 5)$ and $Y_1 \sim N(1, 2)$ and X_1 and Y_1 are independent, then $3X_1 - 4Y_1$ is ____________________.

m) 95% confidence intervals are "naturally connected" with fixed α-level two-tailed tests, where α is equal to ____________________.

n) If the ____________________ is as small or smaller than α, we then typically say that the data are statistically significant at level α.

o) The standard error of a statistic is also called the ____________________ of the statistic.

p) The variance of a t-distribution with k degrees- of-freedom converges to ____________________ as k goes to infinity.

q) A continuity correction is sometimes used when the probabilities from a binomial distribution are approximated by probabilities from a(n) ____________________.

XI. I have a test statistic T that has the following distribution under H_0 and H_1:

T	Under H_0	Under H_1
1	1/8	2/8
2	2/8	2/8
3	2/8	2/8
4	2/8	1/8
5	1/8	1/8

If you have a decision rule that says to reject H_0 if $T = 1$, 2, or 3, find:

a) Type I error:

b) Type II error:

c) Power: