Statistics Winter 2007 - Final Exam

NAME: ____

You have 3 hours.
Like the quizes, it is open book, you can use a calculator, but no computer.

The contents of this exam are confidential until March 18, 2007.
Do not discuss the exam with anyone!!!!

I pledge my honor that I have not violated the Honor Code during this examination.

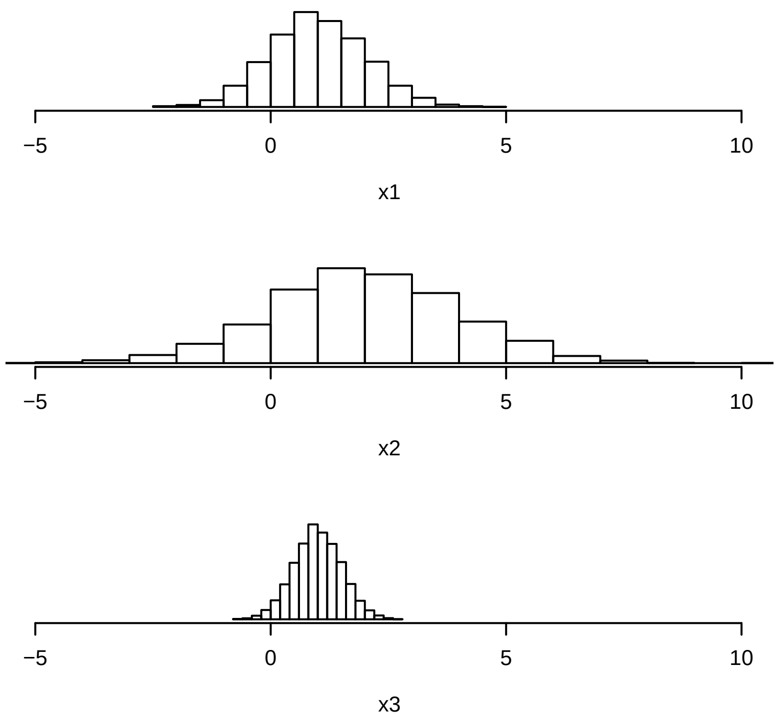
SIGNATURE: ____

There are 11 questions.
Each part of each question is worth two points, except the true and false (question 9) where each part is worth one point.

question 1, 5 parts, 10 points: _____
question 2, 6 parts, 12 points: _____
question 3, 10 parts, 20 points: _____
question 4, 8 parts, 16 points: _____
question 5, 5 parts, 10 points: _____
question 6, 4 parts, 8 points: _____
question 7, 4 parts, 8 points: _____
question 8, 9 parts, 18 points: _____
question 9, 20 parts, 20 points: _____
question 10, 10 parts, 20 points: _____
question 11, 3 parts, 6 points: _____

total, 148 points _____

Question 1



The above histograms depict the numeric values of variables x_1, x_2, and x_3.
Both x_1 and x_2 have sample mean and standard deviation equal to either 1 or 2.

(a) What is the sample mean of x_1?

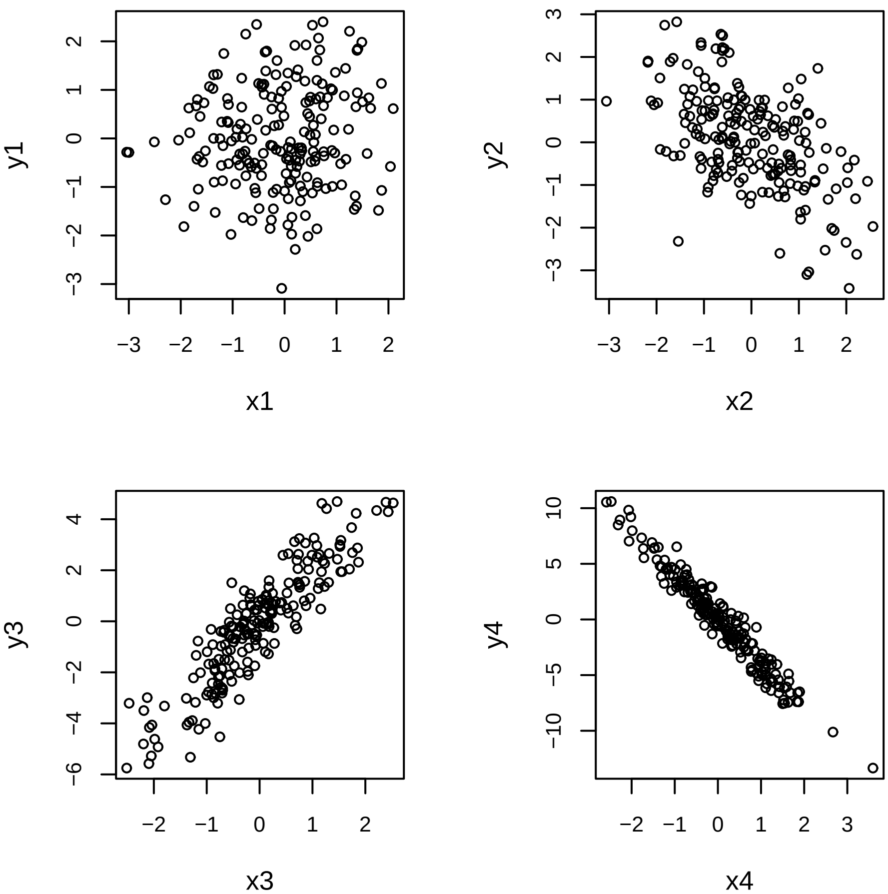
(b) What is the sample standard deviation of x_2 ?

(c) Give an interval which should include roughly 95% of the x_1 values.

(d) Guess a reasonable value for the variance of x_3 ?

(e) In part (c), what assumptions are you making? Do they seem reasonable?

Question 2



The four correlations corresponding to the plots above are -.54, -.97, .11, and .90 (*not given in correct order !!*).

(a) The correlation between x_1 and y_1 is __________ .

(b) The correlation between x_2 and y_2 is __________ .

(c) The correlation between x_3 and y_3 is __________ .

(d) The correlation between x_4 and y_4 is __________ .

(e) The variances of x_3 and y_3 are both either about 1 or 5.
What is the variance of y_3?

(f) What is the covariance between x_3 and y_3 ?

Question 3

The table below gives the joint distribution of the random variables S and F.

For a randomly chosen person from the population:
S=1 if the person reports that they "specially choose to watch the Simpson's"
S=2 if the person reports that they watch the Simpson's if "nothing better is on"
S=3 if the person report that the don't watch the Simpson's.

F means the same thing for the tv show Friends.

		S		
		1	2	3
	1	.15	.02	.08
F	2	.03	.01	.02
	3	.12	.03	.54

(a) What is the joint probability $P(S = 3, F = 3)$?

(b) What is the marginal distribution of F (remember, we want the values and probs for each value) ?

(c) What is the conditional probability $P(S = 3 \mid F = 3)$?

(d) Are F and S identically distributed?

(e) Are F and S independent ?

(f) What is $E(S)$, the expected value of S ?

(g) What is the expected value of S given $F = 3$?

(h) Which is bigger, the variance of F, or the variance of S ?

(i) What is the sign of the covariance between F and S ?

(j) What is the probability that a randomly chosen person does not watch the Simpson's *or* does not watch Friends ?
That is, what is $P(S = 3 \;\; or \;\; F = 3)$.

Question 4

Suppose we are uncertain about the return on an asset and feel
comfortable describing our uncertainty by $R \sim N(.1, .01)$, where R denotes the return.

(a) What is the standard deviation of R ?

(b) Give an interval such that there is a 95% probability that R will end up in it.

(c) What is the probability the return will be negative ?

(d) Suppose there is an asset which gives a certain return of .02.
You form a portfolio which puts 60% of your wealth into the risk-free asset and 40% into R.
Write the return on your portfolio as a linear function of R.

(e) What is the expected value of the portfolio in part (d) ?

(f) What is the distribution of your end of period wealth if you invest 50 dollars
in the porfolio in part(d) ?

Now suppose we are comfortable assuming the returns R_t (return on the asset in future period t) will be iid $N(.1, .01)$ for the forseeable future.

(g) Using the central limit theorem, what is the normal distribution which describes the number of positive returns we will get in the next 100 periods.

(h) Using the central limit theorem, what is the normal distribution which describes the fraction of positive returns we will get in the next 100 periods.

Question 5

Supppose $R_1 \sim N(.2, .04)$ and $R_2 \sim N(.15, .01)$.
The correlation between R_1 and R_2 is .8.

Let $P = .2\,R_1 + .8\,R_2$.

(a) What is $E(P)$?

(b) What is the covariance between R_1 and R_2 ?

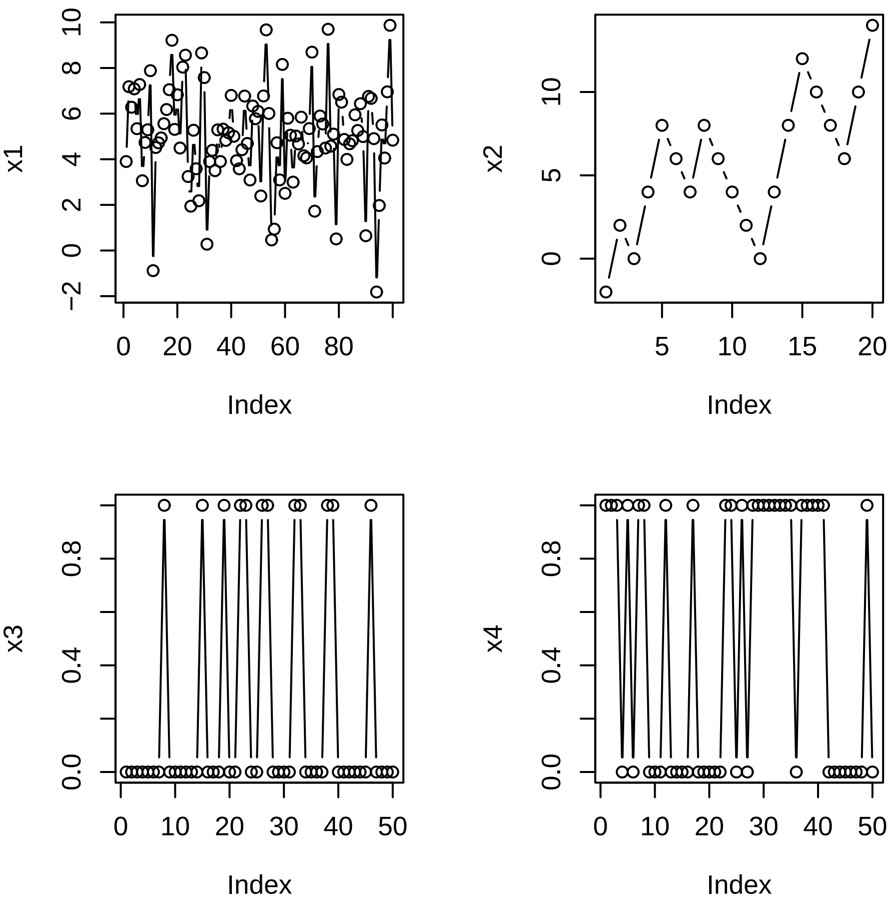
(c) What is the variance of P ?

Suppose $R_3 \sim N(.3, .09)$, $\mathrm{cov}(R_1, R_3) = .048$, $\mathrm{cov}(R_2, R_3) = 0$.

(d) What is the expected value of $Q = .5\,P + .5\,R_3$?

(e) What is the variance of Q?

Question 6



For each of the four series plotted above, eye-ball a probabilistic model for the process.
You will have to eye-ball parameter values from the graphs. Anything "reasonable" is acceptable.
For example, if you say $x_1 \sim$ Bernoulli(.5), iid, that would be of the correct form for an answer, but deemed unreasonable.

(a) Specify a probability model for x_1.

(b) Specify a probability model for x_2.

(c) Specify a probability model for x_3.

(d) Specify a probability model for x_4.

Question 7

In a recent poll, 1000 voters were randomly selected.
700 responded yes when given the choice yes or no to the question:

> "Do you have deep concerns about the future of the country?".

(a) What is the unbiased estimate of the true proportion of voters that would answer yes ?

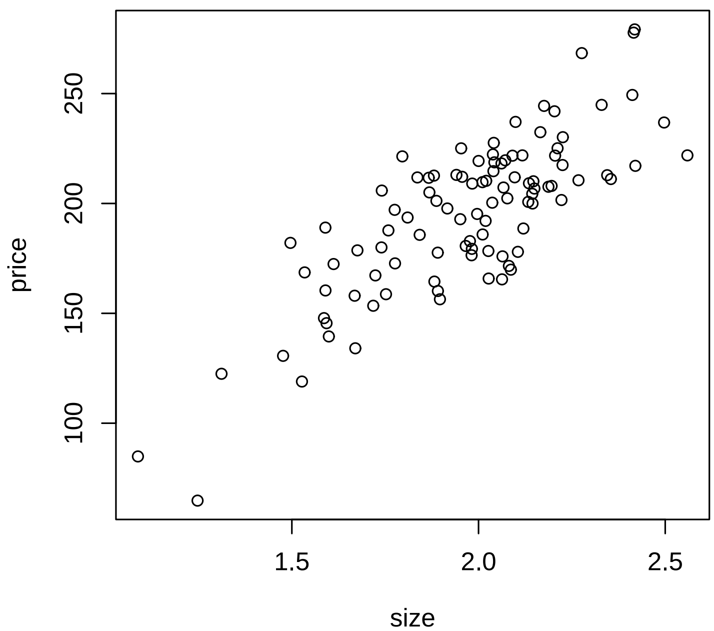
(b) Give an (approximate) 95% confidence interval for the true proportion of voters that would answer yes.

(c) Test the null-hypothesis that the true proportion is equal to .5 at level 5%.

(d) For the test in part (c), what, approximately, is the corresponding p-value?

Question 8

Below is the plot of size (thousands of square feet) vs. price (thousands of dollars) for 100 houses randomly selected from a certain neighborhood.



Here is the regression output. It is from the package "R", but you should be able to guess what everything is.

```
Call:
lm(formula = price ~ size, data = data.frame(price, size))

Residuals:
    Min      1Q  Median      3Q     Max
-54.494 -14.438  -2.013  14.933  44.479

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -11.922     15.805  -0.754    0.452
size         105.165      7.954  13.221   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 21.39 on 98 degrees of freedom
Multiple R-Squared: 0.6408,Adjusted R-squared: 0.6371
F-statistic: 174.8 on 1 and 98 DF,  p-value: < 2.2e-16
```

(a) Does it look like the “size - price process” can reasonably be thought about using the simple linear regression model?

(b) Which number on the output is the estimate of σ ?

(c) Give the 95% confidence interval for the true slope.

(d) Test the null-hypothesis that the true slope is equal to 0 at level .05.

(e) Test the null-hypothesis that the true slope is equal to 100 at level .05.

(f) For the tests in (d) and (e) what (approximately) are the corresponding p-values?

(g) Give the plug-in predictive interval for a house with 2.5 thousand square feet.

(h) What is the correlation between size and price ?

(i) What is the correlation between the fitted values and price?

Question 9

True or False. For each part, circle either T or F.

(1) The p-value is the probability that the null-hypothesis is true.

T F

(2) X and Y are independent if the joint is the product of the marginal times the conditional.

T F

(3) For a fair coin, the probability of 10 heads in a row is $.5^{10}$.

T F

(4) The covariance between two random variables cannot be bigger than the product of the standard deviations.

T F

(5) The estimated slope in the regression of y on x can be positive when the sample correlation between y and x is zero.

T F

(6) If you add 7 to each entry in a list, that adds 49 to the sample variance.

T F

(7) The p-value is useful because it gives us the probability of rejecting the null when the alternative is true.

T F

(8) By forming a two-asset portfolio using positive weights we can hope to obtain a sample mean which is strictly smaller than the two sample means of the returns on the original assets.

T F

(9) If $X = X_1 + X_2 + \cdots + X_n$ for $X_i \sim \text{Bernoulli}(p), \text{iid}$, then $E(X^2) = np(1-p) + n^2p^2$.

T F

(10) From the joint distribution (table) of two discrete random variables we can *always* recover the two marginal distributions.

T F

(11) From the marginal distributions of two discrete random variables we can *always* recover their joint distribution.

T F

(12) If $\sum_{i=1}^{n} X_i$ is the sum of n independent random variables whose common variance is σ^2, then the variance of the sum is $\frac{\sigma^2}{n}$.

T F

(13) The expected value of the average of n identically distributed random variables does not depend on the sample size n.

T F

(14) If $X \sim N(4, 1)$, then $Y = -2X + 3 \sim N(-5, 7)$.

T F

(15) If $Y \sim N(0, 4)$, then $E(Y^2) = 4$.

T F

(16) A large p-value tells us that the null-hypothesis is probably true.

T F

(17) In the random walk model, the next value is independent of the past values given the current value.

T F

(18) The expected value of a random variable is the most likely outcome.

T F

(19) $.25 \geq p(1-p)$ for p between 0 and 1.

T F

(20) Given the sample, there is a 95% chance that the true value is in a 95% confidence interval.

T F

Question 10

Each outcome corresponds to a murder trial.
Let $D = 1$ if the murderer gets the death sentence and 0 otherwise.
Let $M = 1$ if the murderer is white and 0 otherwise.
Let $V = 1$ if the victim is white and 0 otherwise.

Let $p(M = 1) = .5$.
Let $p(V = 1 \mid M = 1) = .8$.
Let $p(V = 1 \mid M = 0) = .2$.
Let $p(D = 1 \mid M = 1, V = 1) = .6$.
Let $p(D = 1 \mid M = 1, V = 0) = .4$.
Let $p(D = 1 \mid M = 0, V = 1) = .6$.
Let $p(D = 1 \mid M = 0, V = 0) = .4$.

These probabilities define the joint distribution of (M, V, D)

HINT: Make the diagram depicting M, $V|M$, $D|V, M$.

(a) Who is more likely to get the death sentence?
That is, what are $p(D = 1 \mid M = 1)$ and $p(D = 1 \mid M = 0)$?

(b) Find the joint distribution of (M, D) (give the two-way table).

(c) Are M and D independent ?

(d) Find the joint distribution of (M, D) *given* $V = 1$ (give the two-way table).

(e) Given $V = 1$, are M and D independent?

(f) Find the joint distribution of (M, D) given $V = 0$ (give the two-way table).

(g) Given $V = 0$, are M and D independent?

(h) Are M and V independent?

(i) Are D and V independent?

(j) How is D related to M and V?
If correct for some population, what do these probabilities
say about the impact of race on whether or not a murder is sentenced to death?

Question 11

Suppose,

$Y = 2X_1 + 3X_2 + \epsilon.$

where X_1, X_2 and ϵ are iid $N(0, 1)$.

(a) What are the mean and variance of Y ?

(b) Given $X_2 = 2$, what are the mean and variance of Y ?

(c) Given $X_2 = 2$ and $X_1 = 1$, what are the mean and variance of Y ?