Statistical Methods
Winter 2004

Text numbers:
Q1) 2.20 Q2) 2.41 Q3) 2.61
Q4) 3.9
Q5) 4.10 Q6) 4.21 Q7) 4.23 Q8) 4.31 Q9) 4.67 Q10) 4.71
Q11) Algebraically, establish the following equalities

$$\Sigma(x_i - c) = \Sigma x_i - nc$$
$$\Sigma(x_i - c)^2 = \Sigma x_i^2 - 2c\Sigma x_i + nc^2$$
$$\Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2/n$$

Q12) Illustrate the formulae in (11) by selecting 20 numbers and performing the calculations in (11) numerically using EXCEL

Q13) Draw a bar chart (what kind should you use?) to see what effect the full moon has on admissions rates into mental health clinics. (Taken from: Blackman, Sheldon and Don Catalina. "The Moon and the Emergency Room." *Perceptual and Motor Skills*, 37 (1973), 624-26.)

Admission Rates (Patients/Day) at Virginia Mental Health Clinic (1971-2)

Month	Before Full Moon	During Full Moon	After Full Moon
August	6.4	5.0	5.8
September	7.1	13.0	9.2
October	6.5	14.0	7.9
November	8.6	12.0	7.7
December	8.1	6.0	11.0
January	10.4	9.0	12.9
February	11.5	13.0	13.5
March	13.8	16.0	13.1
April	15.4	25.0	15.8
May	15.7	13.0	13.3
June	11.7	14.0	12.8
July	15.8	20.0	11.4

Demonic frenzy, moping melancholy
And moon-struck madness. "Lost in Paradise" (Milton)

What other calculations might help you infer whether or not the moon plays games with the mind? What assumptions are you making about admissions at each of the phases (before/during/after)? Do *you* think there is such a thing as the Transylvania affect?

Q14) Read "Filling The World's Belly." The Economist. December 11, 2003. Do the charts used in the article conform to the textbook's guidelines for proper usage of charts?

Statistical Methods
Winter 2004

Assignment # 2 *Due: Feb. 6, by 17:00 (5pm) in mailbox*

Text numbers from chapter 6: 6.7; 6.9; 6.11; 6.13; 6.17; 6.25; 6.45; 6.51; 6.59; 6.80; 6.82.

1A.The Flintstones and the Rubbles are in Rock Vegas vacationing. While Betty and Wilma are off playing the slot machines, Fred and Barney are trying their luck at "seven." The game seven is much like craps, only simpler. Two dice are rolled, and the roller wins if the sum of the digits on the dice equals 7. Draw the sample space and highlight the winning combinations. Assuming that the dice is fair what are the odds of winning?

1B. Now suppose that Betty and Wilma are at a table where "Krazy Joe" is offering the following game:

> A fair coin is flipped until the first head appears. The ladies win 2 "rocks" if it appears on the first toss; 4 "rocks" if it appears on the second toss; 8 "rocks" if it appears on the third toss. In general, they get 2^k "rocks" if the first head appears on the k^{th} toss.

How much do the ladies need to pay – ante – to play this game in order for the game to be fair? (A fair game is one in which the ante and the expected payoff are the same.)

2. How many handshakes do the leaders of the G8 (Canada, France, Germany, Italy, Japan, Russian federation, United Kingdom, and Unites States) countries exchange if each leader shakes hands with every other politician exactly once? (Not withstanding that a "Shawinagin handshake" might not be enjoyed by some of the leaders.)

3. What is the probability that two people in a class of 40 will have the same birthday? Hint: Consider the opposite case – that all 40 people have different birthdays. The probability is much higher than you think. For simplicity assume that birthdays are spread randomly over 365 days, i.e., each person has an equal chance of being born on any given day. Note that the number of ways that k people can have different birthdays is $365(364)\cdots(365\text{-}k+1) = (365)!/(365\text{-}k)!$ Lastly, there are 365^k sequences in the sample space when we have k people.

4. The median, y^M, of a continuous probability distribution is the value for which $\Pr(Y < y^M) = \Pr(Y > y^M) = 0.5$. Thus for a random variable with p.d.f.

$$f_Y(y) = 3y^2 \qquad 0 \le y \le 1$$

find the median, y^M.

* Challenge (Bonus)

A chessboard has the northwest and southeast squares deleted. On the modified deleted board is it possible to perfectly cover the remaining squares with domino tiles? (Domino tiles consist of two squares – each square fitting one square on the chessboard.)

Statistical Methods
Winter 2004

Assignment # 3 ***Due: Feb. 13, by 17:00 (5pm) in mailbox***

I highly recommend that you work in groups; two to four in a group is ideal. If you work in groups you should note the names of your co-authors. In any case, everyone must hand in individually written/typed solutions. Please place your completed HW2 (stapled) in my mailbox in ER level 100 by the date noted above. “Je ne suis pas votre secrétaire.”

Text numbers from chapter 6: 6.77; 6.87; 6.89.
Text numbers from chapter 7: 7.38; 7.55; 7.105; 7.109; 7.125; 7.127.
Text numbers from chapter 8: 8.5; 8.14; 8.36; 8.45; 8.57; 8.94.

1. Show that the Poisson distribution

$$\frac{e^{-\mu}\mu^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots$$

is the limiting distribution (i.e., as n approaches infinity) of the binomial distribution

$$C_x^n p^x (1-p)^{n-x} \qquad \text{where } C_x^n \equiv \binom{n}{x} \equiv \frac{n!}{(n-x)!\,x!}$$

Hint: Let $np = \mu$ so that $p = \mu/n$ and note that $\lim_{n\to\infty}(1 - y/n)^n = e^{-y}$

2. In the 1983 NBA draft two teams were competing to get the first overall draft pick. The procedure to determine which of the two teams would get the first overall pick was determined by a series of coin flips. The first coin toss was to determine who would call the second. The second would be to decide who had the first overall pick. Houston the Rockets won the first toss and then correctly called “heads” on the second call. In the end, Houston used its first pick to draft Ralph Sampson, who by all accounts, never lived up to his billing. In any case, was the procedure to determine the draft fair? More importantly, from the perspective of a statistician, did the procedure make sense?

3. Evaluate the following comment:

> According to the CIA Factbook GDP per capita in the United States in 2002 was \$36,300 while that of Canada was only \$29,300 (both expressed in US dollars). Thus clearly the average American is much wealthier (in money terms) than the average Canadian.

4. Let A and B be any two events defined in S. Using Venn diagrams show that

(i) $(A \cap B)^C = A^C \cup B^C$
(ii) $(A \cup B)^C = A^C \cap B^C$

These rules are known as DeMorgan’s Laws.

5. Some have remarked that the most primal nature of man is to engage in war. Between 1500 and 1931 war broke out some where in the world 299 times (at least those that historians are able to verify). Fill in the values for expected frequency using a Poisson distribution, where μ = expected number of wars in a given year (treat 4+ as simply 4).

OUTBREAK OF WAR

Number of Wars	*Observed Frequency*	*Expected Frequency*
0	223	?
1	142	?
2	48	?
3	15	?
4+	4	?

Challenge (Bonus)

You have probably often heard the assertion that there are infinitely many prime numbers. (If you haven't, now you have.) Indeed, every now and then a person working in a math lab with a really powerful computer finds a "new" prime number and people pop champagne and jumps for joy. In any case, the finding shouldn't be surprising given that there are infinitely many of them; nonetheless, it is still a discovery to behold. Can you show formally that there are an infinite number of prime numbers?

Statistical Methods

Winter 2004

Assignment # 4 *Due: March 16, by 17:00 (5pm) in mailbox*

I highly recommend that you work in groups; two to four in a group is ideal. If you work in groups you should note the names of your co-authors. In any case, everyone must hand in individually written/typed solutions. Please place your completed HW2 (stapled) in my mailbox in ER level 100 by the date noted above. "Je ne suis pas votre secrétaire." Also, as a little bit of advice to you, if you do your homework neatly you are likely to get a better grade – this is just human nature. Your grader will feel less inclined to give you the benefit of the doubt if she has a difficult time reading your work.

The first ten questions are from your textbook.

9.9; 9.11; 9.13; 9.15; 9.17; 9.19; 9.27; 9.29; 9.33; 9.35

I also want you to do some of the problems (of which two are from the practice midterm).

P1. Consider the following probability table below:

	$X = 0$	$X = 1$
$Y = 0$	0.20	0.20
$Y = 1$	0.10	0.25
$Y = 2$	0.15	0.10

(a) Find the expected value of X: E(X).
(b) Find the expected value of X given that Y = 1: E($X|Y$=1).
(c) Find the variance of X: Var(X).

P2. Suppose that the grades for an examination are (approximately) distributed normal with mean 50 and variance 49. That is, $X \sim N(50,7^2)$.

a. The professor wants to raise the class average to 75 by multiplying all test scores by a factor of 3/2. What happens to the variance under this procedure?
b. What if instead she raised the grades by adding a factor of 25 to each test score? That is, how does the variance change under this procedure?
c. With the unadjusted test scores, what cut offs need she make in the raw scores so that only 10 percent get a grade of A and 30 percent a grade of B?

P3. Using the law of expected values, show that Var(X–Y) = Var(X) + Var(Y) – 2Cov(X,Y). That is, use first principles: $\text{Var}(W) \equiv E[W-E(W)]^2 = E(W^2)-[E(W)]^2$ and $\text{Cov}(X,Y) \equiv E(XY) - E(X)E(Y)$.

Statistical Methods

Winter 2004

Assignment # 5 *Due: March 23, by 17:00 (5pm) in mailbox*

I highly recommend that you work in groups; two to four in a group is ideal. If you work in groups you should note the names of your co-authors. In any case, everyone must hand in individually written/typed solutions. Please place your completed HW2 (stapled) in my mailbox in ER level 100 by the date noted above. "Je ne suis pas votre secrétaire." Also, as a little bit of advice to you, if you do your homework neatly you are likely to get a better grade – this is just human nature. Your grader will feel less inclined to give you the benefit of the doubt if she has a difficult time reading your work.

The first eight questions are from your textbook.

10.21; 10.23; 10.25; 10.27; 10.31; 10.35; 10.47; 10.49.

Also please do the following two questions below.

1. Consider the following probability table below:

	$X = 0$	$X = 1$
$Y = 0$	0.20	0.20
$Y = 1$	0.10	0.25
$Y = 2$	0.15	0.10

(d) Find the expected value of X^2: $E(X^2)$.
(e) Find $\text{Cov}(X,Y)$
(f) Find $\text{Var}(3X–2Y)$

2. You are trying to estimate the mean weight of Canadian females. You know that weight is distributed normally: $X \sim N(\mu,10^2)$. In five different samples (each of sample size 100) you get the following sample mean values (in kg): 53.3, 59.4, 57.4, 58.3, 59.7. How likely is it that one of your 95 percent confidence intervals, based on your sample means above, does not cover the true mean?

Challenge: Suppose that you are at a wedding banquet and you want to estimate the size of the gathering. You have no clue how big the banquet hall is, but you know that the bride and groom are having a series of raffles for some really cheesy prizes (except, of course, that ever indispensable *Ronco Rotisserie* – "Just set it, and forget it!"). The raffle tickets are numbered sequentially from 1 through N, where N is the number of guests who have shown up for the festivities. Suppose that the following raffle ticket numbers have been drawn for a series of the prizes: 312, 35, 112, 54, 205, 182, 89, 223. (Don't worry, the *Ronco Rotisserie* is still available!) Can you come up with an unbiased estimate for the number of people in the party?

Statistical Methods

Winter 2004

Assignment # 6 *Due: April 5, by 17:00 (5pm) in mailbox*

I highly recommend that you work in groups; two to four in a group is ideal. If you work in groups you should note the names of your co-authors. In any case, everyone must hand in individually written/typed solutions. Please place your completed HW2 (stapled) in my mailbox in ER level 100 by the date noted above. "Je ne suis pas votre secrétaire." Also, as a little bit of advice to you, if you do your homework neatly you are likely to get a better grade – this is just human nature. Your grader will feel less inclined to give you the benefit of the doubt if she has a difficult time reading your work.

The first eight questions are from your textbook.

Textbook: *11.39; 11.41; 11.43; 11.45; 11.57; 11.61; 12.32; 12.33*

Also please do the following two questions below.

1. As you know from pg. 367 of your textbook: $\chi^2_{n-1} = (n-1)s^2/\sigma^2$, where n-1 are the degrees of freedom of the chi-square random variable. Armed with this piece of knowledge find $E(\chi^2_{n-1})$.

2. A random sample of size 2 is drawn from a uniform pdf defined over the interval $[0,\theta]$. We wish to test

$$H_0: \theta = 2$$

versus

$$H_1: \theta < 2$$

By rejecting H_0 when $y_1 + y_2 \leq k$. Find the value of k that gives a level of significance 0.05.

Challenge: You have thirty black balls and thirty white balls and two urns presented before you. You are to decide how many, and in which urns, to place the black and white balls (you have to use them all). A stranger will then randomly draw a ball from one of the urns. If the ball he draws is black you win a million dollars. Otherwise you get a handshake and a pat on the back. If you want to maximise your chances of winning the million dollars how should you go about allocating the balls between the two urns?