STAT 554: HW #1
due Feb. 17, 2004

I want you to turn in neat easy-to-grade homework solutions. (*Helpful hint*: Papers tend to be easier-to-grade if the answers are correct and easily found.) For some assignments, you are only to write your final answers on an answer sheet that I will supply, and submit the answer sheet to me. For other assignments, you can show all of your work, or you can choose to present condensed solutions where *some* of the grubby work is not shown. In either case, *show adequate supporting work* which is carefully documented, and in no case should you merely just give an answer without some sort of justification, unless the solution is truly trivial. *Be sure to draw boxes around your final answers.* (Marking your answers with a highlighter or a red pen will be greatly appreciated, but is not a requirement). You should choose one answer for each part (i.e., don't hedge by giving two or more answers). It should be noted that when there is an answer sheet, supporting work is not desired unless it is explicitly requested.

You can discuss any of the problems with others, except for the occassional extra credit problems that will be assigned, but you should not turn in any work that is directly copied from someone else, and you are not allowed to turn in any computer output which you did not generate yourself. It will be considered to be a violation of the George Mason University honor code if you turn in work which is not your own.

I will typically give you at least two weeks to do each assignment. If you don't have your paper ready to hand in the Tuesday that it is due, you won't be penalized if you can get it to me by 9 PM the following Saturday. But I won't grade papers received later than that Saturday night deadline unless I haven't started grading the others, and also you should realize that you're taking a risk unless you give me your paper in class on Tuesday or otherwise hand it directly to me. In over sixteen years at GMU, to my knowledge I've never failed to successfully get a paper that someone has placed under my office door or faxed to me, but you should be aware that something might go amiss, and if for some reason I don't get your paper, I can't give you any credit. If you bring your paper by to drop it off for me, the best thing is to put it under my office door (as opposed to my mailbox). If the glass doors of my department prevent you from reaching my office door, then slip it under the glass door that is the *main entrance* to my department. My advice is that you send me an e-mail message to notify me if you fax your paper to me, or if you drop your paper off for me at my office. (The fax number is (703) 993-1700, but many people share the same fax machine, so you should clearly indicate that your fax is for me on a cover sheet.) For a variety of reasons, *don't attempt to send me your homework via e-mail.* Of course, the easiest and safest scheme will be to simply bring your papers to class on the Tuesday that they are due. In addition, I'll stipulate that at the end of the semester, *I won't take late papers after the final exam.*

The maximum score on each homework assignment will range from about 10 to 40 points, but the points for the various parts may total more than the stated maximum score. In such a case you don't have to worry yourself sick if you can't get one or two of the parts. (In return for this generousity on my part, I'll ask that you don't argue with me about the amount of partial credit you receive for an incorrect answer. Of course, you should bring the matter to my attention if I make an obvious mistake like adding up your points incorrectly. But as far as partial credit goes, I'll work very hard to grade everyone's papers consistently, and so it won't be fair if I change the partial credit on your paper unless I make sure that I change the partial credit on everyone else's paper who made the same mistake.)

By the end of the semester you will have been given about 120 points worth of HW problems (not counting the extra credit problems and "bonus points" — if we count those, then it'll be more than 130), and I'll truncate your homework total at 100 points. So, all in all, you can miss quite a few parts and still get a perfect HW score to average into your overall course score. I'll warn you now that a lot of the points will be offered to you towards the end of the semester, and so it may be wise to strive for good scores on the first 3 assignments, even though they won't count as much, so that you'll give yourself a bit of a cushion at the end when you may not have time to do a careful job on all of the problems.

Please follow the following specific instructions pertaining to the presentation of your homework solutions for which you do not submit answers on an answer sheet.

- All homework should be on paper which is approximately 8.5 inches by 11 inches.
- All pages should be *stapled* together in the upper *left* corner. (Paper clips, folders, and binders annoy me.)

- Put your name (first name first, last name last) in the upper right corner of the first page. (It is not necessary to include your ID number.)
- I do *not* want a cover sheet or an executive summary.
- Be sure that your solutions are presented in the proper order (and that all final answers are clearly indicated by drawing boxes around them).
- Unless I specify otherwise, report answers using two significant digits whenever it is not unreasonable to do so. (**Note:** The following numbers have two significant digits: 0.23, 0.023, 0.0023, and 0.00023.)
- Whenever reasonable, unless I specify otherwise, you should report values (possibly rounded) obtained from an exact method (as opposed to using approximations that may result in inaccurate answers). In cases for which it is unreasonable for me to expect you to use an exact method, you should use whatever approximation scheme that I promote in class if I fail to indicate that a specific method be used. If you listen to me carefully, and carefully read material that I supply you with, then there should be little doubt as to how I want you to obtain your answers. But if you're not sure at some point, then you should ask me.
- Under no circumstances should you turn in undocumented computer output. (If the final answer is on the computer output, it should be clearly marked as your answer.)
- With regard to computer output, you can cut out any pertinent computer output and insert each section of output into the most appropriate place of your solution, or you can include 8.5 by 11 pages of documented computer output, as long as **everything is presented in the proper order**. *Do not* direct me to look in an "appendix" of computer output for any of your answers or justification.

Note: For this assignment, parts 1(d), 1(g), 1(i), 1(j), 2(a), 2(b), 2(f), and 3(b) are worth 0 points since I'm supplying you with the answers to them (in order to help you make sure that you understand this assignment). All of the other parts are each worth 1 point. The maximum score for this assignment is 12 (even though there are 14 points worth of problems).

1) When paying roulette, in order for a bet on the outcome black to have a positive expected net profit, the probability of the outcome black must exceed 0.5. This problem deals with hypothesis tests of $H_0 : p \leq 0.5$ against $H_1 : p > 0.5$, where p is the probability of a black outcome, and iid Bernoulli trials are to be assumed. Letting y be the observed number of black outcomes in n trials, consider tests which reject the null hypothesis in favor of the alternative whenever y is sufficiently large.

(a) For the case of $n = 30$, *give the largest upper-tail rejection region which is possible for a level 0.05 test.* (Such a rejection region results in the most powerful nonrandomized level 0.05 test.)

(b) *Give the size of the test corresponding to rejection region called for in part (a).* (*Note*: Obviously, since we are dealing with a level 0.05 test, the size cannot exceed 0.05.)

(c) *Give the power of the test corresponding to rejection region called for in part (a) if* $p = 0.6$. (*Note*: I'll talk about *power* during the third class, but you can read about it ahead of time in Sec. 1.4 of the class notes. The desired power is just the probability of *not* making a Type II error if $p = 0.6$.)

(d) *Give the power of the test corresponding to rejection region called for in part (a) if* $p = 0.7$.

** For parts (e) through (h), repeat parts (a) through (d) using $n = 60$ (instead of $n = 30$). (Note that for a given value of p, the power of the level 0.05 test increases as n is increased.)

** For parts (i) through (l), repeat parts (a) through (d) using $n = 120$ (instead of $n = 30$).

(m) For the case of $n = 60$, *describe the rejection criterion for a randomized size 0.05 upper-tail test.* (*Note*: You can examine my answer to part (n), which is given on the answer sheet for this assignment, to see what type of an answer I'm looking for here.)

(n) For the case of $n = 120$, *describe the rejection criterion for a randomized size 0.05 upper-tail test.*

2) Suppose that you are one of the fortunate few who receive an invitation to Cliff's Casino Night, a gala event where I get out all of my gambling paraphernalia and transform my humble home into a lavishly equipped betting parlor having keno, craps, blackjack, and roulette. If you are a cautious and even only a slightly suspicious person, you should decline the complimentary rotgut drinks and decide to carefully observe my activities.

Since I have noticed in the past that more people tend to bet on red than on black, I have altered my roulette wheel so that red will occur with a smaller probability than would be the case than if it were an "ideal wheel" (thus letting me enjoy an expected profit which exceeds the usual house advantage — if I were to notice that a particular crowd favored betting on black, I can spill food onto the initial roulette wheel and replace it with one which favors red over black). If you observe only 6 outcomes of red in 19 trials, would you be able to claim that there is statistically significant evidence to reject $H_0 : p \geq 9/19$ in favor of $H_1 : p < 9/19$, where p is the probability of observing the outcome red?

(a) Respond to this query by *reporting the p-value which results from the appropriate test.*

Note that there is no need to use the normal approximation for part (a) since Minitab or similar software can be used to obtain an exact answer based on the binomial distribution. Nevertheless, if one does use the approximation, with the continuity correction, you'll see that it does okay, yielding an approximate p-value which matches the exact p-value when both are rounded to three significant digits.

(b) *Approximate the p-value requested in part (a) using the normal approximation with the continuity correction.*

(c) *Approximate the p-value using the normal approximation without the continuity correction*, and note that this approximation isn't as good in this case (not even matching the exact value when both are rounded to the nearest hundredth). (*Note*: Approximations are commonly used in statistics, but it's often important to use the best possible approximations when an exact computation isn't feasible. It's also important to check to see that any guidelines or rules of thumb for the approximations are satisfied. In part (e), even the approximation using the continuity correction fails to produce an answer with even one correct significant digit even though the sample size is 76, because 76 just isn't large enough given that we are working with a rather small value of p.)

I will also have another roulette wheel that has 38 numbers on it $(00, 0, 1, 2, \ldots, 35, 36)$. Since I have observed a tendency for people not to bet on the number 13, I have rigged this wheel so that 13 comes up more often than the other numbers do. If you observe 5 outcomes of 13 in 76 trials, would you be able to claim that there is statistically significant evidence to reject $H_0 : p \leq 1/38$ in favor of $H_1 : p > 1/38$, where p is the probability of observing the outcome 13?

(d) Respond to this query by *reporting the p-value which results from the appropriate test.* (*Note*: You should get a value which does not lead to a statistically significant implication that I'm a crook (but it's *very* close), if a significance level of 0.05 is employed.)

(e) *Approximate the p-value requested in part (d) using the normal approximation with the continuity correction.*

(f) *Approximate the p-value using the normal approximation without the continuity correction.*

3) In order to hopefully get a better understanding of the concept of a consistent estimator, consider iid Bernoulli(p) random variables $X_1, X_2, \ldots, X_n$, and determine the value of $P(|\overline{X} - p| < 0.02)$ for the case of $p = 0.2$ and the values of n indicated below. (*Note*: $\overline{X}$ is a consistent estimator of p, and the probability of it assuming a value within 0.02 of the estimand increases as n increases. Also note that the stated interest here is the probability of the absolute difference being *strictly less than* 0.02 (and not less than or equal to 0.02).)

(a) $n = 100$

(b) $n = 200$

(c) $n = 400$

A just-for-fun problem (not to be turned in for credit): Let Y be a binomial(n, p) random variable. Show that $P(Y \geq k)$ is an increasing function of p. *Hint*: Differentiate $P(Y \geq k)$ w.r.t. p. Consider two cases: $k \geq np$ and $k < np$. In the first case, the derivative is clearly nonnegative, while in the second case it is no smaller than

$$\frac{1}{p(1-p)} \sum_{j=0}^{n} (j - np) \binom{n}{j} p^j (1-p)^{n-j} = \frac{E(Y - np)}{p(1-p)} = 0.$$

A just-for-fun problem (not to be turned in for credit): Show that S^2 is an unbiased estimator for the variance of any distribution (for which the variance exists).

Extra Credit Problem

For the case of normal distributions, find an unbiased estimator for the standard deviation. (You might consider modifying S by multiplying it by a constant (which might be a function of n).)

Note: This semester, extra credit problems will be worth one point apiece, unless it is stated otherwise. Some of these problems are a tad tricky and may require that you use results which are not explicitly given in class (but all of them can be done using material from a good probability class, the class notes for STAT 554, and possibly some indicated parts of Miller's *Beyond ANOVA*.). Since the maximum score on each homework assignment is typically less than the total number of points for the regular problems on the assignment, working the extra credit problems will only benefit your grade if you screw up on some of the regular problems of the assignment (and my guess is that most people who are able to do the extra credit problems should be able to get an A in STAT 554 even if they don't turn in any extra credit problems). With regard to your grade in STAT 554, you'll probably be better off if you spend time checking over the regular parts of the assignments and not spend a great deal of time worrying about the extra credit problems. But if you plan to take additional courses in statistics, any effort you put into these extra credit problems may pay off later on. *If you plan to turn in any extra credit problems for credit, you should solve them entirely on your own. (I might give you a tiny hint on some of the problems, but otherwise the honor code applies.) Always put any extra credit solutions at the end of your paper, starting them on a new sheet that has your name on it in case I have to tear these problems off in order to save them to grade at a later time.*