m-s-1. Working with logits and logs of proportions
In order to have a sampling distribution that is closer to Gaussian (sample proportions, odds, and ratios of them tend to have nasty sampling distributions), we often transform from the (0,1) π i.e., proportion, scale to the $(-\infty, 0)$ $\log[\pi]$ scale, or the $(-\infty, -\infty)$ $\log[\pi/(1-\pi)]$ scale. The latter transformation is called the *logit* transform.

Thus, we do all our inference (SE calculations, CI's, tests) on the log or logit scale, then transform back to the proportion or odds or ratio scale.

i. Suppose $y \sim$ Binomial(n, π) and that $p = y/n$. Derive the variance for the random variables $\log[p]$ and logit$[p] = log[p/(1-p)]$. Ignore the possibility of obtaining $y = 0/n$ or $y = n/n$: people often add 0.5 to y and 1 to n to avoid such complications.

ii. The variance of p is largest when $\pi = 0.5$. At what value of π is the variance of logit$[p]$ largest?

m-s-2. Greenwood's formula for the SE of an estimated Survival Probability
In survival analysis, we often estimate the surviving proportion S after a fixed number k of time intervals as a product of (estimated) conditional probabilities, ie $\hat{S} = \prod_1^k \hat{S}_i$. The i-th element is the conditional probability of surviving the i-th interval, given that one survived the previous intervals.

For inference regarding S, we need SE$[\hat{S}]$. To derive this, it is easier to work in the $\log[S]$ scale, so that $\log \hat{S} = \sum_1^k \log[\hat{S}_i]$, to calculate the SE and CI in this scale, and then transform back to the (0,1) S scale.

Exercise: Treat $\hat{S}_i \sim (1/n_i) \times$ Binomial(n_i, S_i), with n_i fixed (in practice, the n_i's are random, but there are good reasons to treat them as fixed for the variance calculation). Derive the variance for $\log[\hat{S}]$, and from this the variance for $\hat{S}$.

m-s-3. Link between exact tail areas of Binomial and F distributions
Re-write the proof in your own notation (see Fisher 1935, under Resources)

1. Clusters of Miscarriages [based on article by L Abenhaim]

Assume that:

- 15% of all pregnancies end in a recognized spontaneous abortion (miscarriage)
- Across North America, there are 1,000 large companies. In each of them, 10 females who work all day with computer terminals become pregnant within the course of a year [the number who get pregnant would vary, but assume for the sake of this exercise that it is exactly 10 in each company].
- There is no relationship between working with computers and the risk of miscarriage.
- a "cluster" of miscarriages is defined as "at least 5 of 10 females in the same company suffering a miscarriage within a year"

Exercise: Calculate the number of "clusters" of miscarriages one would expect in the 1,000 companies. Hint: begin with the probability of a cluster.

2. "Prone-ness" to Miscarriages ?
Some studies suggest that the chance of a pregnancy ending in a spontaneous abortion is approximately 30%.

i. On this basis, if a woman becomes pregnant 4 times, what does the binomial distribution give as her chance of having 0,1,2,3 or 4 spontaneous abortions?

ii. On this basis, if 70 women each become pregnant 4 times, what number of them would you expect to suffer 0,1,2,3 or 4 spontaneous abortions? (Think of the answers in (i) as proportions of women).

iii. Compare these theoretically expected numbers out of 70 with the following observed data on 70 women, each of whom had 4 pregnancies:

No. of spontaneous abortions:	0	1	2	3	4
No. of women with this many abortions:	23	28	7	6	6

iv. Why don't the expected numbers agree very well with the observed numbers? i.e. which assumption(s) of the Binomial Distribution are possibly being violated? (Note that the overall rate of spontaneous abortions in the observed data is in fact 84 out of 280 pregnancies or 30%).

3. Automated Chemistries (from Ingelfinger et al)
At the Beth Israel Hospital in Boston, an automated clinical chemistry analyzer is used to give 18 routinely ordered chemical determinations on one order (glucose, BUN, creatinine, ..., iron). The "normal" values for these 18 tests were established by the concentrations of these chemicals in the sera of a large sample of healthy volunteers. The normal range was defined so that an average of 3% of the values found in these healthy subjects fell outside.

i. Using the binomial formula, compute the probability that a healthy subject will have normal values on all 18 tests. Also calculate the probability of 2 or more abnormal values.

ii. Which of the requirements for the binomial distribution are definitely satisfied, and which ones may not be?

iii. Among 82 normal employees at the hospital, 52/82 (64%) had all normal tests, 19/82 (23%) had 1 abnormal test and 11/82 (13%) had 2 or more abnormal tests. Compare these observed percentages with the theoretical distribution obtained from calculations using the binomial distribution. Comment on the closeness of the fit.

4. Binomial or Opportunistic? Capitalization on chance... multiple looks at data. Q from Ingelfinger et al.)
Mrs A has mild diabetes controlled by diet. Her morning urine sugar test is negative 80% of the time and positive (+) 20% of the time [It is never graded higher than +].

i. At her regular visit she tells her physician that the test has been + on each of the last 5 days. What is the probability that this would occur if her condition has remained unchanged? Does this observation give reason to think that her condition has changed?

ii. Is the situation different if she observes, between visits, that the test is positive on 5 successive days and phones to express her concern?

5. Can one influence sex of baby?
These data atre taken from an artcle in the NEJM 300:1445-1448, 1979.

i. Consider a binomial variable with $n = 145$ and $\pi = 0.528$. Calculate the SD of, and therefore a measure of the variation in, the proportions that one would observe in different samples of 145 if $\pi = 0.528$.

ii. Then consider the following, abstracted from the NEJM article: and answer the question that follows the excerpt.

   The baby's sex was studied in births to Jewish women who observed the orthodox ritual of sexual separation each month and who resumed intercourse within two days of ovulation. The proportion of male babies was 95/145 or 65.5% (!!) in the offspring of those women who resumed intercourse two days after ovulation (the overall percentage of male babies born to the 3658 women who had resumed intercourse within two days of ovulation [i.e. days -2, -1, 0, 1 and 2] was 52.8%)".

iii. How does the SD you calculated above help you judge the findings?

6. It's the 3rd week of the course: it must be Binomial
In which of the following would Y not have a Binomial distribution? Why?

i. The pool of potential jurors for a murder case contains 100 persons chosen at random from the adult residents of a large city. Each person in the pool is asked whether he or she opposes the death penalty; Y is the number who say "Yes."

ii. Y = number of women listed in different random samples of size 20 from the 1990 directory of statisticians.

iii. Y = number of occasions, out of a randomly selected sample of 100 occasions during the year, in which you were indoors. (One might use this design to estimate what proportion of time you spend indoors)

iv. Y = number of months of the year in which it snows in Montral.

v. Y = Number, out of 60 occupants of 30 randomly chosen cars, wearing seatbelts.

vi. Y = Number, out of 60 occupants of 60 randomly chosen cars, wearing seatbelts.

vii. Y = Number, out of a department's 10 microcomputers and 4 printers, that are going to fail in their first year.

viii. Y = Number, out of simple random sample of 100 individuals, that are left-handed.

ix. Y = Number, out of 5000 randomly selected from mothers giving birth each month in Quebec, who will test HIV positive.

x. You observe the sex of the next 50 children born at a local hospital; Y is the number of girls among them.

xi. A couple decides to continue to have children until their first girl is born; Y is the total number of children the couple has.

xii. You want to know what percent of married people believe that mothers of young children should not be employed outside the home. You plan to interview 50 people, and for the sake of convenience you decide to interview both the husband and the wife in 25 married couples. The random variable Y is the number among the 50 persons interviewed who think mothers should not be employed.

7. Tests of intuition

i. A coin will be tossed either 2 times or 20 times. You will win $2.00 if the number of heads is equal to the number of tails, no more and no less. Which is correct? (i) 2 tosses is better. (ii) 100 tosses is better. (iii) Both offer the same chance of winning.

ii. Hospital A has 100 births a year, hospital B has 2500. In which hospital is it more that at least 55% of births in one year will be boys.

8. Test of a proposed mosquito repellent

An entomologist carried out the following experiment as a test of a proposed mosquito repellent. Thirty-five volunteers had one forearm treated with a small amount of repellent and the other with a control solution. The subjects did not know on which forearm the repellent had been used. At dusk the volunteers exposed themselves to mosquitoes and reported which forearm was bitten first. In 10/35, the arm with the repellent was bitten first.

i. Make a statistical report on the findings.

ii. How would you analyze the results if: (a) some arms were not bitten at all? (b) some people were not bitten at all?

9. Triangle Taste test

In its 1974 manual "Laboratory Methods for Sensory Evaluation of Food", Agriculture Canada described tests (the triangle test, the simple paired comparisons test,...) to determine a difference between samples

> In the triangle test, the panelist receives 3 coded samples and is told that 2 of the samples are the same and 1 is different and is asked to identify the add sample. This method is very useful in quality control work to ensure that samples from different production lots are the same. It is also used to determine if ingredient substitution or some other change in manufacturing results in a detectable difference in the product. The triangle test is often used for selecting panelists.
>
> Analysis of the results of triangle tests is based on the probability that - IF THERE IS NO DETECTABLE DIFFERENCE - the odd sample will be selected by chance one-third of the time. Tables for rapid analysis of triangle test data are given below. As the number of judgements increases, the percentage of correct responses required for significance decreases. For this reason, when only a small number of panelists are available, they should perform the triangle test more than once in order to obtain more judgements.
>
> The results of a test indicate whether or not there is a detectable difference between the samples. Higher levels of significance do not indicate that the difference is greater but that there is less probability of saying there is a difference when in fact there is none.

Chart: Triangle test difference analysis [Table starts at $n = 7$ and ends at $n = 2000$; selected entries shown here]

Number of correct answers necessary to establish...

	level of significance		
No. Tasters	5%	1%	0.1%
7	5	6	7
10	7	8	9
12	8	9	10
30	16	17	19
60	28	30	33
100	43	46	49
1000	363	372	383

i. Show how one arrives at the numbers 7, 8 and 9 of correct answers necessary to establish the stated levels of significance for the case of n=10 tasters. Hint: you can work them out from the `BINOMDIST` function in Excel or [since we are only interested in the principles involved, and not in getting answers correct to several decimal places] you should be able to interpolate them from probability distributions tabulated in the text [the setup here is similar to the therapeutic touch study, but with $\pi = 1/3$ rather than $\pi = 1/2$].

ii. Calculate the exact 90, 98 and 99.8 percent 2-sided CI's for the proportions 7/10, 8/10 and 9/10 respectively, and from these limits verify that indeed 7/10, 8/10 and 9/10 are significantly greater than 0.33, at the stated levels of significance .(I am presuming that their Ha is 1-sided, ie. 0.33 vs. > 0.33

You can obtain these CI's from the spreadsheet "CI for a proportion", under Resources for Ch 8.

iii. Show how one arrives at the numbers 43, 46 and 49 of correct answers necessary to establish the levels of significance for the case of 100 tasters. Hint: you should be able to use a large-sample approximation.

iv. How well would this large-sample approximation method have done for the case of $n = 10$?

v. If you set the α at 0.05 (1-sided), what number of tasters is required to have 80 percent power to 'detect' a 'shift' from $H_0 : \pi = 1/3$ to (i) $H_a : \pi = 1/2$ (ii) $H_a : \pi = 2/3$? Use the sample size formula in section 8.1 of the notes.

Notes: See worked example 2 in notes on Chapter 8.1. This is an good example where a one-sided alternative is more easily justified, so with $\alpha = 0.05$ 1-sided, $Z_a = 1.645$. Note that power of 80 percent means that $Z_\beta = -0.84$. The Z_β is always one-sided, since one cannot be on both sides of H_0 simultaneously!

10. Variability of, and trends in, proportions

The following data are the proportion of Canadian adults responding YES to the question "Have you yourself smoked any cigarettes in the past week?" in Gallup Polls for the years 1974 to 1985.

	1974	'75	'76	'77	'78	'79	'80	'81	'82	'83	'84	85
%	52	47	.	45	47	44	41	45	42*	41	39	39

. question not asked in 1976;
* question worded "occasionally or regularly" in 1982.

Results are based on approximately 1050 personal in-home interviews each year with adults 18 years and over.

i. Plot these percentages along with their 95 confidence intervals.

ii. Is there clear evidence that the trend is downward? To answer this, try to draw a straight line through all (or most of) the confidence intervals and ask can the straight line have a slope of zero i.e. be parallel to the horizontal axis. You might call this a "poor-person's test of trend."

11 A Close Look at Therapeutic Touch

[Rosa L et al., JAMA. 1998;279:1005-1010; for those interested, there is considerable follow-up correspondence]

See the full article under Resources..

In the last paragraph of Methods the authors state (italics by JH):

"The odds of getting 8 of 10 trials correct by chance alone is *45 of 1024 (P=.04)*, a level considered significant in many clinical trials. We decided in advance that an individual would "pass" by making *8 or more correct selections* and that those passing the test would be retested, although the retest results would not be included in the group analysis."

i. Use statistical software, or Table C of M & M3, or first principles, to verify that the probability of getting exactly 8 of 10 correct is indeed 45 of 1024.

ii. In the next sentence the authors state that in fact they used "8 or more correct" as their criterion. Explain why this definition of "evidence for the therapeutic touch" (or, if you prefer, "against the skeptic's null hypothesis"') is more logical than the "exactly 8" for which they calculate the P=0.04 [Hint: See the second half of the first paragraph (about specific outcomes) under P-values in M & M page 457. In our context, imagine that there were 400 trials: then the probability of – by chance alone – getting exactly 320 is indeed, in Dr. Arbuthnot's words, "vanishingly small." but the probability of getting specifically 200 (a value that provides no evidence against H_0, is also small (0.04)]

iii. Calculate – under the "null" hypothesis, the probability of "8 or more correct". Is it indeed less than the arbitrary "level considered significant" of 0.05? If not, then what would the criterion need to be so that the probability – again calculated under "H0" – of reaching this criterion is ¡ 0.05.

iv. Figure 2 shows the scores of the 28 subjects. Multiply the set of Binomial probabilities with n=10 and p = 0.5 (i.e., p[0/10 correct — p = 0.5] to p[10/10 correct — p = 0.5] by 28 to obtain theoretical frequencies. These are the numbers of subjects, out of 28, one would expect to get 0/10, 1/10, ... 10/10 trials correct if all they were doing in each trial was guessing. Compare the theoretical frequencies of subjects with the observed "No. of subjects" with each score. Comment. Ignore for the moment the fact that the 28 people tested were really only 21 distinct people – 14 tested once (10 trials each) and 7 tested twice (10 trials, twice)