STT 315 F06

Preparation for Exam 2 (exam scheduled for 10-5-06 in your recitation).

Solutions of the questions below will be discussed in lectures M 10-2-06; W 10-4-06.

The MW lecture notes will be the only key posted for these questions.

Ch. 4. All of this chapter plus sections 3-5 and 3-9 (binomial and Poisson) and the normal approximation of the Poisson distribution. This material is about using probability models to approximate distributions (typically, but not always, distributions of sums or averages) by normal distributions. Since any normal distribution is completely known once its mean and sd are specified, our task is to know which mean and sd apply in each context.

1. **Binomial**. 28% of our customers pay cash.

a. What is the distribution of $X$ = the number of cash paying customers in a random sample of 300 customers?

b. What are the numerical values of the mean (i.e. $E\,X$) and the sd (i.e. sd $X$)?

c. By the CLT, $X$ is approximately normally distributed. Sketch the normal approximate distribution of $X$ with its mean and sd clearly identified as recognizable features of your sketch.

d. Use the binomial formula (see section (3-5)) to evaluate the probability $p(70)$ that there will be exactly 70 cash paying customers in our sample of 300.

e. Use the normal approximation of the binomial, with 0.5 continuity correction, to approximate $p(70)$ and compare with (d).

f. Use the normal approximation of the binomial and continuity correction to approximate $P(X < 71)$.

**2. Poisson.** Our business averages around 4.2 instances of process stoppage in one month. The Poisson is thought to apply.

a. Let X denote the number of stoppages in one months' time. Use the Poisson formula (see section (3-7)) to determine p(5), the probability that we experience exactly 5 stoppages in one month.

b. Determine the numerical values of E X and sd X.

c. Sketch the approximating normal distribution for X clearly identifying the numerical mean and sd from (b) in your sketch.

d. Determine the normal approximation with continuity correction of p(5) and compare with (a).

e. The Poisson having mean np approximates the binomial if n is large and p is small. The batter from which a cookie is made is randomly mixed, makes 4000 cookies, and has 15,500 raisins.

e1. What is the expected number of raisins in the cookie.

e2. What is the Poisson probability that the cookie gets exactly 3 raisins.

**3. Normal table.** No continuity correction unless requested.

a. Determine $P(Z > 1.23)$.

b. Determine $P(Z \text{ in } [-1.31, 2.53])$.

c. IQ is normal with mean 100 and sd 15.

c1. Determine $P(IQ > 100 + 1.23 \cdot 15)$.

c2. Determine $P(IQ > 137)$.

d. In IQ, give the standard score (i.e. z-score) of $IQ = 116$.

e. Determine z with $P(0 < Z < z) = 0.32$ (use closest table entry in the body of the table).

f. Determine z with $P(Z < z) = 0.57$ (use closest table entry).

g. Determine z with $P(Z < z) = 0.28$.

h. Determine the 76th percentile of IQ.

## 4. Normal properties.

a. Incomes X, Y from two business activities are modeled as independent random variables, each normal, with

E X = 23          E Y = 52
sd X = 11          sd Y = 17

Total profit is profit = 0.1 X + 0.06 Y – 0.4.

a. Determine E profit. Does your calculation require the assumed independence?

b. Determine sd profit. Do you require the independence?

c. Sketch the approximate distribution of profit (linear forms in independent normal r.v. are again normal).

## 5. Normal approximate distribution of xBAR and pHAT.

a. 54% of voters favor a proposal. A sample of 100 voters will be selected with replacement. Sketch the normal approximation of the distribution of pHAT (i.e. of the fraction of the 100 who favor the proposal). Calculate and clearly identify the numerical values of E pHAT and sd pHAT in your sketch.

b. In (a), if the population consists of N = 800 voters re-draw the sketch for a sample of n = 100 without replacement.

c. Each member of a population of business clients is scored with x = the amount they would purchase if contacted by an "automatic salesperson" instead of a real one. We propose to "test the waters" by sampling 50 of our clients with replacement. To get an idea of what a sample of 100 might be capable of, the boss asks us to examine what it would do if the population mean of x is mu = 45.7 with population sd of x sigma = 18.6. Sketch the approximate normal distribution of xBAR (the sample mean of 100) indicating the numerical values of E xBAR and sd xBAR as identifiable elements of your sketch.

d. Redraw (e) if instead the sample of 100 is without replacement and the population of clients numbers 800.

## 6. t-based exact CI for population mean mu based on independent samples of any number n > 1 from a normal population.

a. Calculate sample sd for the sample {3, 7, 2.3}.

b. If the sample (a) is presumed to come from a normally distributed population (i.e. "in control" population) give the 90% t-based CI (confidence interval) for the mean mu of the population. Indicate how you get DF (degrees of freedom).

c. If you desire a 90% CI of the form xBAR +/- 0.8 (i.e. given precision B = 0.8) you can get it provided you re willing to pay the price of additional samples. You must sample to $nFINAL = (t\,s\,/\,0.4)^2$ (round up to the next integer) where nFINAL is the sample mean score of all samples out to nFINAL and both t and s refer to their values from the initial sample of n (n is 3 in this example). From (a) (b) determine nFINAL. If you could afford to continue sampling until the number nFINAL samples have been made, and if we suppose the xBARfinal average of all of them is xBARfinal = 3.97, give the resulting 90% CI: xBARfinal +/- 0.8 (as desired). Although nFINAL may be large we effectively use the n = 3 values of t and s.

**7.  z-based CI for mu or p when n is large (the population need not be normal).**

a.  A with replacement sample of 400 sales finds 64 in which the customer says they would prefer a heavier paper bag be offered at checkout.  Give a 95% z-based CI for the unknown fraction p (of all of our thousands of customers) who would likewise prefer that a heavier bag be offered.

b.  Suppose that in (a) we wish a 95% CI of the form pHATfinal +/- 0.02.  We can have that if we continue to a new nFINAL sample size determined from our initial pHAT

$$nFINAL = (z \; root(pHAT \; qHAT) / B)^2 = 1.96 \; root(.16 \;\; .84) / 0.02)^2$$

Note:  It is possible that nFINAL will be less than our intitial sample size.  This just means we have already achieved the accuracy desired.

Determine nFINAL.  Give the desired CI if, out of all those sampled to nFINAL, there are 200 who say they would prefer that a heavier bag be offered at checkout.

c.  We have sampled 50 sales with replacement finding xBAR = $5.78 with sample sd s = $1.79, for score x = sales amount.  Determine the 90% z-based CI for mu = population mean sales amount.

d.  For (c) give the ME (margin of error).

e.  Determine a sample size nFINAL sufficient to achieve a 90% CI xBARfinal +/- $0.20.  What is the CI if xBARfinal = $591?

**8. z-based CI for mu or p when a stratified sample is used.** We need the number sampled in each stratum to be large enough for the z-approximation in each stratum.

Strata i = 1 to k with respective weights (fractions of population count) $W_i$ are sampled proportionally with $n_i \sim W_i\, n$ (with repl in each stratum).

a. z-based CI for overall population mean mu, in terms of sample sizes $n_i$, sample means xBAR$_i$ and sample sd $S_i$ for each of strata i = 1 to k.

$$\sum\nolimits_{i=1}^{k} W_i\, \bar{X}_i \pm z \sqrt{\sum\nolimits_{i=1}^{k} W_i^2\, \frac{S_i^2}{n_i}}$$

b. z-based CI for overall population proportion p, in terms of sample sizes $n_i$, strata sample proportions pHAT$_i$ for each of strata i = 1 to k.

$$\sum\nolimits_{i=1}^{k} W_i\, \hat{P}_i \pm z \sqrt{\sum\nolimits_{i=1}^{k} W_i^2\, \frac{\hat{P}_i\, \hat{q}_i}{n_i}}$$

c. What is the claim made for estimates and CI based upon proportionally stratified method vs regular with replacement sampling?