[redacted]

Recitation Assignment due at the close of recitation 2 -16-10.

1. Below is a record of tests performed on a random sample of electronic parts. The population of parts has some unknown fraction that are defective. We use "Y" to indicate a defective part. As you can see, the first three sample parts test defective, not defective, defective.

{Y, N, Y, N, Y, Y, N, N, Y, N, N, Y, N, N, N, Y, N, Y, Y, Y, N, Y, N, Y, N, N, Y, N, N, Y, N, N, N, N, N, N, N, Y, Y, N, N, Y, Y, Y, N, N, N, Y, N, N, Y, N, N, Y, N, N, N, N, N, N}

a. What is the sample size n? $n = 60$

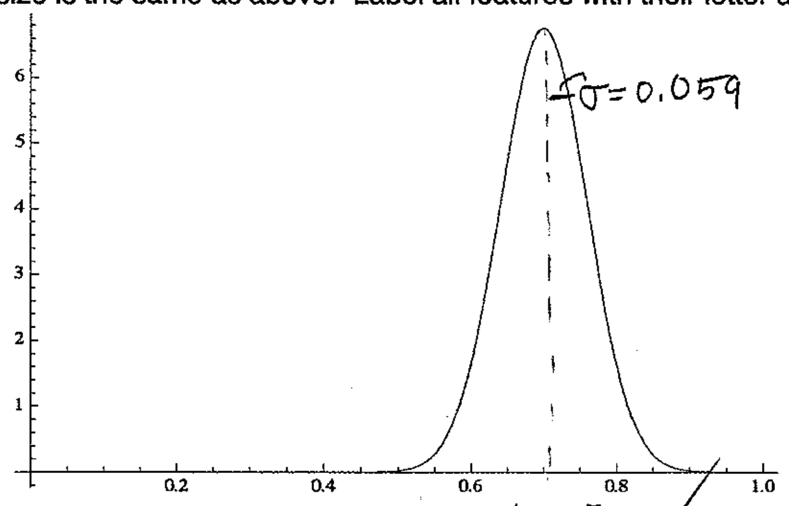
b. What is the point estimate $\hat{p}$ of p for this data? $\hat{p} = \frac{x}{n} = \frac{22}{60} = 0.367$

c. Give the likely size of $P(\hat{p} = p) \sim \boxed{0}$, i.e. the probability that our particular sample will have hit the unknown population defectives rate p dead on? one single point

d. What is the value of (E $\hat{p}$) in terms of p? $E(\hat{p}) = p$

e. What is the value of $\sigma_{\hat{p}}$ in terms of n, p (i.e. the standard deviation of $\hat{p}$)? See pg. 487 above the SE display. $\sigma_{\hat{p}} = \sqrt{pq/n}$

f. Sketch the normal approximation of the distribution of $\hat{p}$ if the actual value of p is p = 0.7 and the sample size is the same as above. Label all features with their letter and numerical values.



$\mu = E(\hat{p}) = p = 0.7$

$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.7 \times 0.3}{60}}$

$= 0.059$

g. What is $\hat{\sigma}_{\hat{p}}$ (i.e. the estimated value of $\sigma_{\hat{p}}$, also called the standard error or SE) for this data? See pg. 487 SE.

$\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n} = \sqrt{\frac{22}{60} \cdot \frac{38}{60} / 60} = 0.0622 = 6.22\%$

Generally, we don't use percentage to express st. dev. and st. error.

h. Give a 68% CI (i.e. confidence interval) for p based on this data.

$(\hat{p} \pm 1\sqrt{\hat{p}\hat{q}/n}) = \frac{22}{60} \pm 1\sqrt{\frac{62}{60} \cdot \frac{38}{60} / 60} = [0.3645, 0.4289]$ ✓

i. The "performance property of the 68% CI method" is that

P(68% CI for p encloses the true value p) ~ $\boxed{0.68}$ ✓

j. The true value of p is actually 0.35. Was the normal approximation leading to the 68% CI for p justified according to our rule of thumb (i.e. are np and nq each at least 10)? Has the 68% CI covered p = 0.35?

$np = 60 \times 0.35 = 21 > 10$ (ok!) Yes it covered $p = 0.35$
$nq = 60 \times 0.35 = 39 > 10$

k. Based on the data, using $n\hat{p}$ instead of np and using $n\hat{q}$ instead of nq would it have appeared that the normal approximation leading to the 68% CI was justified?

$n\hat{p} = n \cdot \frac{x}{n} = 22 > 10$ $\quad n\hat{q} = n - x = 60 - 22 = 38 > 10$ (ok!)

l. The probability that a 68% CI encloses (covers) the true value of p, being an approximation based on the normal, is not generally precisely equal to 0.68. On page 498 there is presented a refinement due to Agresti-Coull (A-C) that tends to achieve coverage probability closer to 0.68, at least for the majority of p and n. Especially, the method is recommended for p nearer 0 or 1 where the normal approximation is less accurate. The A-C refinement is to increase each of the Y and N counts by 2 (thus increasing the overall n by 4). The samples of n+4 are no longer independent samples of the population since the extra 4 are 2 Y and 2 N just dropped into the sample. It is nonetheless true that in a performance comparison the refinement seems to perform better. Give the A-C refined 68% CI for p for this data and compare it with the regular CI. See pg. 498.

$\hat{p}_{AC} = \tilde{p} = \frac{x+2}{n+2} = \frac{24}{64}$ $\quad P(p \text{ is in } \tilde{p} \pm 1\sqrt{\tilde{p}\tilde{q}/(n+4)}) = \frac{24}{64} \pm 1\sqrt{\frac{24}{64} \cdot \frac{40}{64} / 64} = 0.375 \pm 0.0605$

$\hat{q}_{AC} = \tilde{q} = \frac{y+2}{n+2} = \frac{40}{64}$ $\quad [0.3145, 0.4355)$ ✓

m. When the sample size n is not so very small relative to the population size N and we sample without replacement (e.g. don't sample the same voter twice) we have to be concerned about our assumption of independent samples. Fortunately, as long as the sampling gives equal probability to every population member, sampling without replacement can readily be adjusted for. In the CI simply replace $\hat{p} \pm 1\sqrt{\hat{p}\,\hat{q}\,/\,n}$ by $\hat{p} \pm 1\sqrt{\hat{p}\,\hat{q}\,/\,n}\,\sqrt{\frac{N-n}{N-1}}$. If the population size (of parts) is N = 400 what is the applicable CI? How does it compare with the ordinary CI? **DO NOT CONFUSE**

$N = 400$
$n = 60$

$FPC = \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{400-60}{400-1}} = 0.9231$ ✓

The applicable CI is CI w/FPC

$\hat{p} \pm 1\sqrt{\hat{p}\hat{q}/n} \cdot (0.923) = 0.3667 \pm 0.0622 \times 0.9231$
$= 0.3667 \pm \cancel{0.0574} \; 0.05742$

$[0.30928, 0.42412]$ ✓

THIS USE OF N FOR THE POPULATION SIZE WITH ITS USE AS "N" AS IN "Y" OR "N."

The additional factor $\sqrt{\frac{N-n}{N-1}}$ is called the Finite Population Correction (FPC).

2. The following data are from a sample of parts. The population size is N = 1000. $FPC=\sqrt{\frac{N-n}{N-1}}=\sqrt{\frac{1000-80}{1000-1}}=0.9596$

{N, N, Y, N, N, Y, Y, N, N, Y, N, N, N, Y, N, N, N, N, N, N, Y, Y, N, Y, N, N, N, N, Y, N, N, N, Y, N, N, N, N, N, N, Y, N, N, N, N, N, N, N, N, N, Y, N, Y, N, N, N, Y, N, N, N, N, N, N, N, N, N, N, N, N, Y, N, N, N, N, N, N, N, Y, N, N}

$n=80 \quad x=16$ $\quad \hat{p}=\frac{x}{n}=\frac{16}{80}=0.2$, $\sigma_{\hat{p}}=\sqrt{\hat{p}\hat{q}/n}=\sqrt{\frac{0.2\times0.8}{80}}=0.0447$

a. Determine a **68%** CI for the population fraction p of defective parts ("Y" indicates defective). Use the FPC but not A-C. From the data does the nnormal approximation seem applicable?

$\hat{p}\pm(1\cdot\hat{\sigma}_{\hat{p}}\cdot FPC)=0.2\pm0.0447\times0.9596=0.2\pm0.0429$ $[0.1571, 0.2429]$

b. Same as (a) but use AC **(we will use the very same FPC defined in (1m), not substituting n+4 for n in the FPC).**

$\hat{p}_{AC}=\frac{x+2}{n+4}=\frac{16+2}{80+4}=\frac{18}{84}=0.2143$ CI $\hat{p}_{AC}\pm\left(1\sqrt{\frac{\hat{p}_{AC}\hat{q}_{AC}}{n+4}}\cdot FPC\right)$

$=0.2143\pm\sqrt{\frac{0.2143\times0.7837}{80+4}}\cdot0.9596$

$=0.2143\pm0.0430$

$=[0.1713, 0.2573]$

$n\hat{p}=x=16>10$
$n\hat{q}=n-x$
$=80-16=64$
>10 (OK!)

c. Same as (a) but 95% CI.

$95\%CI=\hat{p}\pm(1.96\cdot\sqrt{\hat{p}\hat{q}/n}\cdot FPC)$

$=0.2\pm\left(1.96\sqrt{\frac{0.2\times0.8}{80}}\cdot0.9596\right.$
$=0.2\pm(1.96\cdot0.0429)=0.2\pm0.08408$
$[0.1159, 0.2841]$

d. Same as (b) but 95% CI.

$95\%CI=\hat{p}_{AC}\pm\left(1.96\cdot\sqrt{\frac{\hat{p}_{AC}\hat{q}_{AC}}{n}}\cdot FPC\right)$

$=0.2143\pm\left(1.96\cdot\sqrt{\frac{0.2143\times0.7837}{80+4}}\cdot0.9596\right)$

$=0.2143\pm(1.96\cdot0.0430)$

$=0.2143\pm0.0843$

$[0.1300, 0.2986]$

4 Very Good!