E370 Second Exam Answer Key (2010) Prof. W. E. Becker

I. (33 PTS) STATE WHETHER THE STATEMENT IS TRUE OR FALSE AND GIVE A SHORT REASON TO JUSTIFY YOUR ANSWER.

A. If the random variable *X* has a normal distribution, then at least theoretically, the sampling distribution of $\bar{X}$ has an approximate normal distribution as long as the sample size is not too small.
TRUE - CENTRAL LIMIT THEOREM.

B. The standard deviation of the sample, the standard deviation of the sample mean, and the standard error of the mean are three names representing the same measure of dispersion.
NO - ST DEV OF SAMPLE IS S; ST DEV OF MEAN ANS ST ERROR IS SIGMA/SQRT(n)

C. $P(\bar{X} < 10) = P(P(\bar{X} \leq 10)$.
YES - BECASUE X BAR IS CONTINUOUS

D. The distribution of a point estimator is a density function (or histogram) of the random values used to calculate the estimate from an observed sample of size n.
YES - FOR EXAMPLE X BAR DISTRIBUTION

E. The distribution of the point estimator of the population proportion ($p=X/n$) is approximately normal if X is normal, for small *n*.
NO - X IS BINOMIAL SO X/n COULD ONLY BE APPROCIMATLY NORMAL FOR LARGE n.

F. If *h* is an unbiased estimator of parameter *H*, then the expected value of *H* is *h* so that $H = h$.
NO - the estimator *h* is a random variable while *H* is one number.

G. A point estimator can be thought of as a zero percent confidence interval estimator.
YES, FOR EXAMPLE, PROB(XBAR = MU)=0.

H. If the null hypothesis can be rejected at the 0.01 Type I error level then it can also be rejected at the 0.10 level as well.
YES, 0.01 IMPLIES AN ESTIMATE FARTHER INTO THE TAIL THAN 0.10.

I. If the Alpha risk is reduced, then the Beta risk rises because Alpha equals one minus Beta.
NO, ALPA DOES NOT EQUAL 1 MINUS BETA, ALTHOUGH BETA RISES AS ALPHA IS REDUCED.

J. If one moves from a 95 percent to a 90 percent confidence interval estimator, then the length or width of the interval gets larger.
NO, THE LENGTH GETS SHORTER BECAUSE Z GOES FROM 1.96 TO 1.64.

K. The p value provides the limiting Alpha level for which the null hypothesis can be rejected.
YES, AS LONG A P VALUE IS GREATER THAN ALPHA DO NOT REJECT NULL.

II. (67 PTS) ANSWER EACH AND ALL OF THESE THREE MULTIPLE-PART QUESTIONS.

A. An article on delays in insurance payments, in the *Wall Street Journal* cited Deborah Bohren, a spokesperson for Empire Blue Cross, saying "Empire is among the most efficient Blues plans in the country, with an average claims-processing time for claims with complete information of 4.6 days." The article went on to report that 88 percent of claims are settled within 2 weeks (10 business days).

1. (24 pts) Assuming processing time is normally distributed, what is the standard deviation of processing time?

Z = (10 - 4.6)/SIGMA, WHERE Z = NORMSINV(0.88) = 1.175; THUS, SIGMA = 4.596 DAYS

2. What is the likelihood of a claim requiring positive time to settle given your answer in part A.1?

APPROXIMATELY 84 PERCENT (WHICH IS RIDICULOUS BECAUSE THIS IMPLIES THAT THERE IS 16 PERCENT CHANCE OF NEGATIVE TIME REQUIRED AS USED IN PART 3).

3. Is it reasonable to assume processing time is normally distributed? Explain in terms of your answer to A.1 and the information from the *WSJ*.

NO, BECAUSE NORMALITY IMPLIES THAT ABOUT 16 PERCENT OF THE DISTRIBUTION SHOULD BE BELOW ZERO DAYS AND THIS IS IMPOSSIBLE SINCE DAYS CANNOT BE NEGATIVE.

4. What alternative assumption about the distribution of claims might be appropriate? Show why your alternative distribution is or is not preferred to assuming claims are normally distributed

THE DISTRIBUTION OF CLAIMS IS RIGHT SKEWED; THUS, A LOG TRANFORMATION MAY MAKE THE TRANSFORMED DISTIRIBUTION LOGNORMAL, BECAUSE THE LOG WILL HAVE A GREATER EFFECT ON LARGER CLAIM VALUES THAN ON SMALLER CLAIM VALUES:

2.302585	=LN(10)	
1.526056	=LN(4.6)	
0.660883	=(A4-A5)/A1	= St Dev of LN(claims)

B. (25 pts) One hundred and fifty Empire Blue Cross claims (as described in II A) are randomly selected to check on the claim that the mean time for processing is 4.6days. For this sample of 150, the mean was 10.6 days with standard deviation of 5.1 days.

1. What is the approximate standard error of the mean?

$5.1/\sqrt{150}=5.1/12.247=0.42$

2. What is the appropriate sample statistic to use for probability calculations about sample mean time?

CAN USE THE STANDARD NORMAL Z BECAUSE SAMPLE SIZE IS RELATIVELY LARGE

$z \approx (5.1-4.6)/0.42=1.2$

3. What is the probability of getting a sample mean of at least 5.1 days if the true mean is 4.6 days? 0.115

4. What if anything can you conclude from the probability in part 3?

BECAUSE THE PROBABILITY OF A SAMPLE OF 150 CLAIMS WITH A MEAN OF AT LEAST 5.1 DAYS IS NOT UNUSUAL THE CLAIM OF A MEAN OF 4.6 DAYS IS NOT UNREASONABLE.

0.416413	=5.1/SQRT(150)
1.20073	=(5.1-4.6)/A1
0.885072	=NORMSDIST(A2)
0.114928	=1-A3
0.115881	=TDIST(A2,149,1)

C. (18 pts) The following note appeared in *Sailing Anarchy* (07/24/09, www.sailinganarchy.com/index_page1.php):

> " Given the large size of this sample, we can be 95% confident that the data is representative, with less than a 3% margin of error."

1. Does a 95 percent confidence interval imply that the data are representative of the population?

The confidence interval estimator is based on random sampling, which says nothing about whether any specific individual sample of size *n* contains the true mean or is representative of the population in anyway. A 95 percent confidence interval says that prior to sample there is a 0.95 probability that the sample to be drawn randomly will result in a confidence interval that contains the true population mean.

2. How was the "3% margin of error" calculated?

Assuming a 0.05 probability of any response to a question, and an approximate 95 percent confidence margin of error (where the critical z = 1.96 but the approximate z = 2 is used), the margin for error is ±2*sqrt(.5*.5/n) = ±sqrt(1/n). Thus,

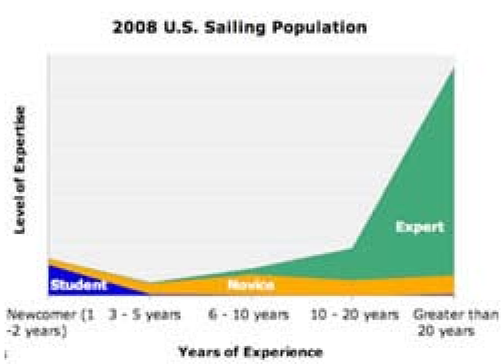
1200 = n

0.028868 =1/SQRT(A1)

3. With regard to the statistical attributes of an estimator, why is a large sample preferred to a small one?

The larger the sample size (*n*) the smaller the margin of error. That is, the larger the sample the smaller the standard deviation of the estimator. For example, the sample mean is used to estimate the population mean. The standard deviation of the distribution of the sample mean is the population standard deviation divided by the square root of *n*. Thus, as *n* goes up, this standard error goes down.

Old and Experienced



From the book Saving Sailing by Nicholas Hayes. Short but compelling. Order up to get a copy.

The online excerpt of Saving Sailing cites some basic age demographics of U.S. sailors. The book will offer many more statistics and trends about sailing and free time, and of course, how and why we might Save Sailing in the future. In the meantime, here is a simple tidbit that tells a lot about the current state of sailing affairs. Among almost 1200 sailors interviewed, 59% said that they have over 20 years of experience and call themselves experts (I see no reason to deny their claims -- if you believe that years of practice count for something).

The downside? Where are the newcomers and why don't they seem to stick around? Given all this experience, you'd think we'd be doing a better job passing it on. Note: Given the large size of this sample, we can be 95% confident that the data is representative, with less than a 3% margin of error.

FORMULA SHEETS COPIED FROM WILLIAM BECKER, STATISTICS FOR BUSINESS AND ECONOMICS: USING MICROSOFT EXCEL (SRB PUBLISHING 1997) :   99-101. pp. 172-173 p. 200. and p. 270.