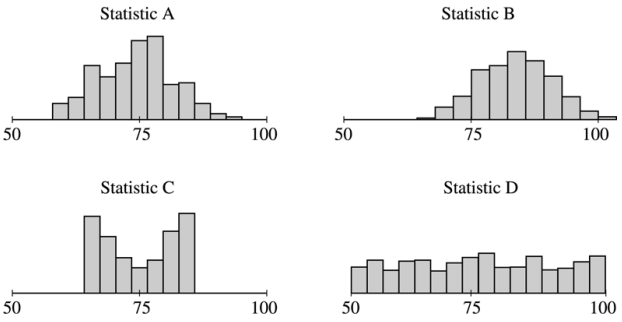
Chapter 9

What does it mean to be considered a biased or unbiased estimator? (08b, 2; M7, 33)

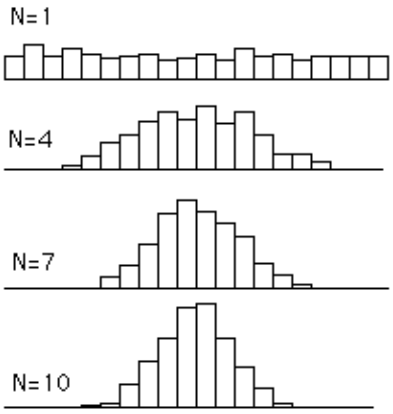
Four different statistics have been proposed as estimators of a population parameter. To investigate the behavior of these estimators, 500 random samples are selected from a known population and each statistic is calculated for each sample. The true value of the population parameter is 75. The graphs below show the distribution of values for each statistic.



Which of the statistics appear to be unbiased estimators of the population parameter?How can you tell?

Statistics A, C, and D appear to be unbiased. This is indicated by the fact that the mean of the estimated sampling distribution for each of these statistics is about 75, the value of the true population parameter.

Know that when you are referring to a standard deviation of a sampling distribution of the sample mean, you do not have to divide by rad n, it has all ready been done, (07, 3b)



Definitions

Sampling Distribution: This is the distribution of the original data taken individually (n =1). This is what would be created if you collected every person's weight in school.

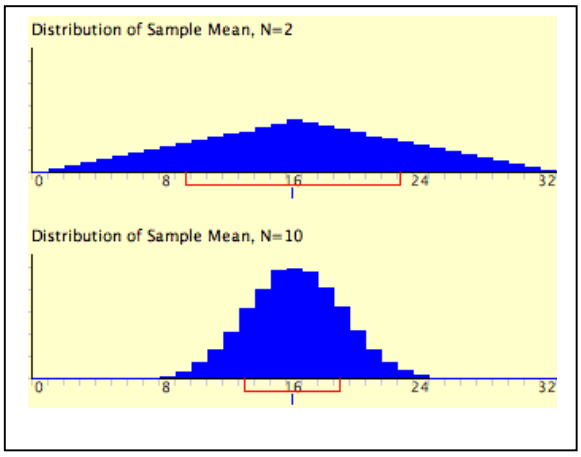
The parameters for a sampling distribution would be μ and σ

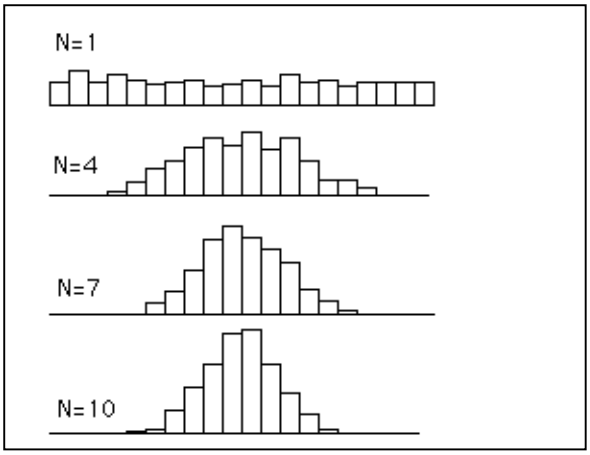
Sampling Distrubution of sample means: This is the distribution of average taken from samples of different sizes (n=4, n=7, or n=10). This is what would be created if 100 people surveyed 4 people randomly and collected their weight. We would then take those 100 average and create a graph, as is done in N = 4 above.

The parameters for a sampling distribution of sample means would be $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

How is a sampling distribution of a sample mean different than a distribution of individual observations? (10, 2a; M7, 23). Know that when you have a large sample size, even the most nonnormal distributions will have a sampling distribution of the sample mean close to a normal distribution by the Central Limit Theorem, (07, 3c; 07b, 2c)

Sampling Distributions of Sample Means





As your sample size become larger, these are things that you you notice:

1) It becomes more normal (But will never be normal, only approximately normal)—(This is the Central Limit Theorem)
   - **a.** If something is all ready symmetric, then you only have to make a few observations for the sampling distribution of the sample means to become approximately normal (notice the picture on the right)
   - **b.** If something is not at all symmetric, then you need about 30 observations for the sampling distribution of the sample means to become approximately normal
2) The average remains the same (If the overall $\mu = 16$ then $\mu_{\bar{x}} = 16$ no matter how large the sample size)
3) It becomes less spread (If the overall $\sigma = 7$ then $\sigma_{\bar{x}} = \frac{7}{\sqrt{n}}$ no matter what n is)

This is true even if the sampling distribution of the sample means is not normal and if it is normal. This means that these formulas apply to all of the distributions pictured in the figure on the right. It is true for n =1, n = 4, n = 7, n= 10.

Know how sampling distributions of sample means change as your sample size gets larger, and how that affects the likelihood of extreme events occurring, (07, 3a; M2, 9)

Big Town Fisheries recently stocked a new lake in a city park with 2,000 fish of various sizes. The distribution of the lengths of these fish is approximately normal.

(a)Big Town Fisheries claims that the mean length of the fish is 8 inches. If the claim is true, which of the following would be more likely? Justify your answer.
• A random sample of 15 fish having a mean length that is greater than 10 inches or
• A random sample of 50 fish having a mean length that is greater than 10 inches

We would expect it to be more difficult to get a random sample of 50 fish to be greater than 10, because of the difference in variability. Whether your sample size is n =15 or n =50, they both should be centered around 8 inches, because they came from the same population (this is true whether it was approximately normal or not). Even though they have the same center, though, we will expect a difference in variability. It is more difficult to have a group of 50 be far away from average than to have a group of 15 be far away from the average. Mathematically speaking, no matter what the standard deviation is, $\frac{\sigma}{\sqrt{50}} < \frac{\sigma}{\sqrt{15}}$. So, there is more variability in a group of 15 than a group of 50.

Know how to find the probability of averaging a certain amount in a Normal distribution ($P(\bar{x} > 10)$)? (10, 2b; 09, 2c; 07, 3b; 06, 3c; 04b, 3c). Know the difference between finding the probability that a group of 10 will average more than 150 lbs and a group of ten will all weight more than 150 lbs, (05b, 6c). Know how to find the average and standard deviation of a sampling distribution of a sample mean if you know the average and standard deviation of the original data, (11b, 6c; 07b, 2c; M2, 30)

A group of 10,000 people have a normal distribution of weights with an average of 145 and a standard deviation of 20.

1) What is the probability that a randomly chosen person will weigh more than 150lb?

N(145, 20)

$$P(X > 150)$$
$$= P\left(Z > \frac{150 - 145}{20}\right)$$
$$= P(Z > 0.25)$$
$$= 0.4013$$

2) What is the probability of 10 random people averaging more than 150lb?

Using the information, we know that the average and standard deviation of the sampling distribution of sample means follows N(145, $\frac{20}{\sqrt{10}}$)

$$P(\bar{X} > 150)$$
$$= P\left(Z > \frac{150 - 145}{\frac{20}{\sqrt{10}}}\right)$$
$$= P(Z > 0.79)$$
$$= 0.2148$$

3) What is the probability that 10 random people will all weigh more than 150?

Since the probability that one individual will weigh more than 150 is 0.4013, the probability that 10 individuals will all weigh more than 150 is…

$$(0.4013)^{10}$$
$$= 0.000108$$

Know that when you have a large sample size, even the most nonnormal distributions will have a sampling distribution of the sample mean close to a normal distribution by the Central Limit Theorem, (07, 3c; 07b, 2c)

Suppose the distribution of weights in the previous problem was was nonnormal but had the same mean and standard deviation. Would it still be appropriate to use the normal distribution to compare the probability in question (2)? Justify your answer.

In this case, it would not be appropriate, because a sample size of 10 would not guarantee that the distribution of sample means would be approximately normal. If, though, we took a sample great than 30, we would be able to say that the distribution of sample means is approximately normal by the Central Limit Theorem.

Note: If you said normal, it would be partially correct. You must say APPROXIMATELY normal.

Know how to find the standard deviation of a sampling distribution of sample proportions, (08, 4c)

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

1) A consequence of the Central Limit Theorem is that for nu sufficiently large (n > 30), if all samples of size n are taken, the mean of the sample means $\mu_{\bar{x}}$ is equal to the population μ. Since the mean of the sampling distribution is equal to the population mean, $\bar{x}$ is referred to as
    a. A biased estimator
    b. An unbiased estimator
    c. A random estimator
    d. A controlled variable
    e. A parameter
2) Billie decides to participate in one of two games of chance involving dice. In the first game each player gets a chance to throw the dice 100 times and wins if the dice show a sum of 7 between 15% and 20% of the time. In the second game, he gets to toss the dice 125 times and wins if his dice show a sum of 6 or 8 between 30% and 35% of the time. Billie decides to play the first game because he has a better chance of winning in that game, according to
    a. The law of large numbers
    b. The law of averages
    c. The central limit theorem
    d. Chebyshev's theorem
    e. Bayes' theorem
3) A normal population has a mean of 100 and a standard deviation of 16. If samples of size 64 are drawn from the population, what proportion of the samples should have mean values between 91 and 103?
    a. 0.0668
    b. 0.4987
    c. 0.9050
    d. 0.9319
    e. 0.9332
4) The variance of a random variable is 100 and a sample size of 25 is taken. The standard error of the sample means is
    a. 2
    b. 4
    c. 5
    d. 10
    e. 100
5) The Central Limit Theorem of statistics states which of the following for a sample of size n, (n > 1)
    a. The shape of the sampling distribution of sample means approaches normal as the n gets larger
    b. The mean of the set of sample means is always less than the mean of the population
    c. The standard deviation of the set of sample means is equal to the standard deviation of the population
    d. The mean of the set of sample means is smaller than the mean of the population when n is large
    e. The standard deviation of the set of sample means is greater than the standard deviation of the population

1) B
2) A
3) E. normalcdf(91, 103, 100, $\frac{16}{\sqrt{64}}$) $= 0.933189$
4) A. If the variance $\sigma^2 = 100$, then $\sigma = 10$. So the standard deviation of the sample means would be $\frac{10}{\sqrt{25}} = 2$.
5) A