Chp 11 | Sampling Distributions

* statistical inference ⇒ we use information from a sample to infer something about the population.

Vocab

- <u>parameter</u>: # that describes population (like μ); most often unknown
- <u>statistic</u>: # that can be computed from the sample (like $\bar{x}$); we often use a statistic to estimate a parameter

statistics from samples
parameters from population

★ $\bar{x}$ rarely $= \mu$ for only one sample; but if we keep taking larger samples, $\bar{x}$ is guaranteed to approach μ.

Law of Large Numbers

As we increase sample size, $\bar{x}$ of observed values gets closer & closer to μ of population.

Ex 1

Insurance. The idea of insurance is that we all face risks that are unlikely but carry high cost. Think of a fire destroying your home. Insurance spreads the risk: we all pay a small amount, and the insurance policy pays a large amount to those few of us whose homes burn down. An insurance company looks at the records for millions of homeowners and sees that the mean loss from fire in a year is $\mu = \$250$ per person. (Most of us have no loss, but a few lose their homes. The \$250 is the average loss.) The company plans to sell fire insurance for \$250 plus enough to cover its costs and profit. Explain clearly why it would be unwise to sell only 12 policies. Then explain why selling thousands of such policies is a safe business.

Chp 11 (cont)

Population distribution	vs.	sampling distribution
distribution of values of the variable among all individuals of a population (describes individuals)		distribution of values taken by statistic in all possible samples of same size from same population (describes how statistic varies in many samples)

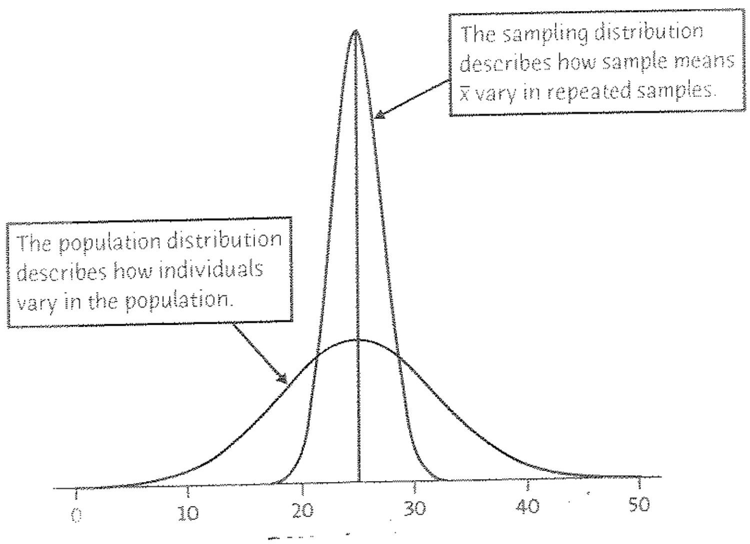
Ex 2

Sampling distribution versus population distribution. During World War II, 12,000 able-bodied male undergraduates at the University of Illnois participated in required physical training. Each student ran a timed mile. Their times followed the Normal distribution with mean 7.11 minutes and standard deviation 0.74 minute. An SRS of 100 of these students has mean time $\bar{x} = 7.15$ minutes. A second SRS of size 100 has mean $\bar{x} = 6.97$ minutes. After many SRSs, the many values of the sample mean $\bar{x}$ follow the Normal distribution with mean 7.11 minutes and standard deviation 0.074 minute.

(a) What is the population? What values does the population distribution describe? What is this distribution?

(b) What values does the sampling distribution of $\bar{x}$ describe? What is the sampling distribution?

Chp 11 (cont)



- $\bar{x}$ = mean of SRS of size n drawn from large population w/ mean μ & s.d. σ. Then sampling distribution of $\bar{x}$ has mean μ & s.d. $\sigma/\sqrt{n}$.
- $\bar{x}$ is unbiased estimator of μ because average of all $\bar{x}$ is μ.
- averages are less variable than individual observations i.e. $\sigma/\sqrt{n} < \sigma$
- the results of large samples are less variable than for small samples, i.e. $\sigma/\sqrt{n} \downarrow$ as $n \uparrow$

Sampling Distribution of Sample mean ($\bar{x}$)

If individual observations have $N(\mu, \sigma)$ distribution, then $\bar{x}$ of an SRS size n has $N(\mu, \sigma/\sqrt{n})$ distribution.

Ex 3

Larger sample, more accurate estimate. Suppose that in fact the blood cholesterol level of all men aged 20 to 34 follows the Normal distribution with mean $\mu = 188$ milligrams per deciliter (mg/dl) and standard deviation $\sigma = 41$ mg/dl.

(a) Choose an SRS of 100 men from this population. What is the sampling distribution of $\bar{x}$? What is the probability that $\bar{x}$ takes a value between 185 and 191 mg/dl? This is the probability that $\bar{x}$ estimates μ within ± 3 mg/dl.

(b) Choose an SRS of 1000 men from this population. Now what is the probability that $\bar{x}$ falls within ± 3 mg/dl of μ? The larger sample is much more likely to give an accurate estimate of μ.

Chp 11 (cont)

Ex 4

Measurements in the lab. Juan makes a measurement in a chemistry laboratory and records the result in his lab report. The standard deviation of students' lab measurements is $\sigma = 10$ milligrams. Juan repeats the measurement 3 times and records the mean $\bar{x}$ of his 3 measurements.

(a) What is the standard deviation of Juan's mean result? (That is, if Juan kept on making 3 measurements and averaging them, what would be the standard deviation of all his $\bar{x}$'s?)

(b) How many times must Juan repeat the measurement to reduce the standard deviation of $\bar{x}$ to 5? Explain to someone who knows no statistics the advantage of reporting the average of several measurements rather than the result of a single measurement.

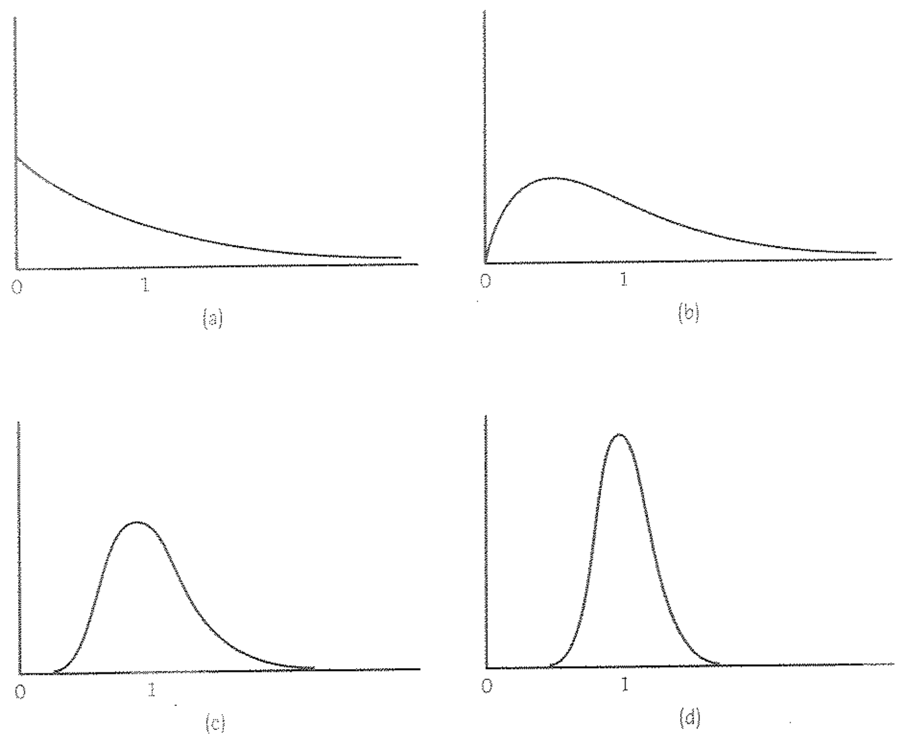
Central Limit Theorem

(Draw SRS of size n from population w/ mean μ and finite s.d. σ.) When n is large, sampling distribution of $\bar{x}$ is approximately $N(\mu, \sigma/\sqrt{n})$.

(true even when population distribution is not normal!)

★ how large does n need to be? it depends on whether the population distribution is close to normal.

Chp 11 (cont)



note:
(a) is from exponential distribution (which is definitely not normal)

FIGURE 11.5

The central limit theorem in action, for Example 11.7. The distribution of sample means $\bar{x}$ from a strongly non-Normal population becomes more Normal as the sample size increases. **(a)** The distribution of 1 observation. **(b)** The distribution of $\bar{x}$ for 2 observations. **(c)** The distribution of $\bar{x}$ for 10 observations. **(d)** The distribution of $\bar{x}$ for 25 observations.

Ex5

More on insurance. An insurance company knows that in the entire population of millions of homeowners, the mean annual loss from fire is $\mu = \$250$ and the standard deviation of the loss is $\sigma = \$1000$. The distribution of losses is strongly right-skewed: most policies have \$0 loss, but a few have large losses. If the company sells 10,000 policies, can it safely base its rates on the assumption that its average loss will be no greater than \$275?