

Sampling & Sampling Distributions

Population: the complete collection of all items of interest (e.g., all STAT 401 F students)

Parameter: a numerical characteristic of the population

- μ usually denotes the population mean (e.g., mean/average height of students)
- σ usually denotes the population standard deviation (e.g., how spread out are individual heights in class from typical/average height μ)

1

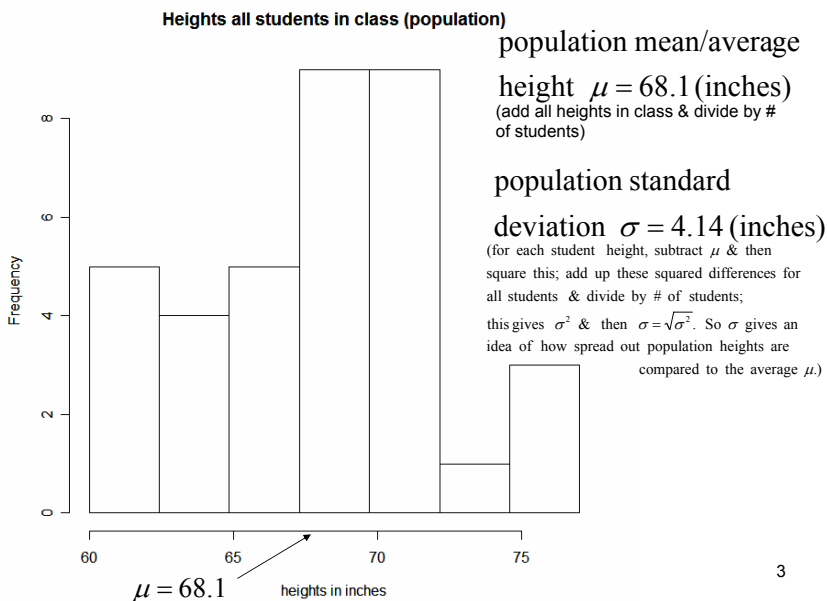
Sample: the subset of the population that is observed (e.g., 10 students selected at random)

Statistic: a numerical characteristic of the sample (e.g., average height of 10 students)

We estimate a population parameter by computing a sample statistic.

- The average of a sample \bar{Y} is used to estimate the mean of a population μ .
- The standard deviation of a sample s is used to estimate the standard deviation of a population σ .

2



3

Sampling Distribution of Sample Average \bar{Y}

Usually we take one random sample, perhaps compute a sample average and standard deviation, and that's our information to work with. But it's important to keep in mind that if we did take another random sample, we would most likely get a different sample and a different value for the statistic

If you start to think about all the possible samples that you might see when you randomly sample and all the possible values of statistics (sample averages) resulting from these samples, you will find that some values of statistics are more likely to occur than others.

Statistics like \bar{Y} have a so-called **Sampling Distribution** (which values of \bar{Y} might occur in a random sample and how likely certain values are.)

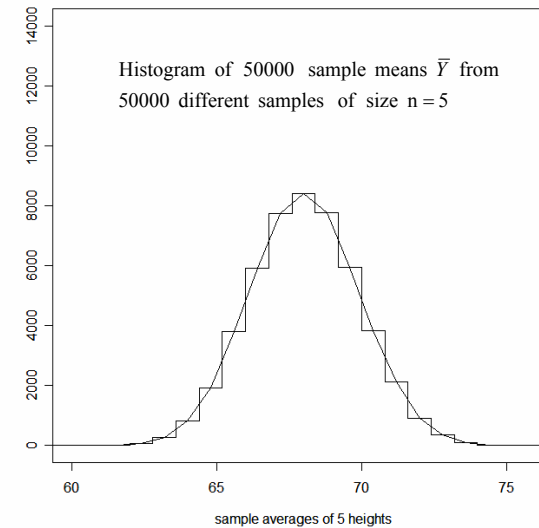
4

Suppose \bar{Y} is the average of a simple random sample of n items drawn from a population with mean μ and standard deviation σ . Imagine that we write down the value of the sample average for each possible sample.

1. Typical/average value of \bar{Y} equals the population mean μ (same as mean/typical value of *individual* observations in pop.)
2. standard deviation of \bar{Y} is $\frac{\sigma}{\sqrt{n}}$ (since \bar{Y} values may vary from sample to sample, this measures the spread or variability in possible \bar{Y} values)
3. shape of distribution of \bar{Y} becomes "normal" (symmetric around μ) as sample size n increases (Central Limit Theorem)

5

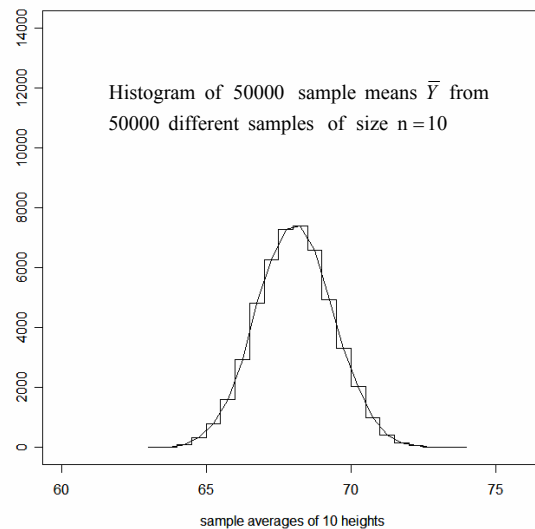
Histogram of averages from 50000 simple random samples of size n=5



Note: different samples can produce different sample means \bar{Y} but there is a typical/central value for \bar{Y}

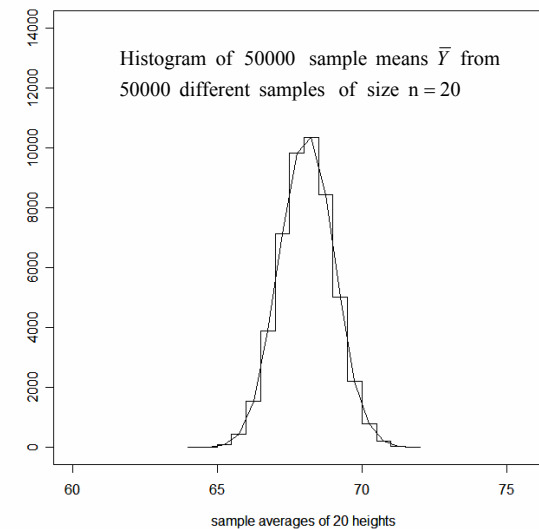
6

Histogram of averages from 50000 simple random samples of size n=10



7

Histogram of averages from 50000 simple random samples of size n=20



8

standard error: the *estimated* standard deviation of a statistic (e.g., sample average)

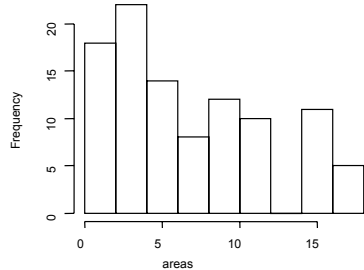
- The standard deviation of \bar{Y} is $\frac{\sigma}{\sqrt{n}}$
- The **standard error** of \bar{Y} is $\frac{s}{\sqrt{n}}$

The standard error estimates how spread out are all possible values of \bar{Y} away from their central/typical value μ (measures how close \bar{Y} is to what its trying to estimate – the population mean). The fact that we can get an estimate of a statistic's standard deviation by seeing only one sample is really quite amazing

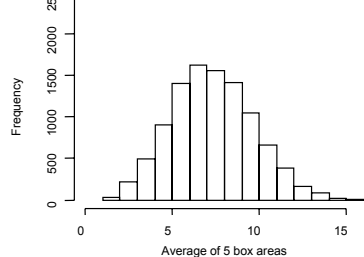
One of the great things about statistics is that we can use the information in a random sample to compute

1. statistics that estimate population parameters, AND
2. **standard errors** that tell us how far from the population parameters the statistics are likely to be (how precise are estimates)

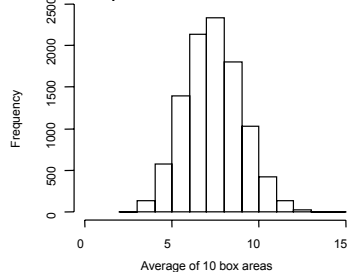
Histogram of areas of 100 boxes (population)



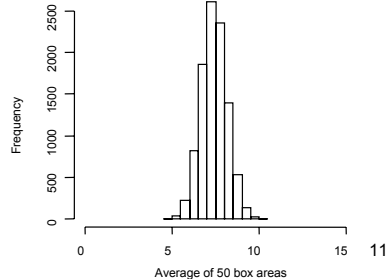
Histogram of Averages from 10000 random samples of size n=5



Histogram of Averages from 10000 random samples of size n=10



Histogram of Averages from 10000 random samples of size n=50



95% Confidence Intervals from Sixty Separate Samples, each of size n=10

μ is known to be 7.42 & 57 out of 60 intervals (57/60 = 95%) contain μ

lower	upper	μ inside?	lower	upper	μ inside?	lower	upper	μ inside?
2.79	11.81	YES	5.41	15.59	YES	3.38	12.02	YES
4.45	10.75	YES	3.73	12.27	YES	3.34	11.06	YES
5.59	10.81	YES	6.15	12.05	YES	5.97	15.03	YES
3.23	11.17	YES	3.73	11.87	YES	3.34	12.86	YES
2.38	11.02	YES	2.68	11.52	YES	7.22	14.38	YES
6.11	14.69	YES	4.14	12.26	YES	2.19	9.21	YES
4.06	11.74	YES	3.16	9.64	YES	2.99	12.21	YES
4.7	11.7	YES	4.47	13.13	YES	4.12	11.48	YES
3.35	8.85	YES	4.61	10.79	YES	3.29	10.11	YES
1.67	8.93	YES	5.66	15.74	YES	6.11	11.29	YES
3.86	12.74	YES	1.42	7.58	YES	3.96	11.44	YES
1.2	6.8	NO	3.93	10.07	YES	5.53	13.87	YES
2.17	7.43	YES	3.72	11.68	YES	7.56	15.64	NO
2.33	8.87	YES	3.4	10.6	YES	3.5	10.3	YES
4.84	13.76	YES	3.02	8.38	YES	3.23	12.37	YES
4.01	10.99	YES	2.64	11.96	YES	5.36	12.24	YES
3.69	10.11	YES	3.58	8.22	YES	2.95	10.85	YES
4.85	12.95	YES	1.4	8.0	YES	3.85	13.15	YES
2.42	9.58	YES	7	17	YES	4.67	11.73	YES ₁₂
7.94	15.66	NO	3.91	11.69	YES	3.39	11.81	YES