# Chapter 5: Statistical Inference (in General)

Shiwen Shen

University of South Carolina

2017 Summer

# Motivation

- In chapter 3, we learn the discrete probability distributions, including Bernoulli, Binomial, Geometric, Negative Binomial, Hypergeometric, and Poisson.
- In chapter 4, we learn the continuous probability distributions, including Exponential, Weibull, and Normal.
- In chapter 3 and 4, we always **assume** that we know the **parameter** of the distribution. For example, we know the mean $\mu$ and variance $\sigma^2$ for a normal distributed random varialbe, so that we can calculate all kinds of probabilities with them.

# Motivation

- For example, suppose we know the height of 18-year-old US male follows $N(\mu = 176.4, \sigma^2 = 9)$ in centimeters.
- Let $Y =$ the height of one 18-year-old US male.
- We can calculate $P(Y > 180) = 1-$pnorm(180, 176.4, 3)$=$ 0.115.
- However, it is natural that we do NOT know the population mean $\mu$ and population variance $\sigma^2$ in reality. What should we do?
- We use **statistical inference**!
- **Statistical inference** deals with making (probabilistic) statements about a population of individuals based on information that is contained in a sample taken from the population.

# Terminology: population/sample

- A **population** refers to the entire group of "individuals" (e.g., people, parts, batteries, etc.) about which we would like to make a statement (e.g., height probability, median weight, defective proportion, mean lifetime, etc.).
  - Problem: Population can not be measured (generally)
  - Solution: We observe a **sample** of individuals from the population to draw inference
  - We denote a random sample of observations by

$$Y_1, Y_2, \ldots, Y_n$$

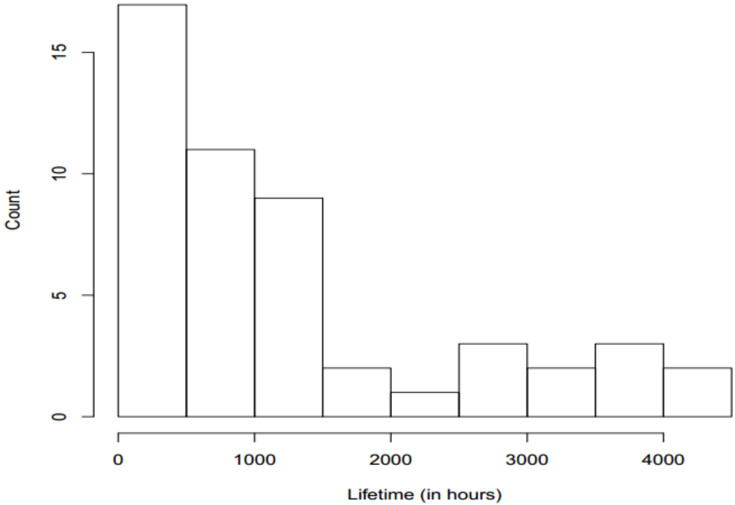
  - $n$ is the **sample size**
  - Denote $y_1, y_2, \ldots, y_n$ to be one realization of $Y_1, Y_2, \ldots, Y_n$.

# Example

*BATTERY DATA:* Consider the following random sample of $n = 50$ battery lifetimes $y_1, y_2, \ldots, y_{50}$ (measured in hours):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4285 | 2066 | 2584 | 1009 | 318 | 1429 | 981 | 1402 | 1137 | 414 |
| 564 | 604 | 14 | 4152 | 737 | 852 | 1560 | 1786 | 520 | 396 |
| 1278 | 209 | 349 | 478 | 3032 | 1461 | 701 | 1406 | 261 | 83 |
| 205 | 602 | 3770 | 726 | 3894 | 2662 | 497 | 35 | 2778 | 1379 |
| 3920 | 1379 | 99 | 510 | 582 | 308 | 3367 | 99 | 373 | 454 |

# A histogram of battery lifetime data



# Cont'd on battery lifetime data

The (empirical) distribution of the battery lifetimes is skewed towards the high side

- Which continuous probability distribution seems to display the same type of pattern that we see in histogram?
- An exponential($\lambda$) models seems reasonable here (based in the histogram shape). What is $\lambda$?
- In this example, $\lambda$ is called a (population) **parameter** (generally unknown). It describes the theoretical distribution which is used to model the entire population of battery lifetimes.

# Terminology: parameter

- A **parameter** is a numerical quantity that describes a *population*. In general, population parameters are unknown.
- Some very common examples are:
  - $\mu$ = population mean
  - $\sigma^2$ = population variance
  - $\sigma$ = population standard deviation
  - $p$ = population proportion
- Connection: all of the probability distributions that we talked about in previous chapter are indexed by population parameters.

# Terminology: statistics

- A **statistic** is a numerical quantity that can be calculated from a sample of data.
- Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from a population, some very common examples are:
  - **sample mean**:
    $$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$
  - **sample variance**:
    $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$
  - **sample standard deviation**: $S = \sqrt{S^2}$
  - **sample proportion**: $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ if $Y_i's$ are binary.

# Back to battery lifetime data

With the battery lifetime data (a random sample of $n = 50$ lifetimes),

$$\bar{y} = 1274.14 \text{ hours}$$
$$s^2 = 1505156 \text{ (hours)}^2$$
$$s \approx 1226.85 \text{ hours}$$

R code:

```
> mean(battery) ## sample mean
 [1] 1274.14
> var(battery) ## sample variance
 [1] 1505156
> sd(battery) ## sample standard deviation
 [1] 1226.848
```

# Parameters and Statistics Cont'd

*SUMMARY:* The table below succinctly summarizes the salient differences between a population and a sample (a parameter and a statistic):

| Comparison between parameters and statistics | |
|---|---|
| *Statistics* | *Parameters* |
| • Describes a sample | • Describes a population |
| • Always known | • Usually unknown |
| • Random, changes upon repeated sampling | • Fixed |
| • Ex: $\bar{X}, S^2, S$ | • Ex: $\mu, \sigma^2, \sigma$ |

# Statistical Inference

**Statistical inference** deals with making (probabilistic) statements about a population of individuals based on information that is contained in a sample taken from the population. We do this by

- **estimating** unknown poopulation parameters with sample statistics.
- quantifying the **uncertainty** (variability) that arises in the estimation process.

# Point estimators and sampling distributions

- Let $\theta$ denote a population parameter.
- A **point estimator** $\hat{\theta}$ is a statistic that is used to estimate a population parameter $\theta$.
- Common examples of point estimators are:
  - $\widehat{\theta} = \overline{Y} \longrightarrow$ a point estimator for $\theta = \mu$
  - $\widehat{\theta} = S^2 \longrightarrow$ a point estimator for $\theta = \sigma^2$
  - $\widehat{\theta} = S \longrightarrow$ a point estimator for $\theta = \sigma$
- Remark: In general, $\widehat{\theta}$ is a statistic, the value of $\widehat{\theta}$ will vary from sample to sample.

# Terminology: sampling distribution

- The distribution of an estimator $\widehat{\theta}$ is called its **sampling distribution**.
- A sampling distribution describes mathematically how $\widehat{\theta}$ would vary in repeated sampling.
- What is a good estimator? And good in what sense?

# Evaluate an estimator

- **Accuracy:** We say that $\widehat{\theta}$ is an **unbiased estimator** of $\theta$ if and only if
$$E(\widehat{\theta}) = \theta$$
- *RESULT:* When $Y_1, \ldots, Y_n$ is a random sample,
$$E(\overline{Y}) = \mu$$
$$E(S^2) = \sigma^2$$
- **Precision:** Suppose that $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are unbiased estimators of $\theta$. We would like to pick the estimator with smaller variance, since it is more likely to produce an estimate close to the true value $\theta$.

# Evaluate an estimator: cont'd

- *SUMMARY:* We desire point estimators $\widehat{\theta}$ which are **unbiased** (perfectly accurate) and have **small variance** (highly precise).
- *TERMINOLOGY:* The **standard error** of a point estimator $\widehat{\theta}$ is equal to
$$se(\widehat{\theta}) = \sqrt{\text{var}(\widehat{\theta})}.$$
- Note:
$$\text{smaller } se(\widehat{\theta}) \iff \widehat{\theta} \text{ more precise.}$$

# Evaluate an estimator: cont'd

Which estimator is better? Why?