Inference part 1

STAT-S520

Arturo Valdivia

02-28-23

- These slides complement material from ISI Chapter 9

So far, we have a random sample $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} \mathbb{P}$ with $EX_i = \mu$ and $VarX_i = \sigma^2$ for $i = 1, \ldots, n$ and

$$Y = \sum_{i=1}^{n} X_i$$

and

$$\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- If we know $\mathbb{P}$, we know everything about Y and $\bar{X}_n$.

The inference problem (in terms of μ)

- Statistical inference describes any procedure for extracting information about a probability distribution (population) from an observed sample. Let's focus, for example, in the population mean, μ
- We want to determine whether or not a claim about μ is correct or estimate as accurately as possible the value of μ

Types of statistical inference

- Point estimation
- Hypothesis testing
- Set estimation

Examples

- We believe drivers are speeding near a school zone. We take a random sample of the speed of cars passing by to determine what is their average speed limit and use it to conclude whether or not drivers are speeding on the school zone.
- We do not believe on psychic powers but someone claims to be a psychic. To test this, we use a sign that randomly shows left or right, and ask the prospective psychic to guess whether the sign is showing left of right. We do this 20 times and use this information to determine whether the individual is a psychic.
- I would like to know what time the entire population of STAT-S 520 students go to sleep, on average. I randomly select a sample of students and find the sample average of sleeping times.

Point Estimation: Terminology

↳ Let's show this for μ

(μ population mean)
(parameter)

This is μ

- Estimand: The value we are trying to estimate
- Estimator: The method or process used to estimate a parameter (μ)
- Estimate: The observed value obtained using the method

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(This is a random variable)

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Point Estimation

Point estimation is a method used to obtained a single value as estimator or estimate

Point estimators for μ

$\bar{X}_n$ is one estimator of μ

X_1 is another estimator of μ

Properties of Estimators

- **Unbiasedness** An estimator is unbiased if its expected value is equal to the estimand

$$E(\bar{X}_n) = \mu$$ (shown last week)

$\hookrightarrow$ so $\bar{X}_n$ is unbiased.

$$E(X_1) = \mu$$ X_1 is also unbiased

- **Consistency** An estimator is consistent if it converges in probability to the estimand

$$\bar{X}_n \xrightarrow{p} \mu$$ (by definition of the WLLN)

X_1 is not consistent.

Example 2

- The plug-in estimate for the variance is

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x}_n)^2}{n}$$

The plug-in estimator is biased:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] = \frac{n-1}{n}\sigma^2 < \sigma^2$$

- The sample variance is the estimate

s^2 is unbiased.

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

with corresponding estimator

$$E\left[S_n^2\right] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] = \cancel{\frac{n-1}{n-1}}\,\sigma^2 = \sigma^2$$

Example 2 Simulation

σ^2 estimator

```
mu = 0; sigma2 = 25; sigma = sqrt(sigma2); n = 60
plug.sd = function(x) sqrt(1/n*sum((x - mean(x))^2))
set.seed(1003)
vec_plug_sd =replicate(10^4, plug.sd(rnorm(n, mu, sigma)))
mean(vec_plug_sd)
```

```
#> [1] 4.949723
```

```
set.seed(1003)
vec_sample_sd =replicate(10^4, sd(rnorm(n, mu, sigma)))
mean(vec_sample_sd)
```

```
#> [1] 4.991494
```