Chapter 12 - Confidence Interval of a Mean

Laura Saba

July 12, 2017

What this chapter covers

- Interpreting a CI of a mean
- What values determine the CI of a mean
- Assumptions: CI of a mean
- How to calculate the CI of a mean
- More about confidence intervals

INTERPRETING A CI OF A MEAN

- A **confidence interval** is a range of values in which we think the value of a population parameter falls.
- The **confidence level** of a confidence interval is the probability that confidence intervals built using an appropriate method will contain the population parameter.
  - This probability is not for a specific interval but for the intervals in general.
  - A specific interval will either contain the population parameter or it will not – the probability is 0 or 1.
  - The totality of intervals containing the population parameter will match the confidence level.

Illustrative example for interpreting CI

- Suppose that the true mean amount of sperm released by male Japanese quail is 15 million sperm.
- Suppose we take a random sample of male Japanese quail and measure the amount of sperm they release and construct a 95% confidence interval.
- Say the interval ranges from 12 million to 22 million sperm. This interval contains the true mean value.
- Suppose we repeat the experiment 49 more times and construct 95% confidence intervals.
- About 95% of the intervals (2 to 3 of the 50 intervals) will contain the true mean number of sperm and about 5% will not.

Illustrative example for interpreting CI

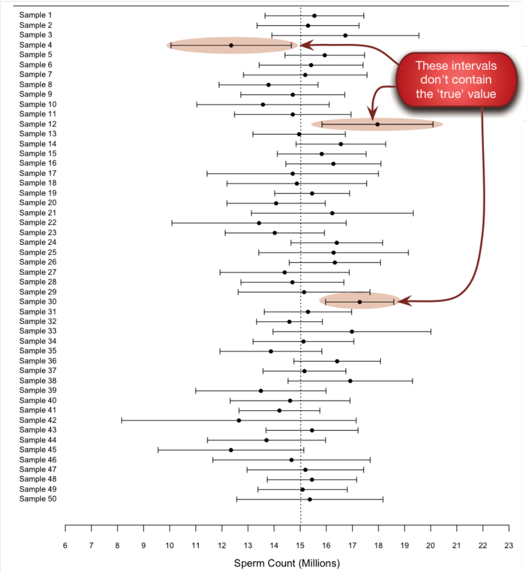


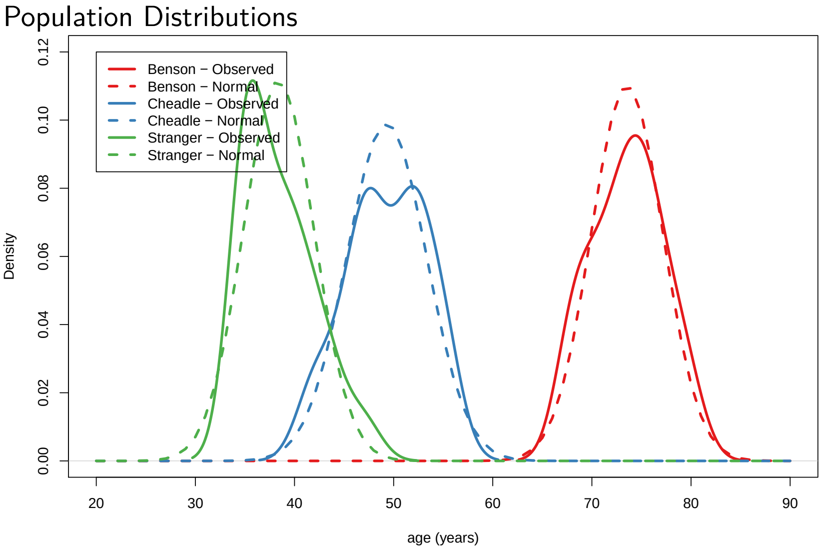
Figure 1: Interpreting CI

R code for calculating confidence interval

```r
values = c(29, 35, 32, 41, 30)
t.test(values)
```

```
##
##  One Sample t-test
##
## data:  values
## t = 15.472, df = 4, p-value = 0.0001018
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  27.40648 39.39352
## sample estimates:
## mean of x
##      33.4
```

Age Guessing Example



Age Guessing Example

```
summary(ageGuess$Don.Cheadle)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    42.0    47.0    49.0    49.3    52.0    55.0
```

```
summary(ageGuess$Bruce.Benson)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   68.00   70.75   73.50   73.55   76.00   80.00
```

```
summary(ageGuess$Stranger)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   35.00   35.00   37.50   38.40   40.25   47.00
```

Calculating many confidence intervals

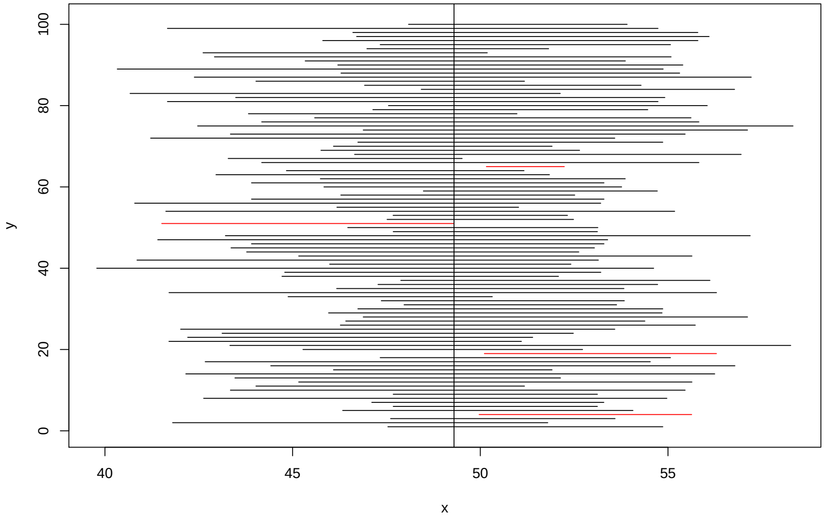
```
set.seed(8392789)
num_random = 100
pop_mean = mean(ageGuess$Don.Cheadle)

all.ci = data.frame()
for(i in 1:num_random){
  random_sample = sample(ageGuess$Don.Cheadle,5,
                         replace=TRUE)
  ci = t.test(random_sample)$conf.int
  all.ci = rbind(all.ci,ci)
}

all.ci$contains_pop_mean = all.ci[,1]<pop_mean &
  all.ci[,2]>pop_mean
sum(all.ci$contains_pop_mean)/nrow(all.ci)
```

```
## [1] 0.96
```

Plot All CI

```r
for (i in 1:nrow(all.ci)){
  x = as.numeric(all.ci[i,1:2])
  y = c(i,i)
  color = c("red","black")[all.ci$contains_pop_mean[i]+1]
  if(i==1) plot(x,y,
                type="l",
                ylim=c(0,nrow(all.ci)+1),
                xlim = c(min(all.ci[,1]),max(all.ci[,2])),
                col=color)
  if(i!=1) points(x,y,
                  type="l",
                  col=color)
}
abline(v=pop_mean)
```

Plot All CI



WHAT VALUES DETERMINE THE CI OF A MEAN

1. **Sample Mean**. Estimate of population mean and center of CI
2. **SD**. The SD is proportional to the width of the CI
3. **Sample Size**. The sample size is inversely proportional to the width of the CI
4. **Degree of Confidence**. The confidence level of the interval, usually 95%. As the degree of confidence increase, the width of the CI increases

ASSUMPTIONS: CI OF A MEAN

1. Random (or representative) sample
2. Independent observations
3. Accurate data
4. Assessing an event that you really care about
5. The population is distributed in a Gaussian manner, at least approximately

What if an assumption is violated

- In many situations, these assumptions are not strictly true.
- If an assumption is violated, the CI will probably be too optimistic (too narrow).

HOW TO CALCULATE THE CI OF A MEAN

W = margin of error of the CI

$t*$ = constant from the t distribution

s = sample standard deviation

n = number of values in the sample

m = sample mean

$W = \frac{t* \times s}{\sqrt{n}}$

CI: $m - W$ to $m + W$

Demonstration of CI formula in R

```
m = mean(ageGuess$Don.Cheadle)
s = sd(ageGuess$Don.Cheadle)
n = length(ageGuess$Don.Cheadle)
cl = 0.95

#Constant from the t distribution
t_star = qt(p=1 - (1-cl)/2,df=n-1)

#Margin of error of the CI
W = t_star*s/sqrt(n)

#CI
c(m-W,m+W)
```

```
## [1] 47.41505 51.18495
```

Confidence interval from t.test function

```r
t.test(ageGuess$Don.Cheadle)
```

```
##
##  One Sample t-test
##
## data:  ageGuess$Don.Cheadle
## t = 54.742, df = 19, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  47.41505 51.18495
## sample estimates:
## mean of x
##      49.3
```

MORE ABOUT CONFIDENCE INTERVALS

One-sided CIs:

```
m = mean(ageGuess$Don.Cheadle)
s = sd(ageGuess$Don.Cheadle)
n = length(ageGuess$Don.Cheadle)
cl = 0.95

#Constant from the t distribution
t_star = qt(p=cl,df=n-1)
#Margin of error of the CI
W = t_star*s/sqrt(n)
#CI
m+W
```

```
## [1] 50.85723
```

95% confidence that the population mean is less than 50.9

Other Types of CI

CI of the SD

- A CI value can be determined for nearly any value you calculate from a sample of data.
- The CI of the SD can be calculated using the Chi-Square distribution

CI of the Geometric Mean

- Calculate confidence interval on the logged data
- Take the antilog of the estimated CI

The CI and the SEM

Standard error of the mean (SEM) $= s/\sqrt(n)$

Margin of error of CI (W) $= t * \times SEM$

What did we learn

1. A confidence interval of the mean shows you how precisely you have determined the population mean.
2. If you compute 95% confidence intervals from many samples, you expect that 95% will include the true population mean and 5% will not.
3. The CI of the mean is computed from the sample mean, the sample SD, and the sample size.
4. The CI does NOT display the scatter of the data.
5. As the confidence level increases, the CI will be larger.
6. Larger samples have narrower confidence itnervals than smaller samples with the same SD.