Confidence intervals for means

Boniface Karia

February 17, 2016

Learning objectives

1. Describe the population mean and standard deviation
2. Calculate the mean and standard deviation from the sample data
3. Describe the standard error of the mean
4. Calculate confidence intervals for sample means

Situation

1. Want to estimate the actual population mean
2. But can only get the sample mean
3. Find a range of values L and U, $L < \mu < U$, that we can be really confident contains μ
4. This range of values is called a "confidence interval"

Calculating the standard error of a mean

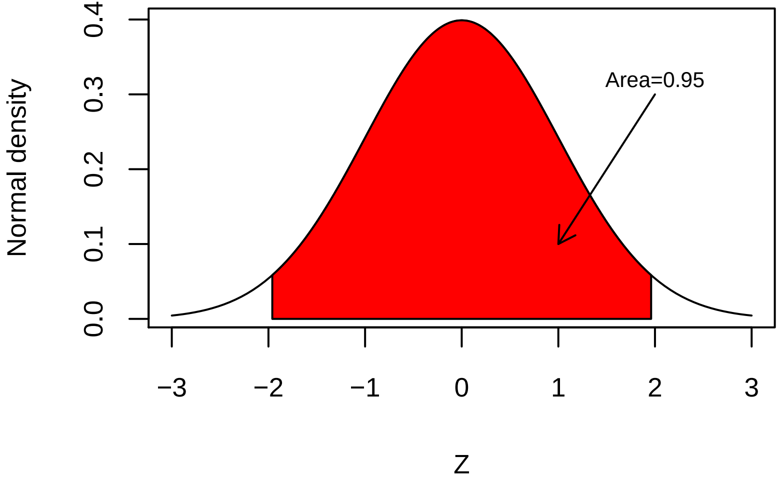
1. Sampling distribution, why is it required?
2. Standard error of the mean

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}}$$

Where,

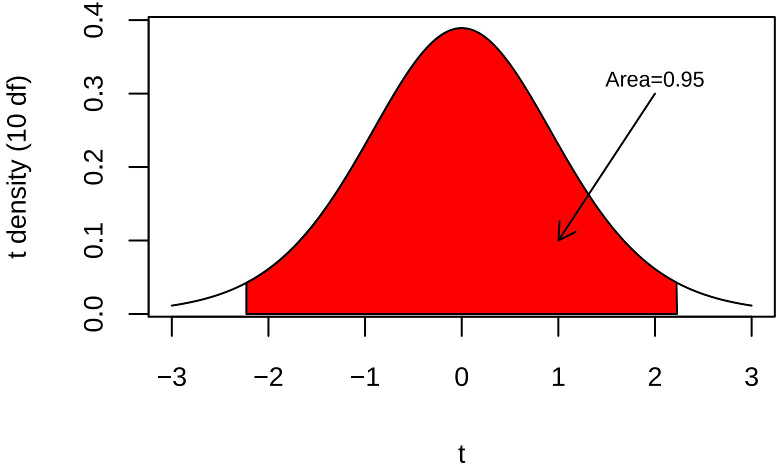
- σ_μ is the standard error of the mean
- σ is the standard deviation of the original distribution
- n is the sample size

Calculating the CI of a mean when σ is known



- Lower limit $= \bar{x} - Z_{\frac{\alpha}{2}} \times \sigma_{\mu}$, Upper limit $= \bar{x} + Z_{\frac{\alpha}{2}} \times \sigma_{\mu}$
- For a 95% CI (i.e $\alpha = 0.05$) we have $\bar{x} \pm 1.96 \times \sigma_{\mu}$

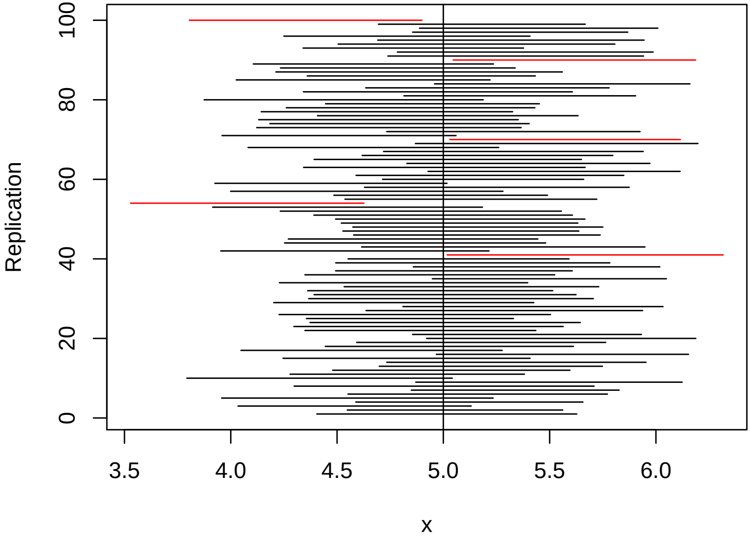
Calculating the CI of a mean when σ is unknown



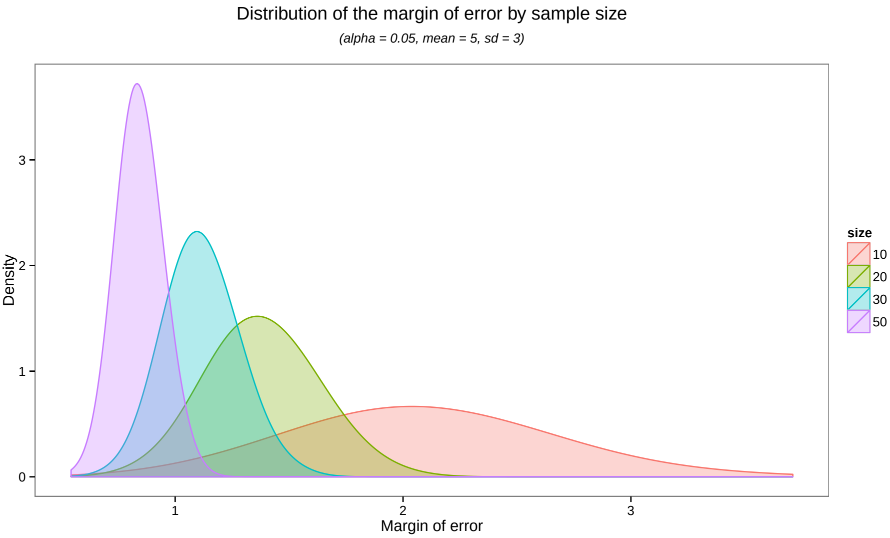
Lower limit $= \bar{x} - t_{(\frac{\alpha}{2},n-1)} \times \sigma_\mu$, Upper limit $= \bar{x} + t_{(\frac{\alpha}{2},n-1)} \times s_\mu$

For a 95% CI (i.e $\alpha = 0.05$) we have $\bar{x} \pm t_{(0.025,n-1)} \times s_\mu$

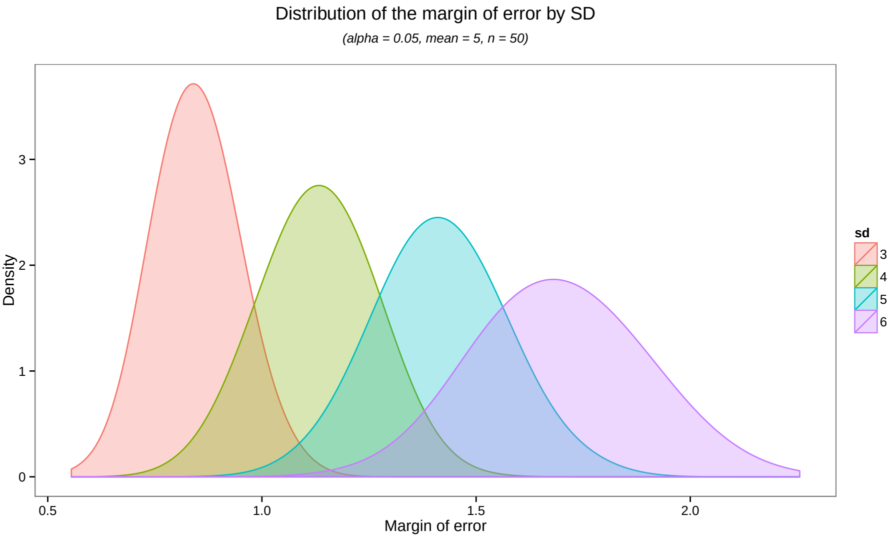
Interpretation of the CI



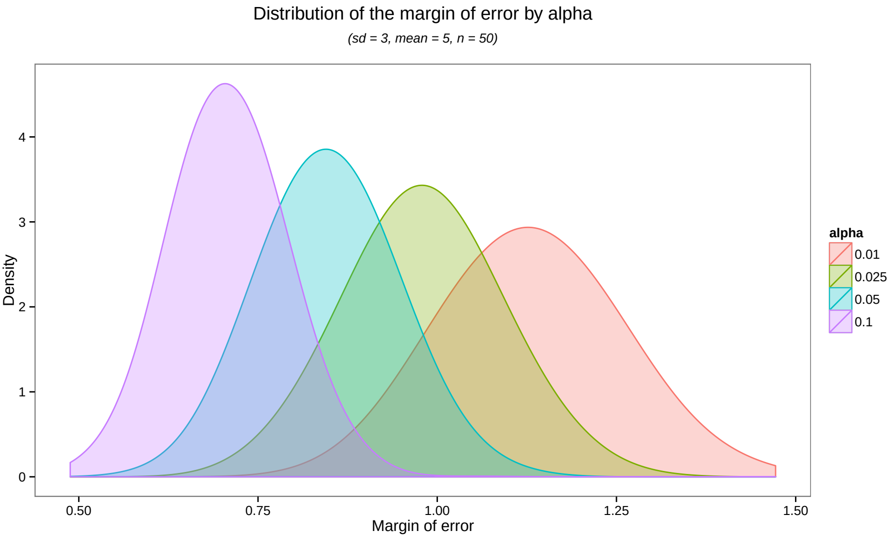
Sample size and the CI



Standard deviation and the CI



Confidence level and the CI



Summary

- The confidence interval expresses the uncertainty in sample estimates of means, proportions, treatment efficacy etc.
- Larger the sample, the narrower the CI
- Can improve the probability of including the population mean by calculating 99% interval but at a cost of having a wider interval and thus greater uncertainty
- Should design studies to yield CI's that are narrow enough to draw conclusions
- Interpretation depends on the assumption that the sample was representative of the population

Part 2: Confidence intervals for means - Practical

1. Suppose we have a random sample of 10 recently graduated students who we asked about their mothly salary in Kenyan shillings. Imagine that this is the data we see, 94617,70606, 47594, 27026, 11078, 38898, 45033, 87151, 120514, and 40000. Estimate the mean salary of the graduated children. Find a 90 and 95 % confidence interval for the mean under the following two scenarios:

- **Setting 1:** Assume that incomes are normally distributed with unknown mean and SD = Ksh 30,000.
- **Setting 2:** Same problem, only now we do not know the value for the standard deviation.

1. An incomplete R code is provided below, complete the code (commented lines with ??) based on the knowledge you have gathered about calculation of confidence intervals for means during the lecture.

Calculating Confidence Interval Using R

```
#--------------------------------
#  Setting 1
#--------------------------------

# input salaries
salaries <-c(94617,70606, 47594, 27026, 11078, 38898, 45033, 8
n=length(salaries)  # sample size

# calculate mean salary
mean.salary <- mean(salaries)
mean.salary
```

```
## [1] 58251.7
```

```r
# 95% CI
#----------------

# significance level for 95% CI
alpha <- 0.05

# calculate standard error of mean salary
se.mean.salary <- 30000/sqrt(n)

# calculate the margin of error for the 95% CI
error.margin <- abs(qnorm(alpha/2))*se.mean.salary

# calculate 95% CI
ci <- mean.salary + c(-error.margin, error.margin)
ci
```

```
## [1] 39657.85 76845.55
```

```
# 90% CI
#----------------

# alpha <- ??
# se.mean.salary <- ??
# error.margin <- abs(qnorm(alpha/2))*se.mean.salary
# ci  <- mean.salary + c(-error.margin, error.margin)
# ci
```

2. Report and interpret both the 90% and 95% confidence intervals that you obtain from setting 1

```r
#-----------------------------------
#  Setting 2
#-----------------------------------
# input salaries
salaries <-c(94617,70606, 47594, 27026, 11078, 38898, 45033, 8
n=length(salaries)  # sample size

# calculate mean salary
mean.salary <- mean(salaries)
mean.salary
```

```
## [1] 58251.7
```

```r
# 95% CI
#----------------

# significance level for 95% CI
alpha <- 0.05

# calculate standard error of mean salary
# se.mean.salary <- ??

# calculate the margin of error for the 95% CI
# error.margin <- abs(qt(alpha/2, df=n-1))*se.mean.salary
# calculate 95% CI
# ci <- mean.salary + c(-error.margin, error.margin)
# ci
```

```r
# 90% CI
#---------------
# alpha <- ??
# se.mean.salary <- ??
# error.margin <- ??
# ci <- mean.salary + c(-error.margin, error.margin)
# ci
```

3. Report and interpret both the 90% and 95% confidence intervals that you obtain from setting 1
4. Which if the 95% confidence intervals between setting 1 and 2 is wider and why do you think is the reason for this difference.