

Chapter 8: Confidence Intervals

Wong Shujia

Contents

1	Confidence Intervals for Population Means	2
1.1	Normal Distribution with Known Population Variance	3
1.2	Normal Distribution with Unknown Population Variance	5
1.3	Confidence Interval for the Difference in Population Means	7
2	Confidence Intervals for Population Variances	10
2.1	Confidence Intervals for Variance	10
2.2	Confidence Intervals for the Ratio of Population Variances	11
3	Confidence Intervals Based on Large Samples	11
3.1	Confidence Interval for the Population Proportion	11
3.2	Confidence Interval for a Difference in Population Proportions	12
3.3	Confidence Interval for the Mean of a Poisson Distribution	12

1 Confidence Intervals for Population Means

Introduction

- A plausible range of values for the population parameter is called a *confidence interval*.
- Using only a point estimate to estimate a parameter is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net.



Point Estimate



Interval Estimate

Definition 1 (Confidence Interval). Let $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ be a random sample from a population $X \sim f(x|\theta)$, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ be the observed values, and θ be an unknown parameter.

Suppose that we can find $L(\mathbf{X})$ and $U(\mathbf{X})$ such that

$$P(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})) = 1 - \alpha$$

Then $[L(\mathbf{x}), U(\mathbf{x})]$ is called a **confidence interval** for θ , $(1 - \alpha)\%$ is called the **confidence level**.

- $\alpha = 0.05$ is a standard 95% confidence interval.
- Interpret: the *random interval* will overlap the parameter θ 95% of the time.
- If $P(L(\mathbf{X}) \leq \theta) = 1 - \alpha$ or $P(\theta \geq U(\mathbf{X})) = 1 - \alpha$, they are called **one-sided** confidence intervals.

Common misconceptions

1. The probability that a confidence interval $[L(\mathbf{x}), U(\mathbf{x})]$ contains the true population parameter is $1 - \alpha$.
 - This is incorrect, CIs are part of the frequentist paradigm and as such the population parameter is fixed but unknown. Consequently, the probability any given CI contains the true value must be 0 or 1 (it does or does not).

2. A narrower confidence interval is always better.
 - This is incorrect since the width is a function of both the confidence level and the standard error.
3. A wider interval means less confidence.
 - This is incorrect since it is possible to make very precise statements with very little confidence.

1.1 Normal Distribution with Known Population Variance

Constructing CIs for μ

Let $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ (σ^2 is known), find the CI for μ .

To obtain a confidence interval for μ , note that

$$Q(\mathbf{X}, \mu) = \frac{\bar{X} - \mu}{\text{SE}(\bar{X})} \sim N(0, 1)$$

where $\text{SE}(\bar{X}) = \sigma/\sqrt{n}$ is the *standard error* of \bar{X} .

$$P(-z_{1-\alpha/2} \leq Q(\mathbf{X}, \mu) \leq z_{1-\alpha/2}) = 1 - \alpha$$

$$P(\bar{X} - z_{1-\alpha/2}\text{SE}(\bar{X}) \leq \mu \leq \bar{X} + z_{1-\alpha/2}\text{SE}(\bar{X})) = 1 - \alpha$$

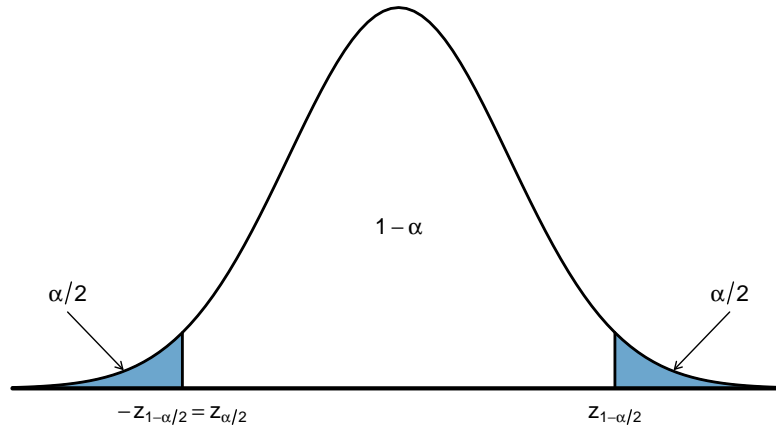
Hence, the $(1 - \alpha)$ CI for μ is

$$CI_{1-\alpha}(\mu) = \left[\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Pivotal Quantity

A random variable $Q(\mathbf{X}, \theta)$ is a **pivotal quantity** or **pivot** if the distribution of Q is independent of the parameter θ .

$$Q = (\bar{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$$



Confidence Interval for μ : Grocery Spending

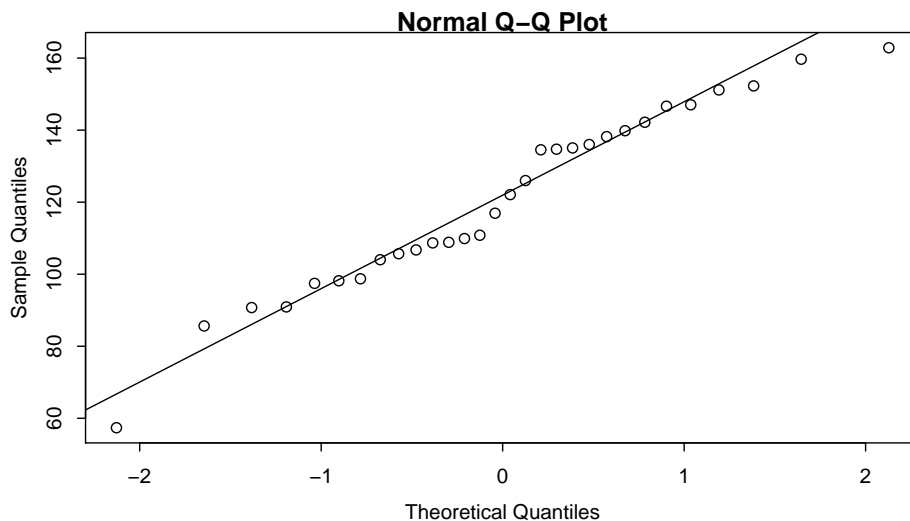
The consumer expenditure survey, created by the U.S. Department of Labor, was administered to 30 households in Watauga County, North Carolina. The amount of money each household spent per week on groceries is stored in the data frame GROCERY.

1. Construct a 97% confidence interval for the true mean weekly grocery expenditure for Watauga County households. Historical records indicate that the variance for grocery expenditure per household in Watauga County is 900 dollars.
2. A grocery chain is considering building a new grocery store in Watauga County. However, it will only do so if it is 99% confident the average amount spent on groceries each week is at least \$105. Does a $UCI_{0.99}(\mu)$ include \$105? If so, what does that imply?

Verify the assumption of normality

Checking normality assumption: QQ-plot, qqnorm()

```
> library("PASWR2")
> with(data = GROCERY, qqnorm(amount))
> with(data = GROCERY, qqline(amount))
```



Construct a 97% confidence interval for the true mean

Compute directly or by use of R function: z.test()

$$CI_{1-\alpha}(\mu) = \left[\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

```
> xbar <- mean(GROCERY$amount); z <- qnorm(0.985)
> CI1 <- xbar + c(-1, 1) * z * sqrt(900)/sqrt(30); CI1
## [1] 108.7473 132.5194
> CI2 <- z.test(GROCERY$amount, sigma.x = sqrt(900), n.x = 30,
+               conf.level = .97)$conf
> CI2
```

```
## [1] 108.7473 132.5194
## attr(,"conf.level")
## [1] 0.97
```

Q2: one-sided 99% CI

```
> z.test(GROCERY$amount, sigma.x = sqrt(900), n.x = 30,
+         conf.level = .99, alternative = "greater")$conf

## [1] 107.8914      Inf
## attr(,"conf.level")
## [1] 0.99
```

Determining Required Sample Size

How to determine the minimum required sample size to be within a given distance of μ when estimating the population mean with known variance σ^2 .

$$P(|\bar{X} - \mu| \leq z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Denote B as the *margin of error* (or *the bound on the error*) such that

$$P(|\bar{X} - \mu| \leq B) = 1 - \alpha$$

$$B = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

hence

$$n = \left(\frac{z_{1-\alpha/2} \sigma}{B} \right)^2$$

Example: Determine the required sample size

Determine the required sample size to estimate the true value of μ within ± 0.02 with a confidence level of 95% when sampling from a normal distribution with ($\sigma = 0.1$).

Solution: $B = 0.02$

$$n = \left(\frac{z_{1-\alpha/2} \sigma}{B} \right)^2 = \left(\frac{1.96 \times 0.1}{0.02} \right)^2 = 96.04$$

The margin of error is no more than 0.02 is $n = 97$.

1.2 Normal Distribution with Unknown Population Variance

Constructing CIs for μ

Let $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ (σ^2 is unknown), find the CI for μ .

To obtain a confidence interval for μ , note that

$$Q(\mathbf{X}, \mu) = \frac{\bar{X} - \mu}{\text{SE}(\bar{X})} \sim t(n-1)$$

where $\hat{SE}(\bar{X}) = S/\sqrt{n}$ is an estimator of $SE(\bar{X}) = \sigma/\sqrt{n}$.

$$P(\bar{X} - t_{1-\alpha/2}(n-1)\hat{SE}(\bar{X}) \leq \mu \leq \bar{X} + t_{1-\alpha/2}(n-1)\hat{SE}(\bar{X})) = 1 - \alpha$$

Hence, the $(1 - \alpha)$ CI for μ is

$$CI_{1-\alpha}(\mu) = \left[\bar{x} - t_{1-\alpha/2}(n-1) \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2}(n-1) \frac{s}{\sqrt{n}} \right]$$

R function: `t.test()`\$conf

Prediction Intervals

1. A *confidence interval* estimates the mean: $CI_{1-\alpha}(\mu)$
2. A *prediction interval* answers the question:

What range of values are plausible for a single future observation?

100(1 - α)% of the data lie in the interval

$$\bar{x} \pm t_{1-\alpha/2}(n-1)s\sqrt{1 + \frac{1}{n}}$$

CIs for CEO Compensation

USA Today publishes annual data on S&P 500 companies CEO compensation. 2010 dataset: ceo.txt.

Total compensation: "Total" = "Salary" + "Bonus" + "Stock".

- Find the 95% confidence interval for the CEO average compensation.

```
> ceo<-read.csv("R/ceo.txt")
> mean(ceo$Total)

## [1] 10531441

> t.test(ceo$Total)$conf

## [1] 9141027 11921855
## attr(,"conf.level")
## [1] 0.95
```

CIs for Salary proportion

Calculate the 95% CI for proportion of total compensation due to salary, bonus and stock .

```
> attach(ceo)
> Salary_ratio<-Salary/Total; t.test(Salary_ratio)$conf.int

## [1] 0.1359473 0.1841172
## attr(,"conf.level")
## [1] 0.95

> Bonus_ratio<-Bonus/Total; t.test(Bonus_ratio)$conf.int
```

```
## [1] 0.258932 0.307346
## attr(,"conf.level")
## [1] 0.95

> Stock_ratio<-Stock/Total; t.test(Stock_ratio)$conf.int

## [1] 0.5516663 0.6011062
## attr(,"conf.level")
## [1] 0.95

> detach(ceo)
```

1.3 Confidence Interval for the Difference in Population Means

Goals in this section

1. Consider $X_1, X_2, \dots, X_{n_X} \sim N(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, \dots, Y_{n_Y} \sim N(\mu_Y, \sigma_Y^2)$, We want to construct a *confidence interval* for $\mu_X - \mu_Y$ when
 - σ_X^2, σ_Y^2 known (equal or not equal)
 - σ_X^2, σ_Y^2 unknown but assumed equal
 - σ_X^2, σ_Y^2 unknown and not equal
2. CIs for mean difference $D = \mu_X - \mu_Y$ with paired data: $(X_1, Y_1), (X_2, Y_2), \dots$

CIs for $(\mu_X - \mu_Y)$ with Known Variances

A confidence interval for $\mu_X - \mu_Y$ is easily derived using a pivotal quantity

$$Q = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\text{SE}(\bar{X} - \bar{Y})} \sim N(0, 1)$$

where

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

The $(1 - \alpha)$ CI for $\mu_X - \mu_Y$ is given by

$$CI_{1-\alpha}(\mu_X - \mu_Y | \sigma_X, \sigma_Y \text{ is known}) = [(\bar{x} - \bar{y}) \pm z_{1-\alpha/2} \text{SE}(\bar{X} - \bar{Y})]$$

R function: `z.test(x=,y=,sigma.x=,sigma.y=)$conf`

Example: CI for Calculus Difference: Variance Known

Data frame CALCULUS provide the mathematical assessment scores for 36 students enrolled in a biostatistics course according to whether or not the students had successfully completed a calculus course prior to enrolling in the biostatistics course.

- Construct a 95% confidence interval for the difference in the means of the mathematical assessment scores for students who had successfully completed a calculus course (X) and of those who had not (Y). Assume the distributions for X and Y have variances of 25 and 144, respectively.
- Determine if it is advantageous to take calculus prior to taking the biostatistics course.

R code: Using CI formula

```
> str(CALCULUS)

## 'data.frame': 36 obs. of 2 variables:
## $ score : int 82 90 85 87 86 79 85 92 89 82 ...
## $ calculus: Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...

> MEANS <- tapply(CALCULUS$score, CALCULUS$calculus, mean)
> MEANS

##      No      Yes
## 62.61111 86.94444

> pe <- MEANS[2] - MEANS[1]
> z <- qnorm(0.975)
> pe + c(-1, 1)*z*sqrt(25/18 + 144/18)

## [1] 18.32775 30.33892
```

R code: Using z.test()

```
> ScoreYesCalc <- subset(CALCULUS, select = score,
+   subset = calculus == "Yes", drop = TRUE)
> ScoreNoCalc <- subset(CALCULUS, select = score,
+   subset = calculus == "No", drop = TRUE)
> CI <- z.test(x = ScoreYesCalc, y = ScoreNoCalc, sigma.x = sqrt(25),
+   sigma.y = sqrt(144), conf.level = 0.95)$conf
> CI

## [1] 18.32775 30.33892
## attr(,"conf.level")
## [1] 0.95
```

CIs for $(\mu_X - \mu_Y)$ with Variances Unknown but Assumed Equal

The pivotal quantity

$$Q = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\hat{SE}(\bar{X} - \bar{Y})} \sim t_{1-\alpha/2}(n_X + n_Y - 2)$$

$$\hat{SE}(\bar{X} - \bar{Y}) = S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \text{ and } S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$$

The $(1 - \alpha)$ CI for $\mu_X - \mu_Y$ is given by

$$CI_{1-\alpha}(\mu_X - \mu_Y | \sigma_X = \sigma_Y \text{ but unknown}) \\ = [(\bar{x} - \bar{y}) \pm t_{1-\alpha/2}(n_X + n_Y - 2)\hat{SE}(\bar{x} - \bar{y})]$$

Culculus Difference: Variance Equal But Unknown

`t.test(x=,y=,var.equal=T)$conf` or `t.test(z~w,data,var.equal=T)`, where z is numeric and w is dichotomous.


```

> t.test(score ~ calculus,data=CALCULUS,var.equal=TRUE)$conf

## [1] -30.99621 -17.67046
## attr(,"conf.level")
## [1] 0.95

> ScoreYesCalc <- subset(CALCULUS, select = score,
+                          subset = calculus == "Yes", drop = TRUE)
> ScoreNoCalc <- subset(CALCULUS, select = score,
+                        subset = calculus == "No", drop = TRUE)
> t.test(x = ScoreYesCalc, y = ScoreNoCalc, var.equal=TRUE)$conf

## [1] 17.67046 30.99621
## attr(,"conf.level")
## [1] 0.95

```

CIs for $(\mu_X - \mu_Y)$ with Variances Unknown and Unequal Welch's approximation

$$Q = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\hat{\text{SE}}(\bar{X} - \bar{Y})} \sim t(\nu)$$

$$\hat{\text{SE}}(\bar{X} - \bar{Y}) = \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} \text{ and } \nu = \frac{\left(\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}\right)^2}{\frac{(S_X^2/n_X)^2}{n_X-1} + \frac{(S_Y^2/n_Y)^2}{n_Y-1}}$$

The $(1 - \alpha)$ CI for $\mu_X - \mu_Y$ is given by

$$\begin{aligned} CI_{1-\alpha}(\mu_X - \mu_Y | \sigma_X, \sigma_Y \text{ unknown and unequal}) \\ = [(\bar{x} - \bar{y}) \pm t_{1-\alpha/2}(\nu) \hat{\text{SE}}(\bar{x} - \bar{y})] \end{aligned}$$

Calculus Difference: Variances Unknown and Unequal

`t.test(x=y)$conf` or `t.test(z~w,data)`, where z is numeric and w is dichotomous.

```

> CALCULUS$calculus<-factor(CALCULUS$calculus,levels=c("Yes","No"))
> t.test(score ~ calculus,data=CALCULUS)$conf

## [1] 17.50677 31.15990
## attr(,"conf.level")
## [1] 0.95

> t.test(score ~ calculus,data=CALCULUS) # Two sample t-test

##
## Welch Two Sample t-test
##
## data: score by calculus
## t = 7.4219, df = 20.585, p-value = 3.04e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 17.50677 31.15990
## sample estimates:
## mean in group Yes mean in group No
## 86.94444 62.61111

```

CIs for $(\mu_X - \mu_Y)$ with Paired Samples

CIs for mean difference $\mu_D = \mu_X - \mu_Y$ with paired data: $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Assume $D_i = X_i - Y_i \sim N(\mu_D, \sigma_D^2) (i = 1, 2, \dots, n)$,

The $(1 - \alpha)$ CI for the mean difference $\mu_D = \mu_X - \mu_Y$ is

- If σ_D is known,

$$CI_{1-\alpha}(\mu_D) = \left[\bar{d} - z_{1-\alpha/2} \frac{\sigma_D}{\sqrt{n}}, \bar{d} + z_{1-\alpha/2} \frac{\sigma_D}{\sqrt{n}} \right]$$

- If σ_D is unknown

$$CI_{1-\alpha}(\mu_D) = \left[\bar{d} - t_{1-\alpha/2}(n-1) \frac{s_D}{\sqrt{n}}, \bar{d} + t_{1-\alpha/2}(n-1) \frac{s_D}{\sqrt{n}} \right]$$

R function: `t.test(x, y, paired = TRUE)$conf`

2 Confidence Intervals for Population Variances

2.1 Confidence Intervals for Variance

Constructing CIs for σ^2

Let $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ (μ and σ^2 are unknown), find the CI for σ^2 . The pivotal quantity

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$P(\chi_{\alpha/2}^2(n-1) \leq Q \leq \chi_{1-\alpha/2}^2(n-1)) = 1 - \alpha$$

The $(1 - \alpha)$ CI for σ^2 is

$$CI_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2(n-1)}, \frac{(n-1)s^2}{\chi_{\alpha/2}^2(n-1)} \right]$$

Example: CI for variance: Calculus data

```
> ScoreYesCalc <- subset(CALCULUS, select = score,
+                          subset = calculus == "Yes", drop = TRUE)
> n1 <- length(ScoreYesCalc); var1 <- var(ScoreYesCalc)
> L1 <- round((n1-1)*var1/qchisq(0.975, n1), 2)
> U1 <- round((n1-1)*var1/qchisq(0.025, n1), 2)
> paste("CI for variance of YesCalc: (", L1, ",", U1, ")")

## [1] "CI for variance of YesCalc: ( 10.05 , 38.51 )"
```

```
> ScoreNoCalc <- subset(CALCULUS, select = score,
+                       subset = calculus == "No", drop = TRUE)
> n2 <- length(ScoreNoCalc); var2 <- var(ScoreNoCalc)
> L2 <- round((n2-1)*var2/qchisq(0.975, n2), 2)
> U2 <- round((n2-1)*var2/qchisq(0.025, n2), 2)
> paste("CI for variance of NoCalc: (", L2, ",", U2, ")")

## [1] "CI for variance of NoCalc: ( 94.28 , 361.12 )"
```

2.2 Confidence Intervals for the Ratio of Population Variances

Constructing CIs for the Ratio of Variances

Consider $X_1, X_2, \dots, X_{n_X} \sim N(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, \dots, Y_{n_Y} \sim N(\mu_Y, \sigma_Y^2)$, We want to construct a *confidence interval* for $V = \sigma_X^2/\sigma_Y^2$.

The pivotal quantity

$$Q = \frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2} \sim F(n_Y-1, n_X-1)$$

The $(1 - \alpha)$ CI for $V = \sigma_X^2/\sigma_Y^2$ is

$$CI_{1-\alpha}(V) = [F_{\alpha/2}(n_Y-1, n_X-1) \frac{S_X^2}{S_Y^2}, F_{1-\alpha/2}(n_Y-1, n_X-1) \frac{S_X^2}{S_Y^2}]$$

Example: CI for variance Ratio: Calculus data

$$[F_{\alpha/2}(n_Y-1, n_X-1) \frac{S_X^2}{S_Y^2}, F_{1-\alpha/2}(n_Y-1, n_X-1) \frac{S_X^2}{S_Y^2}]$$

```
> VR <- var1/var2
> CI <- c(qf(0.025, n1-1, n2-1) * VR, qf(0.975, n1-1, n2-1) * VR)
> CI
## [1] 0.03988834 0.28506343
```

3 Confidence Intervals Based on Large Samples

3.1 Confidence Interval for the Population Proportion

Construct the CI for the Population Proportion

Let $X_1, X_2, \dots, X_n \sim \text{Bin}(1, \pi)$, find the CI for π .

Denote $P = \frac{1}{n} \sum X_i$ is the sample proportion and p is its observed value.

From the Central Limit Theorem, the pivotal quantity

$$Q = \frac{P - \pi}{\sqrt{P(1-P)/n}} \sim N(0, 1) \text{ as } n \rightarrow \infty$$

The $(1 - \alpha)$ CI for π is

$$CI_{1-\alpha}(\pi) = [p - z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}]$$

R function: `prop.test(x,n)$conf` or `binom.confint(x,n,method="")`

[fragile]Example: Biogen Idec, Inc

Biogen Idec, Inc is a global biotechnology company based in Cambridge, Massachusetts.

“Avonex delivers the highest rate of satisfaction: 95% among patients”

- The U.S. Food and Drug Administration (FDA) on October 30th, 2002 informed the biotech company Biogen to stop publishing misleading promotions for its drug Avonex.

- The FDA found that in a random sample of 75 patients surveyed, only 60% said they were "very satisfied" with Avonex.
- Who is right?

```
> prop.test(45,75)$conf
## [1] 0.4802630 0.7094213
## attr(,"conf.level")
## [1] 0.95
```

3.2 Confidence Interval for a Difference in Population Proportions

[fragile]Two Sample Test for Proportions

- A 95% confidence interval for a difference in proportions

$$CI_{1-\alpha}(\pi_1 - \pi_2) = (p_1 - p_2) \pm z_{1-\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- Two sample t -test for populations: `prop.test()`

Example: Pfizer

- During 1998 of the 6 million Viagra users 77 died from coronary problems such as heart attacks.
- Pfizer claimed that this rate is no more than the general population.
- A clinical study found 11 out of 1, 500, 000 men who were not on Viagra died of coronary problems during the same length of time.

```
> prop.test(x=c(11,77),n=c(1500000,6000000),alternative="greater",
+ conf.level=.95)$conf
## [1] -1.027715e-05 1.000000e+00
## attr(,"conf.level")
## [1] 0.95
```

3.3 Confidence Interval for the Mean of a Poisson Distribution

Construct the CI for the Mean of Poisson

Let $X_1, X_2, \dots, X_n \sim \text{Pois}(\lambda)$, find the CI for λ .

The MLE of λ is $\hat{\lambda} = \bar{X}$, with $E(\bar{X}) = \lambda$ and $\text{Var}(\bar{X}) = \lambda/n$. From the Central Limit Theorem,

$$Q = \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \sim N(0, 1)$$

The $(1 - \alpha)$ CI for λ is

$$CI_{1-\alpha}(\lambda) = \left[\bar{x} - z_{1-\alpha/2} \sqrt{\frac{\bar{x}}{n}}, \bar{x} + z_{1-\alpha/2} \sqrt{\frac{\bar{x}}{n}} \right]$$

Example: Average Goals in a Game

```
> n <- sum(!is.na(SOCCER$game)) # number of games
> xbar <- mean(SOCCER$goals, na.rm = TRUE)
> z <- qnorm(0.95) # z_{0.95}
> CI <- xbar + c(-1, 1) * z * sqrt(xbar/n)
> CI

## [1] 2.308439 2.648458
```