chap 02

Chris Parrish

July 1, 2016

Contents

chap 02

reference:
- ARM chapter 02, github

```
library(ggplot2)
```

ARM 2.3

CI for continuous data

```
y <- c(35, 34, 38, 35, 37)
n <- length(y)
estimate <- mean(y)
se <- sd(y) / sqrt(n)
int.50 <- estimate + qt(c(.25, .75), n - 1) * se
int.95 <- estimate + qt(c(.025, .975), n - 1) * se
```

CI for proportions

```
y <- 700
n <- 1000
estimate <- y / n
se <- sqrt (estimate * (1 - estimate) / n)
int.95 <- estimate + qnorm(c(.025, .975)) * se
```

CI for discrete data

```
y <- rep(c(0, 1, 2, 3, 4), c(600, 300, 50, 30, 20))
n <- length(y)
estimate <- mean(y)
se <- sd(y) / sqrt(n)
int.50 <- estimate + qt(c(.25, .75), n - 1) * se
int.95 <- estimate + qt(c(.025, .975), n - 1) * se
```

plot figure 2.3

polls data file

```
# 1 polls.dat - Gallup poll data support for death penalty by year - fig 2.3
#   5 columns per row
#   c1 : year
#   c2 : month
#   c3 : supports death penalty
#   c4 : opposed to death penalty
#   c5 : no opinion
```

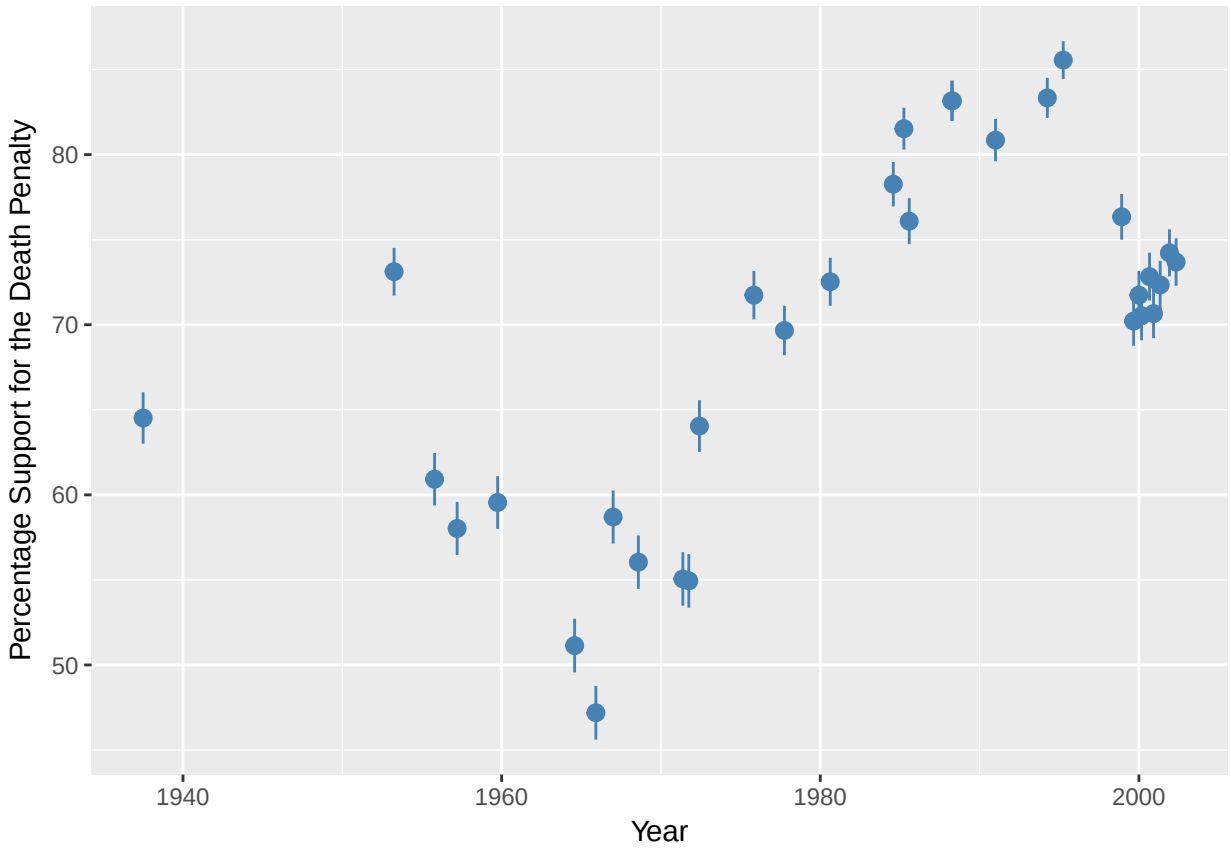
polls data

```
polls <- matrix (scan("polls.txt"), ncol = 5, byrow = TRUE)
support <- polls[, 3] / (polls[, 3] + polls[, 4])
year <-  polls[, 1] + (polls[, 2] - 6) / 12
y.se <- sqrt(support * (1 - support) / 1000)
y.max <- 100 * (support + y.se)
y.min <- 100 * (support - y.se)
limits <- aes(ymax = y.max, ymin = y.min)
frame1 = data.frame(year = year, support = support * 100)
str(frame1)
```

```
## 'data.frame':    32 obs. of  2 variables:
##  $ year   : num  2002 2002 2001 2001 2001 ...
##  $ support: num  73.7 74.2 72.3 70.7 72.8 ...
```

fig 2.3

```
ggplot(frame1, aes(x = year, y = support)) +
  geom_point() +
  scale_y_continuous("Percentage Support for the Death Penalty") +
  scale_x_continuous("Year") +
  theme_gray() +
  geom_pointrange(limits, color = "steelblue")
```



Weighted averages

```
N <- c(65633200, 80523700, 59685200) # population sizes FR, DE, IT
p <- c(0.55, 0.61, 0.38)  # estimated proportions of Yes responses
se <- c(0.02, 0.03, 0.03)

w.avg <- sum(N * p) / sum(N)
se.w.avg <- sqrt (sum ((N * se / sum(N)) ^ 2))
int.95 <- w.avg + c(-2, 2) * se.w.avg
```

CI using simulations

```
n.men <- 500
p.hat.men <- 0.75
se.men <- sqrt (p.hat.men * (1 - p.hat.men) / n.men)

n.women <- 500
p.hat.women <- 0.65
se.women <- sqrt (p.hat.women * (1 - p.hat.women) / n.women)

p.hat.men / p.hat.women
```

```
## [1] 1.153846
```

```r
n.sims <- 10000
p.men <- rnorm (n.sims, p.hat.men, se.men)
p.women <- rnorm (n.sims, p.hat.women, se.women)
ratio <- p.men / p.women
int.95 <- quantile (ratio, c(.025, .975))
int.95
```

```
##     2.5%    97.5%
## 1.062456 1.253227
```