Worksheet8

STAT414

2024-11-09

Month	Well 1 (Background)	Well 2 (Background)	Well 3 (Compliance)
1	4.2	5.2	9.4
2	5.8	6.4	10.9
3	11.3	11.2	14.5
4	7.0	11.5	16.1
5	7.3	10.1	21.5
6	8.2	9.7	17.6

Table 3.14 Copper data (ppb) from groundwater monitoring wells (USEPA, 1992c, p. 47)

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

```
library(EnvStats)
```

```
##
## Attaching package: 'EnvStats'

## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```
# 1. Consider the copper data shown in Table 3.14 of Exercise 3.7 in Chapter 3.
# Combine the data from Wells 1 and 2, and compute the pooled estimate of standard
# deviation, allowing for different means for the background and compliance well data.
Well1 = c(4.2, 5.8, 11.3, 7.0, 7.3, 8.2, 5.2, 6.4, 11.2, 11.5, 10.1, 9.7) #Background
Well2 = c(9.4, 10.9, 14.5, 16.1, 21.5, 17.6)  #Compliance

cat("
      n1 = well1 length
      n2 = well2 length
      s1 = well1 sd
      s2 = well2 sd")
```

```
##
##       n1 = well1 length
##       n2 = well2 length
##       s1 = well1 sd
##       s2 = well2 sd
```

```
n1 <- length(Well1)
n2 <- length(Well2)
s1 <- sd(Well1)
s2 <- sd(Well2)

pooled_sd <- sqrt(
  (((n1 - 1) * s1^2) + ((n2 - 1) * s2^2)) / (n1 + n2 - 2)
)
pooled_sd
```

```
## [1] 3.261797
```

```
# a. Based on this estimate of standard deviation, how many samples are required
# at the background and compliance wells in order to achieve a confidence interval
# half-width of 5 ppb?

#sample estimation: n >= (Z/h)^2 * (sigma/mu)^2
#since we are given half-width = 5 (MOE), use n >= (Z*s/halfwidth)^2

#choose alpha = 0.05
halfwidth <- 5
zscore <- 1.96 #alpha = .05

x1 <- zscore * pooled_sd
n_total <- (x1/halfwidth)^2
n_total <- ceiling(n_total)
n_total
```

```
## [1] 2
```

```
#I choose the result of CiNormN because it should be theoretically the same
#as my manual approach, but there is a discrepancy in my results. My manual
#sample size is n=2, while ciNormN fron EnvStats returns n=5.
sample_size <- ciNormN(half.width = halfwidth, sigma.hat = pooled_sd, conf.level = 0.95)
sample_size
```

```
## [1] 5
```

```
cat("Need", sample_size, "samples for the background and compliance wells to
    achieve a confidence interval with a MOE with a half-width of 5ppb.")
```

```
## Need 5 samples for the background and compliance wells to
##     achieve a confidence interval with a MOE with a half-width of 5ppb.
```

```r
# b. Repeat part a above, but assume the background well sample size is fixed
# at n2 = 12
sample_size2 <- ciNormN(half.width = halfwidth, sigma.hat = pooled_sd, conf.level = 0.95,
                        n2=12)
sample_size2
```

```
## $n1
## [1] 3
##
## $n2
## [1] 12
```

```r
cat("If n2 is fixed at 12, a sample size of", sample_size2$n1,"for n1 is needed
    to achieve a confidence interval with a half-width of 5ppb.")
```

```
## If n2 is fixed at 12, a sample size of 3 for n1 is needed
##     to achieve a confidence interval with a half-width of 5ppb.
```

```r
# 2. Sometimes when the results of an opinion poll are reported, the results are
# qualified by a statement like the following: "53% of those polled said they were
# willing to pay higher taxes to enforce stricter environmental standards. These
# results are accurate to within three percentage points." Assuming this statement
# means that the 95% confidence interval for the estimated proportion has a
# half-width of three percentage points, determine how many people must have been
# polled. Assume the estimated proportion is about 50%.

z <- 1.96 #alpha = .05, CL = 95%
p <- .5 #estimated proportion
hw <- .03 #half-width of 3 percentage points

#sample size= ((z * sqrt(p(1-p)) / E) ^2
x1 <- sqrt(p * (1-p))
x2 <- z * x1
x3 <- x2 / hw
n = ceiling(x3^2)

cat("At least", n, "people must have been polled")
```

```
## At least 1068 people must have been polled
```