Assignment 1

Sven Bergmann

September 26, 2023

Chapter 5

5.5

Problem

In 2003, the lab of Human Computer Interaction and Health Care Informatics at the Georgia Institute of Technology conducted empirical research on the performance of patients with diabetic retinopathy. The experiment included 29 participants placed either in the control group (without diabetic retinopathy) or the treatment group (with diabetic retinopathy). The visual acuity data of all participants are listed below. Normal visual acuity is 20/20, and 20/60 means a person sees at 20 ft what a normal person sees at 60 ft:

20/20, 20/20, 20/20, 20/25, 20/15, 20/30, 20/25, 20/20, 20/25, 20/80, 20/30, 20/25, 20/30, 20/50, 20/30, 20/20, 20/15, 20/20, 20/25, 20/16, 20/30, 20/15, 20/15, 20/25

The data of five participants were excluded from the table due to their failure to meet the requirement of the experiment, so 24 participants are counted in all. To verify if the data can represent the visual acuity of the general population, a 90 upper tolerance bound for 80 of the population is calculated.

Solution

We want to have a 90% upper tolerance bound for the 80th quantile. The upper bound is going to be X_r for some r. So

$$n = 24, p = 0.8, 1 - \alpha = 0.9 \Leftrightarrow \alpha = 0.1.$$

and we are looking for

$$P(X_{0.8} \leq X_{(r)}) \geq 1 - \alpha$$

with

```
n <- 24
p <- 0.8
alpha <- 0.1
```

The sorted values are:

```
values_sorted <- sort(c(20, 20, 20, 25, 15, 30, 25, 20, 25, 80, 30, 25, 30, 50, 30,
    20, 15, 20, 25, 16, 30, 15, 15, 25))
```

To be able to know the exact value for r, we need to create a vector from 1 to the length of the values:

```r
numbers <- seq_along(values_sorted)
```

Since

$$P(X_{0.8} \leq X_{(r)}) \Leftrightarrow P(T \leq r-1)$$

with

$$T \sim B(X, 0.8),$$

we are going to check for

$$P(T \leq r-1) \geq 0.9.$$

The first element that fulfills $P(T \leq r-1) \geq 0.9$ would be $X_{(r-1)}$, so in order to find $X_{(r)}$, we need to take the second element.

```r
pbinom_vals <- pbinom(numbers, size = n, prob = p)
upper_bound <- values_sorted[which(pbinom_vals >= 1 - alpha)[2]]
cat("The 90% upper tolerance bound is", upper_bound, "\n")
```

```
## The 90% upper tolerance bound is 50
```

5.10

Problem

How large must the sample be in order to have 95% confidence that at least 90% of the population is less than $X_{(n-1):n}$?

Solution

Here we have

$$p = 0.9, 1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05.$$

With the same reasoning as above in Problem 5, we are searching for two values $X_{(n-1)}$ and X_n for n such that

$$P(T \leq r-1) \geq 0.95$$

with

$$T \sim B(X, 0.9).$$

```r
for (n in 1:1000) {
    if (length(which(pbinom(1:n, size = n, prob = 0.9) >= 0.95)) == 3) {
        cat("The sample size must be greater or equal to", n, ".")
        break
    }
}
```

```
## The sample size must be greater or equal to 46 .
```

5.13

Problem

Find a 90% upper tolerance interval for the 99th percentile of a sample of size $n = 1000$.

Solution

With the same reasoning as above we are searching for all the values such that

$$P(T \leq r - 1) \geq 0.9$$

with

$$T \sim (X, 0.99)$$

The interval is half open because we searched for $r - 1$ so leaving the first element out delivers an interval starting from $X_{(r)}$.

```
n <- 1000
p <- 0.99
alpha <- 0.1
cat("The intervall is: (", which(pbinom(q = 1:n, size = n, prob = p) >= 1 - alpha),
    "]")
```

```
## The intervall is: ( 994 995 996 997 998 999 1000 ]
```