Lecture3-Extra

Elizabeth Colantuoni

2/2/2021

In this "handout-extra" I wanted to briefly review a few concepts to be sure we are all on the same page.

1. Population parameters vs. statistics

We are evaluating the Nepali Children's Anthropometry study. Define as the population of interest children who are in their third year of life, i.e. 25 to 36 month old children.

Assume that the arm circumferences of children in this population follow a normal distribution with mean μ and variance σ^2. Both μ and σ^2 are population parameters; scalars that describe features of the distribution of arm circumference in this population.

In general, the populations of interest to us are too large to conduct a census from which we could compute the value of μ and σ^2. Therefore, we estimate μ and σ^2 using a sample of size n drawn from the population. From the sample, we observe Y_i, the arm circumference for the i^{th} child from the sample. Based on the sample of size n, we can compute:

$$\hat{\mu} = n^{-1}\sum_{i=1}^{n} Y_i$$

$$\hat{\sigma}^2 = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2$$

$\hat{\mu}$ and $\hat{\sigma}^2$ are statistics which provide a point estimate for μ and σ^2.

Also, we know that $\hat{\mu}$ and $\hat{\sigma}^2$ depend on the particular sample of size n we observe from the population and n. Fixing n, we can imagine collecting all the possible values of $\hat{\mu}$ and $\hat{\sigma}^2$ from all the possible samples of size n from the population. The variance of these statistics is called the "statistical variance"; but we more often describe the standard deviation of these statistics called the "standard error".

Focusing on μ, we know from 651-2,

$$Var(\hat{\mu}) = n^{-1}\sigma^2$$

$$SE(\hat{\mu}) = \sqrt{n^{-1}\sigma^2}$$

In any given sample of size n, we can estimate $SE(\hat{\mu})$ by plugging in our estimate of σ^2 from the sample. We would refer to this value as $\hat{SE}(\hat{\mu}) = \sqrt{n^{-1}\hat{\sigma}^2}$.

2. Link back to Lecture 3 Handout

Let's pull all these ideas together within the context of a regression model. We will use our population of children within their 3rd year of life to **estimate**:

- the population mean arm circumference, μ
- the population standard deviation for arm circumference, σ
- the standard error for the sample mean arm circumference, $SE(\hat{\mu})$

We will do the estimation using simple calculations and show equivalence in a linear regression model.

Simple calculations

```
load("NepalAnthro.rdata")
d= nepal.anthro %>% select(.,arm,age,num) %>% filter(.,num==1) %>% filter(.,age>=25 & age<=36)
Y <- d[complete.cases(d),1]
# Estimate of mu
mean(Y)
```

```
## [1] 13.89444
```

```
# Estimate of sigma
sqrt(var(Y))
```

```
## [1] 0.8618731
```

```
# Estimate of SE(mu-hat)
sqrt(var(Y)/length(Y))
```

```
## [1] 0.1436455
```

Regression formulation

Define the following regression model:

$$Y_i = \mu + \epsilon_i, \epsilon_i \ N(0, \sigma^2), Cov(\epsilon_i, \epsilon_j) = 0$$

This regression model only includes an intercept. We are defining that distribution of arm circumferences for our population follows a normal distribution with mean μ and variance σ^2.

Fit this model using "lm" and identify the estimates for the three quantities of interest; estimate of μ, estimate of σ, estimate of $SE(\hat{\mu})$. Compare the estimates you obtain from the regression model fit to those you computed by hand.

See output on next page!

```r
summary(lm(Y~1))
```

```
##
## Call:
## lm(formula = Y ~ 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.79444 -0.44444 -0.09444  0.53056  2.00556
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.8944     0.1436   96.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8619 on 35 degrees of freedom
```