Tut2

Mike

2024-02-27

1

A 2.1: a.

$$
\begin{aligned}
cdf(X_{(n)}) =& P(X_{max} < x) \\
=& \prod_{i=1}^{n} P(X_i < x)
\end{aligned}
$$

so

$$
cdf(x) = \begin{cases} 0\ \ , & \text{for } x < 0 \\ (\frac{x}{\theta})^n, & \text{for } x \in [0, \theta] \\ 1\ \ , & \text{for } x > \theta \end{cases}
$$

therefore,

$$
pdf(x) = \begin{cases} n(\frac{x}{\theta})^{n-1}, & \text{for } x \in (0, \theta) \\ 0\ \ , & \text{o.w.} \end{cases}
$$

b. obviously, $E(2\bar{X}) = \theta$ and $Var(2\bar{X}) = \frac{4\theta^2}{3n}$ is almost zero if n is large which means the estimator $2\bar{X}$ will converges in probability to θ;

coefficient is 1/3

$E(\frac{n+1}{n}X_{(n)}) = \int_0^\theta \theta(n+1)(\frac{x}{\theta})^n = \theta$ and $Var(\frac{n+1}{n}X_{(n)}) = \int_0^\theta \theta^2 \frac{(n+1)^2}{n}(\frac{x}{\theta})^{n+1} - \theta^2 = \frac{\theta^2}{n^2+2n}$ is almost zero if n is large which means the estimator $2\bar{X}$ will converges in probability to θ; The variances of $\frac{n+1}{n}X_{(n)}$ is smaller

A 2.2: Accoridng to $\theta = E(\widetilde{\theta}) = (k_1 + k_2)\theta$, to make $\widetilde{\theta}$ unbaised, we must have $k_1 + k_2 = 1$. $Var(\widetilde{\theta}) = k_1^2\sigma_1^2 + k_2^2\sigma_2^2 = k_1^2(\sigma_1^2 + \sigma_2^2) - 2k_1\sigma_2^2 + \sigma_2^2$ so when $k_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ and $k_2 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$ $\widetilde{\theta}$ is unbaised and has smallest variance.

```
gdp=read.csv("GDP2020.csv");
gdp[,5]=as.numeric(gdp[,5]);
```

```
## Warning: NAs introduced by coercion
```

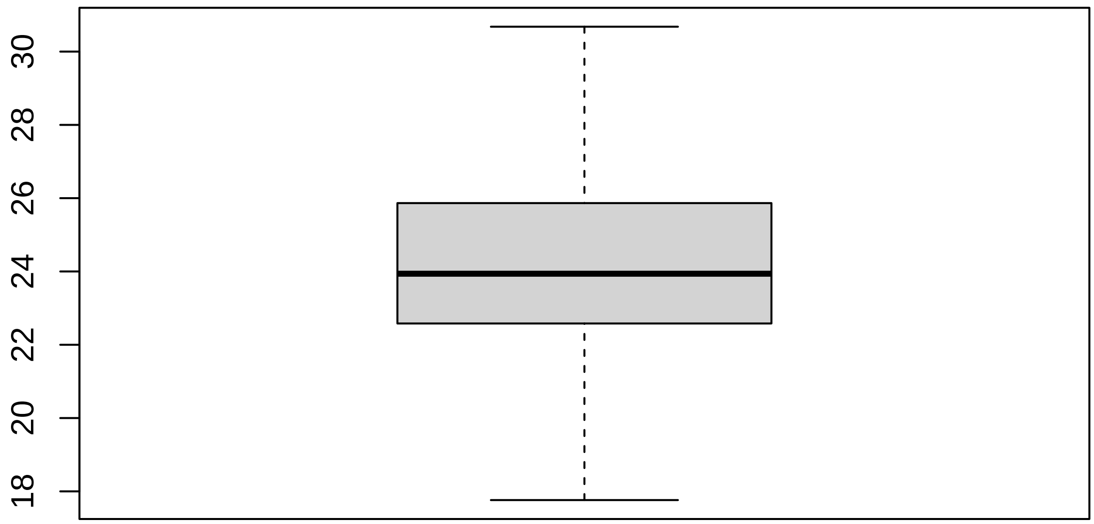
```
gdp=na.omit(gdp);
gdp[,5]=log(gdp[,5])
gdplog=gdp[,5]
summary(gdplog)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.76   22.58   23.94   24.11   25.87   30.68
```

```
below=quantile(gdplog,0.25)-1.5*IQR(gdplog)
above=quantile(gdplog,0.75)+1.5*IQR(gdplog)
outliers=gdp[(gdplog<below|gdplog>above),]
outliers
```

```
## [1] Series.Name  Series.Code  Country.Name Country.Code Y2020
## <0 rows> (or 0-length row.names)
```

```r
boxplot(gdplog)
```




```r
mean(gdplog)
```

```
## [1] 24.10863
```

```r
sd(gdplog)
```

```
## [1] 2.364882
```