CSSS/POLS 510 MLE Lab

Lab 4. Prediction and Quantities of Interest

Ramses Llobet

October 18, 2024

Agenda

1. Review HW1
2. Last lab review
3. Quantities of Interest

1. Review HW1: simulation precision

```
dbinom(x=16, size=30, prob=0.49)
```

```
## [1] 0.1293457
```

```
set.seed(12345)

sims <- 100       # 100 simulations

nmen <- rep(NA,sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(c(0,1),
                        30,
                        replace = TRUE,
                        prob = c(0.51, 0.49) ))
}
sum(nmen==16)/length(nmen) # sum of trials with 16 males
```

```
## [1] 0.16
```

1. Review HW1: simulation precision

```r
dbinom(x=16, size=30, prob=0.49)
```

```
## [1] 0.1293457
```

```r
set.seed(12345)

sims <- 1000       # 1000 simulations

nmen <- rep(NA,sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(c(0,1),
                        30,
                        replace = TRUE,
                        prob = c(0.51, 0.49) ))
}
sum(nmen==16)/length(nmen) # sum of trials with 16 males
```

```
## [1] 0.131
```

1. Review HW1: simulation precision

```r
dbinom(x=16, size=30, prob=0.49)
```

```
## [1] 0.1293457
```

```r
set.seed(12345)

sims <- 10000       # 10000 simulations

nmen <- rep(NA,sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(c(0,1),
                        30,
                        replace = TRUE,
                        prob = c(0.51, 0.49) ))
}
sum(nmen==16)/length(nmen) # sum of trials with 16 males
```

```
## [1] 0.1285
```

1. Review HW1: simulation precision

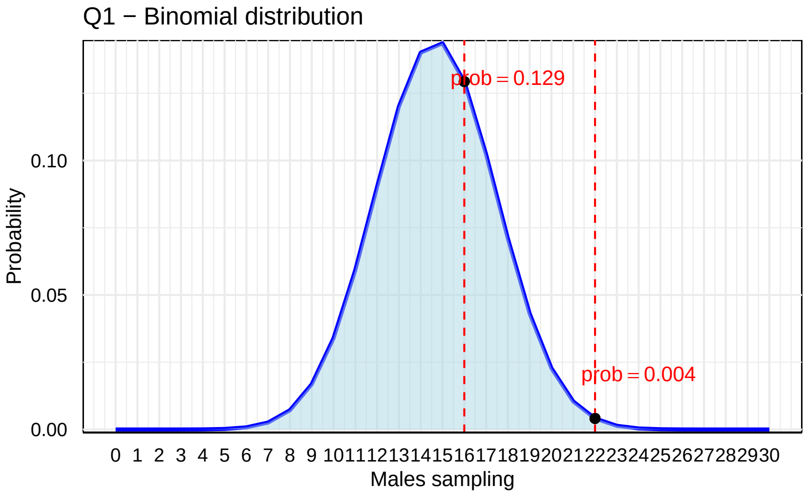
```
dbinom(x=16, size=30, prob=0.49)
```

```
## [1] 0.1293457
```

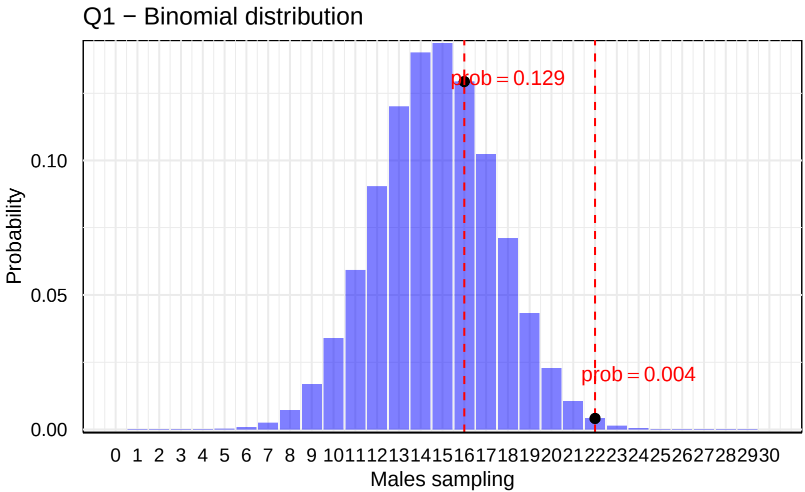
```
set.seed(12345)

sims <- 100000      # 100000 simulations

nmen <- rep(NA,sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(c(0,1),
                      30,
                      replace = TRUE,
                      prob = c(0.51, 0.49) ))
}
sum(nmen==16)/length(nmen) # sum of trials with 16 males
```

```
## [1] 0.1294
```

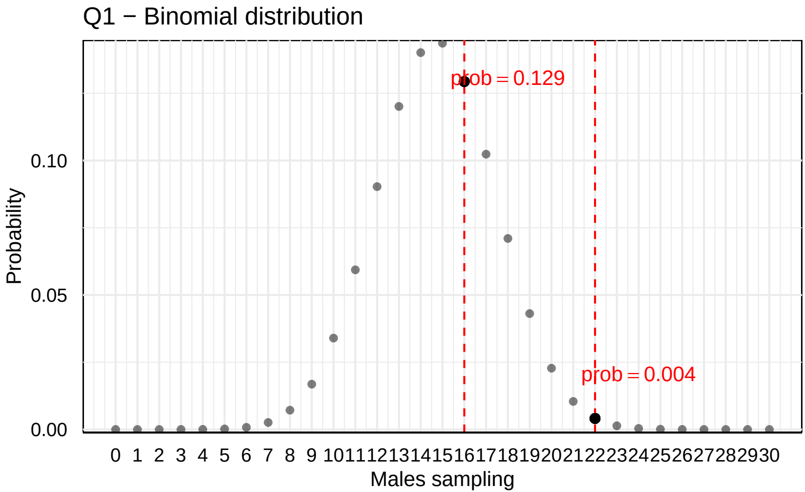
1. Review HW1: Probability Mass Functions



1. Review HW1: Probability Mass Functions



1. Review HW1: Probability Mass Functions



1. Review HW1: summary

- When displaying P**M**Fs, provide visuals with a discrete sample space.
- **Sanity checks**:
  - Use R built-in or packages programs/functions to **double-check**.
  - Increase simulations/sample **size**.
    - Convergence in probability ($N \rightarrow \infty$)

2. Last lab review: MLE

- How do we estimate the MLE?
  1. Define a probability model (PDF): $Y_i \sim N(\mu_i, \sigma^2)$.
  2. Derive the log-likelihood function.
  3. Reduce to sufficient statistics and substitute systematic component.
  4. Use `optim()` or any other function to find the maxima.

2. Normal homoskedastic

Two **different** notations for the **same** model.

LS notation:

$$\begin{aligned} \varepsilon &\sim N(0, \sigma^2) && (\textit{stochastic}) \\ Y_i &= x_i\beta && (\textit{systematic}) \\ Y_i &= x_i\beta + \varepsilon && (\textit{stochastic} + \textit{systematic}) \end{aligned}$$

MLE notation:

$$\begin{aligned} Y_i &\sim N(\mu_i, \sigma^2) && (\textit{stochastic}) \\ \mu_i &= x_i\beta && (\textit{systematic}) \\ Y_i &\sim N(x_i\beta, \sigma^2) && (\textit{stochastic} + \textit{systematic}) \end{aligned}$$

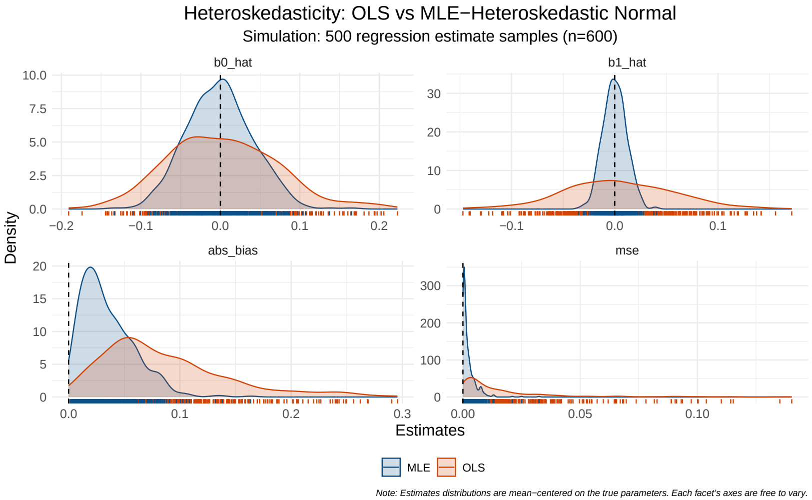
2. MLE general notation

$$
\begin{aligned}
Y_i &\sim f(\theta_i, \alpha) \quad \text{(stochastic)} \\
\theta_i &= g(\boldsymbol{x}_i \boldsymbol{\beta}) \quad \text{(systematic)}
\end{aligned}
$$

where

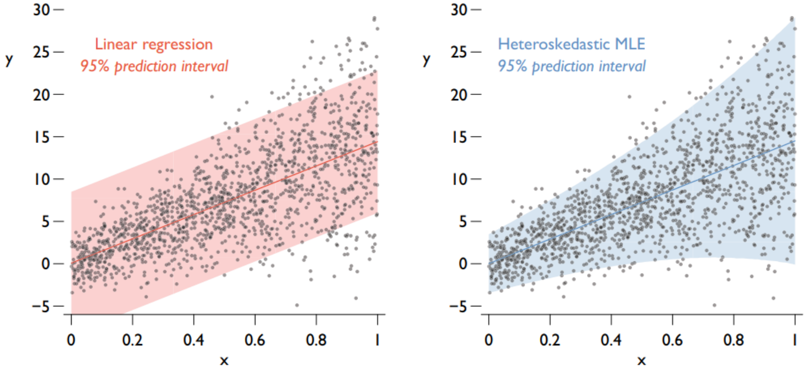
- Y_i is a random outcome variable.
- $f(.)$ is a probability density function.
- θ_i is a systematic feature of the PDF that varies over i.
- α is an ancillary parameter (feature of f that we treat as constant).
- $g(.)$ functional form for reparametrization of the data model.
- x_i explanatory variables vector.
- β vector of effect parameters.

2. Last lab review: MLE - Heteroskedastic normal



3. Quantities of Interest: prediction

Once we have estimated the model, we can compare the predictive performance of each model. Refer to Chris's MLE lecture (slide 80) for more details.



3. Quantities of Interest

Motivation: We want to study how the change in a particular explanatory variable affects the outcome variable, *all else being equal*.

We will focus on how to simulate predictions from *estimands* or **quantities of interest**.

Let's open `RStudio` and Lab4.html file jointly with the Lab4.Rmd file with the code and contents of today's lab!

F I N