

CSSS/POLS 510 MLE Lab

Lab 4. Prediction and Quantities of Interest

Minji Jeong

Housekeeping

- Problem set 1 grades are posted on Canvas
- Problem set 2 is due October 30 3:30pm
 - Please properly present your answers to the questions
 - Don't use `dbinom()` in Problem 1
- Agenda
 - Review Problem set 1
 - Last lab review
 - 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 <- vector(mode = "numeric", length = sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(x = c(0,1),
                      size = 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

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

## [1] 0.1293457
set.seed(12345)

sims <- 1000      # 1000 simulations

nmen <- vector(mode = "numeric", length = sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(x = c(0,1),
                      size = 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

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

## [1] 0.1293457
set.seed(12345)

sims <- 10000      # 10000 simulations

nmen <- vector(mode = "numeric", length = sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(x = c(0,1),
                      size = 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)

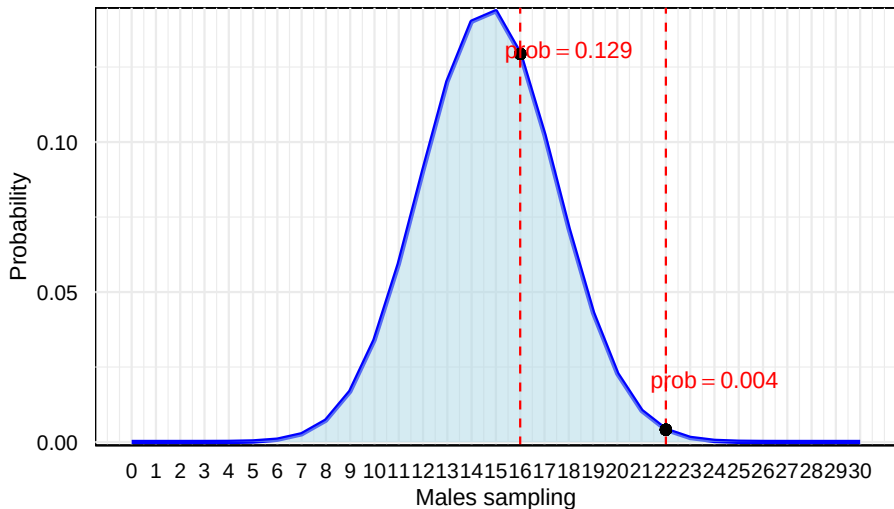
sims <- 100000      # 100000 simulations

nnmen <- vector(mode = "numeric", length = sims)
for (i in 1:sims) {
  nmen[i] <- sum(sample(x = c(0,1),
                        size = 30,
                        replace = TRUE,
                        prob = c(0.51, 0.49) ))
}
sum(nmen==16)/length(nmen) # sum of trials with 16 males

## [1] 0.1294
```

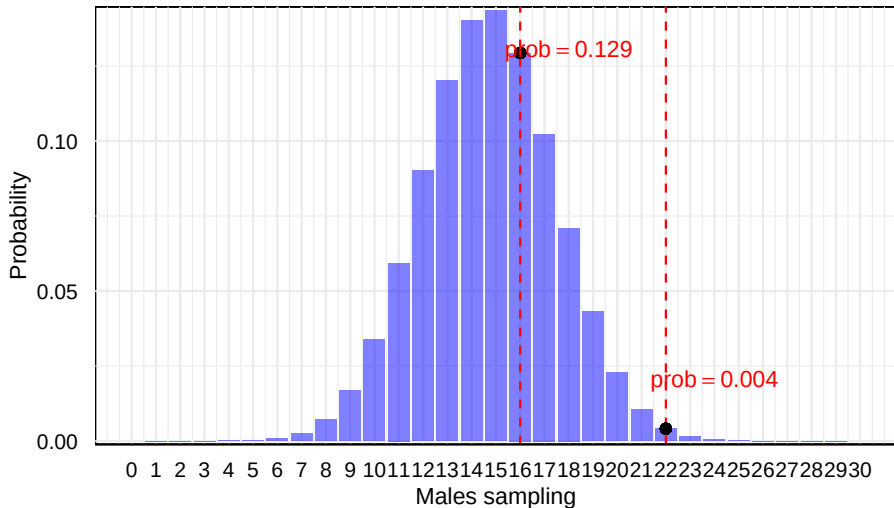
1. Review HW1: Probability Mass Functions

Q1 – Binomial distribution



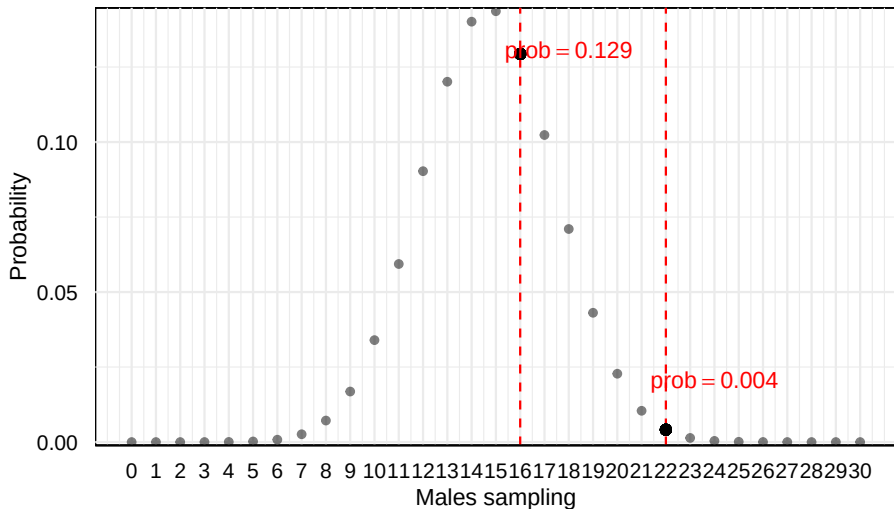
1. Review HW1: Probability Mass Functions

Q1 – Binomial distribution



1. Review HW1: Probability Mass Functions

Q1 – Binomial distribution



1. Review HW1: summary

- When displaying PMFs, 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. Last lab review: MLE general notation

$$Y_i \sim f(\theta_i, \alpha) \quad (\text{stochastic})$$

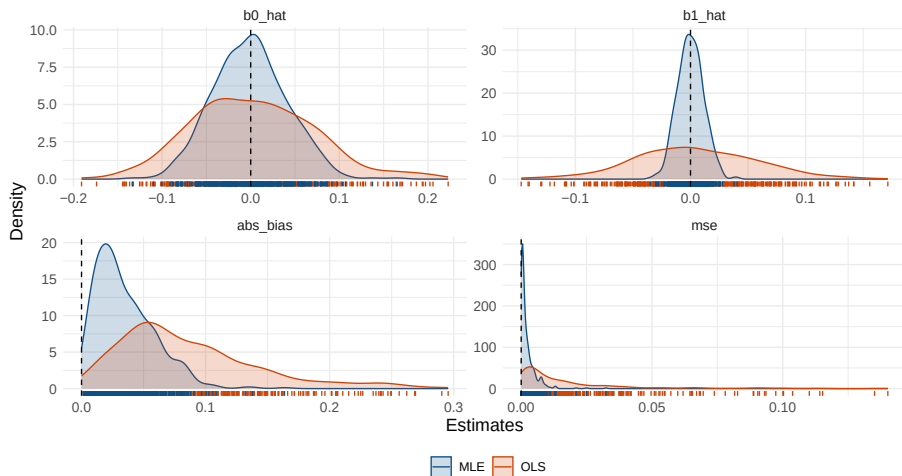
$$\theta_i = g(\mathbf{x}_i\beta) \quad (\text{systematic})$$

where

- Y_i is a random outcome variable.
- $f(\cdot)$ 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).
- $g(\cdot)$ functional form for reparametrization of the data model.
- $\mathbf{x}_i$ explanatory variables vector.
- β vector of effect parameters.

2. Last lab review: MLE - Heteroskedastic normal

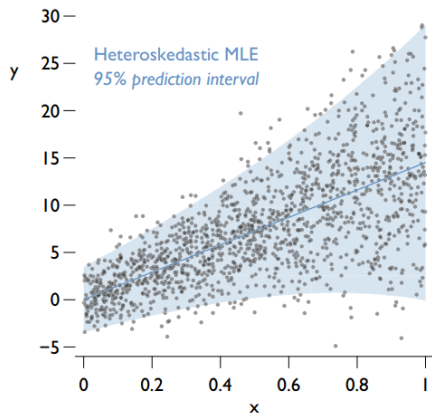
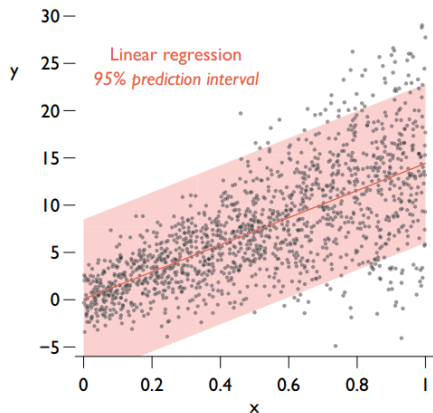
Heteroskedasticity: OLS vs MLE-Heteroskedastic Normal
Simulation: 500 regression estimate samples (n=600)



Note: Estimates distributions are mean-centered on the true parameters. Each facet's axes are free to vary.

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**.

3. Quantities of Interest

In order to use `simcf` to generate quantities of interest, the following steps are needed:

- 1 Estimate: MLE $\hat{\beta}$ and its variance $\hat{V}(\hat{\beta})$
- 2 Simulate estimation uncertainty from a multivariate normal distribution:
Draw $\tilde{\beta} \sim MVN[\hat{\beta}, \hat{V}(\hat{\beta})]$
- 3 Create hypothetical scenarios of your substantive interest:
Choose values of X , X_c

2.1 Simulating QoI

- 4 Calculate expected values:

$$\tilde{\mu}_c = g(X_c, \tilde{\beta})$$

- 5 Simulate fundamental uncertainty:

$$\tilde{y}_c \sim f(\tilde{\mu}_c, \tilde{\alpha})$$

or

- 6 Compute EVs, First Differences or Relative Risks

$$\text{EV: } \mathbb{E}(y|X_{c1})$$

$$\text{FD: } \mathbb{E}(y|X_{c2}) - \mathbb{E}(y|X_{c1})$$

$$\text{RR: } \frac{\mathbb{E}(y|X_{c2})}{\mathbb{E}(y|X_{c1})}$$

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

FIN