Capítulo 2

Santiago Barreda

Chapter 2

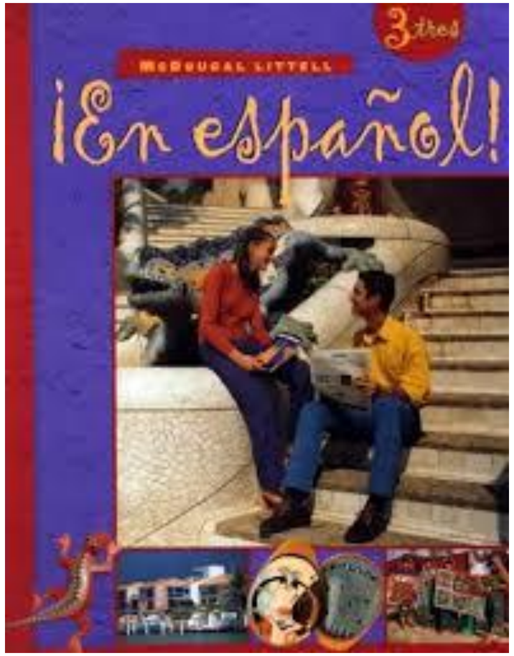
El propósito de este capítulo es:

- Explicación "básica" de la *probabilidad* y la *verosimilitud*.
- Motivar el uso de verosimilitudes en la inferencia científica.
- Hacer que los modelos estadísticos dejen de ser una "caja negra".
- Empezar a acostumbrarnos a la notación matemática.

El propósito de este capítulo <u>no</u> es:

- Comprender todo el contenido del capítulo.

Estadística = Conocimiento Procedimental



Data and Research Questions

Objetivo: Utilizar los datos de nuestro experimento para responder a estas preguntas:

(1) ¿Qué tan alto "suena" el hombre adulto promedio?

(2) ¿Podemos establecer límites a las alturas aparentes creíbles en base a los datos que hemos recopilado?

Why Estimate Variation?

La interpretación de los valores medios se basa en la comprensión de la variación por al menos dos razones:

- Incertidumbre en las mediciones.
- Interpretación de las magnitudes (es decir, si el efecto es grande).

Probability

- Hay muchas maneras de pensar en la probabilidad matemáticamente y filosóficamente.

- Definición simple: La probabilidad de evento es el número de veces que ocurre un resultado en relación con todos los demás resultados posibles.

- La probabilidad de cada resultado es un valor entre 0 y 1.

- La suma de las probabilidades para todos los resultados es 1.

Probability: Example

- Un dado tiene 6 resultados posibles: 1,2,3,4,5 y 6.

$$\mathrm{resultado_i} \qquad\qquad i = \{1,2,3,4,5,6\}$$

- Cada resultado es igualmente probable, una probabilidad de 1/6 (0.167) de ocurrir.

$$P(\mathrm{resultado_i}) = 1/6 = 0.16666\ldots$$

- La suma de los siete resultados es igual a 1.

$$1 = \sum_{i=1}^{6} P(\mathrm{resultado_i})$$

Probability: Not quite so simple.....

- "La probabilidad de que Getafe gane La Liga..."
- "La probabilidad de ver un terremoto más grande que cualquier terremoto que hayamos visto es..."
- "La probabilidad de que el dado saque un 6 es....."

Probability: Quite so simple?

Imagina que tienes una urna que contiene 20 bolas rojas y 80 bolas azules. Sacas al azar dos bolas de la urna. Después de cada sorteo, la bola se reemplaza y la urna se asigna al azar (se sacude).

¿Cuál es la probabilidad de que ambas bolas sean rojas?

- La probabilidad de sacar uno es 20/(20+80) = 20/100 = 0.2
- La probabilidad de sacar dos es 0.2 x 0.2 = 0.04

Probability: Not quite so simple!

"In probability theory there is a very clever trick for handling a problem that becomes too difficult. We just solve it anyway by: making it still harder; redefining what we mean by 'solving' it, so that it becomes something we can do; inventing a dignified and technical-sounding word to describe this procedure, which has the psychological effect of concealing the real nature of what we have done, and making it appear respectable.

In the case of sampling with replacement, we apply this strategy as follows. Suppose that, after tossing the ball in, we shake up the urn. However complicated the problem was initially, it now becomes many orders of magnitude more complicated, because the solution now depends on every detail of the precise way we shake it, in addition to all the factors mentioned above.

We now assert that the shaking has somehow made all these details irrelevant, so that the problem reverts back to the simple one where the Bernoulli urn rule applies. We invent the dignified-sounding word randomization to describe what we have done. This term is, evidently, a euphemism, whose real meaning is: deliberately throwing away relevant information when it becomes too complicated for us to handle."

From: "Probability Theory: The Logic of Science", by ET Jaynes

Probability: Not quite so simple!

“We have described this procedure in laconic terms, because an antidote is needed for the impression created by some writers on probability theory, who attach a kind of mystical significance to it. For some, declaring a problem to be ‘randomized’ is an incantation with the same purpose and effect as those uttered by an exorcist to drive out evil spirits; i.e. it cleanses their subsequent calculations and renders them immune to criticism. We agnostics often envy the True Believer, who thus acquires so easily that sense of security which is forever denied to us.

[…] Shaking does not make the result ‘random’, because that term is basically meaningless as an attribute of the real world; it has no clear definition applicable in the real world. The belief that ‘randomness’ is some kind of real property existing in Nature is a form of the mind projection fallacy which says, in effect, ‘I don’t know the detailed causes – therefore – Nature does not know them.’ What shaking accomplishes is very different. It does not affect Nature’s workings in any way; it only ensures that no human is able to exert any willful influence on the result. Therefore, nobody can be charged with ‘fixing’ the outcome.”

From: “Probability Theory: The Logic of Science”, by ET Jaynes

Probability: Not quite so simple!

- ¿Confundidos? ¡Estás en buena compañía!

- Bertrand Russell: “Probability is the most important concept in modern science, especially as nobody has the slightest notion of what it means.”

- John von Neumann: "Young man, in mathematics you don't *understand* things. You just *get used to them*."

Conditional and Marginal Probabilities

- Una probabilidad marginal es la probabilidad general de algún resultado. Es 'incondicional’.

- Una probabilidad condicional es la probabilidad de que se dé algún resultado o si se cumple alguna condición.

  - "dado" significa "asumiendo que" o "condicionado a".

  - ¿Cuál es la probabilidad de que estés dormido dado que son las 4 de la tarde? ¿Y a las 4 de la mañana?

Conditional and Marginal Probabilities

- La probabilidad marginal de alguna variable se denota por:

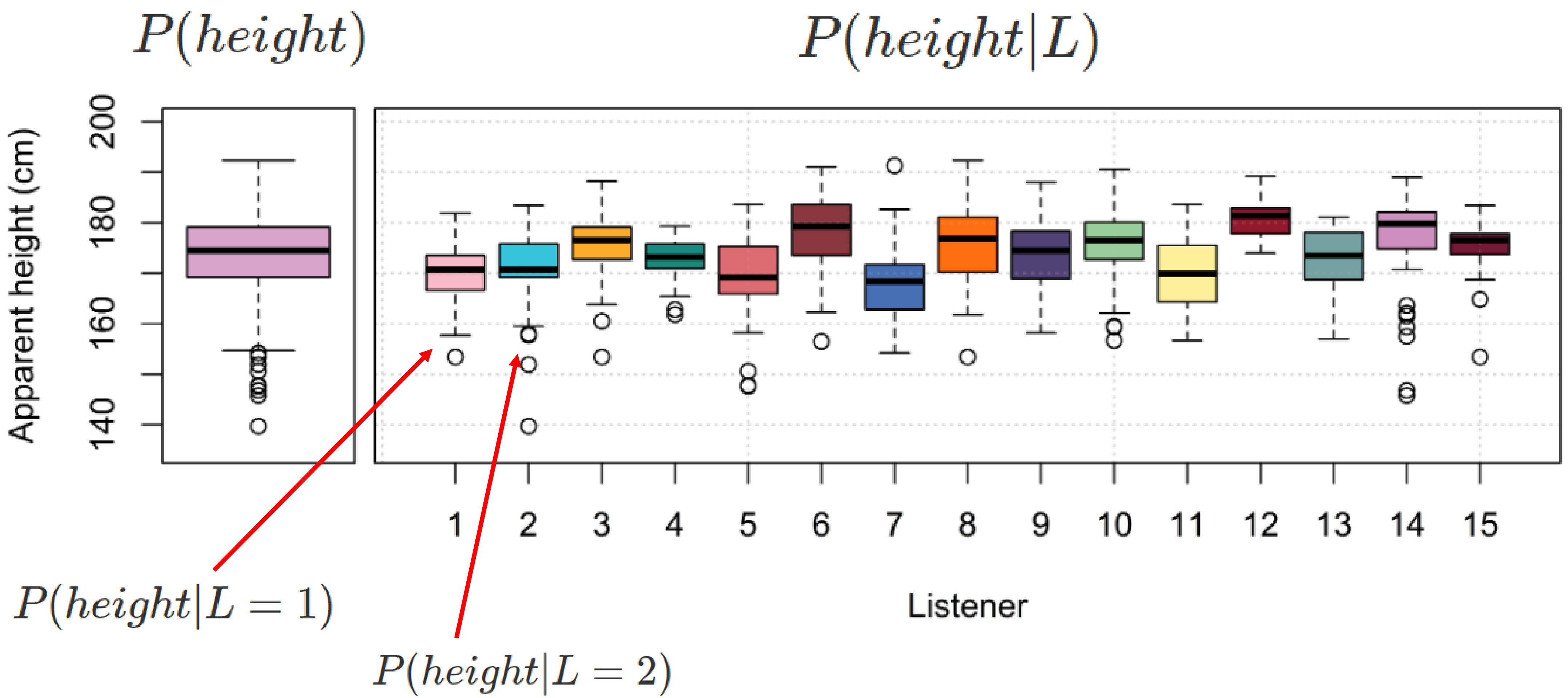
$$P(variable)$$

- La probabilidad de esa misma variable, condicionada a un segundo, se denota por:

$$P(outcome\,variable \mid conditioning\,variable)$$

Esto significa "dado".

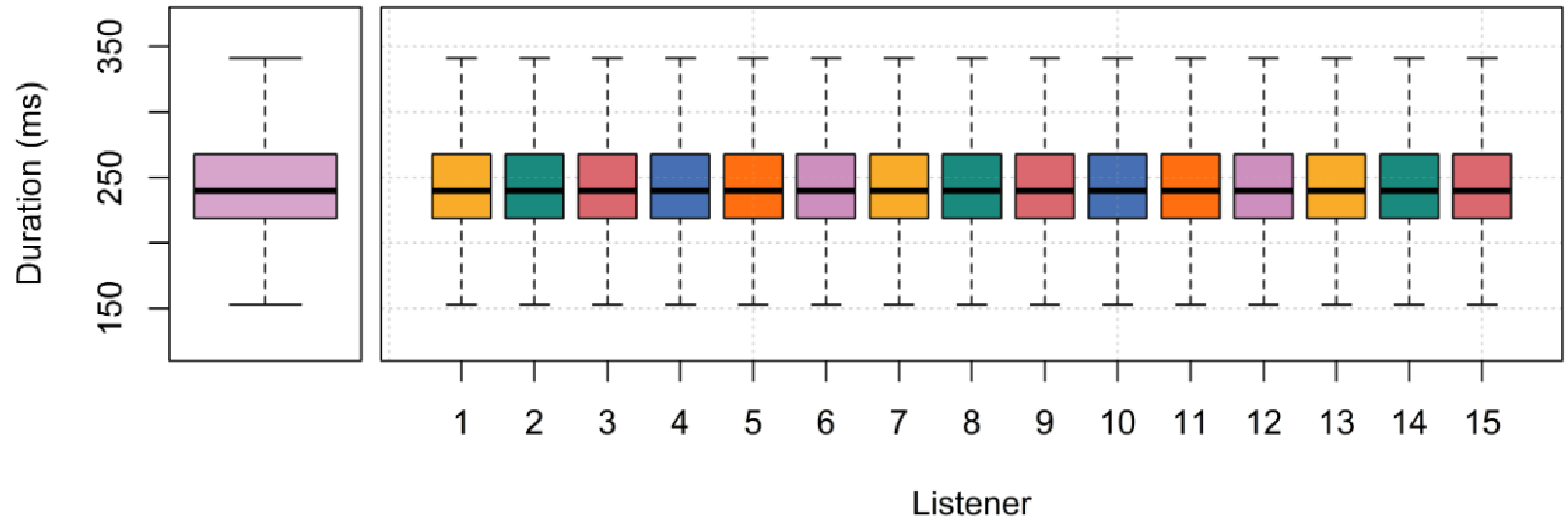
Conditional and Marginal Probabilities



Statistical Independence

- Cuando la distribución marginal de una variable es la misma que su distribución condicional a alguna variable, es <u>independiente</u> de esa variable.

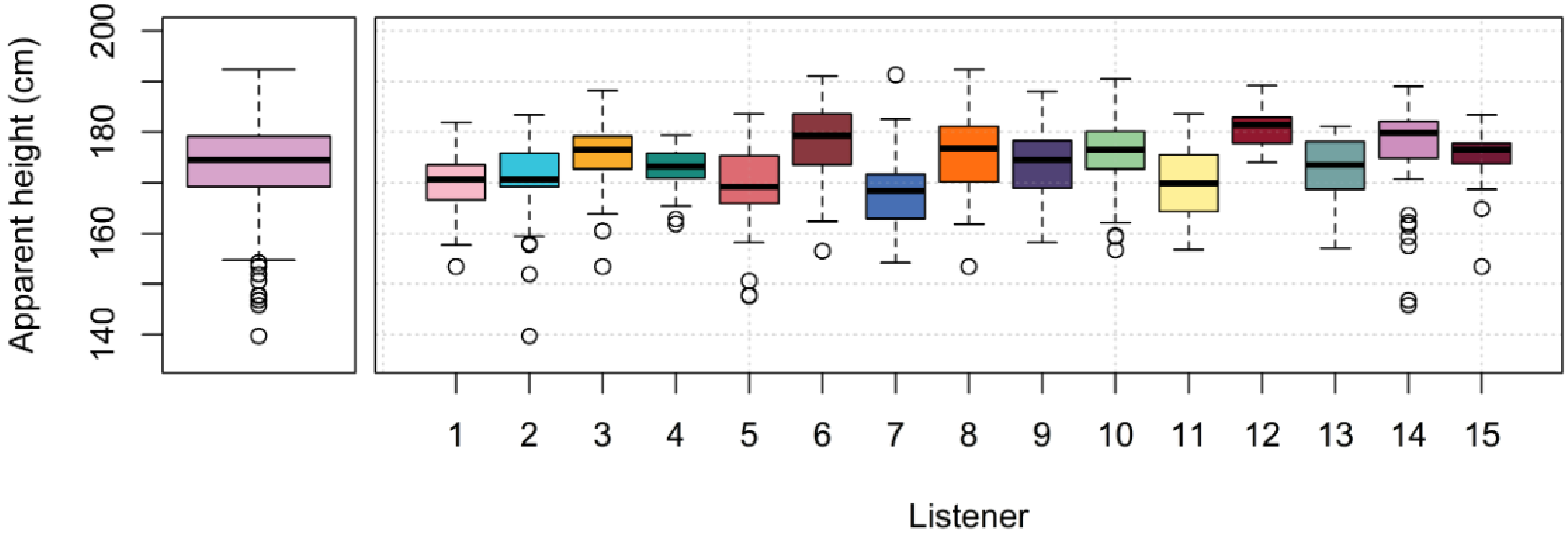
$$P(variable) = P(variable|conditioning\ variable)$$



Statistical Dependence

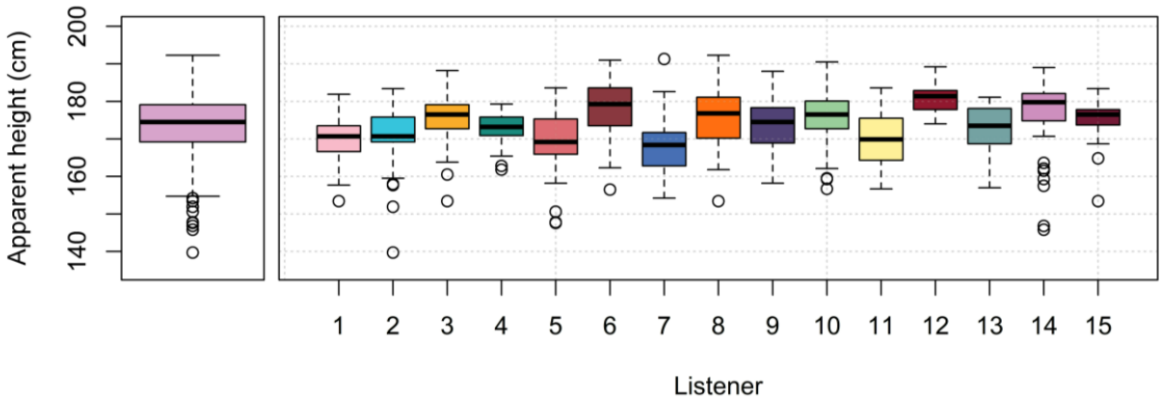
- Cuando la distribución marginal de una variable no es la misma que su distribución condicional a alguna variable, depende de esa variable.

$$P(variable) \neq P(variable|conditioning\, variable)$$

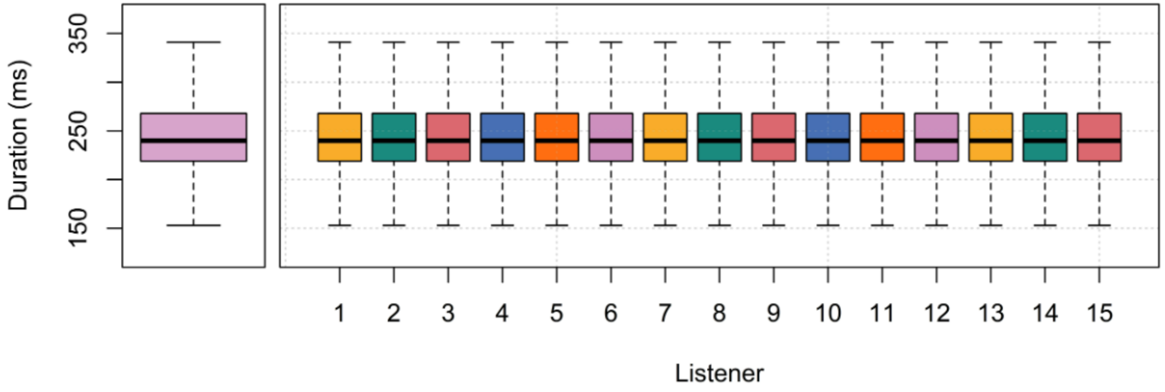


Statistical Dependence and Independence

$$P(variable) \neq P(variable|conditioning\,variable)$$



$$P(variable) = P(variable|conditioning\,variable)$$



Joint Probabilities

- La probabilidad conjunta es la probabilidad de que ocurran varios eventos juntos.

$$P(A \cap B) \text{ or } P(A \& B)$$

- Se calcula multiplicando una probabilidad condicional y una marginal.

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Joint Probabilities and Independence

- Para las variables independientes, la probabilidad condicional es igual a su probabilidad marginal.

$$P(A|B) = P(A)$$

- Por esto, la probabilidad conjunta de eventos independientes es el producto de sus probabilidades marginales.

$$P(A|B) \cdot P(B) = P(A) \cdot P(B)$$

Joint Probabilities and Independence

$$\text{Independent events :}$$
$$P(A \,\&\, B \,\&\, C \,\&\, D) = P(A) \cdot P(B) \cdot P(C) \cdot P(D)$$

$$\text{Dependent events :}$$
$$P(A \,\&\, B \,\&\, C \,\&\, D) = P(A|B, C, D) \cdot P(B|C, D) \cdot P(C|D) \cdot P(D)$$

Probability Distributions

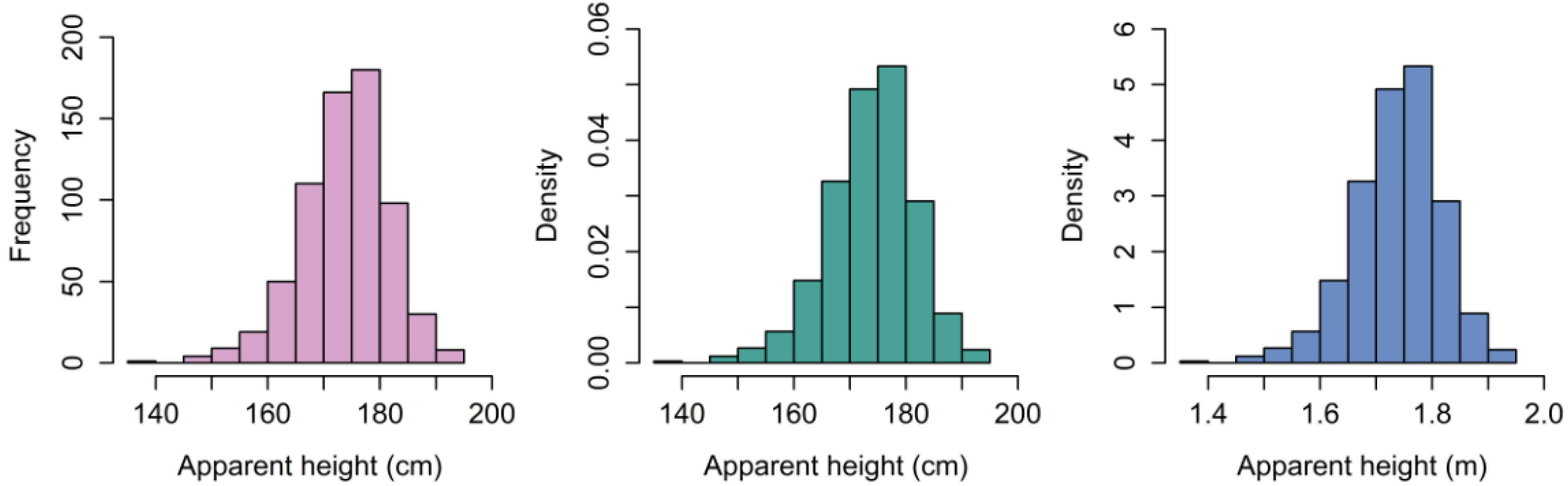
- Formalmente: Función que asigna probabilidades a todos los valores de un espacio muestral.

$$f(x) = p \qquad\qquad P(x) = p$$

- Donde *x* es algún resultado o evento posible.

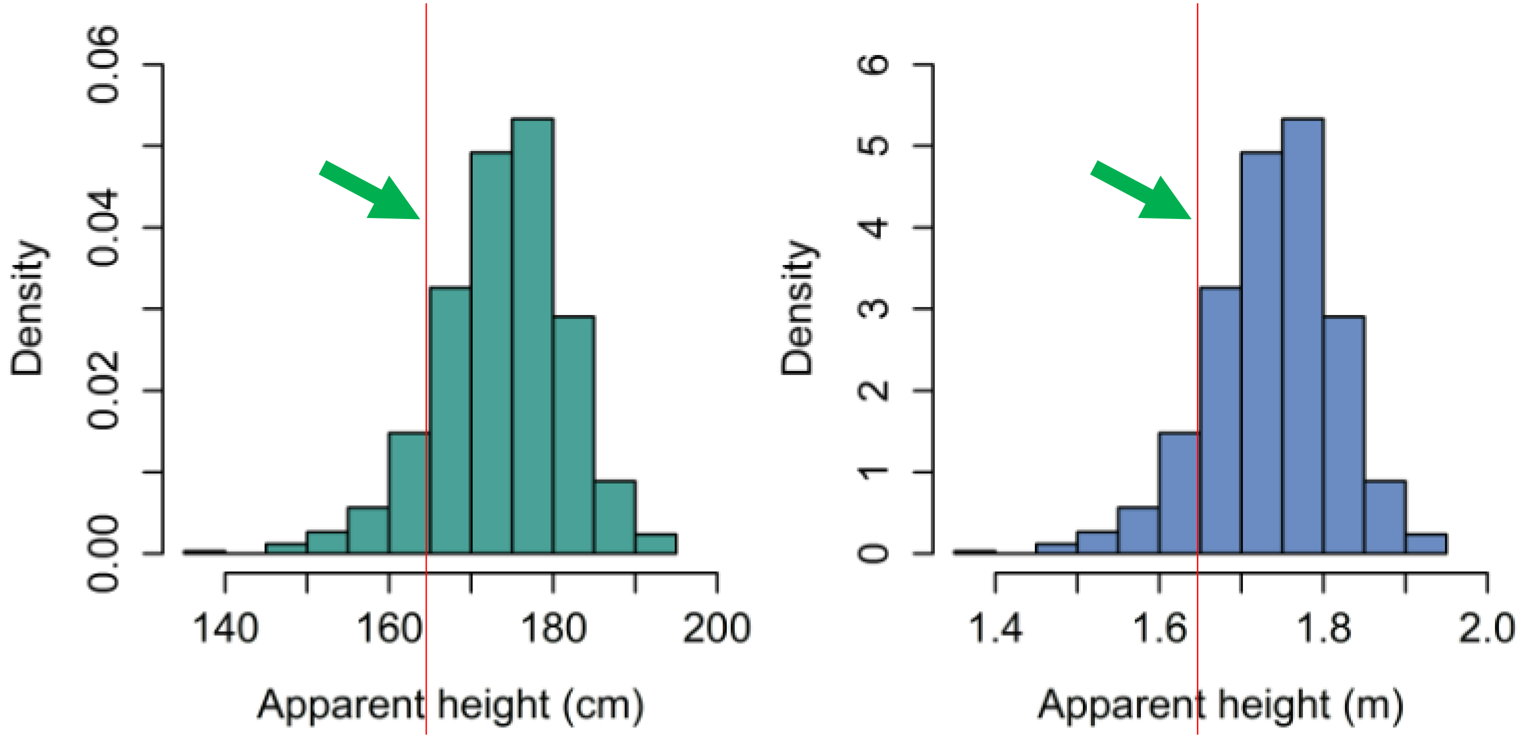
Probability Density

- Se utiliza para visualizar/comprender distribuciones de probabilidad
- Una curva con un área de la curva de 1.



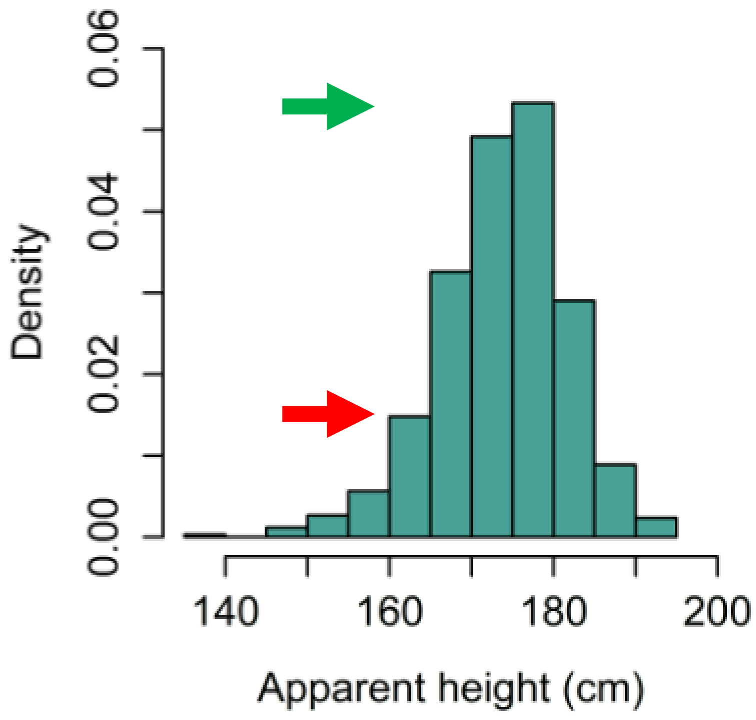
Probability Density: Proportions

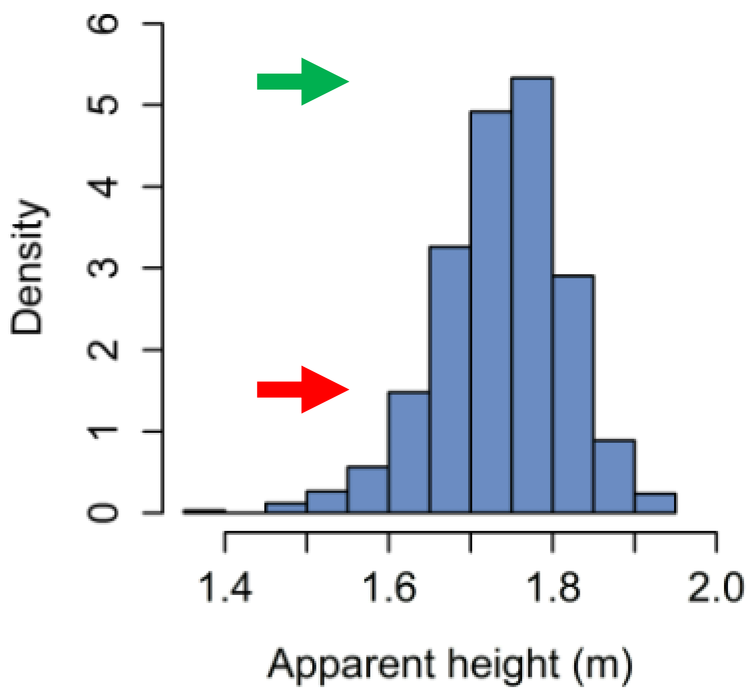
- El área bajo la curva entre dos puntos refleja la probabilidad de que la variable tenga un valor en ese rango.



Probability Density: Relative Values

- Si la densidad es x veces mayor/menor para un valor que para otro, ese valor es x veces mas/menos probable.

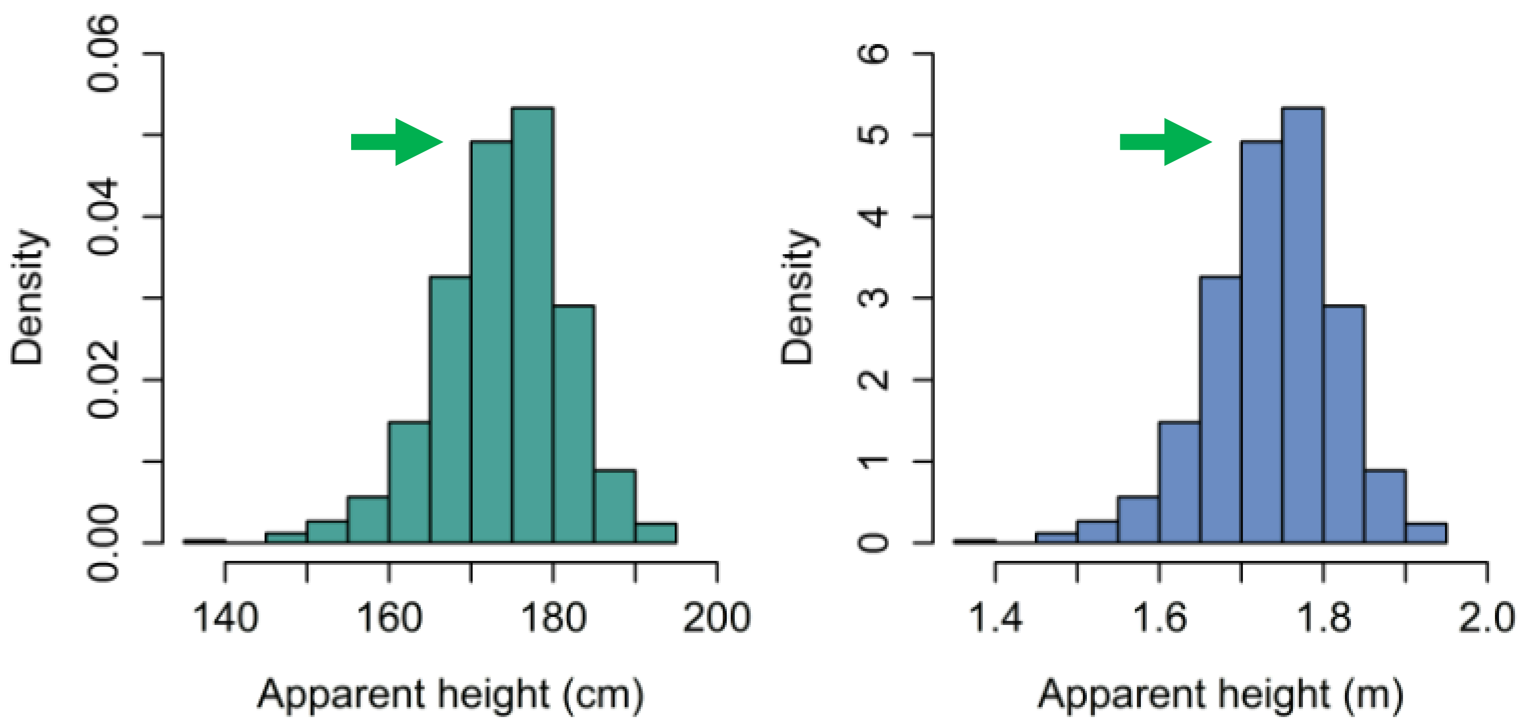




Probability Density: No Absolute Interpretation

- El valor absoluto de la curva en un punto dado <u>no es</u> la probabilidad.

- <u>No debe</u> interpretarse de forma aislada.



Parametric Probability Distributions

- Algunas distribuciones de probabilidad tienen un comportamiento relativamente ‘simple’.

- Este comportamiento se puede predecir mediante un pequeño conjunto de parámetros.

- Piense en ellas como las formas definidas por:

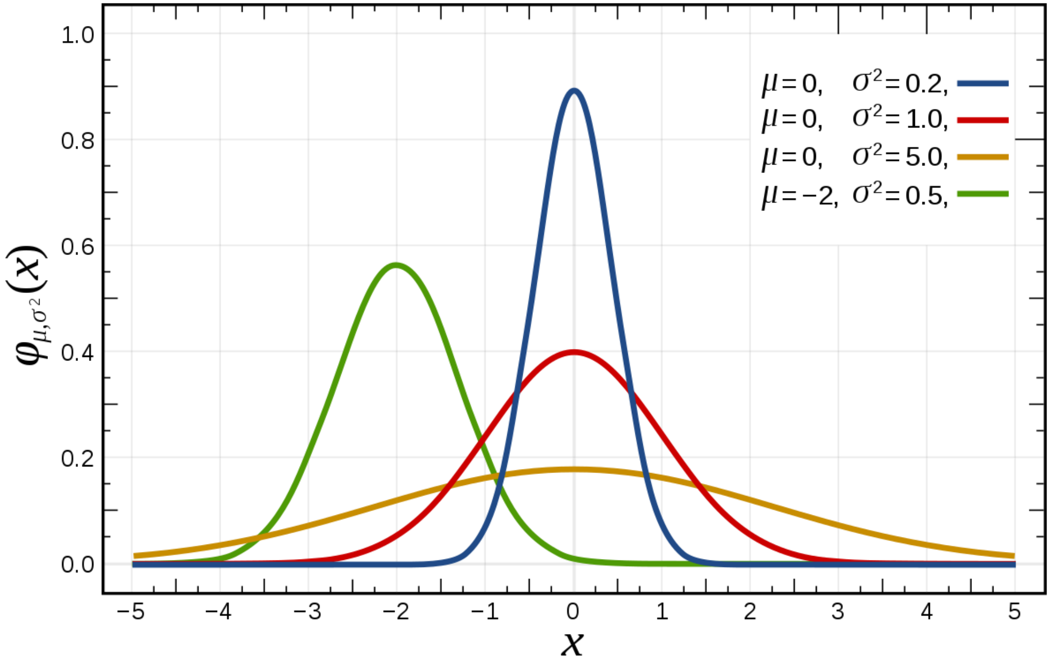
$$y = a + b \cdot x \qquad y = a + b(c - x)^2$$

The Normal Distribution

- Quizás la distribución más conocida y utilizada.

- Dos parámetros: La media(μ) y la desviación estándar (σ).

$$f(x) = \cancel{\frac{1}{\sigma\sqrt{2\pi}}} \cdot \cancel{\exp}\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$



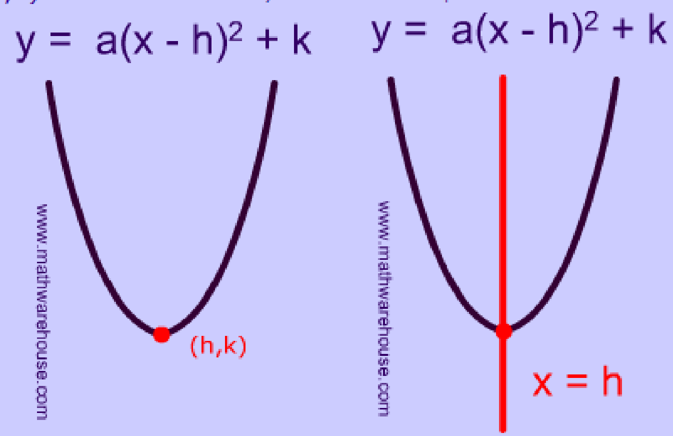
Quadratic Equations

Vertex Form of Equation

The vertex form of a parabola's equation is generally expressed as :

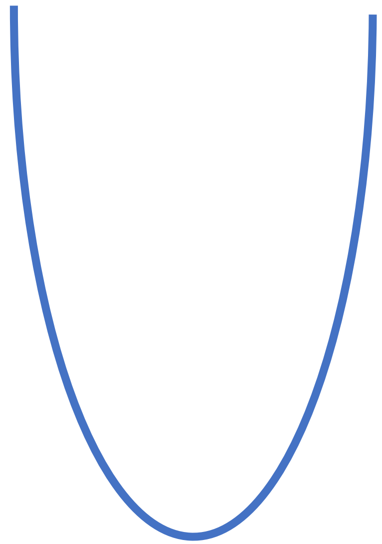
$y = a(x-h)^2+k$

- (h,k) is the vertex as you can see in the picture below



- If a is positive then the parabola opens upwards like a regular "U".
- If a is negative, then the graph opens downwards like an **upside down** "U".

- If |a| < 1, the graph of the parabola widens. This just means that the "U" shape of parabola stretches out sideways . **Explore the way that 'a' works using our interactive parabola grapher** .
- If |a| > 1, the graph of the graph becomes narrower(The effect is the opposite of |a| < 1).



$$-\frac{1}{2\sigma^2}(x-\mu)^2$$

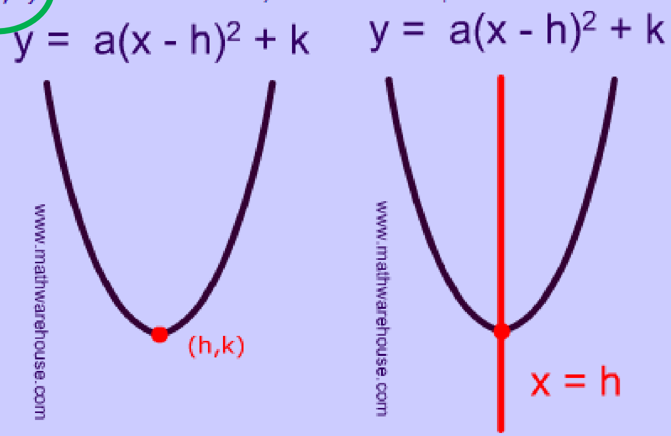
Quadratic Equations

Vertex Form of Equation

The vertex form of a parabola's equation is generally expressed as :

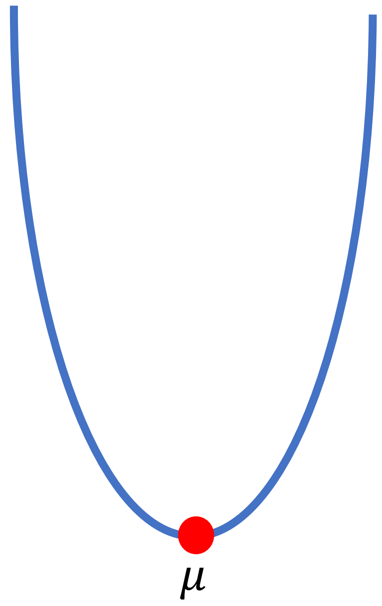
$y= a(x-h)^2+k$

- (h,k) is the vertex as you can see in the picture below



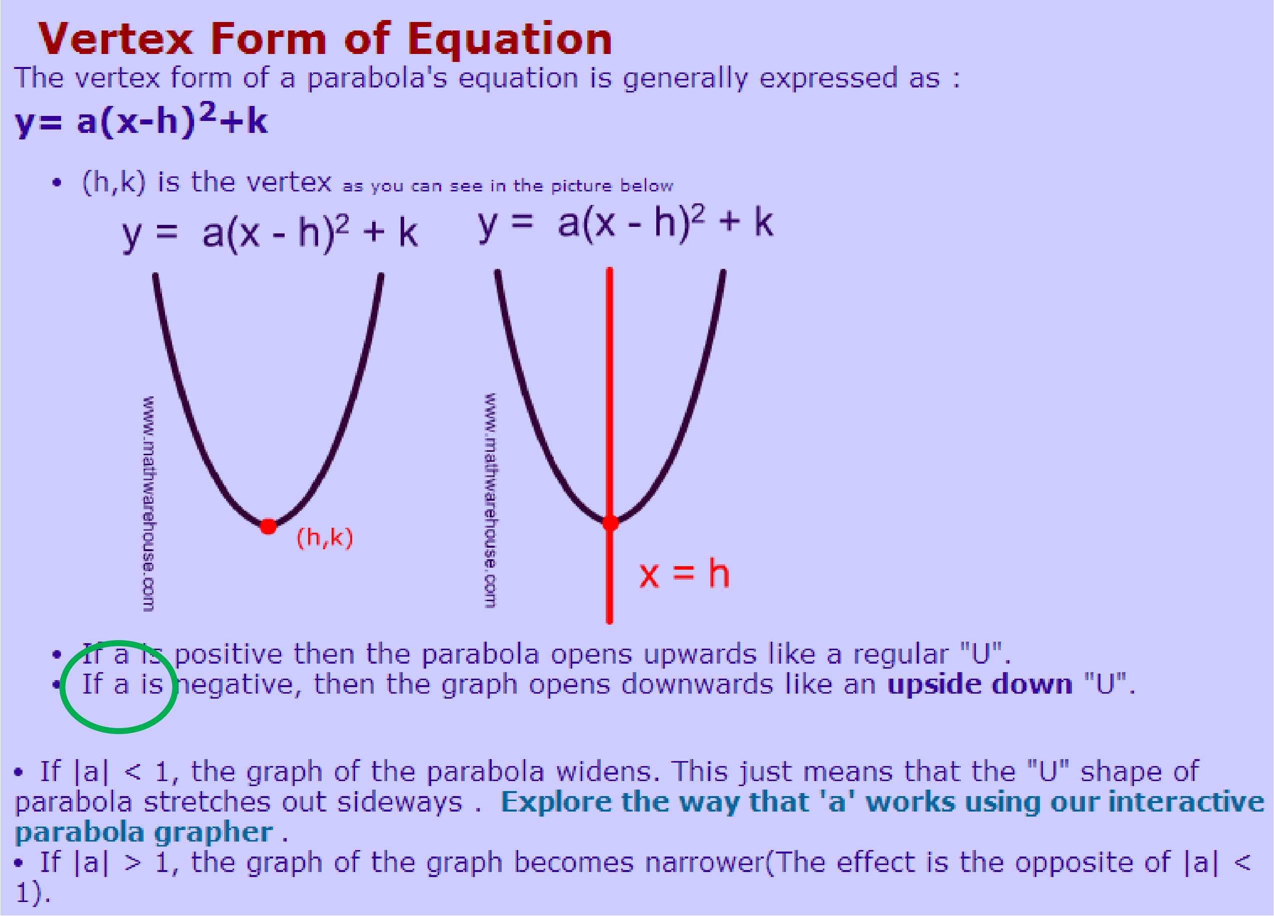
- If a is positive then the parabola opens upwards like a regular "U".
- If a is negative, then the graph opens downwards like an **upside down** "U".

- If |a| < 1, the graph of the parabola widens. This just means that the "U" shape of parabola stretches out sideways . **Explore the way that 'a' works using our interactive parabola grapher** .
- If |a| > 1, the graph of the graph becomes narrower(The effect is the opposite of |a| < 1).



$$-\frac{1}{2\sigma^2}(x-\mu)^2$$

Quadratic Equations



Vertex Form of Equation

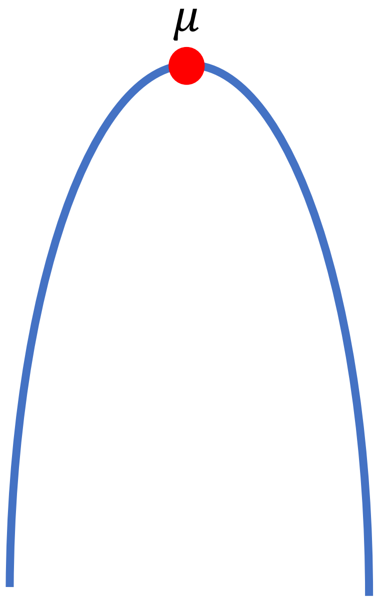
The vertex form of a parabola's equation is generally expressed as :

$y= a(x-h)^2+k$

- (h,k) is the vertex as you can see in the picture below

- If a is positive then the parabola opens upwards like a regular "U".
- If a is negative, then the graph opens downwards like an **upside down** "U".

- If |a| < 1, the graph of the parabola widens. This just means that the "U" shape of parabola stretches out sideways . **Explore the way that 'a' works using our interactive parabola grapher** .
- If |a| > 1, the graph of the graph becomes narrower(The effect is the opposite of |a| < 1).



$$-\frac{1}{2\sigma^2}(x-\mu)^2$$

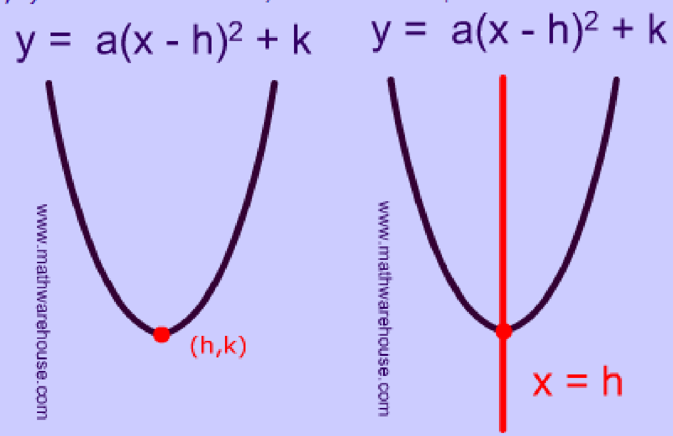
Quadratic Equations

Vertex Form of Equation

The vertex form of a parabola's equation is generally expressed as :

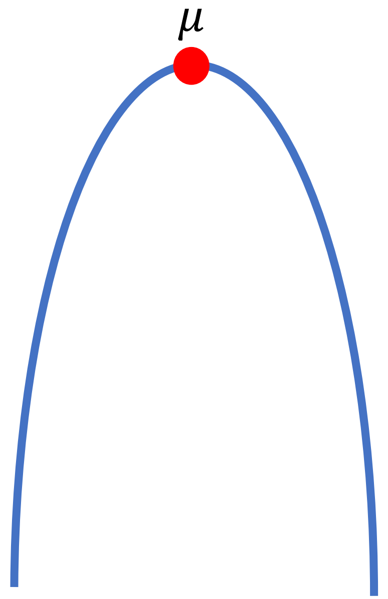
$y = a(x-h)^2 + k$

- (h,k) is the vertex as you can see in the picture below



- If a is positive then the parabola opens upwards like a regular "U".
- If a is negative, then the graph opens downwards like an **upside down** "U".

- If |a| < 1, the graph of the parabola widens. This just means that the "U" shape of parabola stretches out sideways . **Explore the way that 'a' works using our interactive parabola grapher** .
- If |a| > 1, the graph of the graph becomes narrower(The effect is the opposite of |a| < 1).



$$-\frac{1}{2\sigma^2}(x-\mu)^2$$

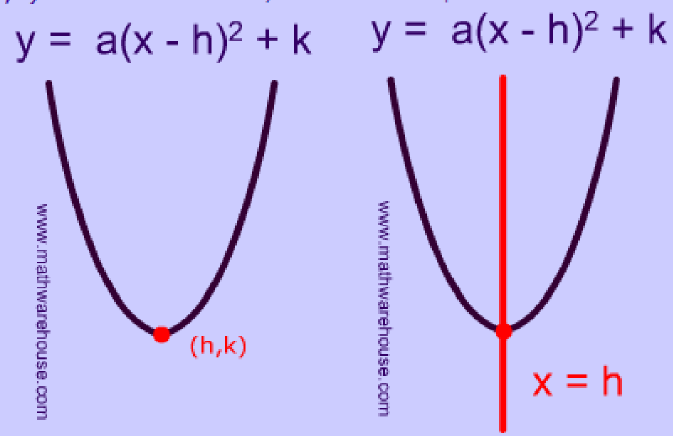
Quadratic Equations

Vertex Form of Equation

The vertex form of a parabola's equation is generally expressed as :

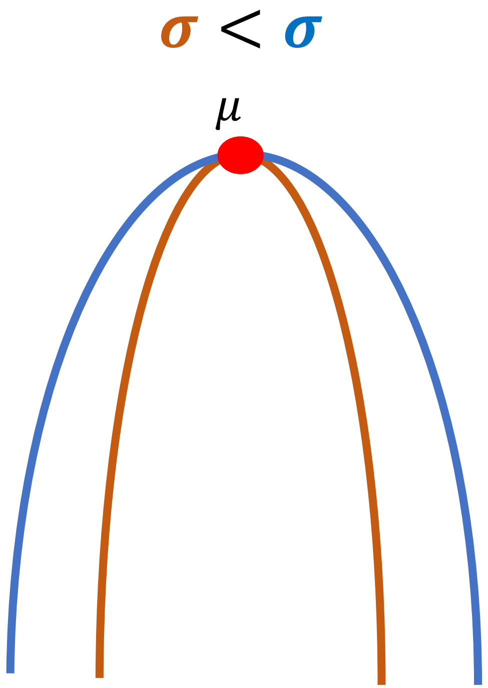
y= a(x-h)^2+k

- (h,k) is the vertex as you can see in the picture below



- If a is positive then the parabola opens upwards like a regular "U".
- If a is negative, then the graph opens downwards like an **upside down** "U".

- If |a| < 1, the graph of the parabola widens. This just means that the "U" shape of parabola stretches out sideways . **Explore the way that 'a' works using our interactive parabola grapher** .
- If |a| > 1, the graph of the graph becomes narrower(The effect is the opposite of |a| < 1).

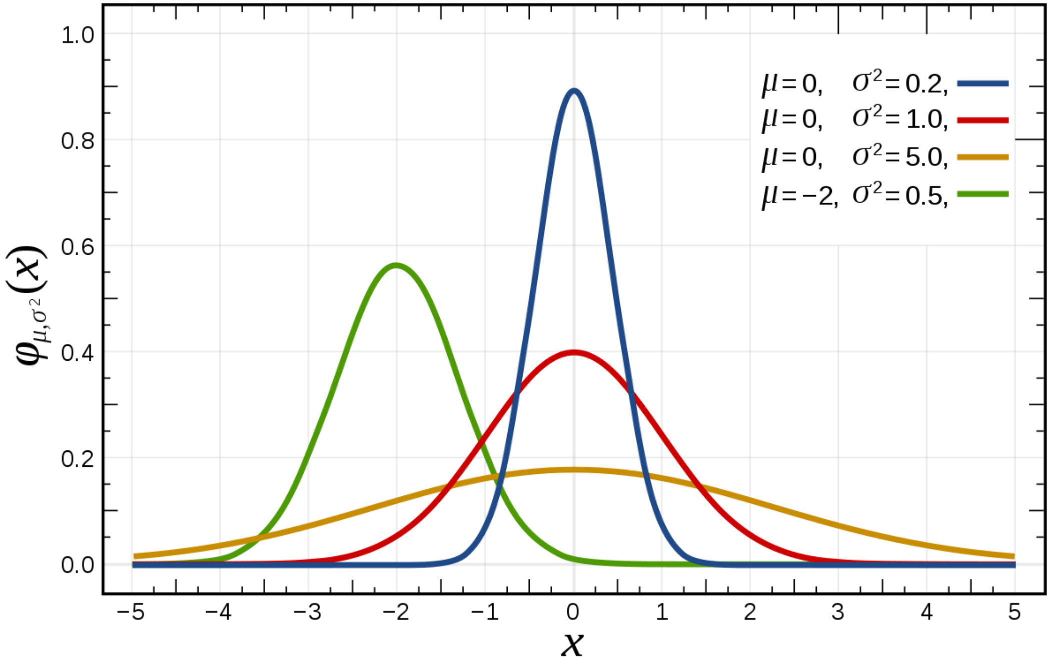


$$-\frac{1}{2\sigma^2}(x-\mu)^2$$

Mean = Location

- La media de la distribución es el valor esperado/promedio.
- Podemos estimar esto usando la media de la muestra.

$$\hat{\mu}_y = \sum_{i=1}^{n} y_{[i]}/n$$



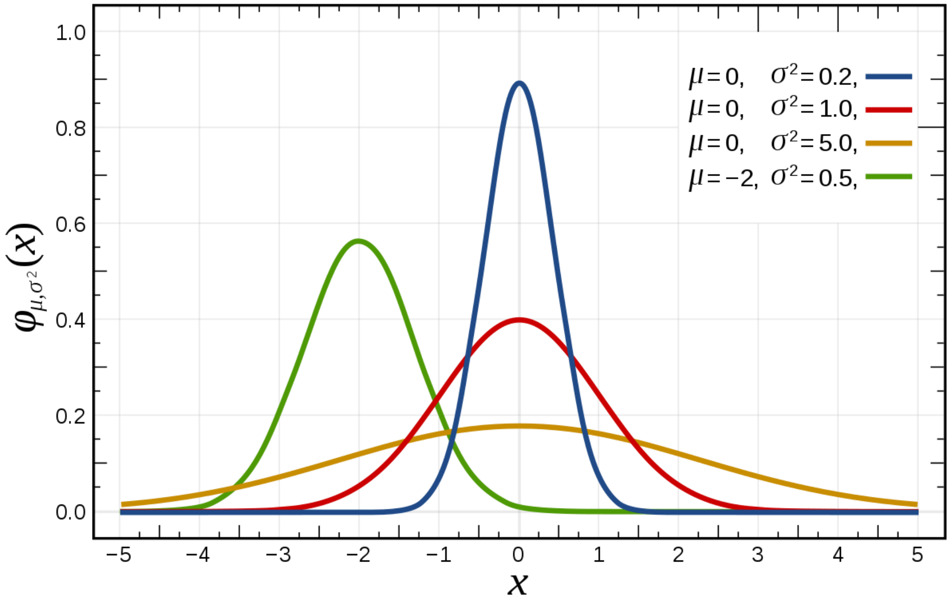
Standard Deviation = Spread

- La varianza de la distribución es el valor esperado/promedio de las desviaciones cuadráticas de la media.

- Podemos estimar esto usando la varianza de la muestra.

$$\hat{\sigma}_y^2 = \sum_{i=1}^{n} (y_{[i]} - \hat{\mu}_y)^2 / (n-1)$$

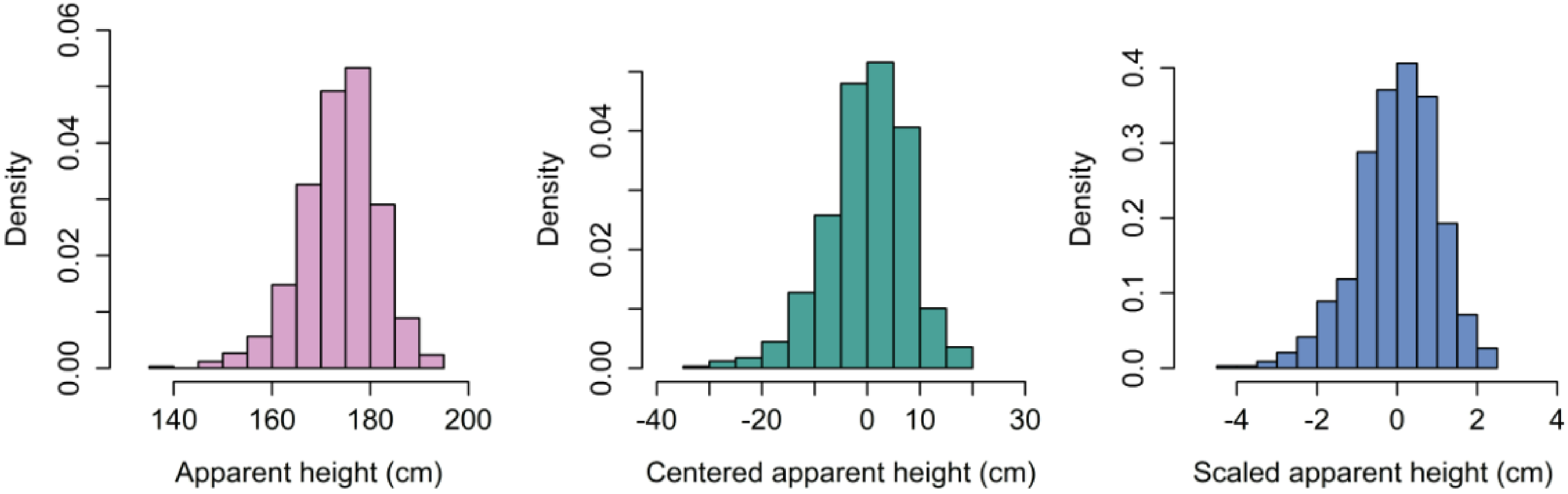
$$\hat{\sigma}_y = \sqrt{\hat{\sigma}_y^2} = \sqrt{\sum_{i=1}^{n} (y_{[i]} - \mu_y)^2 / (n-1)}$$



The Standard Normal Distribution

- Una distribución normal estándar tiene μ=0 y σ=1.

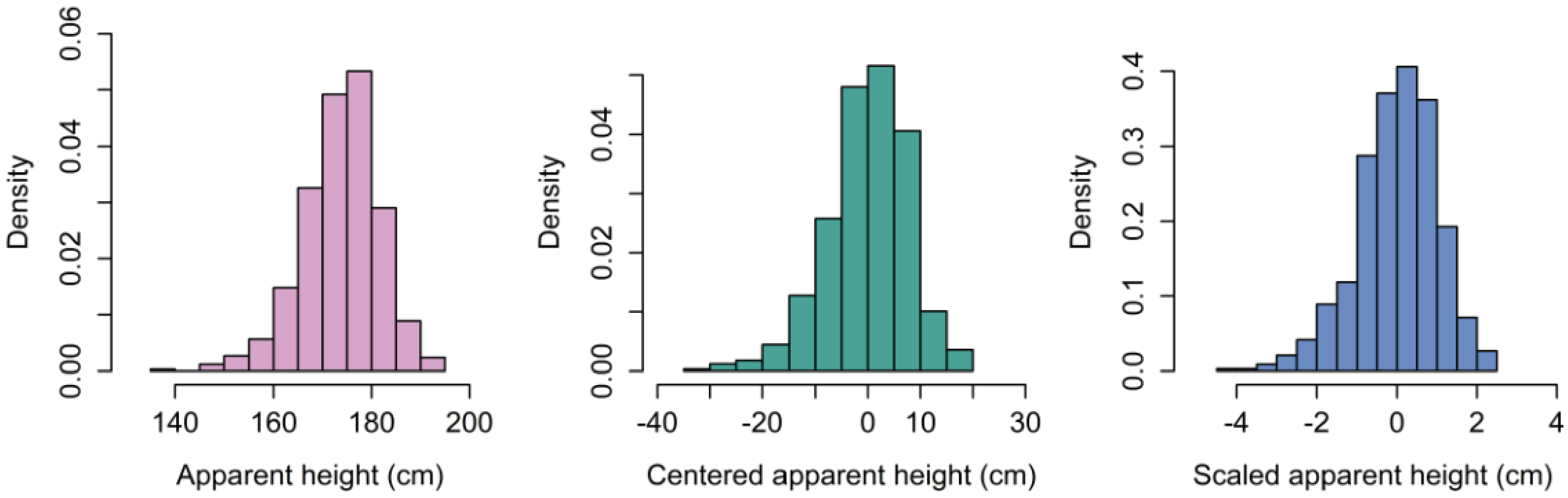
$$z = (x - \mu)/\sigma \qquad x = z \cdot \sigma + \mu$$



Standardization (z scores)

Esto es útil porque:

- Todas las distribuciones normales son iguales cuando se expresan en unidades estandarizadas.
- Las probabilidades de los cuantiles estándar se pueden aprender fácilmente.



Models and Inference

- Modelo exacto: El modelo "real" que conduce a la ‘Verdad’. Se basa en el conocimiento real de todos los procesos y relaciones subyacentes.

- Modelos aproximados: Modelos lo suficientemente buenos para algún propósito. Puede estar equivocado de maneras impredecibles.

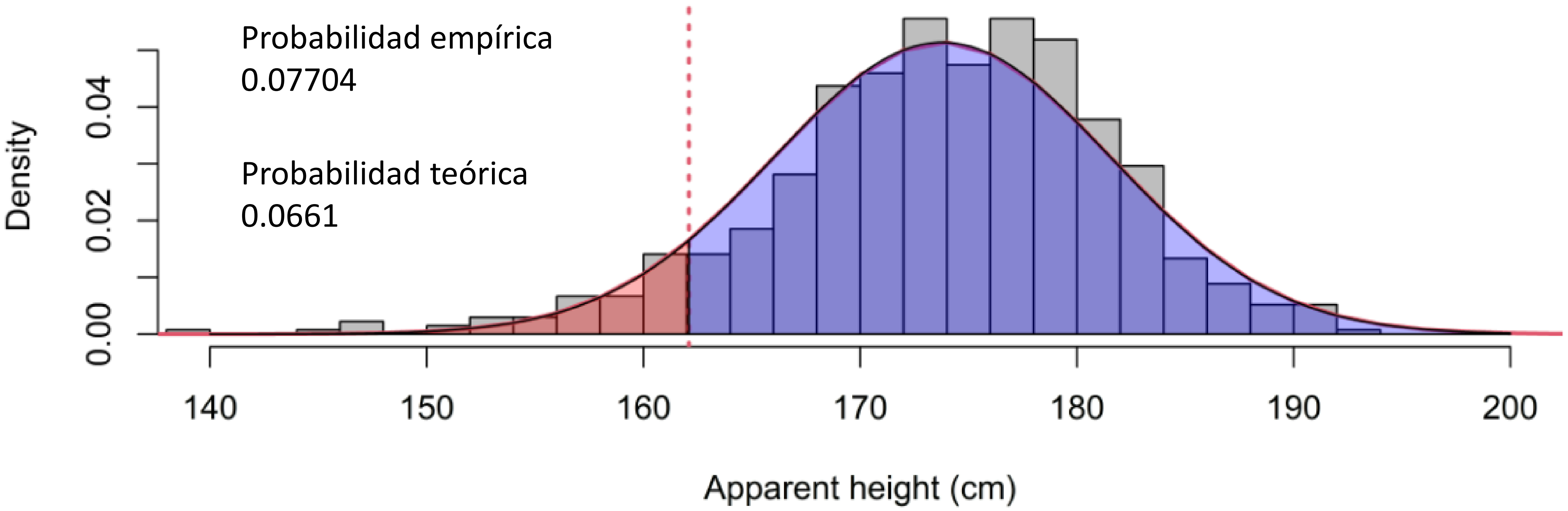
Solo tenemos acceso a estos.

Theoretical Probabilities

- Probabilidades teóricas: basadas en modelos sobre el mundo y no en observaciones.
- Los modelos estadísticos se basan principalmente en probabilidades teóricas.
- Dado que nuestros modelos son solo aproximados y nunca exactos, nuestras probabilidades teóricas también pueden ser solo aproximadas y nunca exactas.

Theoretical Probabilities

- La probabilidad teórica es fiable si y si y si y si....



Verosimilitud

- La probabilidad (densidad) conjunta de observar los datos observados, dados los valores de parámetros específicos.

$$\mathcal{L}_{(\mu|x)} = \mathrm{P}(\mathrm{x}|\mu)$$

- Para varios puntos de datos independientes, podemos simplemente multiplicar la densidad sobre cada punto.

$$\mathcal{L}_{(\mu|x)} = \mathrm{P}(\mathrm{x}_1\, , x_2, \dots, x_n|\mu)$$

$$\mathcal{L}_{(\mu|x)} = \mathrm{P}(\mathrm{x}_1|\mu) \cdot \mathrm{P}(\mathrm{x}_1|\mu) \cdot \dots \cdot \mathrm{P}(\mathrm{x}_n|\mu)$$

Likelihood

- Debajo tenemos la densidad de unión de x_1 y x_2, suponiendo que son independientes.

$$f(x_1, x_2) = [\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x_1-\mu)^2)] \cdot [\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x_2-\mu)^2)] \qquad (2.13)$$

- Si consideramos esto como una función del valor de μ, tenemos la función para la probabilidad de la media.

$$\mathcal{L}_{(\mu|x)} = [\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x_1-\mu)^2)] \cdot [\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x_2-\mu)^2)] \qquad (2.14)$$

Likelihood

- Si asumimos que todos nuestros datos son independientes, la probabilidad se puede calcular multiplicando un montón de probabilidades independientes.

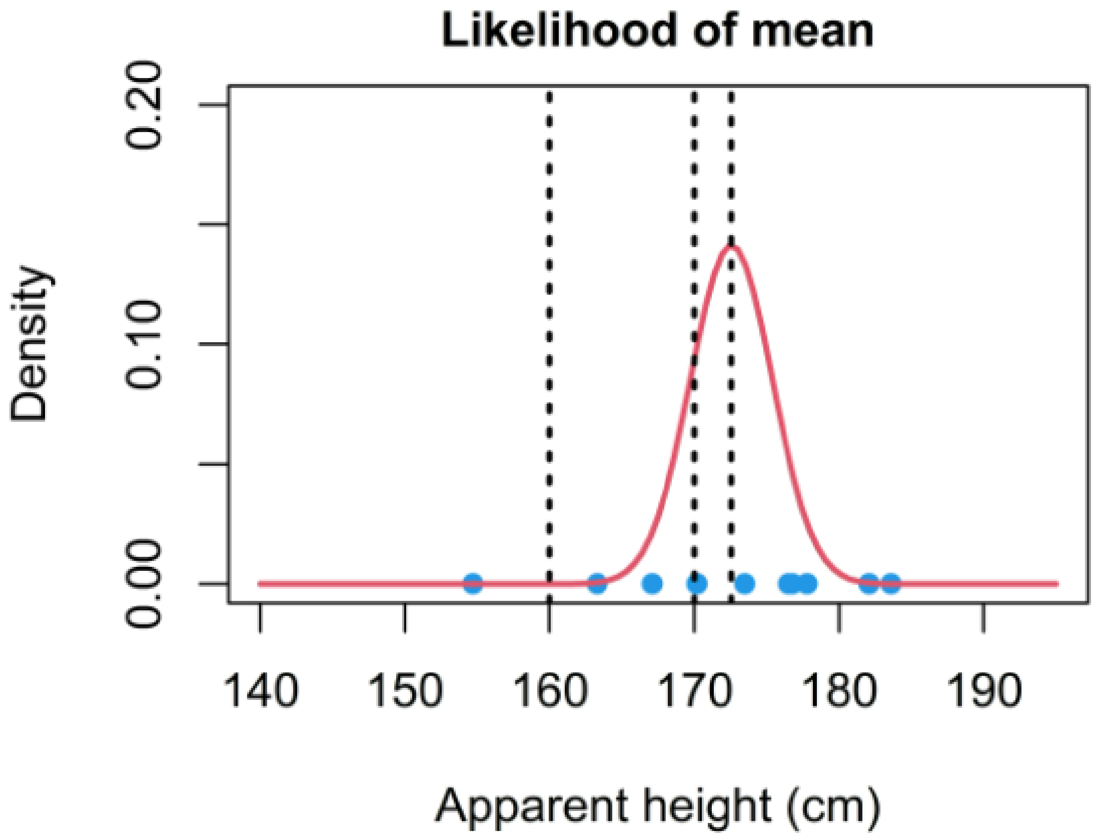
$$\mathcal{L}_{(\mu|x)} = [\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x_1-\mu)^2)] \cdot [\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x_2-\mu)^2)] \qquad (2.14)$$

- Si nuestros datos no son independientes, calcular las probabilidades puede ser mucho más complicado.

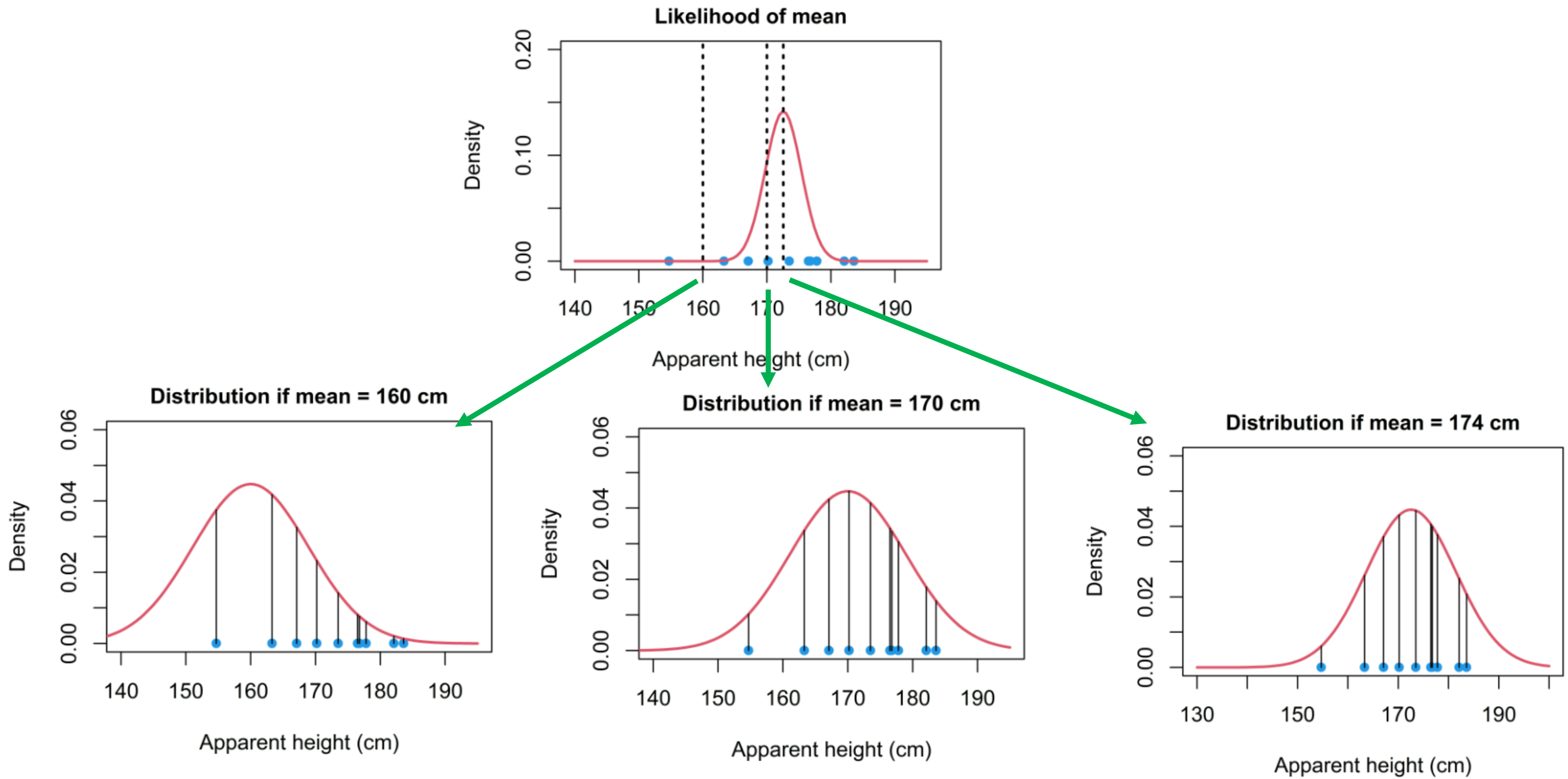
Función de Verosimilitud

- Función que muestra la verosimilitud de diferentes valores de los parámetros de interés.

- Los valores absolutos no son muy útiles, los valores relativos importan más.



Calculating Likelihoods



Probability and Likelihood

- La probabilidad y la verosimilitud son inversas entre sí.
- Las probabilidades expresan la frecuencia relativa esperada de los datos observados, dados algún valor de parámetros.
  - Para las probabilidades, los parámetros son fijos y los datos son variables.
- Las verosimilitudes expresan la credibilidad relativa de los parámetros, dados algunos datos observados.
  - Para las verosimilitudes, los datos son fijos y los parámetros son variables.

Probability and Likelihood: Example

- Ves a tu amigo tirar 100 tiros de tres puntos en Baloncesto y anota 12 de ellos. Afirman que normalmente suelen meter alrededor del 50%.

- ¿Les crees? Probablemente no. ¿Por qué no?

- Si alguien realmente mete el 50% de sus tiros (parámetro fijo), es improbable que solo meta el 12% un día (datos variables).

- Para alguien que metio solo el 12% de sus tiros un día (datos fijos) la habilidad de meter 50% tiene baja verosimilitud (parámetro variable).

Characteristics of Likelihoods

- Mucho detalle en el libro.
- La parte importante es que las probabilidades se reducen bajo dos condiciones:
  - Si aumentas el número de observaciones.
  - Si el error (varianza residual) baja.

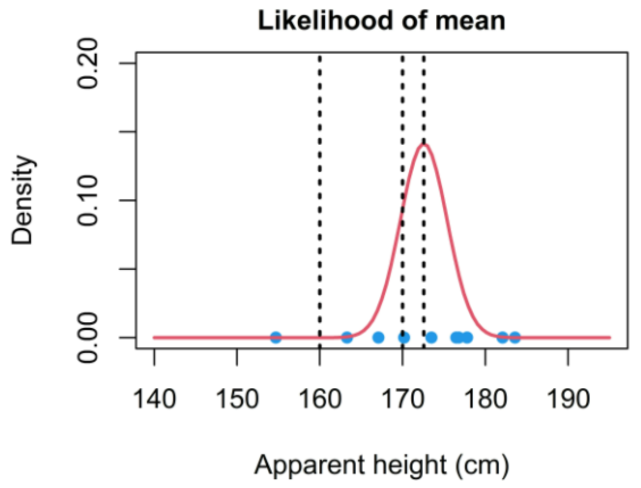
Likelihood and Inference

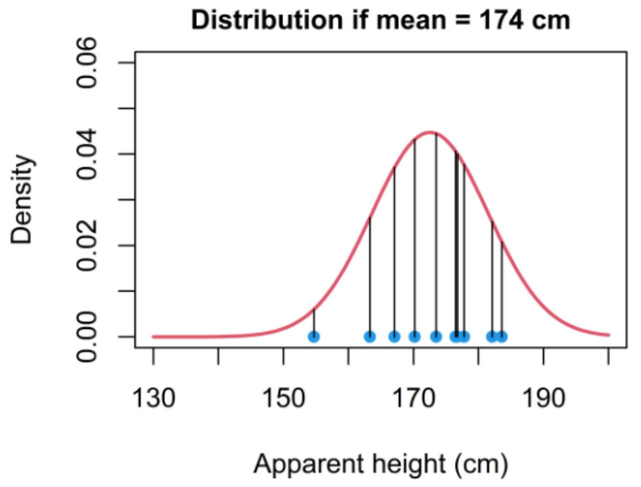
- Por lo general, deseas hacer inferencias sobre los valores de los parámetros (tus hipótesis), no sobre resultados específicos (tus datos).

- Por lo tanto, normalmente se hacen inferencias basadas en verosimilitudes en lugar de probabilidades.

- Los modelos estadísticos se centran en la interpretación de las verosimilitudes de los parámetros.

Calculating the Likelihood for our Data






```
# make candidates for mean parameter
mus = seq (172.5,175, .01)

# easy way to make zero vector of same length as above
log_likelihood = mus*0

# add the log-density of all observations. Notice only the
# mean changes across iterations of the for loop.
for (i in 1:length(mus)) log_likelihood[i] =
  sum (dnorm (mens_height, mus[i], sd(mens_height), log = TRUE))
```

Logarithms

- Convierte la multiplicación en suma y la exponenciación en multiplicación.

- La probabilidad de 675 eventos con P=0.1 es:

$$0.1^{675}$$

- O....

$$\log(0.1) = -2.3$$

$$\log(0.1^{675}) = \log(0.1) \cdot 675 = -1554.25$$

$$\log(e^x) = x$$

$$\log(1) = 0$$

$$\text{if } x < 1, \log(x) < 0$$

$$\text{if } x > 1, \log(x) > 0$$

$$\text{if } x < 0 \ \log(x) = \text{undefined}$$

$$\log(x^y) = \log(x) \cdot y$$

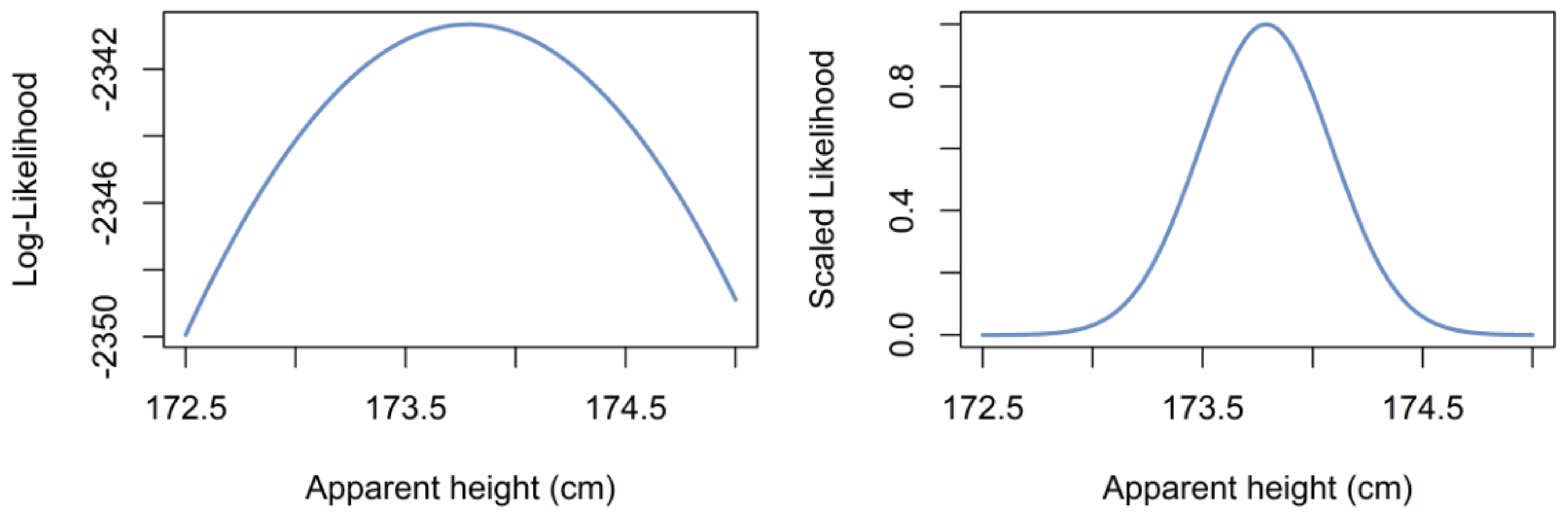
$$\log(\sqrt[y]{x}) = \log(x) / y$$

$$\log(x) + \log(y) = \log(x \cdot y)$$

$$\log(x) - \log(y) = \log(x / y)$$

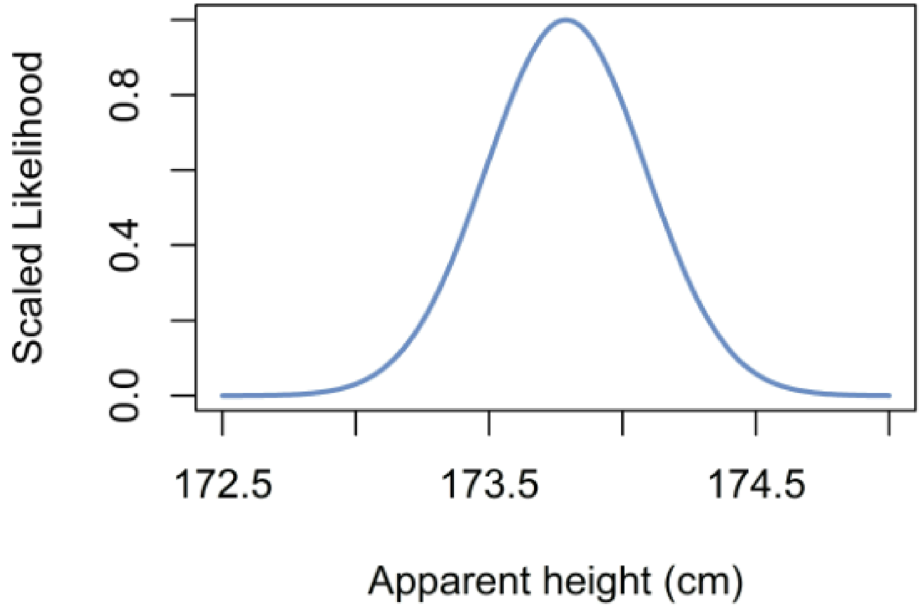
Log-Likelihoods

- En la práctica, las verosimilitudes a menudo deben calcularse como verosimilitudes logarítmicas.



Likelihoods and Inference

- Inferimos los valores más verosímiles de la media dados nuestros datos utilizando su función de verosimilitud.




```r
# find index number of highest values in log-likelihood
maximum = which.max(scaled_log_likelihood)

# print and compare to sample mean
mus[maximum]
## [1] 173.8
mean (mens_height)
## [1] 173.8
```

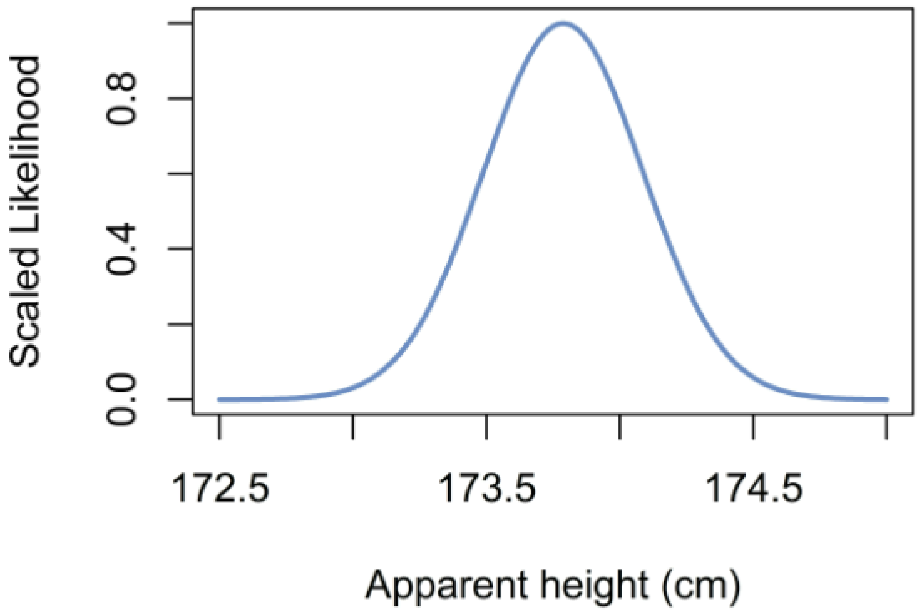
Answering our Research Questions

(1) ¿Qué tan alto "suena" el hombre adulto promedio?

Unos 173.8 cm de altura.

(2) ¿Podemos establecer límites a las alturas aparentes medias creíbles en función de los datos que hemos recopilado?

Sí, entre 173 y 174.5 cm



Exercises

- Utilice las técnicas del capítulo 2 para analizar los valores razonables de la media y la desviación estándar de alguna variable.

- Escriba un informe utilizando un archivo qmd.

- Incluya al menos dos gráficos y describa la información que representan.