{desafío}
latam_

Variables Aleatorias y Gráficos_

Sesión Presencial 2

{desafío}
latam_

Activación de Conceptos

- En la unidad anterior aprendimos sobre probabilidades y algunos gráficos básicos
- ¡Pongamos a prueba nuestros conocimientos!

¿Cómo generamos un histograma con barras azules?

- `histogram(x, color=blue)`
- `df['x'].histogram(color=blue)`
- `plt.hist(x, color="blue")`
- `df['x'].plt(kind = 'histogram')`

¿Cómo recodificamos la siguiente variable?

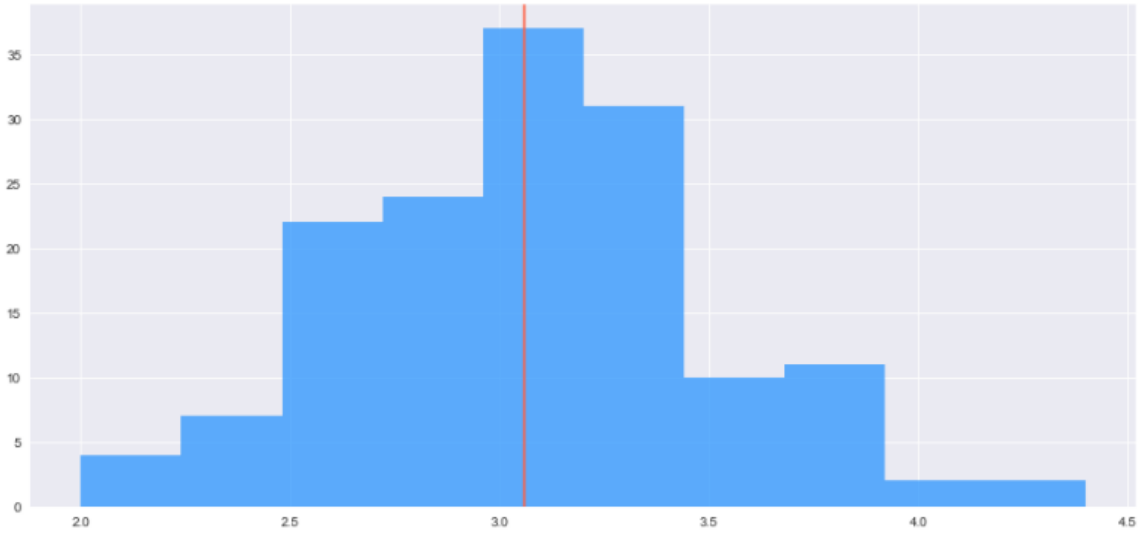
```
df['animales'] = [1, 4, 5, 2, 3, 1]
```

Donde `1` representa Gato, `2` representa Tapir, `3` representa Lemur, `4` representa Perro y `5` representa Cocodrilo.

- `df['animales'].recode(1=>'Gato', 2=>'Tapir', 3=>'Lemur', 4=>'Perro', 5=>'Cocodrilo')`
- `df['animales'].replace([1, 2, 3, 4, 5], ['Gato', 'Tapir', 'Lemur', 'Perro', 'Cocodrilo'])`
- `df['animales'].replace(["1", "2", "3", "4", "5"], ['Gato', 'Tapir', 'Lemur', 'Perro', 'Cocodrilo']`
- `df['animales'].recode([1, 2, 3, 4, 5], ['Gato', 'Tapir', 'Lemur', 'Perro', 'Cocodrilo'])`

¿Qué tan normal es esta variable?

In [13]:
```
import seaborn as sns; df = sns.load_dataset('iris')
_,_ = plt.subplots(figsize = (15,7))
plt.hist(df['sepal_width'], color='dodgerblue', alpha=.7)
plt.axvline(np.mean(df['sepal_width']), color='tomato');
```



Aspectos Avanzados

{desafío}
latam_

Objetivos

- Hoy estudiaremos sobre la distribución normal y su omnipresencia.

- También hablaremos sobre los aspectos *asintóticos* de las distribuciones:
    - Nos permiten aproximar pruebas y regiones de confianza (más sobre ésto la próxima semana).
    - Saber esto nos habilita para generar diagnósticos en casos donde la distribución es intractable o cuando se asume una aproximación.

Distribución normal estandarizada

- Hablamos de estandarizada cuando:

$$X \sim \mathcal{N}(0, 1)$$

- Debe ser la distribución más utilizada. Esto dado que permite capturar diversos fenómenos como:
    - Captura la ausencia de efecto en un estimador.
    - Reescala y centra todos las observaciones de una variable, facilitando la comparación

Algunas características importantes

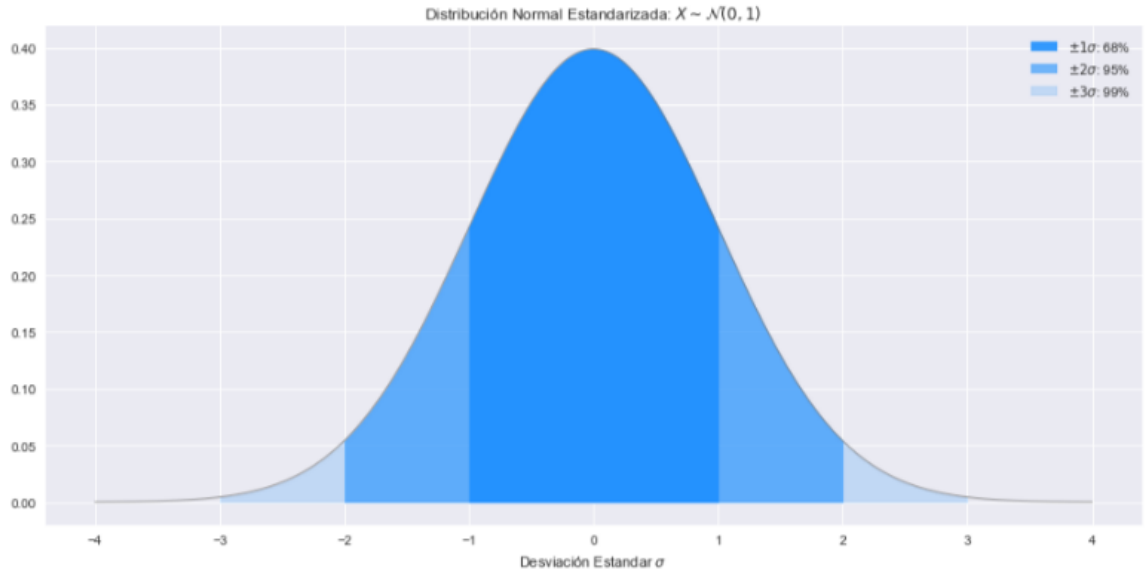
- **Unimodal**: tiene sólo un punto que es el más alto, donde se concentran la mayoría de los datos.
- **Simétrica**: la mayoría de los datos gravitan alrededor de la media.
- **Positiva**: todos los valores (positivos o negativos) tienen una probabilidad $\Pr \geq 0$ de suceder.

Relación entre media y varianza

- Los componentes paramétricos permiten aproximar la cantidad de casos bajo la curva:

Porcentaje	Límites
68%	$\mu \pm 1\sigma$
95%	$\mu \pm 2\sigma$
99%	$\mu \pm 3\sigma$

In [10]:
```
_,_ = plt.subplots(figsize = (15,7))
gfx.normal_distribution_sigma()
```



Puntaje Z

- Con la distribución normal estandarizada podemos reescalar observaciones respecto a la media de una variable. Esto se conoce como la estandarización de una variable.

$$\text{Puntaje Z} = \frac{x_i - \bar{x}}{\sigma}$$

- La reconversión de una observación nos indica **a cuántas desviaciones estándares se encuentra una observación respecto a la media**.

Aspectos Asintóticos

{desafío}
latam_

Ley de los Grandes Números

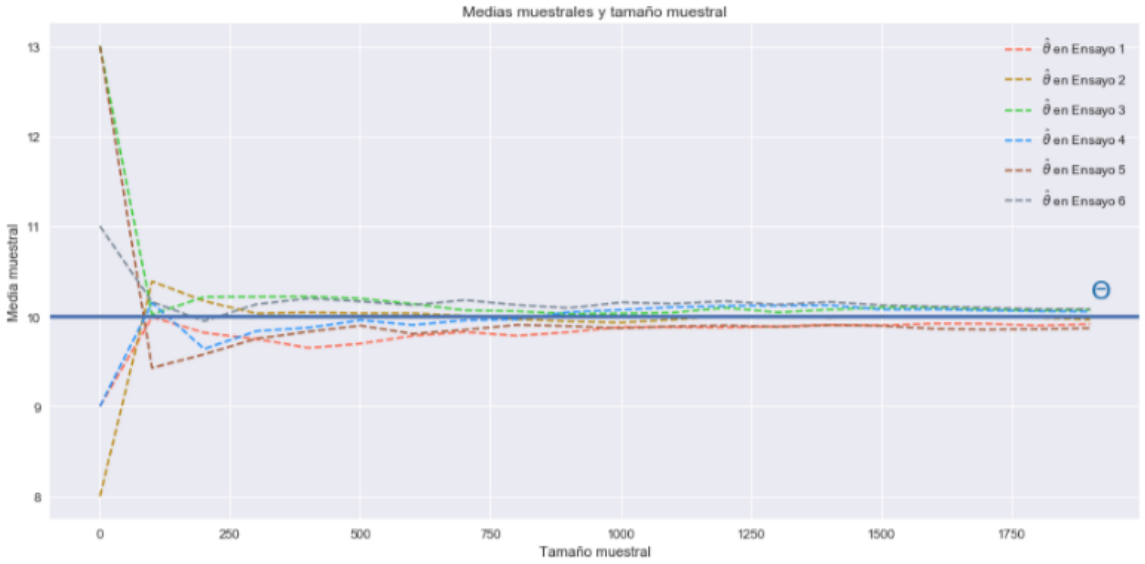
La ley de los grandes números establece que en una sucesión infinita de variables aleatorias i.i.d (independientes e idénticamente distribuídas) con expectativa μ y varianza σ^2, el promedio de la sucesión:

$$\bar{X}_n = (X_1 + X_2 + \ldots + X_n)/n$$

convergerá en probabilidad a μ.

In [11]:

```
_,_ = plt.subplots(figsize = (15,7))
gfx.law_large_numbers()
```



Teorema del Límite Central

Si tenemos una secuencia de variables aleatorias independientes con media μ y varianza finita σ^2,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

donde $\xrightarrow{d}$ significa **Convergencia en distribución**.

In [12]:

```
_,_ = plt.subplots(figsize = (15,7))
gfx.central_limit_theorem()
```