Análisis de Datos para Inteligencia Artificial

Ing. Magdalena Bouza, Dr. Ing. Pablo Briff, Lic. Nicolás Horro

Laboratorio de Sistemas Embebidos - FIUBA

bouza.magdalena@gmail.com, pbriff@fi.uba.ar, nhorro@gmail.com

Mayo de 2021

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tabla de Contenidos I

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tabla de Contenidos II

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Tabla de Contenidos III

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Estimación de Intervalo

- La estimación puntual estima un parámetro θ

Repaso de Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza
- Llamaremos α al nivel de significación

Repaso de Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza
- Llamaremos α al nivel de significación
- Llamaremos a $1-\alpha$ el nivel de confianza

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Estimación de Intervalo

- La estimación puntual estima un parámetro θ
- En la estimación por intervalo queremos estimar el intervalo en el cual está θ con un cierto grado de confianza
- A esto se lo denomina Intervalo de Confianza
- Llamaremos α al nivel de significación
- Llamaremos a $1-\alpha$ el nivel de confianza
- Para estimar el intervalo, vamos a usar la pdf del estimador puntual

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$
- $P(-1.96 < Z < 1.96) = 0.95$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Estimación de Intervalo

- Sea $X = \{x_i\}_{i=1}^{N}$ una muestra de N puntos de una v.a. normal con media μ (desconocida) y varianza σ^2 (conocida)
- Queremos estimar la media μ con la media muestral $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- $\hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$
- Definimos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$
- $P(-1.96 < Z < 1.96) = 0.95$
- En otras palabras, Z está en ese intervalo con una confianza del 95%

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- Reemplazando queda:

- Reemplazando queda:
- $P(-1.96 < \sqrt{N}\frac{\hat{\mu}-\mu}{\sigma} < 1.96) = 0.95$

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N}\frac{\hat{\mu}-\mu}{\sigma} < 1.96) = 0.95$
- Despejando μ:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N}\frac{\hat{\mu}-\mu}{\sigma} < 1.96) = 0.95$
- Despejando μ:
- $P(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}}) = 0.95$

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N}\frac{\hat{\mu}-\mu}{\sigma} < 1.96) = 0.95$
- Despejando μ:
- $P(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}}) = 0.95$
- Este es un intervalo de confianza de dos lados (colas de la normal por exceso y por defecto)

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N}\frac{\hat{\mu}-\mu}{\sigma} < 1.96) = 0.95$
- Despejando μ:
- $P(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}}) = 0.95$
- Este es un intervalo de confianza de dos lados (colas de la normal por exceso y por defecto)
- Se puede calcular el intervalo de confianza para un valor arbitrario dado

Estimación de Intervalo

- Reemplazando queda:
- $P(-1.96 < \sqrt{N}\frac{\hat{\mu}-\mu}{\sigma} < 1.96) = 0.95$
- Despejando μ:
- $P(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}} < \mu < \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}}) = 0.95$
- Este es un intervalo de confianza de dos lados (colas de la normal por exceso y por defecto)
- Se puede calcular el intervalo de confianza para un valor arbitrario dado
- A medida que aumenta la confianza, el intervalo se hace más grande

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \texttt{normcdf}(-1.96) = 0.025 = \alpha/2$

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \texttt{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \texttt{normcdf}(1.96) = 0.975 = 1 - \alpha/2$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \texttt{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \texttt{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \texttt{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \texttt{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \texttt{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \texttt{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$

Intervalo de Confianza Bilateral

- $P(Z < -z_\alpha) = P(Z > z_\alpha) = \alpha, \quad 0 < \alpha < 1$
- Dado que Z es simétrica alrededor de la media, se cumple
- $z_{1-\alpha/2} = -z_{\alpha/2}$
- $P(Z < -z_{\alpha/2}) = P(Z > z_{\alpha/2}) = \alpha/2$
- Para $\alpha = 0.05$, $P(Z < -z_{\alpha/2}) = \texttt{normcdf}(-1.96) = 0.025 = \alpha/2$
- $P(Z > z_{\alpha/2}) = \texttt{normcdf}(1.96) = 0.975 = 1 - \alpha/2$
- Para un intervalo de confianza $1 - \alpha$ se cumple:
- $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$
- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- Entonces al intervalo $\mu \in (\hat{\mu} - 1.96\frac{\sigma}{\sqrt{N}}, \hat{\mu} + 1.96\frac{\sigma}{\sqrt{N}})$ se lo llama *intervalo de confianza de 95% bilateral*

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}} < \mu) = 0.95$

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*
- Podemos definir análogamente el intervalo de confianza de 95% inferior unilateral como $\mu \in (-\infty, \hat{\mu} + 1.64\frac{\sigma}{\sqrt{N}})$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*
- Podemos definir análogamente el intervalo de confianza de 95% inferior unilateral como $\mu \in (-\infty, \hat{\mu} + 1.64\frac{\sigma}{\sqrt{N}})$
- El intervalo de confianza no nos dice nada acerca de la distribución de μ !

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Intervalo de Confianza Unilateral

- $P(\hat{\mu} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < Z < \hat{\mu} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}) = 1 - \alpha$
- El intervalo de confianza porcentual $100 \times (1 - \alpha)$ está dentro de $z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$ unidades de la media muestral $\hat{\mu}$
- Dado que $P(Z < 1.64) = 0.95$, tenemos
- $P(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}} < \mu) = 0.95$
- Entonces al intervalo $(\hat{\mu} - 1.64\frac{\sigma}{\sqrt{N}}, \infty)$ se lo llama *intervalo de confianza de 95% superior unilateral*
- Podemos definir análogamente el intervalo de confianza de 95% inferior unilateral como $\mu \in (-\infty, \hat{\mu} + 1.64\frac{\sigma}{\sqrt{N}})$
- El intervalo de confianza no nos dice nada acerca de la distribución de μ !
- De hecho μ puede ser una constante desconocida

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis
- N.B.: varios textos hablan de *no rechazar* en lugar de *aceptar* la hipótesis

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis
- N.B.: varios textos hablan de *no rechazar* en lugar de *aceptar* la hipótesis
- En caso contrario, rechazamos la hipótesis

Repaso de Test de Hipótesis

- A veces vamos a querer ensayar una hipótesis
- Por ejemplo si la media es mayor o distinta de un valor
- Una hipótesis es el planteo de una pregunta científica en términos de un valor propuesto de un parámetro en un modelo probabilístico
- Si las muestras aleatorias son consistentes con la hipótesis, aceptamos la hipótesis
- N.B.: varios textos hablan de *no rechazar* en lugar de *aceptar* la hipótesis
- En caso contrario, rechazamos la hipótesis
- De todas formas, nunca decimos que la hipótesis es verdadera o falsa, sino que los datos la avalan o no con un cierto nivel de confianza

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2
- Queremos ensayar la hipótesis que $\mu = \mu_0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2
- Queremos ensayar la hipótesis que $\mu = \mu_0$
- La hipótesis nula es $\mathcal{H}_0 : \mu = \mu_0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- En un test de hipótesis definimos una estadística que debe obedecer cierta distribución si la hipótesis es correcta
- Si la estadística calculada tiene muy baja probabilidad de ocurrir bajo esa hipótesis, rechazamos la hipótesis
- Caso contrario, la aceptamos
- Por ejemplo, tenemos muestras de una distribución de la que desconocemos la media μ pero conocemos la varianza σ^2
- Queremos ensayar la hipótesis que $\mu = \mu_0$
- La hipótesis nula es $\mathcal{H}_0 : \mu = \mu_0$
- La hipótesis alternativa es $\mathcal{H}_1 : \mu = \mu_1 \neq \mu_0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula
- Es decir:

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula
- Es decir:
- Aceptamos la hipótesis si $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \in (-z_{\alpha/2}, z_{\alpha/2})$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de Hipótesis

- Vamos a comparar la media muestral con la hipótesis
- Si la media muestral cae fuera de un intervalo de confianza $100 \times (1 - \alpha)$ rechazamos la hipótesis nula
- Es decir:
- Aceptamos la hipótesis si $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \in (-z_{\alpha/2}, z_{\alpha/2})$
- Este es un test bilateral (rechazamos la hipótesis si μ cae fuera del intervalo)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α, típicamente se elige 0.1, 0.05, 0.01

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α, típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α, típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*
- $\beta = P(\text{Tipo II}|\mathcal{H}_1)$

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α, típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*
- $\beta = P(\text{Tipo II}|\mathcal{H}_1)$
- $\beta(\mu) = P_\mu\left(-z_{\alpha/2} < \frac{\sqrt{N}(\hat{\mu}-\mu_0)}{\sigma} < z_{\alpha/2}\right)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tipos de Errores

- Error de tipo I: ocurre cuando rechazamos la hipótesis cuando es correcta, es decir es una *falsa alarma*
- Formalmente, $\alpha = P(\text{Tipo I}|\mathcal{H}_0)$
- La p. de error tipo I es α, típicamente se elige 0.1, 0.05, 0.01
- Error de tipo II: ocurre cuando aceptamos la hipótesis cuando es incorrecta, es decir *no activar la alarma*
- $\beta = P(\text{Tipo II}|\mathcal{H}_1)$
- $\beta(\mu) = P_\mu\left(-z_{\alpha/2} < \frac{\sqrt{N}(\hat{\mu}-\mu_0)}{\sigma} < z_{\alpha/2}\right)$
- Esta p. aumenta a medida que μ se acerca a μ_0

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$
- El intervalo de confianza de nivel $100 \times (1 - \alpha)$ ahora se define en base a un solo lado donde $\hat{\mu}$ debe caer para no rechazar $\mathcal{H}_0$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$
- El intervalo de confianza de nivel $100 \times (1 - \alpha)$ ahora se define en base a un solo lado donde $\hat{\mu}$ debe caer para no rechazar $\mathcal{H}_0$
- Aceptamos $\mathcal{H}_0$ si:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ensayo Unilateral

- Para un ensayo unilateral, tenemos:
- $\mathcal{H}_0 : \mu \leq \mu_0$
- $\mathcal{H}_1 : \mu > \mu_0$
- El intervalo de confianza de nivel $100 \times (1 - \alpha)$ ahora se define en base a un solo lado donde $\hat{\mu}$ debe caer para no rechazar $\mathcal{H}_0$
- Aceptamos $\mathcal{H}_0$ si:
- $\frac{\sqrt{N}(\hat{\mu}-\mu_0)}{\sigma} \in (-\infty, z_\alpha)$

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$
- Para el caso anterior de ensayo bilateral de medias, aceptamos $\mathcal{H}_0$ si

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$
- Para el caso anterior de ensayo bilateral de medias, aceptamos $\mathcal{H}_0$ si
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \in \left(-t_{\alpha/2,(N-1)}, t_{\alpha/2,(N-1)}\right)$

Test de Hipótesis con Varianza Desconocida

- Si la varianza es desconocida, usamos la distribución t de Student
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \sim t_{N-1}$
- Para el caso anterior de ensayo bilateral de medias, aceptamos $\mathcal{H}_0$ si
- $\frac{\sqrt{N}(\hat{\mu}-\mu)}{\sigma} \in \left(-t_{\alpha/2,(N-1)}, t_{\alpha/2,(N-1)}\right)$
- Es decir, este es un ensayo bilateral t

Valor p (p-value) y Puntaje Z (Z-score)

- Anteriormente vimos que en estimación de intervalo usamos la siguiente estadística:

Valor p (p-value) y Puntaje Z (Z-score)

- Anteriormente vimos que en estimación de intervalo usamos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Valor p (p-value) y Puntaje Z (Z-score)

- Anteriormente vimos que en estimación de intervalo usamos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$
- A esto llamamos z-score o Puntaje z

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Valor p (p-value) y Puntaje Z (Z-score)

- Anteriormente vimos que en estimación de intervalo usamos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$
- A esto llamamos z-score o Puntaje z
- Definimos al valor p (p-value) al valor de probabilidad que nos arroja el z-score al calcularlo con las las muestras disponibles ($\hat{\mu}$), asumiendo que la hipótesis es la nula $\mathcal{H}_0$ (es decir $\mu = \mu_0$)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Valor p (p-value) y Puntaje Z (Z-score)

- Anteriormente vimos que en estimación de intervalo usamos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$
- A esto llamamos z-score o Puntaje z
- Definimos al valor p (p-value) al valor de probabilidad que nos arroja el z-score al calcularlo con las las muestras disponibles ($\hat{\mu}$), asumiendo que la hipótesis es la nula $\mathcal{H}_0$ (es decir $\mu = \mu_0$)
- Por ejemplo, en un test unilateral, rechazamos $\mathcal{H}_0$ si el p-value es menor que la significancia α

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Valor p (p-value) y Puntaje Z (Z-score)

- Anteriormente vimos que en estimación de intervalo usamos la siguiente estadística:
- $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$
- A esto llamamos z-score o Puntaje z
- Definimos al valor p (p-value) al valor de probabilidad que nos arroja el z-score al calcularlo con las las muestras disponibles ($\hat{\mu}$), asumiendo que la hipótesis es la nula $\mathcal{H}_0$ (es decir $\mu = \mu_0$)
- Por ejemplo, en un test unilateral, rechazamos $\mathcal{H}_0$ si el p-value es menor que la significancia α
- Esto es porque, basado en $\mathcal{H}_0$, la p. de tener las muestras obtenidas es muy baja (menor que el nivel de significancia que toleramos)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones
- Primero, ingresamos los datos en una tabla de r filas y c columnas

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones
- Primero, ingresamos los datos en una tabla de r filas y c columnas
- Generamos la $\mathcal{H}_0$: las filas y columnas son independientes, es decir:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones
- Primero, ingresamos los datos en una tabla de r filas y c columnas
- Generamos la $\mathcal{H}_0$: las filas y columnas son independientes, es decir:
- La p. de que las filas tengan un valor x_i y las columnas tengan un valor x_j es $\hat{\pi}_{ij} = \pi_i \pi_j$

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones
- Primero, ingresamos los datos en una tabla de r filas y c columnas
- Generamos la $\mathcal{H}_0$: las filas y columnas son independientes, es decir:
- La p. de que las filas tengan un valor x_i y las columnas tengan un valor x_j es $\hat{\pi}_{ij} = \pi_i \pi_j$
- A su vez, el valor esperado es $\hat{E}_{ij} = n\hat{\pi}_{ij}$

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones
- Primero, ingresamos los datos en una tabla de r filas y c columnas
- Generamos la $\mathcal{H}_0$: las filas y columnas son independientes, es decir:
- La p. de que las filas tengan un valor x_i y las columnas tengan un valor x_j es $\hat{\pi}_{ij} = \pi_i \pi_j$
- A su vez, el valor esperado es $\hat{E}_{ij} = n\hat{\pi}_{ij}$
- En este tipo de ensayos, vamos a valernos del valor esperado de cada elemento de la tabla en el caso de que las filas y columnas sean independientes

Test Independencia de Pearson

- Este ensayo se usa para examinar la independencia de dos variables en base a n observaciones
- Primero, ingresamos los datos en una tabla de r filas y c columnas
- Generamos la $\mathcal{H}_0$: las filas y columnas son independientes, es decir:
- La p. de que las filas tengan un valor x_i y las columnas tengan un valor x_j es $\hat{\pi}_{ij} = \pi_i \pi_j$
- A su vez, el valor esperado es $\hat{E}_{ij} = n\hat{\pi}_{ij}$
- En este tipo de ensayos, vamos a valernos del valor esperado de cada elemento de la tabla en el caso de que las filas y columnas sean independientes
- Luego comparamos cada valor con el real obtenido (mediciones)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia de Pearson

- Recordemos, $\mathcal{H}_0$: las filas y columnas son independientes

Test Independencia de Pearson

- Recordemos, $\mathcal{H}_0$: las filas y columnas son independientes
- Si rechazamos $\mathcal{H}_0$, entonces aceptamos $\mathcal{H}_1$: las filas y columnas son dependientes

Test Independencia de Pearson

- Recordemos, $\mathcal{H}_0$: las filas y columnas son independientes
- Si rechazamos $\mathcal{H}_0$, entonces aceptamos $\mathcal{H}_1$: las filas y columnas son dependientes
- Generamos la estadística del ensayo: $T : \chi^2_{(r-1)\times(c-1)} = \sum_{i,j} \frac{(n_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia de Pearson

- Recordemos, $\mathcal{H}_0$: las filas y columnas son independientes
- Si rechazamos $\mathcal{H}_0$, entonces aceptamos $\mathcal{H}_1$: las filas y columnas son dependientes
- Generamos la estadística del ensayo: $T : \chi^2_{(r-1)\times(c-1)} = \sum_{i,j} \frac{(n_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}}$
- donde $\hat{E}_{ij} = n\hat{\pi}_{ij} = n(n_i/n)(n_j/n) = n_i n_j/n$, y n es la suma de los datos en toda la tabla (grand total)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia de Pearson

- Recordemos, $\mathcal{H}_0$: las filas y columnas son independientes
- Si rechazamos $\mathcal{H}_0$, entonces aceptamos $\mathcal{H}_1$: las filas y columnas son dependientes
- Generamos la estadística del ensayo: $T : \chi^2_{(r-1)\times(c-1)} = \sum_{i,j} \frac{(n_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}}$
- donde $\hat{E}_{ij} = n\hat{\pi}_{ij} = n(n_i/n)(n_j/n) = n_i n_j/n$, y n es la suma de los datos en toda la tabla (grand total)
- Calculamos el percentil $\chi^2_{\alpha,(r-1)\times(c-1)}$, inversa de cdf, que corresponde a $1-\alpha$ (tipicamente $\alpha = 0.05$)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia de Pearson

- Recordemos, $\mathcal{H}_0$: las filas y columnas son independientes
- Si rechazamos $\mathcal{H}_0$, entonces aceptamos $\mathcal{H}_1$: las filas y columnas son dependientes
- Generamos la estadística del ensayo: $T : \chi^2_{(r-1)\times(c-1)} = \sum_{i,j} \frac{(n_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}}$
- donde $\hat{E}_{ij} = n\hat{\pi}_{ij} = n(n_i/n)(n_j/n) = n_i n_j/n$, y n es la suma de los datos en toda la tabla (grand total)
- Calculamos el percentil $\chi^2_{\alpha,(r-1)\times(c-1)}$, inversa de cdf, que corresponde a $1-\alpha$ (tipicamente $\alpha = 0.05$)
- Rechazamos $\mathcal{H}_0$ si $T > \chi^2_\alpha$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test Independencia - Ejemplo

- Se quiere saber si algunos genios del fútbol rinden mejor que otros (meten más goles) en sus equipos que en la selección nacional. Usar un test de independencia con significancia de 5% para responder la pregunta.

Genio del Fútbol	Goles Selec. Nacional	Goles Equipos
Maradona	34	320
Messi	71	741

Datos verdaderos al 13 Mayo 2021.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$
- $\mathcal{H}_1 : \mu_{x_1} \neq \mu_{x_2}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$
- $\mathcal{H}_1 : \mu_{x_1} \neq \mu_{x_2}$
- Las hipótesis del test son:

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$
- $\mathcal{H}_1 : \mu_{x_1} \neq \mu_{x_2}$
- Las hipótesis del test son:
- Los dos grupos de datos siguen distribución normal

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$
- $\mathcal{H}_1 : \mu_{x_1} \neq \mu_{x_2}$
- Las hipótesis del test son:
- Los dos grupos de datos siguen distribución normal
- Los dos grupos de datos tienen variabilidades casi idénticas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$
- $\mathcal{H}_1 : \mu_{x_1} \neq \mu_{x_2}$
- Las hipótesis del test son:
- Los dos grupos de datos siguen distribución normal
- Los dos grupos de datos tienen variabilidades casi idénticas
- Cada grupo es una muestra aleatoria de su población

Test de t de Student de 2 Muestras

- Nos interesa comparar la media muestral de dos conjuntos de datos, y sacar conclusiones sobre si estas son distintas
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Para tal fin, usamos el test de t de Student de 2 muestras sobre las medias *reales* μ_{x_1}, μ_{x_2} basado en las muestras
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2}$
- $\mathcal{H}_1 : \mu_{x_1} \neq \mu_{x_2}$
- Las hipótesis del test son:
- Los dos grupos de datos siguen distribución normal
- Los dos grupos de datos tienen variabilidades casi idénticas
- Cada grupo es una muestra aleatoria de su población
- Todas las observaciones son independientes

FACULTAD DE INGENIERIA Universidad de Buenos Aires

- Consideremos dos muestras X_1, X_2 con n_1, n_2 muestras, medias muestrales $\bar{x}_1, \bar{x}_2$ y varianzas muestrales S_1, S_2 respectivamente

Test de t de Student de 2 Muestras

- Consideremos dos muestras X_1, X_2 con n_1, n_2 muestras, medias muestrales $\bar{x}_1, \bar{x}_2$ y varianzas muestrales S_1, S_2 respectivamente
- Estimamos la varianza poblacional *combinada* (de ambas al mismo tiempo, son muy parecidas por hipótesis), como:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- Consideremos dos muestras X_1, X_2 con n_1, n_2 muestras, medias muestrales $\bar{x}_1, \bar{x}_2$ y varianzas muestrales S_1, S_2 respectivamente
- Estimamos la varianza poblacional *combinada* (de ambas al mismo tiempo, son muy parecidas por hipótesis), como:
- $S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

Test de t de Student de 2 Muestras

- Consideremos dos muestras X_1, X_2 con n_1, n_2 muestras, medias muestrales $\bar{x}_1, \bar{x}_2$ y varianzas muestrales S_1, S_2 respectivamente
- Estimamos la varianza poblacional *combinada* (de ambas al mismo tiempo, son muy parecidas por hipótesis), como:
- $S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$
- Para comparar la diferencia entre las medias reales, vamos a definir el error estandar para la diferencia de las medias muestrales como:

Test de t de Student de 2 Muestras

- Consideremos dos muestras X_1, X_2 con n_1, n_2 muestras, medias muestrales $\bar{x}_1, \bar{x}_2$ y varianzas muestrales S_1, S_2 respectivamente
- Estimamos la varianza poblacional *combinada* (de ambas al mismo tiempo, son muy parecidas por hipótesis), como:
- $S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$
- Para comparar la diferencia entre las medias reales, vamos a definir el error estandar para la diferencia de las medias muestrales como:
- $SE_{dif} = S_{\bar{x}_1-\bar{x}_2} = \sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- La diferencia de medias $\bar{x}_1 - \bar{x}_2$ tiene distribucion t de Student con $df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ grados de libertad

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- La diferencia de medias $\bar{x}_1 - \bar{x}_2$ tiene distribucion t de Student con $df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ grados de libertad
- Definimos un intervalo de confianza de porcentaje $100 \times (1 - \alpha)$ deseado como:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- La diferencia de medias $\bar{x}_1 - \bar{x}_2$ tiene distribucion t de Student con $df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ grados de libertad
- Definimos un intervalo de confianza de porcentaje $100 \times (1 - \alpha)$ deseado como:
- $IC = \left((\bar{x}_1 - \bar{x}_2) - t_{1-\alpha/2,df} \times SE_{dif}, (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2,df} \times SE_{dif}\right)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student de 2 Muestras

- La diferencia de medias $\bar{x}_1 - \bar{x}_2$ tiene distribucion t de Student con $df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ grados de libertad
- Definimos un intervalo de confianza de porcentaje $100 \times (1 - \alpha)$ deseado como:
- $IC = \left((\bar{x}_1 - \bar{x}_2) - t_{1-\alpha/2,df} \times SE_{dif}, (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2,df} \times SE_{dif}\right)$
- Si el IC contiene el cero (es decir no hay diferencia entre las medias), entonces no podemos rechazar $\mathcal{H}_0$, caso contrario rechazamos $\mathcal{H}_0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de t de Student - Ejemplo

- Un grupo de amigos discute en un bar si Messi y Maradona rindieron igual de bien en la selección argentina de fútbol. Proponen usar como criterio la cantidad de goles por partido para describir un comportamiento más general del juego de cada jugador en la selección nacional. Usar un test de t de Student con significancia de 5% para responder la duda planteada por el grupo de amigos.

	Maradona	Messi
No. Partidos en Selección	91	142
Goles Promedio en Selección	0.37	0.5
Desvío estándar Goles en Selección	4.4	5.9

Datos verdaderos al 13 Mayo 2021.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos
- Ahora nos interesa comparar las medias de más de 2 conjuntos de datos y sacar conclusiones sobre si son iguales o no

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos
- Ahora nos interesa comparar las medias de más de 2 conjuntos de datos y sacar conclusiones sobre si son iguales o no
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos
- Ahora nos interesa comparar las medias de más de 2 conjuntos de datos y sacar conclusiones sobre si son iguales o no
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Tenemos k grupos de datos con medias reales (desconocidas) $\mu_{x_1}, \mu_{x_2} \ldots, \mu_{x_k}$, medias muestrales $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k$, dispersiones muestrales $S_1, S_2, \ldots, S_k$, y cantidad de muestras $n_1, n_2, \ldots, n_k$

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos
- Ahora nos interesa comparar las medias de más de 2 conjuntos de datos y sacar conclusiones sobre si son iguales o no
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Tenemos k grupos de datos con medias reales (desconocidas) $\mu_{x_1}, \mu_{x_2} \ldots, \mu_{x_k}$, medias muestrales $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k$, dispersiones muestrales $S_1, S_2, \ldots, S_k$, y cantidad de muestras $n_1, n_2, \ldots, n_k$
- Test de hipotesis sobre las medias *reales*

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos
- Ahora nos interesa comparar las medias de más de 2 conjuntos de datos y sacar conclusiones sobre si son iguales o no
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Tenemos k grupos de datos con medias reales (desconocidas) $\mu_{x_1}, \mu_{x_2} \ldots, \mu_{x_k}$, medias muestrales $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k$, dispersiones muestrales $S_1, S_2, \ldots, S_k$, y cantidad de muestras $n_1, n_2, \ldots, n_k$
- Test de hipotesis sobre las medias *reales*
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2} = \ldots = \mu_{x_k}$

Análisis de Varianza (ANOVA) Unidireccional

- Con el test de t de Student de 2 muestras comparamos las medias de 2 conjuntos de datos
- Ahora nos interesa comparar las medias de más de 2 conjuntos de datos y sacar conclusiones sobre si son iguales o no
- En definitiva, queremos saber si las diferencias en medias muestrales son reales o solo debido al muestreo aleatorio
- Tenemos k grupos de datos con medias reales (desconocidas) $\mu_{x_1}, \mu_{x_2} \ldots, \mu_{x_k}$, medias muestrales $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k$, dispersiones muestrales $S_1, S_2, \ldots, S_k$, y cantidad de muestras $n_1, n_2, \ldots, n_k$
- Test de hipotesis sobre las medias *reales*
- $\mathcal{H}_0 : \mu_{x_1} = \mu_{x_2} = \ldots = \mu_{x_k}$
- $\mathcal{H}_1$: las medias no son todas iguales

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$
- Luego calculamos media cuadrática *entre* grupos, una especie de varianza entre grupos:

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$
- Luego calculamos media cuadrática *entre* grupos, una especie de varianza entre grupos:
- $S_e^2 = \frac{\sum_i n_i (\hat{x}_i - \hat{x})^2}{df_e}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$
- Luego calculamos media cuadrática *entre* grupos, una especie de varianza entre grupos:
- $S_e^2 = \frac{\sum_i n_i (\hat{x}_i - \hat{x})^2}{df_e}$
- Grados de libertad entre grupos: $df_e = k - 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$
- Luego calculamos media cuadrática *entre* grupos, una especie de varianza entre grupos:
- $S_e^2 = \frac{\sum_i n_i (\hat{x}_i - \hat{x})^2}{df_e}$
- Grados de libertad entre grupos: $df_e = k - 1$
- Calculamos la media cuadrática *dentro* de los grupos, como:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$
- Luego calculamos media cuadrática *entre* grupos, una especie de varianza entre grupos:
- $S_e^2 = \frac{\sum_i n_i (\hat{x}_i - \hat{x})^2}{df_e}$
- Grados de libertad entre grupos: $df_e = k - 1$
- Calculamos la media cuadrática *dentro* de los grupos, como:
- $S_d^2 = \frac{\sum_i (n_i - 1) S_i^2}{df_d}$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

ANOVA

- Para comparar todas las medias de los k grupos al mismo tiempo, hacemos lo siguiente:
- Primero la media total (*grand mean*):
- $\hat{x} = \frac{\sum_i n_i \hat{x}_i}{\sum_i n_i}$
- Luego calculamos media cuadrática *entre* grupos, una especie de varianza entre grupos:
- $S_e^2 = \frac{\sum_i n_i (\hat{x}_i - \hat{x})^2}{df_e}$
- Grados de libertad entre grupos: $df_e = k - 1$
- Calculamos la media cuadrática *dentro* de los grupos, como:
- $S_d^2 = \frac{\sum_i (n_i - 1) S_i^2}{df_d}$
- Grados de libertad dentro de grupos: $df_d = (\sum_i n_i) - k$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

ANOVA

- Ahora comparamos los dos números encontrados antes, dividiéndolos, y generamos la estadística del test

FACULTAD DE INGENIERIA Universidad de Buenos Aires

ANOVA

- Ahora comparamos los dos números encontrados antes, dividiéndolos, y generamos la estadística del test
- $F = \frac{S_e^2}{S_d^2} \sim$ F-Snedecor

ANOVA

- Ahora comparamos los dos números encontrados antes, dividiéndolos, y generamos la estadística del test
- $F = \frac{S_e^2}{S_d^2} \sim$ F-Snedecor
- Para decidir acerca de $\mathcal{H}_0$, encontramos el percentil (valor crítico) de una distribución F con grados de libertad df_e, df_b y valor de significancia α, $F_{critico}$

ANOVA

- Ahora comparamos los dos números encontrados antes, dividiéndolos, y generamos la estadística del test
- $F = \frac{S_e^2}{S_d^2} \sim$ F-Snedecor
- Para decidir acerca de $\mathcal{H}_0$, encontramos el percentil (valor crítico) de una distribución F con grados de libertad df_e, df_b y valor de significancia α, $F_{critico}$
- Si $F > F_{critico}$, rechazamos $\mathcal{H}_0$, caso contrario la aceptamos

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Test de ANOVA - Ejemplo

- Un grupo de amigos discute en un bar si Messi, Riquelme y Maradona rindieron igual de bien en la selección argentina de fútbol. Proponen usar como criterio la cantidad de goles por partido para describir un comportamiento más general del juego de cada jugador en la selección nacional. Usar un test de ANOVA con significancia de 5% para responder la duda planteada por el grupo de amigos.

	Maradona	Messi	Riquelme
No. Partidos en Selección	91	142	51
Goles Promedio en Selección	0.37	0.5	0.33
Desvío estándar Goles en Selección	4.4	5.9	3.4

Datos verdaderos al 13 Mayo 2021.

Definición de Entropía

- Supongamos que una moneda tiene dos lados *ceca*

Definición de Entropía

- Supongamos que una moneda tiene dos lados *ceca*
- Claramente la aleatoriedad desaparece ya que $P(ceca) = 1$ y $P(cara) = 0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Definición de Entropía

- Supongamos que una moneda tiene dos lados *ceca*
- Claramente la aleatoriedad desaparece ya que $P(ceca) = 1$ y $P(cara) = 0$
- Si alguien nos *anticipa* que va a salir ceca, nos esta aportando alguna información util?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Definición de Entropía

- Supongamos que una moneda tiene dos lados *ceca*
- Claramente la aleatoriedad desaparece ya que $P(ceca) = 1$ y $P(cara) = 0$
- Si alguien nos *anticipa* que va a salir ceca, nos esta aportando alguna información util?
- En este caso decimos que la información que nos otorga esta *fuente de información* es cero

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*
- Cuando usamos $\log_e(\cdot) = \ln(\cdot)$ la información se mide en *nats*

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*
- Cuando usamos $\log_e(\cdot) = \ln(\cdot)$ la información se mide en *nats*
- Definamos la información promedio, o *entropía* $H(X)$, de una v.a. discreta X como:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*
- Cuando usamos $\log_e(\cdot) = \ln(\cdot)$ la información se mide en *nats*
- Definamos la información promedio, o *entropía* $H(X)$, de una v.a. discreta X como:
- $H(X) = E[\log_2(1/p(X))]$

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*
- Cuando usamos $\log_e(\cdot) = \ln(\cdot)$ la información se mide en *nats*
- Definamos la información promedio, o *entropía* $H(X)$, de una v.a. discreta X como:
- $H(X) = E[\log_2(1/p(X))]$
- Expandiendo:

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*
- Cuando usamos $\log_e(\cdot) = \ln(\cdot)$ la información se mide en *nats*
- Definamos la información promedio, o *entropía* $H(X)$, de una v.a. discreta X como:
- $H(X) = E[\log_2(1/p(X))]$
- Expandiendo:
- $H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$

Entropía como Información Promedio

- Definamos a $I = \log_2(1/p) = -\log_2(p)$ como una medida de *información* de cada símbolo emitido por la fuente
- Cuando usamos $\log_2(\cdot)$ la informacion se mide en *bits*
- Cuando usamos $\log_e(\cdot) = \ln(\cdot)$ la información se mide en *nats*
- Definamos la información promedio, o *entropía* $H(X)$, de una v.a. discreta X como:
- $H(X) = E[\log_2(1/p(X))]$
- Expandiendo:
- $H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$
- Como en este caso la esperanza se toma respecto de p, se suele simbolizar $H(X) = E_p[\log_2(1/p(X))]$

Entropía como Información Promedio

- La entropía mide la información promedio, medida en bits o nats, emitida por la fuente, considerando todos los símbolos

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía como Información Promedio

- La entropía mide la información promedio, medida en bits o nats, emitida por la fuente, considerando todos los símbolos
- La entropía es no-negativa, $H(X) \geq 0$, por lo siguiente:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía como Información Promedio

- La entropía mide la información promedio, medida en bits o nats, emitida por la fuente, considerando todos los símbolos
- La entropía es no-negativa, $H(X) \geq 0$, por lo siguiente:
- $H(X) \triangleq E[\log_2(1/p(X))]$, pero $\log_2(1/p(X)) \geq 0$ ya que $0 \leq p(X) \leq 1$, entonces $1 \leq 1/p(X) < +\infty$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía como Información Promedio

- La entropía mide la información promedio, medida en bits o nats, emitida por la fuente, considerando todos los símbolos
- La entropía es no-negativa, $H(X) \geq 0$, por lo siguiente:
- $H(X) \triangleq E[\log_2(1/p(X))]$, pero $\log_2(1/p(X)) \geq 0$ ya que $0 \leq p(X) \leq 1$, entonces $1 \leq 1/p(X) < +\infty$
- Cambio de base logarítmica: $H_b(X) = (\log_b a) H_a(X)$

Entropía como Información Promedio

- La entropía mide la información promedio, medida en bits o nats, emitida por la fuente, considerando todos los símbolos
- La entropía es no-negativa, $H(X) \geq 0$, por lo siguiente:
- $H(X) \triangleq E[\log_2(1/p(X))]$, pero $\log_2(1/p(X)) \geq 0$ ya que $0 \leq p(X) \leq 1$, entonces $1 \leq 1/p(X) < +\infty$
- Cambio de base logarítmica: $H_b(X) = (\log_b a) H_a(X)$
- Cuanto menor es la p. de un evento, mayor es la entropía (resultado más inesperado, i.e. mayor información o *novedad*)

Ejercicio

- Calcular entropía de una v.a. Bernoulli con probabilidad p para los casos:

Ejercicio

- Calcular entropía de una v.a. Bernoulli con probabilidad p para los casos:
- i) $p_1 = 1, p_2 = 0,$

Ejercicio

- Calcular entropía de una v.a. Bernoulli con probabilidad p para los casos:
- i) $p_1 = 1, p_2 = 0,$
- ii) $p_1 = p_2 = 0.5$

Ejercicio

- Calcular entropía de una v.a. Bernoulli con probabilidad p para los casos:
- i) $p_1 = 1, p_2 = 0,$
- ii) $p_1 = p_2 = 0.5$
- Graficar $E(p)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio

- Calcular entropía de un dado de n caras balanceado, donde X es la v.a. para la salida de cada cara del dado:

Ejercicio

- Calcular entropía de un dado de n caras balanceado, donde X es la v.a. para la salida de cada cara del dado:
- $H(X) = -\sum_{i=1}^{n}(1/n)\log_2(1/n) = \log_2 n$

Entropía Conjunta y Condicional

- Definimos la entropía conjunta a partir de la distribución conjunta de dos v.a. X, Y:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Conjunta y Condicional

- Definimos la entropía conjunta a partir de la distribución conjunta de dos v.a. X, Y:
- $H(X, Y) = -\sum_{x,y \in X,Y} p_{X,Y}(x, y) \log p_{X,Y}(x, y)$

- Definimos la entropía conjunta a partir de la distribución conjunta de dos v.a. X, Y:
- $H(X,Y) = -\sum_{x,y \in X,Y} p_{X,Y}(x,y) \log p_{X,Y}(x,y)$
- La entropía condicional es:

$$\begin{aligned} H(Y|X) &= \sum_{x \in X} p(x) H(Y|X=x) \\ &= -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log p(y|x) \\ &= -\sum_{x,y \in X,Y} p(x,y) \log p(y|x) \end{aligned} \tag{1}$$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Entropía Conjunta y Condicional

- Definimos la entropía conjunta a partir de la distribución conjunta de dos v.a. X, Y:
- $H(X,Y) = -\sum_{x,y \in X,Y} p_{X,Y}(x,y) \log p_{X,Y}(x,y)$
- La entropía condicional es:

$$
\begin{aligned}
H(Y|X) &= \sum_{x \in X} p(x) H(Y|X=x) \\
&= -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log p(y|x) \\
&= -\sum_{x,y \in X,Y} p(x,y) \log p(y|x) \qquad (1)
\end{aligned}
$$

- En el último paso hemos usado Bayes

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Divergencia de Kullback Leibler

- Hasta ahora hemos asumido que conocemos exactamente la función de densidad de masa (pmf) de X, es decir $P(X = x_i) = p_i$

- Hasta ahora hemos asumido que conocemos exactamente la función de densidad de masa (pmf) de X, es decir $P(X = x_i) = p_i$
- Dado que este no es siempre el caso (solemos *estimar* p_i)...

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Divergencia de Kullback Leibler

- Hasta ahora hemos asumido que conocemos exactamente la función de densidad de masa (pmf) de X, es decir $P(X = x_i) = p_i$
- Dado que este no es siempre el caso (solemos *estimar* p_i)...
- llamemos p_i a la p. verdadera, mientras que q_i es la p. estimada

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Divergencia de Kullback Leibler

- Hasta ahora hemos asumido que conocemos exactamente la función de densidad de masa (pmf) de X, es decir $P(X = x_i) = p_i$
- Dado que este no es siempre el caso (solemos *estimar* p_i)...
- llamemos p_i a la p. verdadera, mientras que q_i es la p. estimada
- En este caso, definimos la divergencia Kullback Leibler (KL) como:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Divergencia de Kullback Leibler

- Hasta ahora hemos asumido que conocemos exactamente la función de densidad de masa (pmf) de X, es decir $P(X = x_i) = p_i$
- Dado que este no es siempre el caso (solemos *estimar* p_i)...
- llamemos p_i a la p. verdadera, mientras que q_i es la p. estimada
- En este caso, definimos la divergencia Kullback Leibler (KL) como:
- $D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Divergencia de Kullback Leibler

- Hasta ahora hemos asumido que conocemos exactamente la función de densidad de masa (pmf) de X, es decir $P(X = x_i) = p_i$
- Dado que este no es siempre el caso (solemos *estimar* p_i)...
- llamemos p_i a la p. verdadera, mientras que q_i es la p. estimada
- En este caso, definimos la divergencia Kullback Leibler (KL) como:
- $D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$
- $D(p||q)$ mide la ineficiencia de usar $q(x)$ cuando la verdadera distribución es $p(x)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Divergencia KL: Propiedades

- $D(p||q) \geq 0$

Divergencia KL: Propiedades

- $D(p||q) \geq 0$
- Si $p = q$, $D(p||q) = 0$ porque $\log \frac{p(x)}{q(x)} = \log 1 = 0$

Divergencia KL: Propiedades

- $D(p||q) \geq 0$
- Si $p = q$, $D(p||q) = 0$ porque $\log \frac{p(x)}{q(x)} = \log 1 = 0$
- La divergencia KL no es simétrica, es decir $D(p||q) \neq D(q||p)$ a menos que $p = q$ (en cuyo caso $D(p||q) = D(q||p) = 0$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Divergencia KL: Propiedades

- $D(p||q) \geq 0$
- Si $p = q$, $D(p||q) = 0$ porque $\log \frac{p(x)}{q(x)} = \log 1 = 0$
- La divergencia KL no es simétrica, es decir $D(p||q) \neq D(q||p)$ a menos que $p = q$ (en cuyo caso $D(p||q) = D(q||p) = 0$
- $D(p||q)$ no es una distancia porque no obedece la desigualdad de triángulo $\|x + y\| \leq \|x\| + \|y\|$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p,q) = -\sum_x p(x)\log(q(x))$

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p, q) = -\sum_x p(x) \log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p, q) = -\sum_x p(x) \log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x \in X} p(x) \log(p(x))$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p, q) = -\sum_x p(x) \log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x \in X} p(x) \log(p(x))$
- Reescribimos la divergencia KL, $D_{p||q}(X) = \sum_{x \in X}(p(x) \log p(x) - p(x) \log q(x))$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p,q) = -\sum_x p(x)\log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x\in X} p(x)\log(p(x))$
- Reescribimos la divergencia KL, $D_{p||q}(X) = \sum_{x\in X}(p(x)\log p(x) - p(x)\log q(x))$
- Vemos que $D_{p||q}(X) = -H_p(X) + H_{p,q}(X)$

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p,q) = -\sum_x p(x)\log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x \in X} p(x)\log(p(x))$
- Reescribimos la divergencia KL, $D_{p||q}(X) = \sum_{x \in X}(p(x)\log p(x) - p(x)\log q(x))$
- Vemos que $D_{p||q}(X) = -H_p(X) + H_{p,q}(X)$
- Entonces $H_{p,q}(X) = H_p(X) + D_{p||q}(X)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p, q) = -\sum_x p(x) \log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x \in X} p(x) \log(p(x))$
- Reescribimos la divergencia KL, $D_{p||q}(X) = \sum_{x \in X}(p(x) \log p(x) - p(x) \log q(x))$
- Vemos que $D_{p||q}(X) = -H_p(X) + H_{p,q}(X)$
- Entonces $H_{p,q}(X) = H_p(X) + D_{p||q}(X)$
- Vemos que como $D_{p||q}(X) \geq 0$, entonces $H_{p,q}(X) \geq H_p(X)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p, q) = -\sum_x p(x) \log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x \in X} p(x) \log(p(x))$
- Reescribimos la divergencia KL, $D_{p||q}(X) = \sum_{x \in X} (p(x) \log p(x) - p(x) \log q(x))$
- Vemos que $D_{p||q}(X) = -H_p(X) + H_{p,q}(X)$
- Entonces $H_{p,q}(X) = H_p(X) + D_{p||q}(X)$
- Vemos que como $D_{p||q}(X) \geq 0$, entonces $H_{p,q}(X) \geq H_p(X)$
- Una conclusión importante es que minimizar $H_{p,q}(X)$ es equivalente a minimizar $D_{p||q}(X)$ ...

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Entropía Cruzada

- Definimos la entropía cruzada como:
- $H(p,q) = -\sum_x p(x)\log(q(x))$
- Esta es la información promedio que obtenemos cuando usamos $q(x)$ como pmf para reemplazar la verdadera $p(x)$
- Reescribimos la entropía, $H_p(X) = -\sum_{x \in X} p(x)\log(p(x))$
- Reescribimos la divergencia KL, $D_{p||q}(X) = \sum_{x \in X}(p(x)\log p(x) - p(x)\log q(x))$
- Vemos que $D_{p||q}(X) = -H_p(X) + H_{p,q}(X)$
- Entonces $H_{p,q}(X) = H_p(X) + D_{p||q}(X)$
- Vemos que como $D_{p||q}(X) \geq 0$, entonces $H_{p,q}(X) \geq H_p(X)$
- Una conclusión importante es que minimizar $H_{p,q}(X)$ es equivalente a minimizar $D_{p||q}(X)$ ...
- porque dada la pmf de X, $H_p(X)$ es constante

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua

- Definimos información mutua entre X, Y como la información que X tiene acerca de Y, y viceversa

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua

- Definimos información mutua entre X, Y como la información que X tiene acerca de Y, y viceversa
- La expresión está dada por:

Información Mutua

- Definimos información mutua entre X, Y como la información que X tiene acerca de Y, y viceversa
- La expresión está dada por:
- $I(X;Y) = D(p(x,y) \| p(x)p(y)) = \sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y) \log(p(x|y))$

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y) \log(p(x|y))$
- La IM es no-negativa: $I(X;Y) \geq 0$ (es una divergencia, $D(p(x,y) \| p(x)p(y))$)

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y) \log(p(x|y))$
- La IM es no-negativa: $I(X;Y) \geq 0$ (es una divergencia, $D(p(x,y) \| p(x)p(y))$)
- Por este motivo, $H(X) - H(X|Y) \geq 0$ entonces $H(X|Y) \leq H(X)$, i.e. condicionar reduce la entropía (por qué?)

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y) \log(p(x|y))$
- La IM es no-negativa: $I(X;Y) \geq 0$ (es una divergencia, $D(p(x,y) \| p(x)p(y))$)
- Por este motivo, $H(X) - H(X|Y) \geq 0$ entonces $H(X|Y) \leq H(X)$, i.e. condicionar reduce la entropía (por qué?)
- La IM de una v.a. con ella misma es la entropía (i.e. información propia)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y)\log(p(x|y))$
- La IM es no-negativa: $I(X;Y) \geq 0$ (es una divergencia, $D(p(x,y)\|p(x)p(y))$)
- Por este motivo, $H(X) - H(X|Y) \geq 0$ entonces $H(X|Y) \leq H(X)$, i.e. condicionar reduce la entropía (por qué?)
- La IM de una v.a. con ella misma es la entropía (i.e. información propia)
- $I(X;X) = H(X) + H(X|X) = H(X)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y)\log(p(x|y))$
- La IM es no-negativa: $I(X;Y) \geq 0$ (es una divergencia, $D(p(x,y)\|p(x)p(y))$)
- Por este motivo, $H(X) - H(X|Y) \geq 0$ entonces $H(X|Y) \leq H(X)$, i.e. condicionar reduce la entropía (por qué?)
- La IM de una v.a. con ella misma es la entropía (i.e. información propia)
- $I(X;X) = H(X) + H(X|X) = H(X)$
- Demonstrar. Pista: usar $p(x,x) = p(x)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Información Mutua: Propiedades

- La información mutua (IM) es simétrica:
- $I(X;Y) = H(X) - H(X|Y) = I(Y;X) = H(Y) - H(Y|X)$
- con:
- $H(X|Y) = -\sum_{x,y} p(x,y)\log(p(x|y))$
- La IM es no-negativa: $I(X;Y) \geq 0$ (es una divergencia, $D(p(x,y)\|p(x)p(y))$)
- Por este motivo, $H(X) - H(X|Y) \geq 0$ entonces $H(X|Y) \leq H(X)$, i.e. condicionar reduce la entropía (por qué?)
- La IM de una v.a. con ella misma es la entropía (i.e. información propia)
- $I(X;X) = H(X) + H(X|X) = H(X)$
- Demonstrar. Pista: usar $p(x,x) = p(x)$
- La IM y la entropía cruzada se relacionan mediante:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Resumen

- La entropía $H(X)$ mide la información que una v.a. ofrece

Resumen

- La entropía $H(X)$ mide la información que una v.a. ofrece
- La entropía cruzada $H_{p,q}(X)$ mide la información que ofrece usar una pmf aproximada ($q(x)$) en lugar de la pmf verdadera ($p(x)$)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Resumen

- La entropía $H(X)$ mide la información que una v.a. ofrece
- La entropía cruzada $H_{p,q}(X)$ mide la información que ofrece usar una pmf aproximada $(q(x))$ en lugar de la pmf verdadera $(p(x))$
- La divergencia KL mide la distancia entre entropía y la entropía cruzada

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Resumen

- La entropía $H(X)$ mide la información que una v.a. ofrece
- La entropía cruzada $H_{p,q}(X)$ mide la información que ofrece usar una pmf aproximada $(q(x))$ en lugar de la pmf verdadera $(p(x))$
- La divergencia KL mide la distancia entre entropía y la entropía cruzada
- entropía cruzada = divergencia KL + entropía

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Resumen

- La entropía $H(X)$ mide la información que una v.a. ofrece
- La entropía cruzada $H_{p,q}(X)$ mide la información que ofrece usar una pmf aproximada ($q(x)$) en lugar de la pmf verdadera ($p(x)$)
- La divergencia KL mide la distancia entre entropía y la entropía cruzada
- entropía cruzada $=$ divergencia KL $+$ entropía
- Dado que a mayor entropía mayor *aleatoriedad* en los datos, como veremos, una aplicación de *minimizar* la entropía es detectar *features* con alta correlación

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 1

Verificar si las acciones de Apple y Microsoft en los últimos 6 meses se comportaron distinto en cuanto a su precio medio de cierre (como de costumbre, $\alpha = 0.05$).

Ejercicio 2

Simular una v.a. Bernoulli X con $p = 1$ que emite *dos* símbolos, x_1 con probabilidad p, y x_2 con probabilidad $1 - p$.

a. Encontrar la información de cada símbolo.
b. ¿Cómo se interpreta que la información del símbolo x_2 tiende a infinito cuando su p. de ocurrencia es nula?
c. Encontrar la entropía de X.

Bibliografía I

E. Alpaydin, *Introduction to machine learning*.
MIT press, 2020.

R. L. Ott and M. T. Longnecker, *An introduction to statistical methods and data analysis*.
Nelson Education, 2015.

"Indian Institute of Technology: Notes on Entropy, Relative Entropy and Cross Entropy." https://www.iitg.ac.in/cseweb/osint/slides/Anasua_Entropy.pdf.
Accessed: 2021-02-07.

A. F. Siegel, *Statistics and data analysis: an introduction*.
Wiley, 1988.