Tests

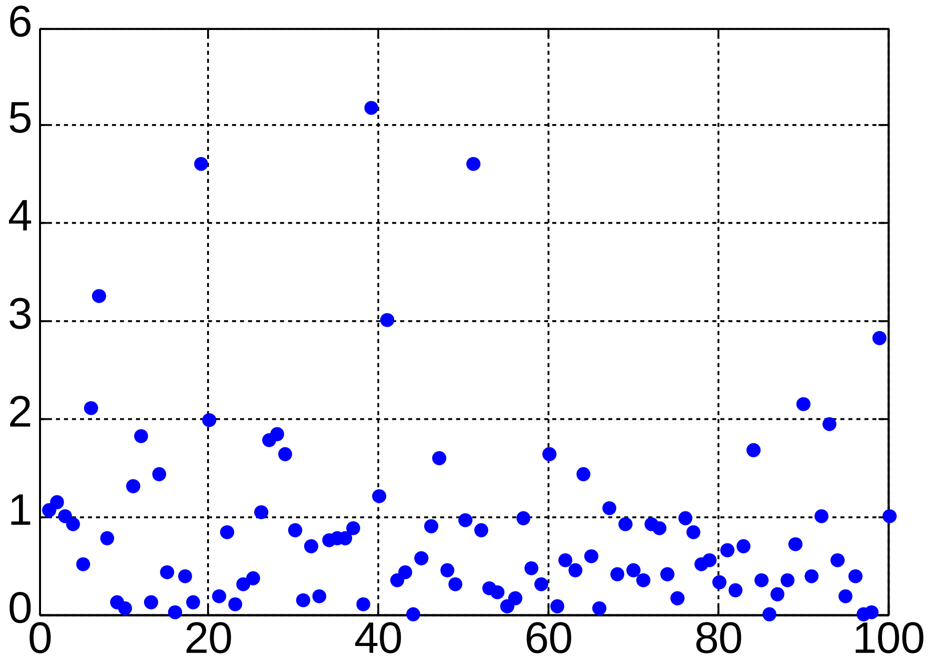
Modelado y Análisis de Redes de Telecomunicaciones

FACULTAD DE INGENIERIA

UNIVERSIDAD DE LA REPUBLICA URUGUAY

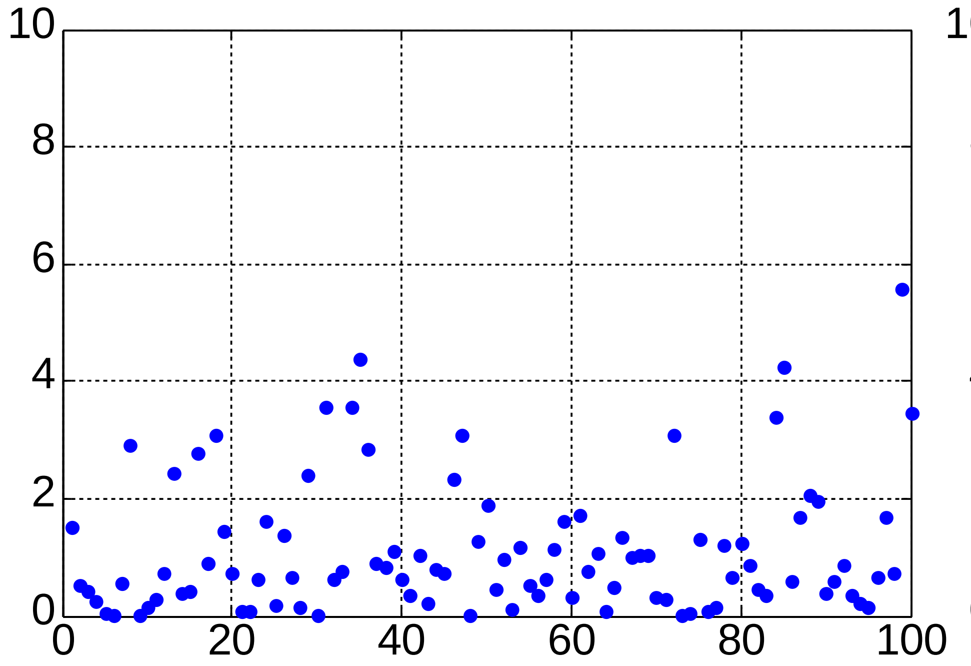
Tests

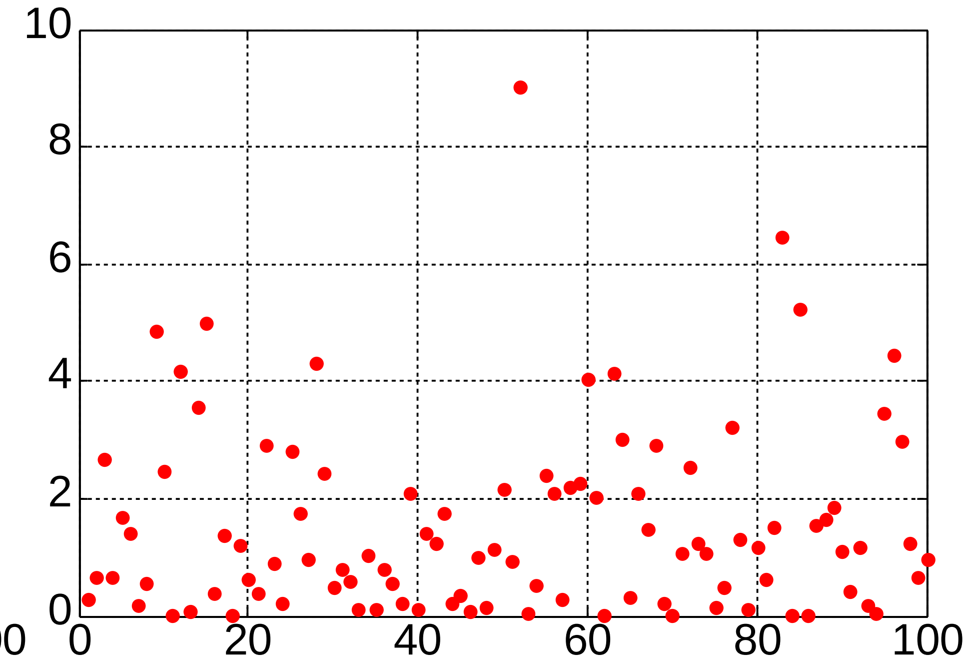
- Los tests son una herramienta para contestar preguntas del tipo:
  - ¿Mis datos son la realización de una V.A. normal?
  - ¿Mis datos son la realización de una V.A. con media nula?



Tests

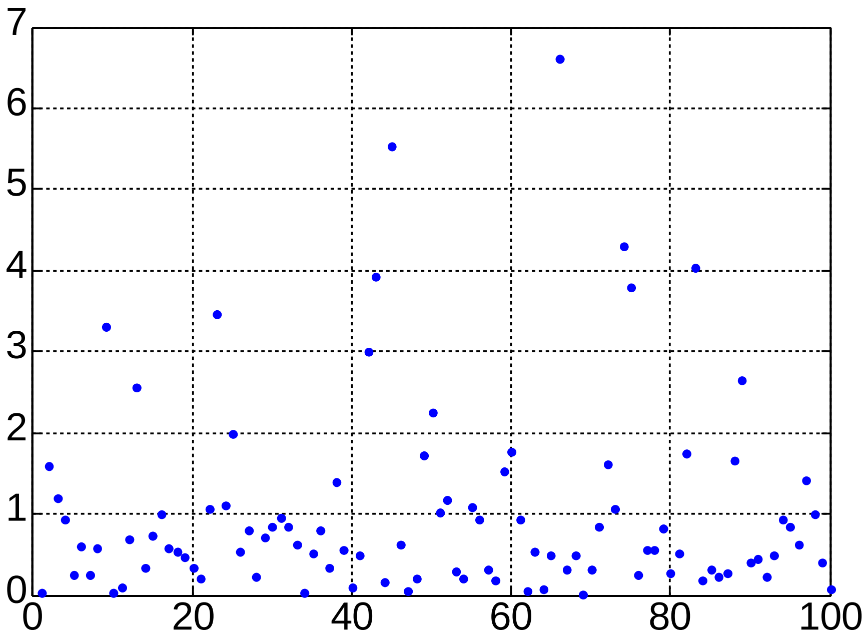
- Los tests son una herramienta para contestar preguntas del tipo:
  - ¿Estos sets de datos tiene la misma distribución?
  - ¿La misma media?

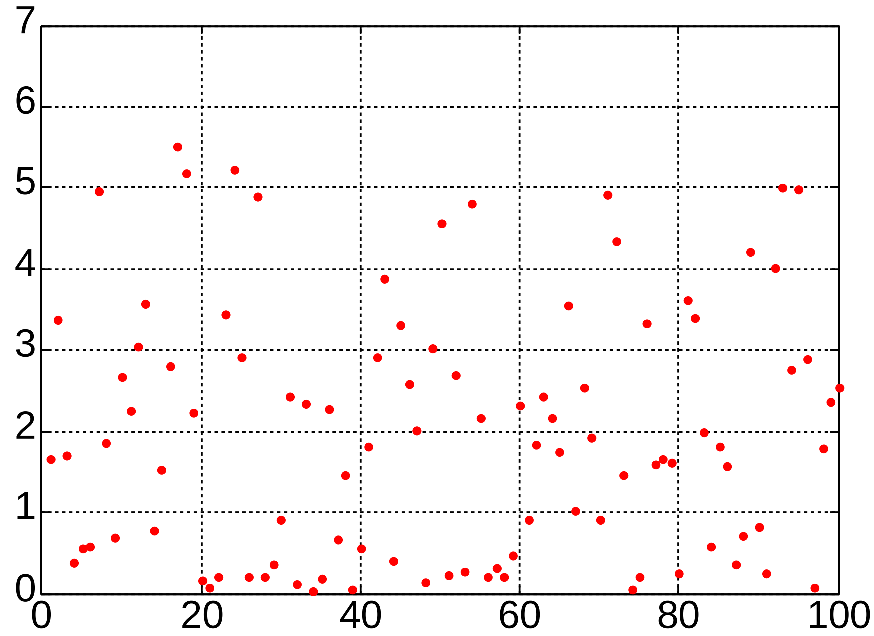




Tests

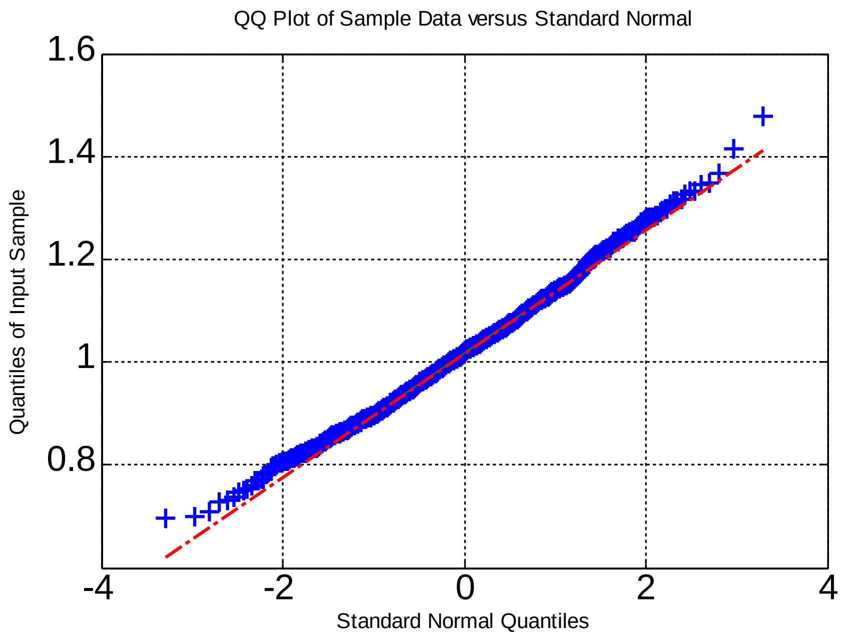
- Con lo que vimos hasta ahora hay ciertos casos que ya sabemos contestar...
- **Ejemplo**: comparación de dos procesadores
  - ¿cuál es más rápido?

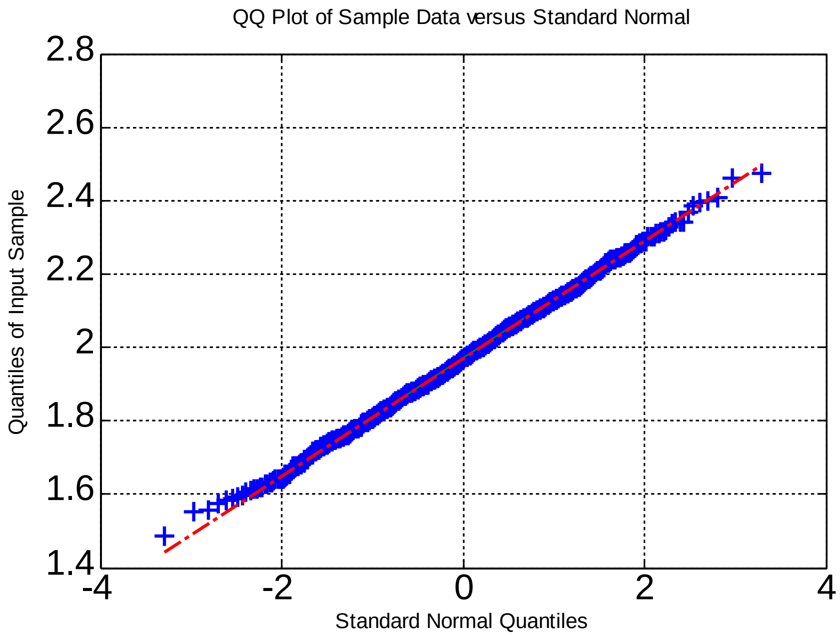




Tests

- QQPlot de la la distribución del promedio (con bootstrap)





- Podemos estimar un intervalo de confianza para la esperanza:
  - I. de C. =[0.78958, 1.2586]
  - I. de C.= [1.6665, 2.2724]
- Con p=0.95 la media del azul es menor que la del rojo (¿faltó testear algo?)

Tests

- Conclusión:
  - Hay ciertos casos que no necesito entrar en la teoría de los tests, sino que basta con calcular intervalos de confianza
- ¿Qué hubiera pasado si los intervalos se solapaban?
- Hay muchas preguntas que no se pueden formular como un intervalo de confianza

Agenda

- Neyman-Pearson
- Likelihood Ratio Tests
- ANOVA
- Resultados asintóticos
- Otros tests...

Contexto

- ¿Qué nos interesa decidir?
  - Tenemos un montón de números $x_1,\ldots,x_n$ y sospechamos que responden a un cierto modelo
  - Nos interesaría saber si el modelo es razonablemente correcto
  - Principio de parsimonia
    - Ante la duda preferir el modelo más sencillo
- Un marco que se adapta a este contexto es el de Neyman-Pearson

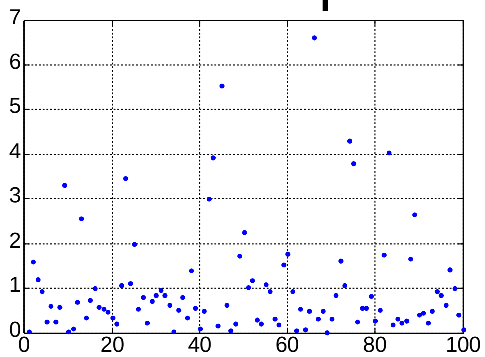
Neyman-Pearson

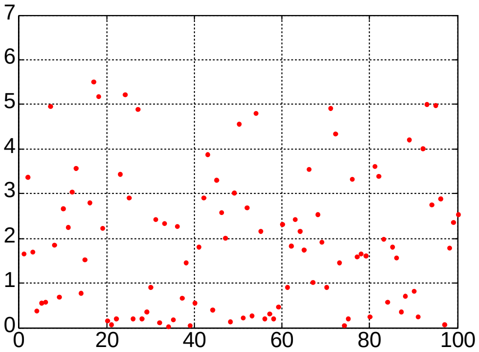
- La hipótesis nula y la alternativa
  - Tenemos la muestra $x_1,\ldots,x_n$ generadas por un modelo desconocido y dos hipótesis
    - H_0: la hipótesis nula
    - H_1: la hipótesis alternativa
  - A partir de la muestra queremos determinar cuál es la correcta
  - En el contexto de Neyman-Pearson ambas hipótesis NO son equivalentes:
    - Preferiría quedarme con H_0 (i.e. rechazo H_0 (acepto H_1) sólo si la evidencia es suficientemente grande)
- En la mayoría de los tests que veremos se asumirá que el modelo desconocido se parametriza por $\theta \in \Theta$ y quiero verificar si θ está en un conjunto dado:

$$\begin{cases} H_0 : \theta \in \Theta_0 \in \Theta \\ H_1 : \theta \in \Theta \setminus \Theta_0 \end{cases}$$

Neyman-Pearson

- **Ejemplo**: comparación de dos procesadores
  - ¿cuál es más rápido?





- En este caso el modelo puede ser el siguiente:
  - Los dos sets de datos fueron generados cada uno por *n* V.A. iid (ambas con distribución exponencial) y quiero verificar si E{*X*}=E{*Y*}

$$\begin{cases} H_0 : E\{X\} = E\{Y\} & \left(\Theta_0 = \{(F(x), F(x)) \text{ con } F(x) = 1 - e^{-\lambda x}; \lambda > 0\right) \\ H_1 : E\{X\} \neq E\{Y\} & \left(\Theta = \{\left(F(x), F\left(\frac{x}{k}\right)\right) \text{ con } F(x) = 1 - e^{-\lambda x}; \lambda > 0, k > 0\}\right) \end{cases}$$

Neyman-Pearson

- Una vez que tengo mi set de datos $x_1,\ldots,x_n$ la decisión del test es determinística
  - Hay familias de tests aleatorias que no consideraremos
- Por lo tanto, se puede definir un conjunto $C \in R^n$ de valores del vector $(x_1,\ldots,x_n)$ tales que:

$$\text{Acepto } H_0 \Leftrightarrow (x_1, \ldots, x_n) \notin C$$

=> *C* define el test y se denomina la **región crítica** del mismo

- De aquí en adelante asumiremos que los datos son la realización de *n* V.A. $X_1,\ldots,X_n$ (no necesariamente iid)
- Posibles errores
  - Tipo 1 (o falso positivo): rechazar H_0 cuando era correcta (su probabilidad la notaremos $P(\text{decidir } H_1 | H_0) = P((X_1,\ldots,X_n) \in C | \theta \in \Theta_0)$)
  - Tipo 2 (o falso negativo): aceptar H_0 cuando era incorrecta ($P(\text{decidir } H_0 | H_1) = P((X_1,\ldots,X_n) \in C^c \mid \theta \in \Theta \backslash \Theta_0)$)

Neyman-Pearson

- Dos parámetros importantes:
  - Tamaño del test:

$$\text{tamaño} = \alpha = \sup_{\theta \in \Theta_0} P\left(\text{decidir } H_1 | H_0\right)$$

    En claro: la máxima probabilidad de falso positivo
  - Potencia del test:

$$\text{potencia} = 1 - \beta = P\left(\text{decidir } H_1 | H_1\right) = 1 - P\left(\text{decidir } H_0 | H_1\right)$$

    En claro: 1 menos la probabilidad de falso negativo en función de $\theta \in \Theta \backslash \Theta_0$
- Aquí se nos plantea el típico problema cuando existen varios indicadores a optimizar
  - ¿Qué tomo como objetivo del test?
  - ¿Qué es un buen test?
  - POSIBLE respuesta: Criterio de Neyman-Pearson

Neyman-Pearson

- Criterio de Neyman-Pearson
  - De aquellos tests con tamaño menor a un cierto umbral, elegir el más potente
- Criterio de Neyman-Pearson en claro:
  - Tengo bajo control la máxima probabilidad de falso positivo y hago lo mejor que puedo con la probabilidad de falso negativo
- Encaja perfectamente en el contexto que estamos trabajando
  - Quiero verificar (prefiero) H_0 => si el test dice H_1, queremos estar relativamente seguros que efectivamente es H_1
  - ¿Si es test dice H_0 que seguridad tenemos?
    - A posteriori habría que calcular la potencia del test

Ejemplo de Test

- Ejemplo ilustrativo
  - Tenemos $x_1,\dots,x_n$ que asumimos generados por n V.A. iid $X_1,\dots,X_n$ normales (con media μ y varianza σ^2 desconocida)
  - Queremos testear: $\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu > 0 \end{cases}$
  - Posible test es: si la media empírica es demasiado grande rechazamos H_0
  - Re-escalando nos queda la siguiente región crítica:

$$C = \left\{ (x_1, \dots, x_n) : \frac{\overline{x}}{s_n/\sqrt{n}} > c \right\}$$

  - El tamaño del test resulta:

$$\alpha = \max_{\sigma>0} P\left(\frac{\overline{X}}{s_n/\sqrt{n}} > c \,\middle|\, \mu = 0, \sigma \right) \underset{n \text{ grande}}{\approx} \max_{\sigma>0} 1 - F_{N_{0,1}}(c) =$$

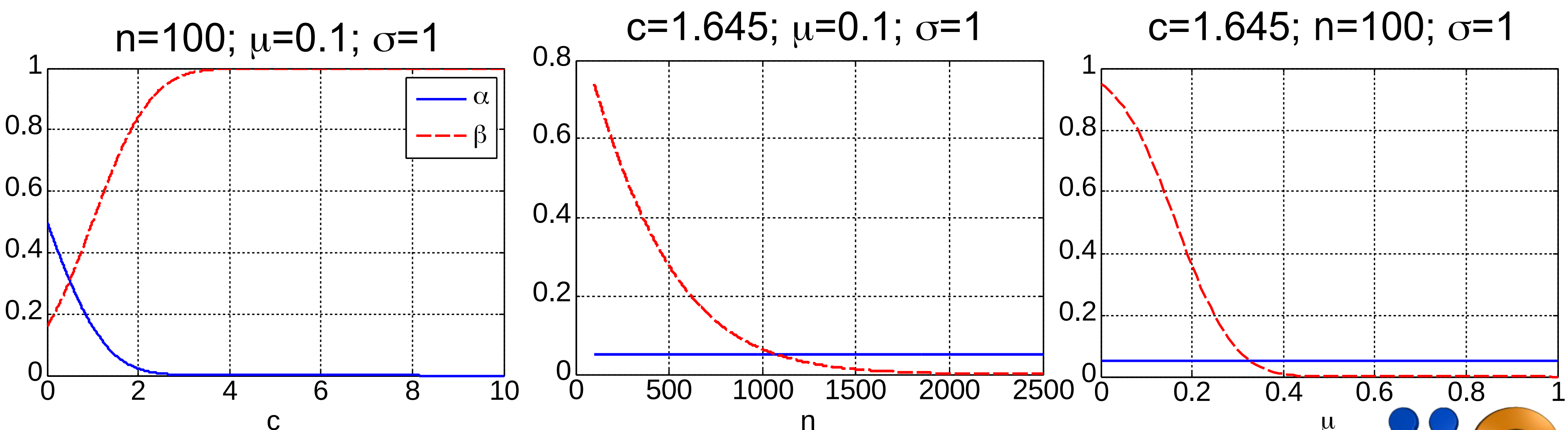
$$= 1 - F_{N_{0,1}}(c)$$

Ejemplo de Test

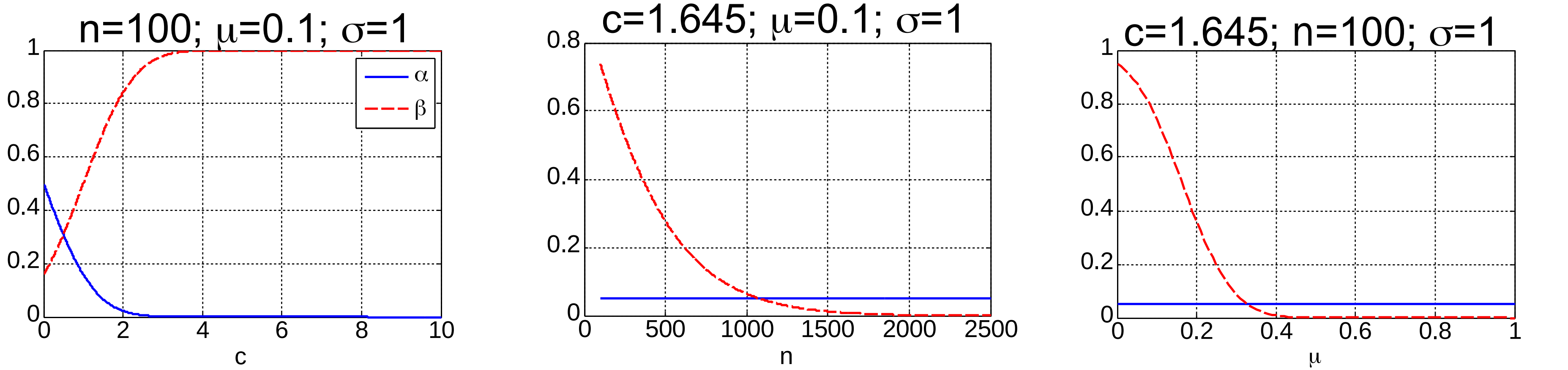
- Ejemplo ilustrativo (cont…)
  - Si quiero un tamaño de test 0.05 => c=1.645
  - Potencia del test:

$$1-\beta = P\left(\frac{\overline{X}}{s_n/\sqrt{n}} > c \,\middle|\, \mu,\sigma>0\right) =$$

$$= P\left(\frac{\overline{X}-\mu}{s_n/\sqrt{n}} > c - \frac{\mu}{s_n/\sqrt{n}} \,\middle|\, \mu,\sigma>0\right) \underset{n \text{ grande}}{\approx} 1 - F_{N_{0,1}}\left(c - \frac{\mu}{\sigma/\sqrt{n}}\right)$$



Ejemplo de Test



- Algunas conclusiones:
  - Dados los datos, si la probabilidad de falso positivo decrece, la probabilidad de falso negativo crece (y viceversa)
  - Cuantos más datos tengo, mejor “resolución” tiene el test
  - Es posible que la probabilidad de falso positivo sea pequeña, y la probabilidad de falso negativo sea grande (e.g. μ pequeño)
    - **REPITO**: Bajo Neyman-Pearson si el test rechaza H_0 estoy bastante seguro que la decisión es la correcta, pero si acepta H_0 no significa que sea verdadera, sino que **no pudo rechazarla** y hay que verificar el valor de β (que por lo general es muy difícil de calcular)

p-valor

- En muchos tests la región crítica se define a partir de una cierta función *T*() de las observaciones $x_1,\ldots,x_n$

$$C = \{(x_1,\ldots,x_n) : T(x_1,\ldots,x_n) > m_0\}$$

  - *T*() se denomina estadístico del test
  - Por ejemplo en nuestro test sencillo

$$T(x_1,\ldots,x_n) = \frac{\overline{x}}{s_n/\sqrt{n}}$$

- Para esos casos se define el p-valor como:

$$p^*(x_1,\ldots,x_n) = \sup_{\theta\in\Theta_0} P\left(T(X_1,\ldots,X_n) > T(x_1,\ldots,x_n) \,\middle|\, \theta\right)$$

  - En claro: la probabilidad de que otro set de datos (bajo H_0) sea “más rechazable” que el que tengo ahora (con el θ que la maximiza)
  - Claramente cuanto menor *p** más razones tengo para rechazar H_0

p-valor

- Existe una relación entre *p** y el tamaño del test α

$$\alpha(m_0) = \sup_{\theta \in \Theta_0} P\left(T(X_1, \ldots, X_n) > m_0 \,\middle|\, \theta\right)$$

$$p^*(x_1, \ldots, x_n) = \sup_{\theta \in \Theta_0} P\left(T(X_1, \ldots, X_n) > T(x_1, \ldots, x_n) \,\middle|\, \theta\right)$$

- $\alpha(m_0)$ es siempre no-creciente en m_0 y por lo general es decreciente
  - En este último caso:

$$p^*(x_1, \ldots, x_n) < \alpha \Leftrightarrow T(x_1, \ldots, x_n) > m_0$$

  => rechazo H_0 si *p es menor que el tamaño del test**

- Ejemplo: En nuestro test sencillo

$$p^*(x_1, \ldots, x_n) = 1 - F_{N_{0,1}}\left(\frac{\overline{x}}{s_n/\sqrt{n}}\right)$$

- Por lo general los tests devuelven *p** (o una estimación) pues nos da más información que un simple aceptar/rechazar

Un comentario....

- Atención con los tests y el resultado que devuelven
  - REPITO UNA VEZ MÁS: Que el test acepte H_0 significa que la evidencia no es suficiente para rechazarla
  - Es importante la distinción entre **significado estadístico** y **significado práctico**
    - Con suficientes datos (*n* suficientemente grande) puedo detectar diferencias en la media de cualquier orden
    - Pero en un test como:

$$\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu > 0 \end{cases}$$

existe un 0 “práctico”

    - E.g. en un enlace larga distancia un retardo menor a 1ms es insignificante (significado práctico) pero con suficientes medidas puedo detectar con gran precisión medias de órdenes menores a 1ms (significado estadístico)

Agenda

- Neyman-Pearson
- Likelihood Ratio Tests
- ANOVA
- Resultados asintóticos
- Otros tests...

Verosimilitud

- Sea X una V.A. con densidad paramétrica en θ $f(x,\theta)$ (el caso discreto es análogo)
- La función de verosimilitud (likelihood) se define como :

$$L(\theta, x_o) = C(x_o) f(x_o, \theta)$$

Con $C(x_0)$ una función positiva acotada que no depende de θ

- Aunque a priori parece más de lo mismo, la idea es que la función $L(\theta,x_0)$ como función de θ con x_0 fijo nos daría una medida de su "correctitud"
  - I.e. Si tengo una muestra de X (x_0) y dos posibles valores de θ (θ_1 y θ_2) tales que $L(\theta_1,x_0)> L(\theta_2,x_0)$ ¿con qué valor de θ me quedo?

Verosimilitud

- De esta idea surge el estimador de máxima verosimilitud
  - A partir de la muestra x_0 de X estimo θ como θ^*:

$$\theta^* = \arg \max_{\theta \in \Theta} L(\theta, x_o)$$

- Ejemplo: $X_1,\ldots,X_n$ iid ~ Exp(λ)

$$L(\theta, x) = \prod_{i=1}^{n} C(x_i)\lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \prod_{i=1}^{n} C(x_i)$$

$$\Rightarrow \frac{\partial L(\theta, x)}{\partial \theta} = \lambda^{n-1} e^{-\lambda \sum_{i=1}^n x_i} \left(n - \lambda \sum_{i=1}^{n} x_i \right) \prod_{i=1}^{n} C(x_i)$$

$$\Rightarrow \lambda^* = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$$

Likelihood Ratio Tests

- La verosimilitud lógicamente también se utiliza en el contexto de tests
- Sean $X=(X_1,\ldots,X_n)$ un vector con n V.A. y sean $f(x,\theta)$ la densidad del vector y x_0 una realización
  - Imaginemos el siguiente test:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

  - Si $L(\theta_1,x_0)/L(\theta_0,x_0)$ es suficientemente grande es natural rechazar H_0
- De esta idea surgen los denominados "likelihood ratio tests" (tests de cociente de verosimilitudes)

Likelihood Ratio Tests

- En nuestro caso más general

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta \setminus \Theta_0 \end{cases}$$

- Se definen las siguientes funciones:

$$l_x(\theta) = \log f_X(x, \theta)$$

$$l_x(H_0) = \sup_{\theta \in \Theta_0} l_x(\theta)$$

$$l_x(H_1) = \sup_{\theta \in \Theta} l_x(\theta)$$

- El likelihood ratio test se define por la siguiente región crítica:

$$C(x) = \{l_x(H_1) - l_x(H_0) > k\}$$

  - Si H_0 es verdad uno esperaría que el supremo de la verosimilitud se dé en Θ_0 ($l_x(H_1)-l_x(H_0)\approx 0$)
  - Si no el supremo de la verosimilitud se da fuera Θ_0 ($l_x(H_1)-l_x(H_0) > 0$)

Likelihood Ratio Tests

- **Lema de Neyman-Pearson**
  - Consideremos el caso sencillo $\Theta=\{\theta_0,\theta_1\}$ y $\Theta_0=\{\theta_0\}$
  - Sea *C* la región crítica definida antes con *k* tal que el tamaño del test resultante es α
  - Sea *S* cualquier región crítica tal que el tamaño del test resultante es también α

$$\Rightarrow P(X \in S \,|\, \theta = \theta_1) \leq P(X \in C \,|\, \theta = \theta_1)$$

  - En claro:
    - De todos los tests de tamaño α el likelihood ratio test es el de mayor potencia (el de menor probabilidad de falso negativo)

Likelihood Ratio Tests

- Volvamos a nuestro ejemplo sencillo …
  - Tenemos $x_1,\dots,x_n$ que asumimos generados por n V.A. iid $X_1,\dots,X_n$ normales (con media μ y varianza σ^2 desconocida)
  - Queremos testear:

$$\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu > 0 \end{cases}$$

  - Construyamos el likelihood ratio test correspondiente:

$$l_x(\theta) = \log\left(f_X(x,\theta)\right) = \log\left(\prod_{i=1}^{n} f_{X_i}(x,\theta)\right) =$$

$$\log\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Likelihood Ratio Tests

- Volvamos a nuestro ejemplo sencillo ...

$$l_x(H_1) = \sup_{\mu>0;\sigma} -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow l_x(H_1) = -\frac{n}{2}\log\left(2\pi\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}^+)^2\right) - \frac{n}{2}$$

$$l_x(H_0) = \sup_{\mu=0;\sigma} -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow l_x(H_0) = -\frac{n}{2}\log\left(2\pi\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - \frac{n}{2}$$

$$\therefore \quad l_x(H_1) - l_x(H_0) = -\frac{n}{2}\log\left(\frac{\sum_{i=1}^{n}(x_i-\overline{x}^+)^2}{\sum_{i=1}^{n}x_i^2}\right)$$

Likelihood Ratio Tests

- Volvamos a nuestro ejemplo sencillo ...
  - La región crítica resulta:

$$C = \left\{ x : -\frac{n}{2} \log \left(\frac{\sum_{i=1}^{n} (x_i - \overline{x}^{+})^2}{\sum_{i=1}^{n} x_i^2} \right) > k \right\} = \left\{ x : \frac{\sum_{i=1}^{n} (x_i - \overline{x}^{+})^2}{\sum_{i=1}^{n} x_i^2} < k' \right\}$$

- Obs: el cociente de la región de exclusión es siempre menor o igual a 1
  => Para que el test tenga sentido k' debe ser menor a 1
  => Si el promedio es negativo, ya sabemos que hay que aceptar H0
  => Asumamos que el promedio es positivo

$$C = \left\{ x : \frac{1/n \sum_{i=1}^{n} (x_i - \overline{x})^2}{\overline{x}^2 + 1/n \sum_{i=1}^{n} (x_i - \overline{x})^2} < k' \right\} = \left\{ x : \frac{\overline{x}^2}{1/n \sum_{i=1}^{n} (x_i - \overline{x})^2} \geq \frac{1 - k'}{k'} \right\} =$$

$$= \left\{ \frac{\overline{x}}{s_n / \sqrt{n}} \geq \sqrt{\frac{n(1-k')}{k'}} \right\}$$

=> El test sencillo que diseñamos era un LRT!

Test de Student

- Tenemos $x_1,\ldots,x_n$ que asumimos generados por n V.A. iid $X_1,\ldots,X_n$ normales (con media μ y varianza σ^2 desconocida)
- Queremos testear:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

- Dejamos como ejercicio demostrar que el likelihood ratio test en este caso es equivalente al test de student cuyo estadístico es:

$$T(x_1,\ldots,x_n) = \frac{\overline{x} - \mu_0}{s_n/\sqrt{n}}$$

Agenda

- Neyman-Pearson
- Likelihood Ratio Tests
- ANOVA
- Resultados asintóticos
- Otros tests…

ANOVA

- Tenemos $x_1,\ldots,x_n$ que asumimos generados por n V.A. iid $X_1,\ldots,X_n$ normales (con medias $\mu_1,\ldots,\mu_n$ y varianza σ^2 desconocida)
- Imaginemos que queremos testear alguna propiedad de las medias (i.e. $\theta=(\mu_1,\ldots,\mu_n,\sigma)$)
  - E.g. si son todas iguales o si hay dos grupos
- En este caso se puede usar los tests denominados ANOVA
- Otras hipótesis necesarias:
  - $\mu_1,\ldots,\mu_n$ responden a un modelo lineal; i.e. $(\mu_1,\ldots,\mu_n)\in M$, donde M es un subespacio lineal de R^n
  - Queremos testear si $(\mu_1,\ldots,\mu_n)\in M_0$, el cual a su vez es un subespacio lineal de M

$$\begin{cases} \Theta = M \times (0,\infty) \\ \Theta_0 = M_0 \times (0,\infty) \end{cases}$$

  - $\Pi_M(x)$ es el proyector ortogonal de x en M (y $\Pi_{M_0}(x)$ en M_0)

ANOVA

- Un ejemplo en concreto:
  - Tenemos $x_1,\ldots,x_k$ e $y_1,\ldots,y_l$ que asumimos generados por k y l V.A. iid $X_1,\ldots,X_k$ y $Y_1,\ldots,Y_l$ normales (con medias μ_x y μ_y respectivamente y varianza σ^2 desconocida)
  - Queremos testear si $\mu_x=\mu_y$ o no
  - Entonces:
    - $M=\{(\mu_x,\ldots,\mu_x,\mu_y,\ldots,\mu_y),\ \mu_x\in R,\ \mu_y\in R\}$
    - $M_0=\{(\mu,\ldots,\mu,\mu,\ldots,\mu),\ \mu\in \mathrm{R}\}$
    - $\Pi_M(x_1,\ldots,x_k,y_1,\ldots,y_l)=((1/k)\sum x_i,\ldots,(1/k)\sum x_i,(1/l)\sum y_i,\ldots,(1/l)\sum y_i)$
    - $\Pi_{Mo}(x_1,\ldots,x_k,y_1,\ldots,y_l)=(1/(k+l)\sum x_i+1/(k+l)\sum y_i,\ldots,1/(k+l)\sum x_i+1/(k+l)\sum y_i)$

ANOVA

- Teorema
  - Sea un test del tipo:

$$\begin{cases} H_0 : (\mu_1, \ldots, \mu_n) \in M_0;\ \sigma > 0 \\ H_1 : (\mu_1, \ldots, \mu_n) \in M \setminus M_0;\ \sigma > 0 \end{cases}$$

  - Entonces el p-valor del likelihood ratio test resultante es:

$$p^* = 1 - F_{k-k_0, n-k}(f)$$

Donde $F_{m,n}(x)$ es la distribución de Fisher con *m* y *n* grados de libertad

$$k = \dim\ M;\ k_0 = \dim\ M_0$$

$$f = \frac{SS2/(k-k_0)}{SS1/(n-k)}$$

$$SS2 = \|\widehat{\mu} - \widehat{\mu}_0\|^2;\ SS1 = \|x - \widehat{\mu}\|^2$$

$$\widehat{\mu}_0 = \Pi_{M_0}(x);\ \widehat{\mu} = \Pi_M(x)$$

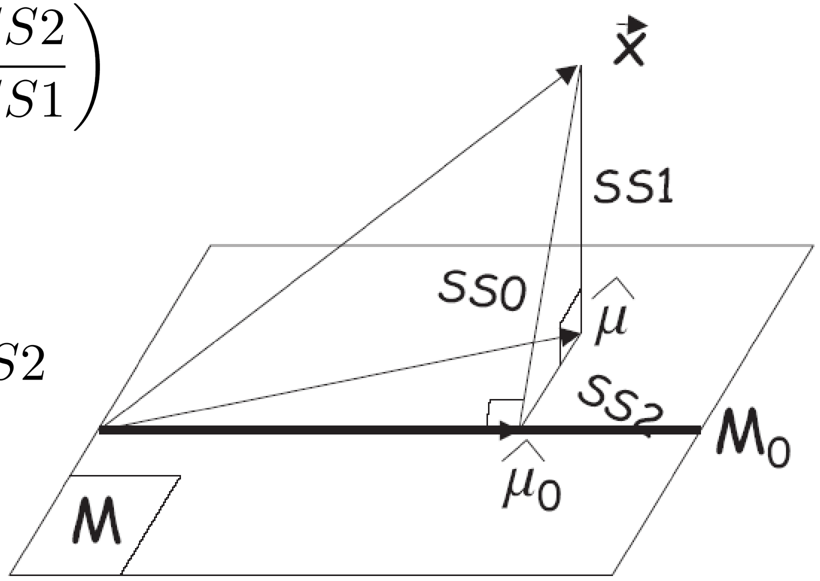
ANOVA

- La prueba es RELATIVAMENTE más sencilla de lo que parece
  - Con algunas cuenta bastante similares a las que ya hicimos se llega a que el estadístico del test es:

$$T(x) = -\frac{n}{2}\log\left(\frac{SS1}{SS0}\right) = \frac{n}{2}\log\left(1+\frac{SS2}{SS1}\right)$$

$$\text{donde } SS0 = \|x-\widehat{\mu}_0\|^2$$

$$= \|\underbrace{x-\widehat{\mu}}_{\perp M}\|^2 + \|\underbrace{\widehat{\mu}-\widehat{\mu}_0}_{\in M}\|^2 = SS1 + SS2$$



  - Y se puede probar que *f* tiene distribución Fisher lo que resulta en el teorema

ANOVA

- La interpretación intuitiva que me parece la más clara es la siguiente:

$$T(x) = \frac{n}{2} \log \left(\frac{SS0}{SS1} \right)$$

ANOVA = ANalysis Of VAriance

$$\left. \begin{aligned} SS0 &= \|x - \widehat{\mu}_0\|^2 \\ SS1 &= \|x - \widehat{\mu}\|^2 \end{aligned} \right\} \Rightarrow \frac{SS0}{SS1} = \frac{s^2_{n(0)}}{s^2_n}$$

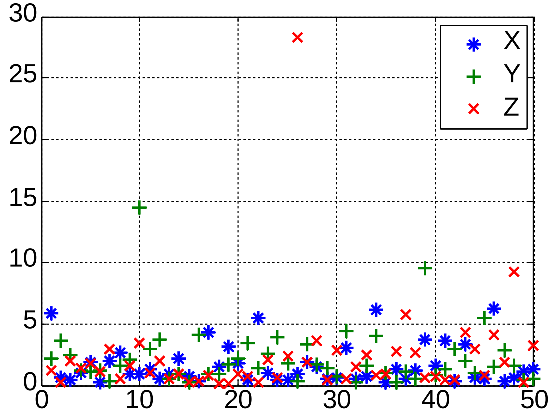
$$\text{donde } s^2_{n(0)} = \frac{1}{n-1}\|x - \widehat{\mu}_0\|^2 = \frac{1}{n-1}\|x - \widehat{\mu}\|^2 + \frac{n}{n-1}\|\widehat{\mu} - \widehat{\mu}_0\|^2$$

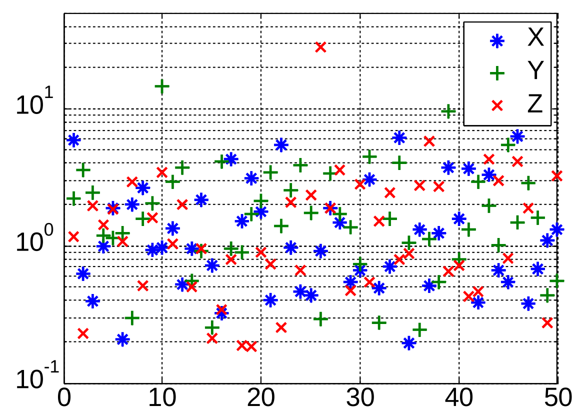
$$s^2_n = \frac{1}{n-1}\|x - \widehat{\mu}\|^2$$

- Sea $\mu=(\mu_1,\ldots,\mu_n)$
  - Si μ está en M_0 entonces SS0 será similar a SS1 (ambas proyecciones deberían ser similares)
  - Si μ no está en M_0 entonces SS0 será grande con respecto a SS1

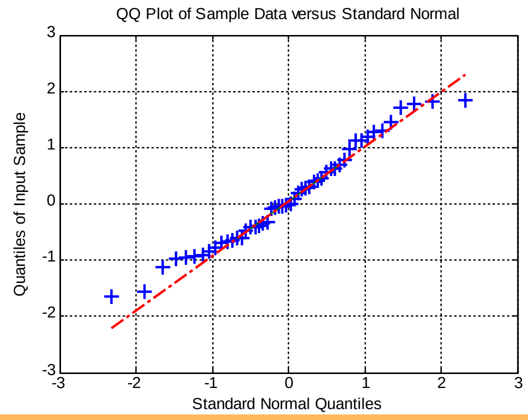
ANOVA - Ejemplo

- Imaginemos que tenemos tres routers, a los cuales se les hacen una serie de experimentos que miden el throughput promedio (el vector con las medidas para cada equipo los notaremos como *x*, *y* y *z*)
  - Queremos saber si los tres son equivalentes



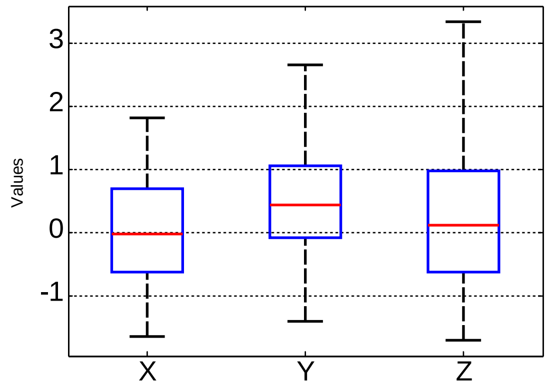


- El logaritmo de los datos parece normal
  - Ejemplo con *x*
- Redefinamos los datos como su logaritmo



ANOVA - Ejemplo

- Empecemos con lo más fácil:
  - Boxplot a ver si la diferencia no es evidente



  - *Y* parece tener una media mayor
- Hagamos un intervalo de confianza de la media
  - I de C (μ_x) = [-0.14152 , 0.37081]
  - I de C (μ_y) = [0.14944 , 0.66886]
  - I de C (μ_z) = [-0.11191 , 0.49079]
  - Sigue pareciendo que la media es mayor para *y* que el resto
    - Pero los I de C se solapan bastante: la diferencia puede ser debida a la aleatoriedad de los datos y no porque las medias sean verdaderamente distintas

ANOVA - Ejemplo

- Apliquemos ANOVA en este caso:

$$M = (\mu_x, \ldots, \mu_x, \mu_y, \ldots, \mu_y, \mu_z, \ldots, \mu_z) \; \mu_x, \mu_y, \mu_z \in R$$

$$M_0 = (\mu, \ldots, \mu) \; \mu \in R$$

$$k = \dim\ M = 3;\ k_0 = \dim\ M_0 = 1$$

$$\widehat{\mu}_0 = \Pi_{M_0}(x, y, z) =$$

$$= \left(\frac{1}{n_x + n_y + n_z} \left(\sum_{i=1}^{n_x} x_i + \sum_{i=1}^{n_y} y_i + \sum_{i=1}^{n_z} z_i \right) \right) * \text{ones}(1, n_x + n_y + n_z)$$

$$\widehat{\mu} = \Pi_M(x, y, z) =$$

$$= \left[\frac{1}{n_x} \left(\sum_{i=1}^{n_x} x_i \right) * \text{ones}(1, n_x)\, ,\ \frac{1}{n_y} \left(\sum_{i=1}^{n_y} y_i \right) * \text{ones}(1, n_y)\, ,\ \frac{1}{n_z} \left(\sum_{i=1}^{n_z} z_i \right) * \text{ones}(1, n_z) \right]$$

$$SS2 = \|\widehat{\mu} - \widehat{\mu}_0\|^2;\ SS1 = \|x - \widehat{\mu}\|^2$$

$$f = \frac{SS2/(k - k_0)}{SS1/(n - k)} \approx 3.25$$

$$p^* = F_{2, n_x + n_y + n_z - 3}(f) \approx 0.669$$

ANOVA - Ejemplo

- $p^*=0.669$ => no puedo rechazar H_0
- Es decir, la diferencia en las medias no es tan grande como para no poder atribuírsela a la aleatoriedad de los datos
- ¿Qué falto testear?
  - Independencia
  - Idénticamente distribuidos en cada grupo
  - Que fueran todos normales
  - Que la varianza fuera igual para todos
    - Es decir, las hipótesis del test
    - Es muy importante “testear el test” pues si no estamos en sus hipótesis el resultado pierde significado

Agenda

- Neyman-Pearson
- Likelihood Ratio Tests
- ANOVA
- Resultados asintóticos
- Otros tests...

Resultados Asintóticos

- Sea un likelihood ratio test; i.e.

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta \setminus \Theta_0 \end{cases} \qquad C(x) = \{x : T(x) = l_x(H_1) - l_x(H_0) > k\}$$

- Más en particular (además de algunas condiciones de regularidad básicas):

$$\Theta = \Theta_1 \times \Theta_2$$

$$\text{con } \Theta_1 \text{ subconjunto abierto de } R^{q_1}$$

$$\text{y } \Theta_2 \text{ subconjunto abierto de } R^{q_2}$$

$$\text{Sea } \theta = (\theta_1, \theta_2) \quad \theta_1 \in \Theta_1,\, \theta_2 \in \Theta_2$$

$$\begin{cases} H_0 : \theta_2 = 0 \\ H_1 : \theta_2 \neq 0 \end{cases}$$

- Para muestras grandes, y bajo H_0, es válida la siguiente aproximación: $2T(X) \sim \chi^2_{q2}$
- Por lo tanto tenemos la siguiente aproximación:

$$p^* \approx 1 - F_{\chi^2_{q_2}}(2T(x))$$

Aplicación 1: Test de Ajuste

- Tenemos $x_1,\ldots,x_n$ que asumimos generados por n V.A. iid $X_1,\ldots,X_n$
- Queremos verificar si la distribución $F()$ de X_i viene de una cierta familia de distribuciones paramétrica en θ (con θ en cierto espacio Θ_0)
- Idea :
  - Hagamos un histograma de k bins de los datos
  - Si la distribución es $F(x,\theta)$ => la probabilidad de que X_i esté en el bin j es $q_j(\theta)=F(M_j,\theta)-F(m_j,\theta)$ (el caso multidimensional es análogo)
  - Si N_j es una V.A. que indica la cantidad de datos que cayeron en el bin j => $N=(N_1,\ldots,N_k)$ sigue una distribución multinomial de parámetros n y $q(\theta)=(q_1(\theta),\ldots, q_k(\theta))$

Aplicación 1: Test de Ajuste

- Por lo tanto el test resultante es:

$$\begin{cases} H_0 : N \sim M_{n,q(\theta)} \text{ con } \theta \in \Theta_0 \\ H_1 : N \sim M_{n,p} \text{ con } p \text{ cualquiera} \end{cases}$$

- Desarrollemos un likelihood ratio test para este caso:

$$P(N_1 = n_1, \ldots, N_k = n_k) = \frac{n!}{n_1! \ldots n_k!} q_1(\theta)^{n_1} \ldots q_k(\theta)^{n_k}$$

$$\Rightarrow l_x(H_0) = \sup_{\theta \in \Theta_0} K + \sum_{i=1}^{k} n_i \log(q_i(\theta)) = K + \sum_{i=1}^{k} n_i \log(q_i(\theta^*))$$

(donde asumimos que sabemos calcular θ^*)

  - Es relativamente fácil probar que

$$l_x(H_1) = C + \sum_{i=1}^{k} n_i \log\left(\frac{n_i}{n}\right)$$

Aplicación 1: Test de Ajuste

- Por lo tanto el test resultante es:

$$\begin{cases} H_0 : N \sim M_{n,q(\theta)} \text{ con } \theta \in \Theta_0 \\ H_1 : N \sim M_{n,p} \text{ con } p \text{ cualquiera} \end{cases}$$

- Desarrollemos un likelihood ratio test para este caso:

$$\Rightarrow l_x(H_1) - l_x(H_0) = T(x) = \sum_{i=1}^{n} n_i \log\left(\frac{n_i}{nq_i(\theta^*)}\right)$$

$$\Rightarrow p^* = \sup_{\theta \in \Theta_0} P\left(\sum_{i=1}^{k} N_i \log\left(\frac{N_i}{nq_i(\theta^*)}\right) > \sum_{i=1}^{k} n_i \log\left(\frac{n_i}{nq_i(\theta^*)}\right)\right)$$

- Para θ fijo ya es difícil de calcular ¿cómo encuentro el supremo?
- Utilicemos el resultado asintótico: $q_2=k\text{-dim}(\Theta_0)\text{-}1$ (¿cuántas dimensiones agrega H_1 con respecto a H_0?)

$$p^* \approx 1 - F_{\chi^2_{k-\dim \Theta_0 - 1}}\left(2\sum_{i=1}^{n} n_i \log\left(\frac{n_i}{nq_i(\theta^*)}\right)\right)$$

Test de Ajuste - Ejemplo

- Hagamos un ejemplo con los datos que teníamos de los routers, para verificar que log(*x*), log(*y*) y log(*z*) son normales
- Ejemplo con log(*x*)
  - $\theta=(\mu,\sigma^2)$
  - $\theta^*=(\mu^*; \sigma^{*2})=(\sum x_i/n_x; \sum(x_i-\mu^*)^2/n_x)$
  - Elegí k=10
  - $q(\theta^*)$ = [0.015 0.040 0.086 0.14 0.19 0.19 0.16 0.097 0.048 0.018]
  - N/n = [0.02 0.06 0.09 0.1 0.19 0.18 0.21 0.1 0.04 0.01]
  - $T(x)$ = 3.72
  - p* = 0.38
- => No puedo rechazar la hipótesis que log(*X*) es normal

Aplicación 2: Test de Indepcia.

- Tenemos $(x_1,y_1),\dots,(x_n,y_n)$ que asumimos generados por n vectores aleatorios iid $Z_1,\dots,Z_n$ con $Z_i=(X_i,Y_i)$
- Queremos saber si X_i y Y_i son independientes entre sí
- La idea es similar a la anterior:
  - Partimos los datos de X [Y] en I [J] bins B_i [C_j]
  - Sea N_{ij} una V.A. que indica cuántos datos cayeron en el contenedor B_i y C_j y N un vector conteniendo los N_{ij}:

$$N_{ij} = \sum_{i=1}^{n} 1_{\{B_i\}}(X_i) 1_{\{C_j\}}(Y_i)$$

  - Sea p_{ij} la probabilidad de que Z_i esté en B_i y C_j y p un vector conteniendo los p_{ij}:

$$p_{ij} = P(X_i \in B_i \text{ y } Y_i \in C_j)$$

  - Como en el caso anterior N es multinomial de parámetros n y p

Aplicación 2: Test de Indepcia.

- El test resulta entonces:

$$\begin{cases} H_0 : N_{ij} \sim M_{n,p} \text{ con } p_{ij} = q_i r_j : \sum_i q_i = 1 \text{ y } \sum_j r_j = 1 \\ H_1 : N_{ij} \sim M_{n,p} \text{ con } p_{ij} \text{ cualquiera} \end{cases}$$

- Desarrollemos un likelihood ratio test para este caso:
  - Queda como ejercicio probar que

$$l_x(H_0) = C + \sum_{i,j} \log\left(\frac{n_i}{n}\frac{n_j}{n}\right) \qquad \left(\text{donde } n_i = \sum_j n_{ij} \text{ y } n_j = \sum_i n_{ij}\right)$$

$$l_x(H_1) = C + \sum_{i,j} \log\left(\frac{n_{ij}}{n}\right)$$

$$\Rightarrow T(n) = \sum_{i,j} n_{ij} \log\left(\frac{n n_{ij}}{n_i n_j}\right)$$

Aplicación 2: Test de Indepcia.

- El test resulta entonces:

$$\begin{cases} H_0 : N_{ij} \sim M_{n,p} \text{ con } p_{ij} = q_i r_j : \sum_i q_i = 1 \text{ y } \sum_j r_j = 1 \\ H_1 : N_{ij} \sim M_{n,p} \text{ con } p_{ij} \text{ cualquiera} \end{cases}$$

- Desarrollemos un likelihood ratio test para este caso:
  - Nuevamente nos vemos obligados a utilizar la aproximación asintótica
  - ¿Cuánto vale q_2?
    - ¿Cuántas dimensiones le agrega H_1 a H_0?
    - dim(θ)=dim(p)
    - Bajo H_1 p lo único que debe cumplir es que sume 1 => dim(p,H_1)=IJ-1
    - Bajo H_0 p es el producto de dos valores que suman 1 => dim(p,H_0)=(I-1)+(J-1)
    - q_2=dim(p,H_1)-dim(p,H_0)=(I-1)(J-1)

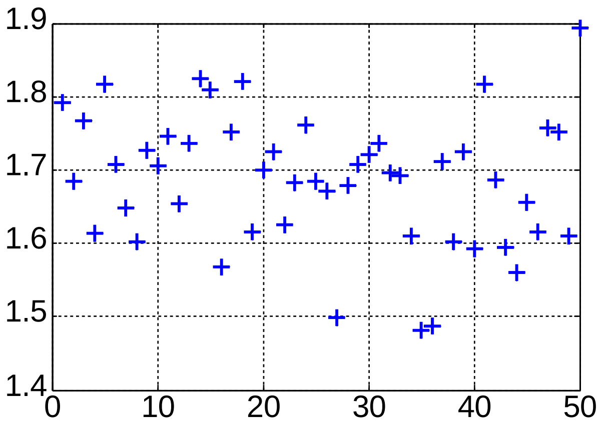
$$\Rightarrow p^* \approx 1 - F_{\chi^2_{(I-1)(J-1)}}\left(2\sum_{i,j} n_{ij}\log\left(\frac{nn_{ij}}{n_i n_j}\right)\right)$$

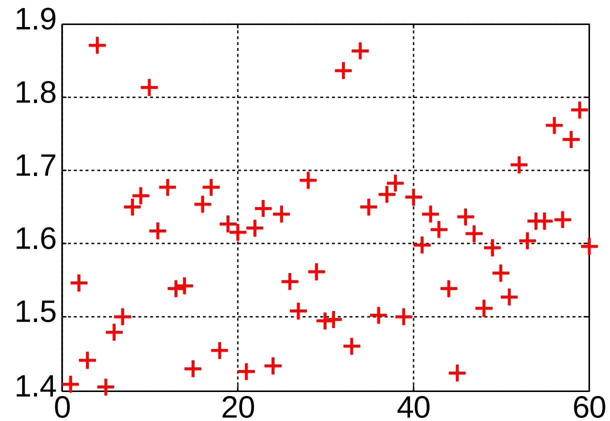
Test de Indepcia: Ejemplo

- Tenemos n datos de altura de personas, de las cuales también conocemos el sexo
  - Asumamos que los datos fueron generados por n vectores aleatorios iid $Z_1,\ldots,Z_n$ con $Z_i=(X_i,Y_i)$
    - X_i vale 1 si la persona es hombre
    - X_i vale 0 si la persona es mujer
    - Y_i indica la altura
- Si la encuesta estuvo bien hecha los datos deberían ser iid
- La pregunta que nos hacemos es:
  - ¿Son la altura y el sexo independientes?

Test de Indepcia: Ejemplo

- Datos:





- Uso cinco bins para las alturas y dos para el sexo:
  - n_{ij} = 3 4 19 18 6
    15 13 24 4 4
  - n_i = 50 60
  - n_j = 18 17 43 22 10
  - $T(n)=11.7$
  - $p^*=0.1\text{x}10^{-3}$

=> Tengo que rechazar H_0, son claramente dependientes!

Agenda

- Neyman-Pearson
- Likelihood Ratio Tests
- ANOVA
- Resultados asintóticos
- Otros tests...

Otros tests

- Obviamente no vimos ni cerca todos los tests posibles
- Por ejemplo, dejamos afuera toda una familia de tests
  - Los tests no paramétricos
    - no se testea sobre un parámetro de una distribución
- Ejemplos son los tests de iid-ismo

$$\begin{cases} H_0 : X_1, \ldots, X_n \text{ son iid} \\ H_1 : X_1, \ldots, X_n \text{ no son iid} \end{cases}$$

  - Vemos un test como este a continuación

Test de Turning Point

- Tenemos $x_1,\ldots,x_n$ que asumimos generados por n V.A. $X_1,\ldots,X_n$
- Se dice que la serie $x_1,\ldots,x_n$ es monótona en i si $x_{i-1} \leq x_i \leq x_{i+1}$ o $x_{i-1} \geq x_i \geq x_{i+1}$
- Se dice que la serie tiene un "turning point" en i si la serie no es monótona en i
- Sea T la cantidad de "turning points"
- El p-valor del test de turning point es para n grande:

$$p^* = 2\left(1 - F_{N_{0,1}}\left(\frac{\left|T - \frac{2n-4}{3}\right|}{\sqrt{\frac{16n-29}{90}}}\right)\right)$$

Bibliografía

- Para los temas “Simulación”, “Presentando los Datos” y “Tests” basé las presentaciones en el libro “Performance Evaluation of Computer and Communication Systems” de Jean-Yves Le Boudec (ed. EPFL Press)
- Para el obligatorio quizás quieran y/o necesiten otros tests o intervalos de confianza
  - El libro “Probabilidad y Estadística Matemática: Un Primer Encuentro” de Gonzalo Perera (editado por la Oficina de Publicaciones del CEI) tiene muchos tests e I. de C.
  - Internet también es una buena fuente (recuerden que Timbó da acceso a los libros editados por Elsevier)
- ATENCIÓN: Siempre verifiquen las hipótesis de los métodos con los que trabajen!