Tipos de variables

* **Numéricas**: valores numéricos, sensibles a operaciones aritméticas
  + **Continuas**: Cualquier valor de los reales (en un intervalo)
  + **Discretas**: Valores específicos, ej: enteros
* **Categóricas**: valores de un conjunto acotado. Cada posible valor se denomina **nivel**
  + **Nominales:** No existe un orden natural entre los niveles. Ej: sexo
  + **Ordinales:** Existe un orden natural. Ej: jerarquía

Estadísticas descriptivas

* **Estimadores puntuales**
* **Distribución**
* **Distribución de frecuencia:** Cuantas veces aparece cada valor para una variable en un conjunto de datos
* **Media muestral**  **media poblacional** **(medidas de tendencia central)**
* **Mediana**
* **Moda: unimodales, bimodales y multimodales**
* **Varianza y desviación estándar**
* **Rango**
* **Percentiles, deciles, quintiles y cuartiles**

Estadísticas descriptivas para datos categóricos

* **Frecuencia**
* **Proporción**
* **Tabla de contingencia -> matriz de confusión o tabla de frecuencias**
* **Histograma -> densidad**
* **Grafico de caja -> extremo inferior y superior son el primer y tercer cuartil. Línea horizontal al interior de la caja es la mediana. Altura es el rango intercuartil (Q3-Q1), es decir, la caja engloba el 50% central de los datos. Las líneas exteriores son bigotes y son a lo mas datos situados 1.5 veces el IQR (rango intercuartil). Puntos alejados son valores atípicos**
* **Para variables categóricas, usar grafico de barras o de torta**
* **Para dos variables, gráficos de dispersión**
* **Para dos variables categóricas gráficos de barras apiladas, agrupadas y estandarizadas**
* **Grafico de mosaico para tablas de contingencias**
* **Para una numérica y una categórica usar grafico de cajas o de tiras si son pocas muestras**

Variables aleatorias

* **Continuas:** variable que puede tomar cualquier valor entre los infinitos valores posibles
* **Discreta:** toma un valor de un conjunto finito de valores
* **Valor esperado E(x):** resultado promedio de una variable aleatoria. Ej: 1/6 para obtener 6 en un dado
* **Función de densidad de probabilidad o simplemente distribución o densidad:** Su área bajo la curva siempre es 1

Distribuciones continuas

* **Distribución normal (media, desviación) (0, 1) -> estandarizada (distribución Z)**
  + **dnorm:** densidad de una distribución normal
  + **pnorm:** percentiles, es decir, función de distribución acumulada. Ej: obtener un 75 o menos/mas
  + **qnorm:** encuentra el percentil para las probabilidades dadas en p, es inversa a pnorm
  + **rnorm:** genera aleatoriamente n observaciones
  + **Grafico Q**-**Q:** permite demostrar normalidad: puntos son observaciones, recta es la distribución normal. **Banda coloreada establece el margen aceptable para suponer la normalidad de datos**
* **Distribución Z (distribución normal estandarizada) (media, desviación) (0, 1)**
* **Distribución chi-cuadrado****:** Sirve para caracterizar valores **siempre** positivos y habitualmente desviados a la derecha. Su único parámetro son los **grados de libertad (v)** que son una estimación de la cantidad de observaciones empleadas para calcular un estimador (Tres grupos para obtener un estimador X, solo dos pueden cambiar, por lo tanto, 2 grados de libertad. Lo mismo aplica para observaciones y media, por ejemplo). Se relaciona con Z
* **Distribución t-Student:** Solamente tiene grados de libertad como parámetro, a medida que los grados de libertad aumentan, se parece cada vez mas a la normal, aunque con colas más gruesas. Se relaciona con Z
* **Distribución F:** Usada para comparar varianzas. Se relaciona con chi-cuadrado donde cada chi-cuadrado tiene sus grados de libertad v1 y v2

Distribuciones discretas

* **Distribución de Bernoulli:** Su variable aleatoria es aquella que en cada intento individual tiene solo **dos resultados posibles: éxito (probabilidad p) y fracaso (probabilidad q = 1-p)**. Los intentos son independientes entre si y la proporción de la muestra es la cantidad de éxitos dividida por la cantidad de intentos.
* **Distribución geométrica:** Describe la cantidad de intentos para obtener un éxito para variables de Bernoulli **independientes e idénticamente distribuidas**, es decir, que no se afectan unas a otras y cada una con igual probabilidad de éxito
* **Distribución binomial:** Describe la probabilidad de tener exactamente **k éxitosen n intentos** independientes de Bernoulli con probabilidad de éxito p
* **Distribución binomial negativa:** Describe la probabilidad de encontrar el k-esimo éxito al n-esimo intento.
* **Distribución de Poisson:** Útil para estimar la cantidad de eventos en una población grande en un lapso dado, ej: cantidad de contagios en una semana.

FIN PARTE NO INFERENCIAL (INTRODUCCIÓN/REPASO)

Error Estándar: desviación estándar de la distribución de un estimador muestral:

Distribución de las medias tiende a ser cercana a la normal, por lo que en dicho caso es posible usar el **modelo normal**, sustentado en el teorema del limite central. Las condiciones que deben cumplirse para usar este modelo y que en consecuencia el error estándar sea preciso son:

1. Las observaciones de la muestra son independientes
2. La muestra es grande (n>=30)
3. La distribución de la muestra no es significativamente asimétrica. Suele además relacionarse con la presencia de valores atípicos. Mientras mayor sea el tamaño de la muestra, mas se puede relajar esta condición

Intervalos de confianza: Alrededor de 95% de las veces el estimador puntual se encontrará en un rango de 2 errores estándar del parámetro. Es decir, al considerar un intervalo de confianza de dos errores estándar, tendremos un 95% de **confianza** de haber capturado el parámetro real (x +- 2\*SEx) -> (x+- z\* SEx) donde z\* es el área bajo la curva (10% -> 5% en cada cola), con esto podemos usar qnorm y sacar la probabilidad del 5%. Este porcentaje (90% como el de arriba) nos dice que, **en promedio, x% de los intervalos de confianza que se construyan en torno a un estadístico con muestras de un tamaño fijo capturaran el verdadero valor del parámetro. NO DECIR QUE SE TIENE UNA PROBABILIDAD DEL X% DE ESTAR ENTRE LOS VALORES DEL INTERVALO CALCULADO!!!**

Prueba de hipótesis

* **H0:** Hipótesis nula, postura escéptica, es decir que no hay cambios, se formula siempre como una igualdad
* **Ha:** Hipótesis alternativa, lo contrario al H0
* **Valor nulo:** Valor del parámetro cuando se cumple la hipótesis nula
* Usar: **Se falla al rechazar H0 o bien se rechaza H0 en favor de Ha. Nunca decir que se acepta H0 como verdadera o correcta.**
* Un estadístico de prueba es un estadístico de resumen que resulta ser especialmente útil para evaluar hipótesis o calcular el valor p.

Errores

* **Error tipo I:** Rechazar H0 en favor de Ha cuando H0 es en realidad verdadera. Para evitar este usar un alfa más pequeña 0.01
* **Error tipo II:** No rechazar H0 cuando en realidad Ha es verdadera. Para evitar este usar un alfa más grande 0.10

**Inferencia con medias**

* **Prueba Z:** Inferir acerca de las medias con una o dos muestras. Por ahora solo el primer caso. Sirve para **asegurar o descartar** que la media de una población tiene cierto **valor hipotético (ej: media de pobla = 30).** Condiciones:

1. La **muestra** debe tener al menos **30 observaciones**. Si la muestra tiene menos de 30 observaciones se debe conocer **la varianza de la población (usamos realmente la desviación estándar)**
2. Las observaciones deben ser **independientes**, es decir, la selección de una observación de la muestra no influye en la selección de otras
3. La **población** de donde se obtuvo la muestra sigue aproximadamente una distribución normal
4. **(Comentario) Hacer prueba Q-Q para determinar si la muestra sigue una distribución normal (también podemos usar la prueba Shapiro-Wilk que es mejor que verlo gráficamente)**

* **Prueba t de Student:** En la práctica, rara vez conocemos la desviación estándar de la población y a menudo tenemos muestras pequeñas, por lo que la prueba Z no es muy utilizada. Esta prueba es la mas empleada para inferir acerca de una o dos medias muestrales.
  + **Prueba t para una muestra:** No opera bajo el supuesto de normalidad, aun así, requiere revisar las siguientes condiciones:

1. Las observaciones son **independientes** entre si
2. Las **observaciones provienen de una distribución cercana a la normal**
   * **Prueba t para dos muestras pareadas:** Acá tenemos datos pareados, es decir, cada observación de un conjunto tiene una correspondencia o conexión especial con exactamente una observación del otro (Por ejemplo, tiempo final y tiempo inicial para una observación). Una forma de uso para examinar estos datos es usar la **diferencia entre cada par** y luego usar t para una muestra.
   * **Prueba t para dos muestras independientes:** En este caso utilizamos la prueba para comparar las medias de dos poblaciones en que las observaciones no tienen ninguna relación con las otras observaciones, ni influyen en su selección, ni en ninguna otra parte. En este caso, la inferencia se hace sobre la diferencia de las medias mu1 – mu2 = d0 donde d0 es un valor hipotético fijo para la diferencia, normalmente se usa d0 = 0. En cuyo caso, las muestras podrían provenir de dos poblaciones distintas con igual media o desde la misma población. Para ello la prueba usa como estimador puntual la diferencia de las medias muestrales (x1-x2). Así nace el **estadístico T. Condiciones:**
3. Cada muestra cumple las **condiciones** para usar t student
4. Las muestras son **independientes** entre si

**Inferencias con proporciones: Inferir acerca de una y dos proporciones**

* **Método de Wald:** En general no conocemos la probabilidad de éxito p de la población, por lo que tenemos que usar el estimador puntual (correspondiente a la proporción de éxito de la muestra), denotado por . Este estimador se distribuye de manera cercana a la normal cuando se cumplen las siguientes condiciones:

1. Las observaciones de la muestra son **independientes**
2. Se cumple la **condición de éxito-fracaso**, que establece que en la muestra se espera encontrar al menos **10 observaciones correspondientes a éxito**s y al menos **10 fracasos**
   * **Método de Wald para una proporción: Espero que no entre en la PEP porque no hay mucha explicación y es solo calculo a papel/algunas funciones difíciles de recordar (hay que aprenderse el procedimiento)**
   * **Método de Wald para dos proporciones: IDEM**

* **Método de Wilson:** Este método opera del mismo modo que el de Wald, introduce ajustes en la estimación de la proporción en la muestra y el error estándar de su distribución muestral cuando se estiman los intervalos de confianza. **Este método utiliza el estadístico chi-cuadrado**

FIN

Nivel de significación **alfa** corresponde a la probabilidad de cometer errores del tipo I, definimos **beta** como la probabilidad de cometer errores de tipo II. Alfa y beta están relacionados: **para un tamaño fijo de la muestra: al reducir beta, alfa aumenta y viceversa.** Esto nos lleva a un nuevo concepto el **poder estadístico** de una prueba de hipótesis, también llamado “potencia estadística” dado por 1 – beta, que se define como **la probabilidad de correctamente rechazar H0 cuando es falsa (probabilidad de no cometer un error tipo II).**

Poder, nivel de significación y tamaño de la muestra

* **Me salte varias partes de acá porque no creo que entren en la prueba, ojo**

Tamaño del efecto

* Al comparar dos medias existe la llamada **d de Cohen**. En términos generales se considera que d = 0.2 es un efecto pequeño, d = 0.5 es mediano y d = 0.8 es grande.
* d = media – medioa teórica / desviación estándar de la muestra con n – 1 grados de libertad
* **La fórmula cambia dependiendo de las muestras, no las anoté acá**

FIN

Inferencia no paramétrica con proporciones **(Pruebas de chi-cuadrado)**

* Pruebas que **no mencionan parámetro alguno. Es más, ninguna de ellas hace alguna suposición sobre la distribución de la población (no paramétricas o libres de distribución)**
  + **Entregan menos información (muestran las mismas proporciones/distintas)**
  + **Ninguna indica cuales son esas proporciones, ni siquiera cual es mayor**
  + **Tienen menor poder estadístico (necesitan muestras de mayor tamaño para ser útiles)**

**Prueba chi-cuadrado de Pearson**

* Sirve para inferir con proporciones cuando disponemos de dos variables categóricas y una de ellas es dicotómica (es decir, tiene solo dos niveles). En este caso, podemos registrar las frecuencias observadas para las **posibles combinaciones** de ambas variables mediante una **tabla de contingencia**. En adelante, estas combinaciones serán llamadas **grupo**
* Condiciones para verificar antes de usar la prueba chi-cuadrado:

1. Las observaciones son **independientes** entre si
2. Debe haber a lo menos **5 observaciones** esperadas en cada grupo

* Tres pruebas diferentes:
  + **Chi-cuadrado de homogeneidad**
  + **Chi-cuadrado de bondad de ajuste**
  + **Chi-cuadrado de independencia**
  + **Solo tienen diferencia conceptual y se relaciona en como se miran las variables y las poblaciones involucradas en el problema**
* **Prueba chi-cuadrado de Homogeneidad**
  + Resulta adecuada para determinar si **dos poblaciones** (la variable dicotómica) **presentan las mismas proporciones en los diferentes niveles de una variable categórica. Ej: Encuesta hacia hombres/mujeres (variable categórica de dos niveles) sobre algo en específico. Hay que armar la tabla de frecuencias, es decir, para una opción poner hombres, mujeres y total y la última columna también de total**
  + **El estadistico chi-cuadrado sigue una distribución chi-cuadrado con v = (m-1)\*(n-1) grados de libertad**
  + H0: programadores hombres y mujeres tienen las mismas preferencias en lenguaje de programación favorito (ambas poblaciones muestras las mismas proporciones para cada lenguaje estudiado).
  + HA: programadores hombres y mujeres tienen preferencias distintas en lenguajes de programación favorito.
* **Prueba chi-cuadrado de bondad de ajuste**
  + **Permite comprobar si una distribución de frecuencias observada se asemeja a una distribución esperada.** Usualmente se usa para comprobar si una muestra es representativa de la población. **Hacer tabla de frecuencia para la población y la muestra en cada fila y cada columna los campos (C, Java, Python, etc)**
  + H0: las proporciones de especialistas en cada lenguaje son las mismas para la nómina y la muestra.
  + HA: las proporciones de especialistas en cada lenguaje son diferentes en la nómina que en la muestra.
* **Prueba chi-cuadrado de independencia**
  + **Determinar si dos variables categóricas, de una misma población, son estadísticamente independientes o si están relacionadas. Hacer una tabla de frecuencias para cada variable categórica, siendo la parte de la izquierda una variable (clase que puede ser comestible/venenoso) y arriba la otra variable (forma del sombrero, que puede ser hartas, no necesariamente dicotómica)**
  + H0: las variables clase y forma del sombrero son independientes.
  + HA: las variables clase y forma del sombrero están relacionadas.

Pruebas para muestras pequeñas

* **Prueba exacta de Fisher**
  + Es una alternativa de la prueba chi-cuadrado de independencia en el caso de que **ambas variables sean dicotómicas. La tabla de frecuencias pasa a ser de 2x2 con una variable arriba y otra a la izquierda**
* **Prueba de mcNemar**
  + Resulta apropiada cuando una misma característica, con respuesta dicotómica (sexo: hombre/mujer) se mide en dos ocasiones diferentes para los mismos sujetos (muestras pareadas) y queremos determinar si se produce o no un cambio significativo entre ambas mediciones. **Nuevamente tabla de frecuencias**
  + H0: no hay cambios significativos en las respuestas.
  + HA: sí hay cambios significativos en las respuestas.
  + Ej: Modelo 1 -> Si/No, Modelo 2 -> Si/No y sus valores correspondientes
* **Prueba Q de Cochran**
  + Extensión de la prueba de mcNemar, adecuada cuando la **variable de respuesta es dicotómica (esto se puede mostrar con 0s y 1s)** y la variable independiente tiene mas de dos observaciones pareadas (si ambas variables son dicotómicas, entonces es mcNemar)
  + H0: la proporción de “éxitos” es la misma para todos los grupos.
  + HA: la proporción de “éxitos” es distinta para al menos un grupo.
  + Condiciones:

1. La **variable de respuesta** es **dicotómica**
2. La **variable independiente** es **categórica**
3. Las observaciones son **independientes** entre si
4. El tamaño de la muestra es lo suficientemente grande. Se sugiere **n\*k>=24 donde n>=4 es el numero de casos cuyas respuestas no son únicamente éxitos o fracasos y k es la cantidad de niveles de la variable independiente**
   * Cochran es **Omnibus, que tienen pruebas post-hoc para encontrar exactamente donde falla la hipótesis nula. Solo se hace un procedimiento post-hoc si la prueba ómnibus rechaza la hipótesis nula. Para Cochran existe pairwiseMcnemar(formula, data, method) method: Bonferroni/Holm**

FIN

ANOVA de una vía para muestras independientes

El método **ANOVA** de **análisis de varianza** sirve para comparar simultáneamente tres o más medias muestrales. De manera similar, existe ANOVA para muestras correlacionadas (cap siguiente). Los procedimientos para ANOVA corresponden **al análisis de varianza de una vía**, pues solo consideran una única variable independiente (de tipo categórica, un factor) cuyos niveles definen los grupos que se están comparando

H0 : el tiempo de ejecución promedio para instancias de tamaño E es igual para los tres algoritmos.

HA : el tiempo de ejecución promedio para instancias de tamaño E es diferente para al menos un algoritmo

ANOVA es **ómnibus** **y paramétrica**

**La hipótesis podría escribirse en términos de las diferencias entre pares de medias.**

**Condiciones para usar ANOVA de una vía para muestras independientes**

1. La **escala** con que se mide **la variable dependiente** tiene las **propiedades de una escala de intervalos iguales** (logarítmica no sirve, por ejemplo) – ver si diferencia numérica es en intervalos iguales
2. Las **k muestras** son obtenidas de manera aleatoria e independiente desde la población de origen – revisar enunciado y ver si es confiable
3. Se puede **suponer** **razonablemente** que las **poblaciones** **de** **origen** siguen una **distribución** **normal – idem**
4. Si las **muestras** **provienen** de **mas** de una **población**, **estas** **tienen** la **misma** **varianza – Para revisar la homogeneidad de las varianzas u homocedasticidad** se puede comprobar que la razón entre la máxima y mínima varianza muestral de los grupos no sea superior a 1.5 var1/var1 <= 1.5

**Procedimiento ANOVA de una vía para muestras independientes**

**El procedimiento ANOVA de una vía para variables independientes puede resumirse en los siguientes pasos:**

**1. Calcular la suma de los cuadrados de las desviaciones para la muestra combinada (SS T ).**

**2. Para cada grupo g, calcular la suma de los cuadrados de las desviaciones dentro de dicho grupo (SSg).**

**3. Calcular la variabilidad entre grupos (SSbg ).**

**4. Calcular la variabilidad al interior de los grupos (SS wg ).**

**5. Calcular los grados de libertad (νT , νbg y νwg ).**

**6. Calcular las medias de las desviaciones cuadradas (MSbg y MS wg ).**

**7. Calcular el estadístico de prueba (F).**

**8. Obtener el valor p.**

**EzANOVA** además de ejecutar la prueba ANOVA como lo hace aov, también realiza la **prueba de homocedasticidad de Levenne** la cual permite ver si k muestras tienen igual varianza sin asumir la normalidad de las poblaciones de origen. Con **ezPlot** graficamos el efecto medido. Esta función tiene los mismos argumentos que ezANOVA con la salvedad del nuevo argumento x que señala la variable que va en el eje horizontal del grafico que, normalmente corresponde al factor que identifica los grupos (la vía) (Algoritmo A,B,C)

**Analisis Post-Hoc**

* **Corrección de Bonferroni y Holm:** pairwise.t.test(x, g, p.adjust.method, pool.sd, paired, alternative, …)
* bonferroni <- pairwise .t. test ( datos [[" tiempo "]] , datos [[" algoritmo "]] , p . adj = " bonferroni ", pool .sd = TRUE , paired = FALSE , conf . level = 1 - alfa )
* Ojalá no confiarse tanto de Bonferroni porque es muy conservadora
* P.adjust.method es Bonferroni o holm
* Un valor menor que alfa indica que hay diferencias y mirando el plot podemos decir cual es mejor/peor **REVISAR BIEN ESTO**

**Prueba HSD de Tukey**

* Es mas poderosa que Bonferroni y Holm. Utiliza el estadístico Q, el cual sigue una distribución de rango estudiantizado.
* TukeyHSD(aov, “algoritmo” (A,B,C), ordered = TRUE, conf.level) Al igual que antes, encontrar el valor p que sea menor a alfa

**Prueba de comparación de Scheffé (me dio paja seguir anotando)**

**ANOVA de una vía para muestras correlacionadas**

**Las condiciones que se deben verificar son:**

**1. La escala con que se mide la variable dependiente tiene las propiedades de una escala de intervalos iguales.**

**2. Las mediciones son independientes al interior de cada grupo.**

**3. Se puede suponer razonablemente que la(s) población(es) de origen sigue(n) una distribución normal.**

**4. La matriz de varianzas-covarianzas es esférica. Como explica Horn, esta condición establece que las varianzas entre los diferentes niveles de las medidas repetidas deben ser iguales (esta es la única que cambia con el otro ANOVA)**

**Prueba de esfericidad de Mauchly (incluida en ezANOVA)**

**1. Calcular la suma de los cuadrados de las desviaciones para la muestra combinada (SS T ).**

**2. Para cada grupo g, calcular la suma de los cuadrados de las desviaciones dentro de dicho grupo (SSg).**

**3. Calcular la variabilidad entre grupos (SSbg ).**

**4. Calcular la variabilidad al interior de los grupos (SS wg ).**

**5. Calcular la variabilidad intra-sujetos y la variabilidad del error (SSsujetos y SSerror ).**

**6. Calcular los grados de libertad relevantes (νT , νefecto = νbg y νerror ).**

**7. Calcular las medias cuadradas (MSefecto = MSbg y MSerror ).**

**8. Calcular el estadístico de prueba (F).**

**9. Obtener el valor p**

ezANOVA ( data = datos , dv = tiempo , **within** = algoritmo , wid = instancia , return \_ aov = TRUE ) within en vez de between

Dos correcciones: **Greenhouse-Geisser** para una esfericidad menor que 0.75 y **Huynd-Feldt** para esfericidad mayor o igual que 0.75 (ambas se obtienen en la corrección). **Si los datos no cumplen la esfericidad, deberíamos considerar p[GG] o p[HF] como valor p de la prueba**

**Bonferroni y Holm con paired TRUE para correlacionadas, no usar ni Tukey ni Scheffe**

**Inferencia no paramétrica con variables numéricas**

**Pruebas para una o dos muestras**

**Prueba de suma de rangos de Wilcoxon**

* Alternativa no paramétrica a la prueba t de Student con muestras independientes. Pese a ser no paramétrica requiere verificar lo siguiente:

1. Las observaciones de ambas muestras son independientes
2. La escala de medición empleada debe ser a lo menos ordinal, de modo que tenga sentido hablar de relaciones orden (igual que, menor que, mayor o igual que, etc)

* H0: no hay diferencia en la usabilidad de ambas interfaces (se distribuyen de igual forma).
* HA: sí hay diferencia en la usabilidad de ambas interfaces (distribuciones distintas).
* Al igual que en el caso de la chi-cuadrado de Pearson, estas hipótesis no hacen referencia a ningún parámetro, es decir, entregan menos información que la prueba paramétrica equivalente

**Prueba de suma de rangos de Wilcoxon para muestras grandes**

* **Cuando ambas muestras tienen tamaño mayor o igual a 5**

**Prueba de suma de rangos de Wilcoxon para muestras pequeñas**

**Prueba de rangos con signo de Wilcoxon**

* En este caso es la alternativa no paramétrica de la prueba t de Student con muestras pareadas. Sus condiciones son:

1. Los pares de observaciones son independientes
2. La escala de medición empleada para las observaciones es intrínsicamente continua
3. La escala de medición empleada para ambas muestras debe ser a lo menos ordinal

* **H0: las mismas personas no perciben diferencia en la usabilidad de ambas interfaces.**
* **HA: las mismas personas consideran que la interfaz A tiene mejor usabilidad que la interfaz B.**

**Pruebas para más de dos muestras**

**Prueba de Kruskal-Wallis**

* **Se parece a ANOVA, pero se usa cuando los tamaños de las muestras difieren.** Sus condiciones son:

1. La variable **independiente** debe tener **a lo menos dos niveles** (aunque, para **dos niveles,** se suele usar la prueba de Wilcoxon)
2. La escala de la variable **dependiente** debe ser a lo menos ordinal
3. Las **observaciones** son **independientes** entre si

* H0: todos los algoritmos son igual de eficientes (o, de manera similar, ningún algoritmo es menos ni más eficiente que los demás).
* HA: al menos uno de los algoritmos presenta una eficiencia diferente a al menos algún otro algoritmo.
* Usar holm y esas cosas para post-hoc, pairwise.wilcox.test

**Prueba de Friedman**

* Parecida a ANOVA, pero no paramétrica y para muestras correlacionadas. Sin embargo, no es exactamente una extensión ya que no considera diferencias relativas entre sujetos (como lo hace ANOVA y la prueba de rangos con signo de Wilcoxon). Condiciones:

1. La variable independiente debe ser categórica y tener a lo menos tres niveles.
2. La escala de la variable dependiente debe ser, a lo menos, ordinal.
3. Los sujetos son una muestra aleatoria e independiente de la población

* **H0: las interfaces tienen preferencias similares.**
* **HA: al menos una interfaz obtiene una preferencia distinta a las demás.**