Grupo de estadística para el estudio del lenguaje│ GESEL

17 de abril de 2019

**Test estadísticos : *t*-test de Student y ANOVA**

**1. *t*-test de Student**

**1.1 La distribución *t***

Los test estadísticos para comparar las medidas de tendencia central de dos grupos definidos por una variable categórica son el *t*-test (que compara medias) y las versiones no paramétricas que comparan las medianas: el Wilcoxon test (para muestras dependientes) y el test de U de Mann-Whitney (para muestras independientes).

Ambos test pueden aplicarse a data dependiente (apareada) o independiente (desapareada). Los datos dependientes son aquellos cuyas observaciones pertenecen a ambos grupos, es decir, datos de pares de valores con unidades estadísticas idénticas (los mismos sujetos o los mismos ítems experimentales). Si las observaciones no están relacionadas o se trata de dos grupos de observaciones, se consideran independientes. Ambas pruebas presentan las versiones con una cola o con dos colas (dependiendo del tipo de hipótesis alternativa, si es direccional o no)

La distribución *t* es unimodal y simétrica, sin embargo, sus colas son más extensas. Eso significa que con esta distribución las observaciones pueden caer bajo más desvíos estándar (más de dos). Es una distribución más conservadora que permite tener margen mayor de dispersión cuando se desconoce sigma. Está centrada en cero y tiene un solo parámetro: los grados de libertad. Los grados de libertad determinan el grosor de las colas. Cuando los gl aumentan, la distribución se acerca más a la normalidad, es decir que la dispersión de la curva disminuye. 

**1.2 Calcular el valor de *t***

La fórmula del *t*-test de una muestra (en comparación con la población) es la diferencia de la media observada menos la media de la población sobre el *SE* de la muestra.

$$t=\frac{x̄-μ}{SE}$$

Cuando comparamos dos medias, el problema reside en el cálculo del SE. Si se asumen varianzas iguales (ver supuestos), el *SE* se calcula promediando las varianzas

$$s^{2}\_{p}=\frac{\left(n\_{a}- 1\right)s^{2}\_{a}+ (n\_{b}-1)s^{2}\_{b}}{\left(n\_{a}- 1\right)+ (n\_{b}-1)\_{ }}$$

La fórmula para calcular el valor de *t* de muestras **independientes**, a partir de este tipo de varianza es

$$t=\frac{x̄\_{a}-x̄\_{b}}{\sqrt{s^{2}\_{p}/(n\_{a}+n\_{b})}}$$

Cuando las varianzas no son similares (ver supuestos), las fórmulas para calcular el valor del SE y el valor de *t* con ese error son:

$$SE=√(\frac{s^{2}\_{a}}{n\_{a}}+\frac{s^{2}\_{b}}{n\_{b}})$$

$$t=\frac{x̄\_{a}-x̄\_{b}}{SE}$$

Por último, para el cálculo de t de muestras **dependientes** se debe formular una nueva variable que incluya las diferencias entre los pares de observaciones. Es importante utilizar esta fórmula cuando las muestras son realmente dependientes, ya que los *t*-test para datos pareados tienen más poder estadístico que los que asumen variabilidad por sujeto o ítem.

$$y\_{i}= x\_{ai}- x\_{bi}$$

Con esos valores, se calcula la media de esa variable de diferencia (*d*) y el desvío de esa media $(s^{2}\_{d})$

$$t=\frac{y}{\sqrt{s^{2}\_{y}}}$$

**1.3 Supuestos del *t*-test**

Los *t*-tests tienen los siguientes supuestos:

1. Las muestras son elegidas aleatoriamente de la población

2. Las observaciones son independientes/dependientes según el test que se aplique

3. Una variable debe ser al menos de intervalo (o razón)

4. La data de las dos muestras debe seguir la distribución normal y/o el número de observaciones debe ser mayor a 30. Si el N es menor a 30 y no sigue la distribución normal se debe hacer la prueba de Shapiro-Wilk. En los test de data dependiente lo que se debe chequear por normalidad es la diferencia entre las muestras, no las muestras en sí.

5. Las varianzas de las muestras deben ser homogéneas (o se aplica el ajuste de Welch que modifica los grados de libertad haciendo la distribución de t más conservadora)

El ajuste de Welch sirve también para hacer pruebas si el n de ambas muestras es diferente (si por ej. faltan datos de algunas observaciones de un grupo).

Para testear la varianza de las muestras, se usa la distribución F (la misma que ANOVA) para calcular la probabilidad de la razón de la varianza. En relación a dos muestras se calcula F. Un resultado de F=1 indica que la H0 es verdadera.

$$F=\frac{s^{2}\_{a}}{s^{2}\_{b}}$$

También, se puede calcular el valor de *p* para un determinado valor de F. Para calcular el valor de *p* en R se usa la función:

pf(F, DFx , DFy , lower.tail=F)

Si estos supuestos no se aplican a nuestros datos, hay que utilizar la versión no paramétrica.

**1.4 Métodos paramétricos y no paramétricos**

Es importante elegir un método correcto según los datos que se estén evaluando. Los métodos paramétricos requieren que se cumplan varios supuestos y surge la pregunta ¿por qué no usar directamente métodos no paramétricos? La respuesta es la siguiente: al modificar una variable de razón o intervalo a variables ordinales se pierde información acerca de las diferencias en esas variables. Los test paramétricos tienen más poder estadístico. Por eso, se recomienda usar excepcionalmente test paramétricos incluso cuando la distribución no es normal y tiene colas cortas, por ejemplo con scores discretos en una escala del 1 al 5 de un juicio de aceptabilidad. Por el contrario, los test no paramétricos tienen más poder cuando las colas son más largas y hay más outliers.

Ejemplo:

Datos experimento sobre procesamiento de compuestos agentivos vs metafóricos. En este experimento se debe chequear que los estímulos no difieran en las variables de control. Es necesario controlar ciertas variables (frecuencia, vecinos ortográficos, longitud, etc.) para que la única diferencia entre palabras sea la estructura argumental. En el primer caso, se testear la frecuencia total de las palabras, en el segundo los vecinos ortográficos:

Sintaxis 1: ttest para frecuencia entre dos grupos

t.test(frecc1log ~ grupo, data=datospalabras, var.equal = TRUE)

[Nota: en R el ajuste de Welch se hace por default, para evitarlo se usa var.equal, para la versión pareada de t se agrega paired = T]

Output 1

Two Sample t-test

data: frecc1log by grupo

t = 1.0788, df = 58, p-value = 0.2851

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -0.1106261 0.3692481

sample estimates:

mean in group 1 mean in group 2

 0.932403 0.803092

Sintaxis 2: ttest para vecinos ortográficos entre dos grupos

t.test(vecinosc1 ~ grupo, data=datospalabras, var.equal = TRUE)

Output 2

Two Sample t-test

data: vecinosc1 by grupo

t = -2.0358, df = 58, p-value = 0.04635

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -9.91628794 -0.08371206

sample estimates:

mean in group 1 mean in group 2

 12.96667 17.96667

**1.5 Interpretar *p* y los intervalos de confianza**

El valor de *p* muestra la probabilidad de obtener un valor dado o uno más extremo, si la hipótesis nula es verdadera. El valor de *p*, dado un determinado valor de *p* disminuye a medida en que los gl aumentan.

Ej:

P(ǀtdf=50ǀ > 2): 0.0509

P(ǀtdf=10ǀ > 2): 0.0734

Un intervalo de confianza de 95% para una media significa que si repetimos el proceso de estimación sobre muchas muestras de la población, habría 95% de probabilidad de que este intervalo de confianza sea uno de los que contenga el verdadero valor del parámetro.

**2. ANOVA**

**2.1 La distribución F**

ANOVA (Analysis of Variance) es un grupo de pruebas que permiten investigar las diferencias sobre cualquier cantidad de grupos (n >2) que se definen por una o más variables categóricas. Es un tipo de regresión lineal. Históricamente ANOVA se utilizó para investigación experimental, mientras que las regresiones se usaron en investigaciones de correlaciones basadas en observaciones (ej: corpus).

Las versiones no paramétricas del test son Kruskal-Wallis o ANOVA no paramétrica basada en permutación o bootstrapping

La distribución sobre la que se basa ANOVA se denomina F y fue creada por Fisher. La distribución F es una familia de distribuciones determinadas por los grados de libertad que moldean la dispersión de la curva. Es una distribución asimétrica que comienza en cero (x=0; y=0) ya que al ser el resultado de una razón no puede ser cero ni menor de cero el valor de F. No hay valores de F negativos. En la figura, la línea de puntos corresponde a una distribución menos densa que la línea sólida, esto se debe a que la primera tiene más grados de libertad.



El valor del parámetro estadístico F (F-ratio) se define como la razón entre el promedio de la variabilidad entre los grupos y la variabilidad intragrupo.

$$F=\frac{varianza entre grupos}{varianza intra grupos}$$

En ANOVA los grados de libertad son tres valores (total, entre e intra) y se calculan a partir del valor *k* (cantidad de grupos) y *n* (cantidad de observaciones)

Grados de libertad total $df\_{1}=n-1$

Grados de libertad entre (grupo) $df\_{2}=k-1$

Grados de libertad intra (error) $df\_{3}=df\_{1}-df\_{2}$

**2.2 Tipos de ANOVA:**

1. **ANOVA de un factor (independiente):** se usa en contextos similares a los de los *t*-test independientes pero para comparación de más de dos grupos

2. **ANOVA factorial (independiente):** cuando hay dos o más variables categóricas independientes

3. **ANOVA de medidas repetidas o mixta**: en los mismos contextos en los que se usan los anteriores pero cuando las observaciones no son independientes (un sujeto es evaluado más de una vez o el estímulo es repetido). Considera la variación individual.

Existen dos tipos de diseño posibles para aplicar ANOVA: diseño entre-sujetos o diseño intra-sujetos. En el primer caso, donde dos grupos de sujetos son sometidos a diferentes condiciones experimentales, se necesita usar un ANOVA de un factor o ANOVA factorial según los tipos de variables categóricas.

En el diseño intra-sujeto, los mismos sujetos son sometidos a condiciones experimentales distintas o se evalúa el mismo ítem experimental varias veces. Si todos los sujetos/ítems son evaluados de manera homogénea en todas las condiciones experimentales, significa que todas las variables son intra-sujetos. Para este caso es necesario usar un modelo de medidas repetidas. Si el diseño experimental incluye variables intra y entre sujetos hay que usar un ANOVA de efectos mixtos.

**2.3 Los supuestos para ANOVA de un factor y ANOVA factorial**

1. Las observaciones son independientes
2. La variable de respuesta es de razón o intervalo
3. Cada muestra se distribuye normalmente y/o el *n* es igual entre las muestras
4. La varianza es homogénea entre grupos

Para el testeo de normalidad se puede usar el test de Shapiro para más de dos grupos.

La homocedasticidad, es decir la igualdad de varianza, se verifica con el test de Levene. En el test de Levene, la hipótesis nula es que los grupos tienen igual varianza. Otra prueba de homocedasticidad es la de Fligner-Killeen.

**2.4 ANOVA de un factor**

Se usa ANOVA de un factor para contrastar la hipótesis nula de que las medias procedentes de las *k* subpoblaciones (K >2) son iguales, dada una variable dependiente de naturaleza razón o intervalo (VD) y una variable independiente factor (VI) que define los grupos o subpoblaciones.

La fórmula de ANOVA representa la razón entre varianza entre los grupos y varianza dentro de los grupos. Se divide la variabilidad entre grupos por la varianza intra sujetos. Esta última varianza, también llamada error o residuales, es la variación atribuida al azar o a los sujetos.

Cuanto mayor es el promedio de la varianza entre grupos, en relación a la varianza intra grupos, mayor va a ser el valor de F.

Ejemplo:

Al igual que en el ejemplo anterior, se necesita comprobar que las medias de frecuencia y otras variables psicoligüísticas no difieran entre grupos. En este caso son tres grupos de compuestos: AG-LOC-MET

Sintaxis 3: Prueba de homogeneidad de varianzas

leveneTest(datosanovapalabras$TOTAL.log\_frq,datosanovapalabras$grupo)

Output 3

Levene's Test for Homogeneity of Variance (center = median)

 Df F value Pr(>F)

group 2 1.968 0.1472

 72

Sintaxis 4: ANOVA para frecuencia de tres grupos

> anovafrecuencia<- aov(TOTAL.log\_frq ~ grupo, data = datosanovapalabras)

> summary(anovafrecuencia)

Output 4

 Df Sum Sq Mean Sq F value Pr(>F)

grupo 2 0.0491 0.02456 2.351 0.103

Residuals 72 0.7524 0.01045

La fórmula para la suma de cuadrados de la varianza total es la sumatoria del cuadrado del valor de la variable respuesta (VD) para cada observación al que se le resta la media total de la variable respuesta.



$$SST(varianza total)=\frac{\sum\_{}^{}\left(x-x̄\right)^{2}}{n-1}$$

Sin embargo, esta variabilidad se divide en la varianza entre grupos y la varianza intra grupos. La varianza entre grupos es igual a la sumatoria del cuadrado de cada observación del grupo menos la media para la variable respuesta de ese grupo por el n de ese grupo. Se hace esto en todos los grupos y se suman.



Por último, la suma de cuadrados del error es la resta entre el SST y el SSG.

$$SSE=SST-SSG$$

Con toda esta información ya se puede calcular la media de varianza de grupos e intra grupos que nos lleva al valor final de F.

$$F=\frac{MSG}{MSE}$$

La primera media que se debe calcular es la varianza entre grupos. La varianza entre grupos es el resultado de la razón entre la suma de los cuadrados de grupo (SSG) y los grados de libertad entre (k-1).

La media de la varianza de los residuales o intra grupos es el resultado de la razón entre la suma de cuadrados de los residuales (SSE) y los grados de libertad intra (n-k)

$$F=\frac{(\frac{0.0491}{2})}{(\frac{0.7524}{72})}=\frac{0.02455}{0.01045}=2.351$$

**2.5 Pruebas post hoc**

El resultado de ANOVA nos brinda información acerca de las diferencias entre los grupos. Calcula si hay alguna diferencia significativa entre las medias, pero no explicita exactamente dónde está esa diferencia. Para poder saber dónde se encuentra el par de grupos que difiere entre sí, se debe hacer una prueba post hoc. La función de las pruebas post hoc es comparar de a pares los grupos definidos por la VI. El problema de comparar múltiples veces es aumentar la posibilidad de cometer un error de Tipo I (rechazar H0 cuando es correcta). Para evitar esto, se utilizan métodos para ajustar el nivel de significancia así se mantiene en las múltiples comparaciones el nivel de alpha elegido. Un método para ajustar el valor “honesto” de *p* es el de Tukey HSD (Honest Significant Diferences). El test no es sensible a la normalidad, pero asume varianzas homogéneas e independencia en las observaciones.

Sintaxis 5: ANOVA para frecuencias del primer constituyente

> anovafrecuenciac1<- aov(log\_frq4 ~ grupo, data = datosanovapalabras)

> summary(anovafrecuenciac1)

Output 5

 Df Sum Sq Mean Sq F value Pr(>F)

grupo 2 6.16 3.0786 3.964 0.0233 \*

Residuals 72 55.92 0.7767

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Output 6

Sintaxis 6: Post hoc para ANOVA de frecuencia c1

TukeyHSD(anovafrecuenciac1)

Output 6

Tukey multiple comparisons of means

 95% family-wise confidence level

Fit: aov(formula = log\_frq4 ~ grupo, data = datosanovapalabras)

$`grupo`

 diff lwr upr p adj

2-1 -0.1220295 -0.71855709 0.4744981 0.8764966

3-1 0.5375319 -0.05899573 1.1340595 0.0858309

3-2 0.6595614 0.06303375 1.2560890 0.0267048

**Anexo: Sintaxis de R para ttest y ANOVA.**

A continuación se presentan las sintaxis para realizar ttest y ANOVA con los siguientes paquetes

library(languageR)

library(MASS)

library(lattice)

library(multcomp)

library(car)

library(tidyverse)

library(broom)

****

Sintaxis 7: ttest y recodificación de variables

Sintaxis 8: ANOVA y post hoc

****

**Referencias**

* Johnson, K. (2007). Phonetics. En *Quantitative Methods in Linguistics* (pp. 70-103 y 104-143). Malden: Blackwell Publishing.

Levshina, N. (2015). Comparing two groups. En *How to do Linguistics with R* (pp. 87-113 y 171-198). Amsterdam/Philadelphia: John Benjamins Publishing Company.

* Winter, B. (2015). The F distribution and the basic principle behind ANOVAs. Author, Birmingham.