TABLAS ESTADÍSTICAS[1]

Para comenzar el análisis de este capítulo, voy a sugerir un repaso de conceptos introductorios que hemos venido brindando en capítulos previos.[2]

1. La Función Normal Estándar (Z)

La distribución Normal Estándar es utilizada en tests de hipótesis que incluyen contrastes sobre medias aisladas, diferencia entre dos medias, y contrastes sobre proporciones. La Normal Estándar tiene 0 como media y desvío típico de 1. Ustedes pueden ver una animación que muestra varias colas (a la izquierda) de esta distribución.[3] Más información sobre la Distribución Normal tal como es usada en el contraste estadístico puede ser hallada repasando algunos conceptos elementales.[4] También hallarán un desarrollo matemático algo más avanzado.[5]

Como se muestra en la Figura 1, los valores de la tabla representan el área por debajo de la curva normal estándar entre 0 y el número relativo z. Por ejemplo, a fin de calcular el área entre 0 y 2.36, hay que buscar la celda de intersección de la fila indicada 2.30 y la columna indicada 0.06.[6] El área resultante es .4909. Para calcular el área entre 0 y un valor negativo, fíjense en la celda de intersección de la fila y la columna que dan como suma el valor absoluto del número en cuestión. Por ejemplo, el área debajo de la curva entre -1.3 y 0 es igual al área debajo de la curva entre 1.3 y 0, de modo que hay que buscar la celda correspondiente a la fila 1.3 y la columna 0.00 (el área es 0.4032).

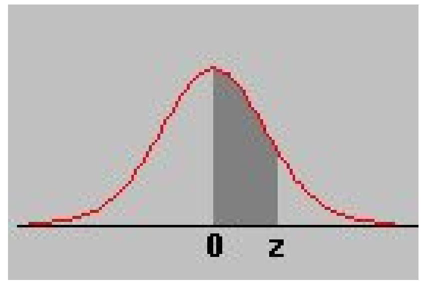

Figura 1. Área entre 0 y z

Ustedes pueden apreciar que la integral A (z) de la distribución normal desde -∞ hasta z (en otros términos, el área de la curva a la izquierda de z) proporciona la probabilidad de que una variable aleatoria normal no sea superior a z desvíos estándar por encima de su media.[7]

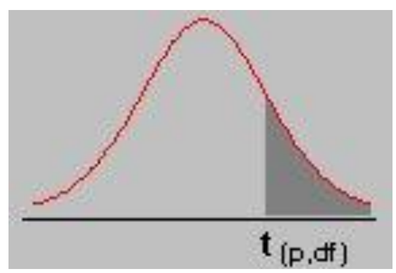

Figura 2

2. La Distribución *t* de Student

La forma de la distribución t de Student queda determinada por el número de grados de libertad. En esta animación,[8] su forma cambia al aumentar el número de grados de libertad. Para ver cómo esta distribución es usada en el contraste de hipótesis, vean el test *t* de muestras independientes[9] y el test *t* de muestras dependientes, en Estadísticos y Tablas Básicas.[10] Pueden leer también algo sobre la Distribución de la *t* de Student.[11] Como indica

---

[1] He utilizado material producido por StatSoft Electronic Statistics Textbook, StatSoft, Inc. (2011). Tulsa, OK, que puede ser bajado de internet en http://www.statsoft.com/textbook/

[2] Getting Started with Statistics Concepts http://www.statsoft.com/textbook/elementary-statistics-concepts/button/1/

[3] http://www.statsoft.com/textbook/graphics/anima2.gif

[4] http://www.statsoft.com/textbook/elementary-concepts-in-statistics/

[5] http://www.statsoft.com/textbook/statistics-glossary/n/button/n/#Normal%20Distribution

[6] http://www.statsoft.com/textbook/distribution-tables/#t

[7] http://www.statsoft.com/textbook/distribution-tables/

[8] http://www.statsoft.com/textbook/graphics/an_tdf.gif

[9] http://www.statsoft.com/textbook/basic-statistics/#t-test%20for%20independent%20samples

[10] http://www.statsoft.com/textbook/basic-statistics/

la Figura 2, el área en la parte de arriba de la tabla es el área de la cola derecha del valor-*t* que está en tabla. Para calcular el valor crítico al 0.05 de la distribución *t* con 6 grados de libertad, hay que fijarse en la columna 0.05 y en la fila 6: $t_{(.05,\ 6)}$ =1.94318.

3. <u>La distribución χ^2 o chi cuadrado</u>

Al igual que la distribución *t* de Student, la forma de la distribución chi cuadrado queda determinada por sus grados de libertad.[12] La animación muestra cómo se transforma la curva de la distribución chi cuadrado a medida que los grados de libertad aumentan (1, 2, 5, 10, 25 y 50). Para ver ejemplos de contrastes de hipótesis que usan esta distribución véase Estadísticos en tablas cruzadas[13] como así también Estimación no Lineal.[14] Como se indica en la Figura 3, los valores de esta tabla son valores críticos de la distribución chi cuadrado con los grados de libertad correspondientes. Para calcular el valor de una distribución chi cuadrado (con determinados grados de libertad) que contiene a cierta área, hay que ir a la columna del área dada y a la fila de los grados deseados de libertad. Por ejemplo, el valor crítico al 25% de una chi cuadrado con 4 grados de libertad es 5.38527. Lo cual significa que el área a la derecha de 5.38527 en una distribución chi cuadrado con 4 grados de libertad es 25%.

4. <u>Grados de Libertad</u>

En este libro el concepto de “grados de libertad” es el de un estimador del número de categorías independientes en un test particular o experimento estadístico. Se encuentran mediante la fórmula n−r, donde n=número de sujetos en la muestra y r es el número de sujetos o grupos estadísticamente dependientes.

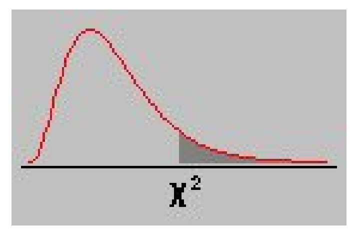

Figura 3. Área de la cola derecha de la distribución chi cuadrado

Cuando uno trata de ajustar un modelo estadístico a un conjunto de datos, los residuos -expresados en forma de vector- se encuentran habitualmente en un espacio de menor dimensión que aquél en el que se encontraban los datos originales. Los grados de libertad del error los determina, justamente, el valor de esta menor dimensión. Por ejemplo, supongan que disponemos de *n* variables aleatorias Y_1, Y_2,... Y_n, que constituyen una muestra de datos con media muestral igual a ‹Y›. La media muestral viene expresada como $\langle Y \rangle = (\sum_{i=1}^{n} Y_i)/n$. Entonces las cantidades $Y_i - \langle Y \rangle$ son los residuos, que pueden ser considerados estimadores de los errores $Y_i - \mu$ (donde μ es la media de la población). La suma de los residuos (a diferencia de la suma de los errores, que no es conocida) es necesariamente 0, pues $\sum_{i=1}^{n} (Y_i - \langle Y \rangle) = \sum_{i=1}^{n} Y_i - n \langle Y \rangle = \langle Y \rangle - \langle Y \rangle = 0$, ya que existen variables con valores superiores e inferiores a la media muestral. Esto también significa que los residuos están restringidos a encontrarse en un espacio de dimensión n − 1 (en este ejemplo, en el caso general a n − r) ya que, si se conoce el valor de n − 1 de estos residuos, la determinación del valor del residuo restante es inmediata. Así, decimos que "el error tiene n − 1 grados de libertad" (o que el error tiene n − r grados de libertad en el caso general).

5. <u>La Distribución F</u>

Usada en teoría de probabilidad y estadística, la distribución F es una distribución de probabilidad continua. También es conocida como distribución F de Snedecor o como distribución F de Fisher-

[11] http://www.statsoft.com/textbook/statistics-glossary/s.aspx?button=s#Student%27s%20t%20Distribution
[12] http://www.statsoft.com/textbook/graphics/an_chi.gif
[13] http://www.statsoft.com/textbook/basic-statistics/#Statistics%20in%20crosstabulation%20tables
[14] http://www.statsoft.com/textbook/nonlinear-estimation/

Snedecor. Esta distribución está sesgada hacia la derecha y es utilizada frecuentemente en lo que se conoce como Análisis de la Varianza (ver ANOVA/MANOVA).[15] Es un cociente de dos distribuciones chi cuadrado, y una distribución F específica se denota indicando los grados de libertad de la chi cuadrado del numerador y los grados de libertad de la chi cuadrado del denominador. Como un ejemplo de la distribución $F_{(10,\ 10)}$ se tiene el indicado en nota al pie.[16] Siempre se pone en primer término la cantidad de grados de libertad del numerador. En las cuatro tablas de la función F,[17] las filas representan los grados de libertad del denominador y las columnas los del numerador. El área de la cola a la derecha le da el nombre a la tabla (p.ej., .025). Ejemplo: para determinar el valor crítico al 5% de una distribución F con 7 y 12 grados de libertad, hay que fijarse en la columna 7 (numerador) y la fila 12 (denominador) para un alfa=0.05. $F_{(0.05,\ 7,\ 12)} = 2.9134$. Las cuatro tablas corresponden a los valores críticos .10, .05, .025 y .01. La Figura 4 representa una función F al 10% con df1 grados de libertad en el numerador y df2 grados de libertad en el denominador. El test F es utilizado para calcular la probabilidad unilateral de la eventualidad de que dos varianzas sean distintas.

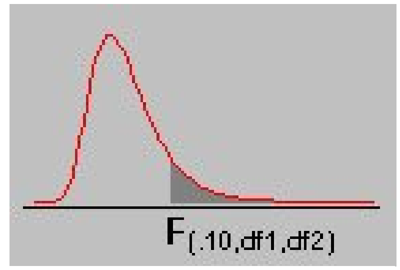

Figura 4. F. de distribución F al 10% y con grados de libertad df1 y df2

[15] http://www.statsoft.com/textbook/anova-manova/
[16] http://www.statsoft.com/textbook/graphics/newan_fdist.gif
[17] http://www.statsoft.com/textbook/distribution-tables/#t