Distribución T-Student

Los problemas de estimación de **n** suponen conocer el valor de σ o que la muestra es tan grande que se justificaba reemplazar σ por su estimación de muestras. Como ilustración de un problema en que σ es desconocida y en que no serían apropiados los métodos de grandes muestras, supóngase que un inspector desea hacer una rápida verificación del peso de los panes producidos en una panadería y toma una muestra aleatoria de 15 panes. Supóngase que el peso medio y la desviación estándar de estos 15 panes sea 15.8 y 0.3% respectivamente, y que el inspector desea un intervalo de 95% de confianza para el peso medio de toda la producción. La muestra es, de manera evidente, demasiado pequeña para proporcionar una buena estimación de σ; así que, con el fin de evitar el error involucrado en la sustitución de σ por s proviene de una muestra tan pequeña, se introduce una nueva variable llamada *variable t de student*. Se define como:

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} \qquad (1)$$

Esta variable se asemeja a la variable normal estándar, sin embargo, se difiere en que t se puede calcular a partir de datos de muestra. Por esta razón puede usarse t sin introducir aproximaciones a parámetros de población.

Si se lleva a cabo un gran número de experimentos de muestreo en los que se tomaran una muestra de tamaño **n** de una población normal y se computase el valor de t para clasificar en un atabla de frecuencia a fin de obtener una buena estimación de la distribución límite o teórica de t. La distribución exacta puede obtenerse, sin embargo, por métodos matemáticos. El caso de la distribución de t depende sólo del valor de **n**, condición de que la variable básica *x* posea una distribución normal. Además, la distribución de t es muy próxima a la distribución de una variable normal estándar z, excepto para valores muy pequeños de **n**. La Figura 1 muestra la gráfica de la distribución de t para **n** = 5 y la gráfica de una variable normal estándar z.

La Tabla 1 da valores de la variable t correspondiente a lo que se llama "grados de libertad" que se designa *ν*, y varias probabilidades. Para el problema considerado aquí, el número de grados de libertad está dado por la fórmula $\nu = n - 1$. Esto corresponde a usar el divisor *n-1* en vez de n al definir la desviación estándar de muestra.

Cualquier encabezado de columna, como 0.25, indica la probabilidad de que *t* exceda el valor de *t* dado a esa altura en la columna. Por simetría e infiere que 0.25 es también la probabilidad de que *t* se encuentre a la izquierda del valor *-t* correspondiente. Sí, para el problema anterior como $n = 15$, se lee el dato en la casilla del renglón que corresponde a 14 grados de libertad y columna encabezada por 0.25, y se encuentra $t = 2.145$. Hay entonces una probabilidad de 0.95 de que *t* satisfaga las desigualdades.

(2) $$-2.145 < t < 2.145$$

Si se aplica esto a (1), se infiere que hay una probabilidad de 0.95 de que

$$-2.145 < \frac{\bar{x} - \mu}{s}\sqrt{n} < 2.145$$

Estas desigualdades se pueden resolver para μ, el resultado es:

(3) $$\bar{x} - 2.145\frac{s}{\sqrt{n}} < \mu < \bar{x} + 2.145\frac{s}{\sqrt{n}}$$

El intervalo de 95% de confianza para μ en el problema discutido se obtiene sustituyendo los valores de muestra $n = 15$, $\bar{x} = 15.8$ y $s = 0.3$ en estas desigualdades. El resultado de esta sustitución es:

$$15.8 - 2.145\,\frac{0.3}{\sqrt{15}} < \mu < 15.8 + 2.145\,\frac{0.3}{\sqrt{15}}$$

Esto simplifica en:

$$15.63 < \mu < 15.97$$

Del análisis precedente resultara que el peso medio de los panes de esta panadería es probablemente muy poco menor que una libra.

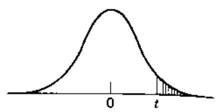
Es importante hacer la distinción entre el nuevo método de encontrar un intervalo de confianza para μ y el método precedente de muestras grandes. Este método no requiere que se aproxime σ por s, como ocurre para el método para grandes muestras y, por tanto, da una solución exacta y no aproximada al problema.

La fórmula (3) no puede usarse para encontrar un intervalo de confianza del 95% para μ a menos que la muestra sea de tamaño 15, porque la Tabla 1 se verá que (2) sólo vale para $v = 14$. Si se usa t_0 para designar el valor de t encontrado en la columna 0.025 de la Tabla 1 puesto a v grados de libertad, entonces una fórmula general para el intervalo de confianza del 95% para μ está dada por las desigualdades:

$$\bar{x} - t_0\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_0\frac{s}{\sqrt{n}}$$

Esta fórmula también vale para porcentajes diferentes del 95 si se emplea el valor correspondiente de t_0. Así, un intervalo de confianza del 90% se obtendría al reemplazar t_0 por el valor de t en la columna 0.05 de la Tabla 1 que se encuentra frente al valor de grados de libertad deseado.

En la primera columna aparece el número de grados de libertad (ν). Los encabezados de las otras columnas dan las probabilidades (P) de que t exceda el valor que aparece en la casilla. Para los valores negativos de t se usa la simetría.



ν \ P	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Tabla 1. La distribución t de Student

Como el método de muestra pequeña basado en la distribución t es un método exacto, parecería que uno debiera usarlo siempre cuando σ es desconocida. Pero la teoría que justifica a la distribución t presupone una distribución normal para la variable x; por tanto, sino se está seguro de que x está al menos de manera aproximada normalmente distribuida, la distribución t puede no ser exacta.

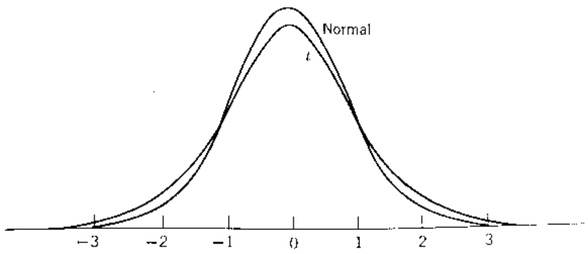


Figura 1. Distribución normal estándar y una distribución t de Student n=5

Referencias:

Mendenhall, W. (2010). Introducción a la Probabilidad y Estadística. Cengage Learning Editores.Hoel, P. G. (1984). Elementary Statistics. John Wiley & Sons.

Kelmansky, D. (2009). Estadística para todos, Estrategias de pensamiento y herramientas para la solución de problemas. Recuperado a partir de: http://www.bnm.me.gov.ar/giga1/documentos/EL001858.pdf