Recursos desarrollados en Tutoría en línea

Editores: Andrés Miniguano Trujillo, Eduardo Arias y Juan Carlos Trujillo
Proyecto CLAVEMAT
Escuela Politécnica Nacional
Email: andres.miniguano@epn.edu.ec y jcto36@gmail.com
Correcciones: Eduardo Arias (ddma1994@gmail.com)
Creación: 9 de febrero de 2017
Publicación: 7 de marzo de 2017
Pregunta: 575657

Un ejercicio completo de cálculo de probailidades con distribución normal

En este documento calcularás la probabilidad de una variable aleatoria que sigue una distribución normal.

Problema: Al analizar el rendimiento de cierta máquina que tiene distribución normal se ha observado que el 10 % de la producción media de 20 días de trabajo es mayor que 800 unidades y el 8 % de la producción media de 20 días es menor que 500 unidades. ¿Qué porcentaje de producción media de 20 días es mayor que 700 unidades?

Comienza analizando el problema, así pregúntate ¿qué mides?

Vas a medir la producción diaria en unidades y a estos puntajes serán tu variable aleatoria que seguirán una distribución normal por día: Y_i. Anotas la producción de cada día y tras pasar 20 decides anotar que ocurrió:

$$\overline{X} = \frac{1}{20}\sum_{i=1}^{20} Y_i.$$

La variable $\overline{X}$ determina la media de producción en periodos de 20 días; además, tienes los siguientes datos:

El 10 % de la producción media en un periodo es mayor a 800 unidades, esto significa que la probabilidad de que $\overline{X}$ sea mayor a 800 es de 0.1 que lo puedes escribir de la siguiente forma:

$$P\left(\overline{X} > 800\right) = 0.1.$$

Y además sabes que el 8 % de la producción media en un periodo es menor a 500 unidades, que significa que la probabilidad de que $\overline{X}$ sea menor a 500 es de 0.08 que lo puedes escribir de la siguiente forma:

$$P\left(\overline{X} < 500\right) = 0.08.$$

Si conocieras al menos un parámetro de la distribución $\mathcal{N}\left(\mu, \sigma\right)$ que siguen las variables aleatorias Y_i, podrías rápidamente decir que $\overline{X}$ sigue una distribución $\mathcal{N}\left(\mu, \sigma/\sqrt{n}\right)$, pero no es el caso. Sólo conoces que $\overline{X}$ se calcula como el promedio de $n = 20$ datos. Esta es una muestra muy pequeña; por lo tanto lo más lógico es emplear la t de Student para estimar los parámetros muestrales.

Así, primero debes recordar lo siguiente:

Si $\{y_i\}_{i \in [1..n]}$ son muestras independientes de una variable aleatoria normal Y y $\overline{x}$ es su promedio, entonces puedes estimar la media y la varianza de Y mediante la media muestral $\overline{x}$ y la varianza muestral $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(y_i - \overline{x})^2$. Si el muestreo se vuelve aleatorio, la variable aleatoria $T_\nu := \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ sigue una distribución t de Student con $\nu := n - 1$ grados de libertad.

La función de densidad de T_ν, para el caso $n = 20$ (es decir $\nu = 19$), viene dada por un número trascendental y un cociente:

$$f_{T_\nu}(t) = \frac{\Gamma\frac{\nu+1}{2}}{\sqrt{\nu\pi}\,\Gamma\frac{\nu}{2}}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \approx 2.41 \times 10^{12}\left(19 + t^2\right)^{-10},$$

El siguiente gráfico te muestra la gráfica de la función $f_{T_\nu}(t)$ y nota que es simétrica con respecto al origen:

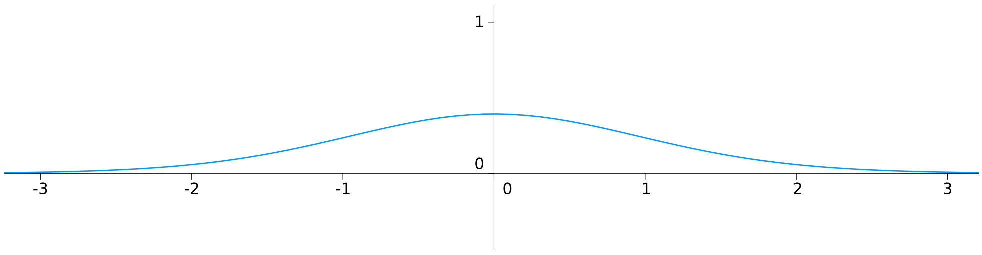

Figura 1: Gráfica de la función $f_{T_\nu}(t)$.

Ahora bien, si vas a usar la variable T para analizar la muestra de la máquina, tienes que conocer dos cosas: la media muestral y la varianza muestral. ¿Las conoces? La verdad es que no, pero las puedes encontrar a partir de los datos. Y para ello vas a formular un sistema de ecuaciones de la siguiente forma:

De la primera propiedad de $\overline{X}$, tienes que:

$$\begin{aligned}
0.1 = P\left(\overline{X} > 800\right) &= P\left(\overline{X} - \mu > 800 - \mu\right) = P\left(\frac{\overline{X} - \mu}{s} > \frac{800 - \mu}{s}\right) \\
&= P\left(\frac{\overline{X} - \mu}{s/\sqrt{20}} > \frac{800 - \mu}{s/\sqrt{20}}\right) = P\left(T_{19} > \frac{800 - \mu}{s/\sqrt{20}}\right).
\end{aligned}$$

Fíjate que en lo anterior lo único que hice fue determinar eventos equivalentes mediante operaciones invertibles. Así, obtienes que

$$P\left(T_{19} \leq \frac{800-\mu}{s/\sqrt{20}}\right) = 0.9.$$

Encontrar el valor de $\frac{800-\mu}{s}$ para que la anterior igualdad tenga sentido se lo puede determinar mediante la inversión de un polinomio que lo puedes verificar en la nota al pie[1], pero puedes usar tablas para aproximar este valor. En este documento usé el valor de $t_{0.9,19} \approx 1.328$, mientras que si usas el lenguaje R puedes obtener una mejor aproximación con $t_{0.9,19} \approx 1.327728$ (para encontrar este valor usé la instrucción qt(0.9,19,0)).

Nota: Usaré tanto la aproximación del documento como la aproximación que R me da, así podrás de paso darte cuenta cómo varía la aproximación de un número cuando truncas ligeramente su expresión decimal (sobre todo si es irracional). En lo que sigue haré los cálculos de forma paralela.

Conociendo el valor de $t_{0.9,19}$, puedes plantear la ecuación:

$$t_{0.9,19}s = \sqrt{20}(800-\mu);$$

y del mismo modo para el segundo dato inicial, donde aproximarás $t_{0.08,19} \approx 1.462$, obtienes la ecuación

$$t_{0.08,19}s = \sqrt{20}(500-\mu).$$

Aproximación de $t_{0.08,19}$:

En tus tablas seguramente no tienes un valor para $t_{0.08,19}$, sino sólo los valores de $t_{0.9,19}$ y de $t_{0.95,19}$, el cual aproximas con 1.729. ¿Cómo obtener el valor de $t_{0.05,19}$ a partir de esta información?

Primero, nota que como te mencioné anteriormente la distribución de T_ν es simétrica respecto al origen, lo cual implica que $t_{a,\nu} = -t_{1-a,\nu}$. Así, bastaría que determines una aproximación de $t_{0.92,19}$ para encontrar una de $t_{0.08,19}$.

Segundo, de tu curso de geometría, sabes que si tienes dos puntos en el espacio, entonces puedes determinar la recta que los une con ecuación $v(u) = mu + b$. De tu curso de vectorial, sabes que si desconoces más información sobre los valores de v que dependen directamente de una fórmula donde interviene u, entonces la recta $mu + b$ es la recta de mejor ajuste entre los datos.

Finalmente, dado que conoces los puntos $(0.9, t_{0.9,19})$ y $(0.95, t_{0.95,19})$ puedes plantear el siguiente sistema de ecuaciones:

$$\begin{cases} t_{0.9,19} = 0.9m + b, \\ t_{0.95,19} = 0.95m + b. \end{cases}$$

La solución está dada por $m = 20(t_{0.95,19} - t_{0.9,19})$ y $b = 19t_{0.9,19} - 18t_{0.95,19}$; tomando la aproximación decimal de arriba tienes que $m \approx 8.02$ y $b \approx -5.89$; de donde si tomas $u = 1 - 0.08$, entonces tienes que $t_{0.92,19}$ es aproximadamente 1.488.

Este valor es similar a lo que obtienes al usar en R: qt(0.92,19,0)≈ 1.462314. Es más, si usarías todos los decimales, el resultado en el mismo R viene a ser

Código 1: Cálculo del cuantíl 0.92 con 19 grados de libertad con R

```
>  (20*(qt(0.95,19,0) - qt(0.9,19,0)))*0.92 + (-18*qt(0.95,19,0) + 19*qt(0.9,19,0))

```

[1] Aquí va el link de wolfram para el cálculo del polinomio

```
[1]   1.48829
```

Para determinar los parámetro μ y s has formulado el siguiente sistema de ecuaciones lineales

$$\begin{cases} 2\sqrt{5}\mu + t_{0.9,19}s = 2\sqrt{5}800, \\ 2\sqrt{5}\mu + t_{0.08,19}s = 2\sqrt{5}500. \end{cases}$$

Donde las variables son precísamente μ y s. La solución de este sistema lineal viene dada por

$$\mu = \frac{100(5t_{0.9,19} + 8t_{0.08,19})}{t_{0.9,19} - t_{0.08,19}} \qquad \text{y} \qquad s = \frac{600\sqrt{5}}{t_{0.9,19} - t_{0.08,19}},$$

te sugiero que lo verifiques.

Calculando estos datos tienes que:

$$\begin{array}{rll} \text{con la aproximación decimal:} & \mu \approx 658.523, & s \approx 476.435 \\ \text{y con la aproximación completa:} & \mu \approx 657.2357, & s \approx 480.8676. \end{array}$$

Fíjate que las aproximaciones de tres dígitos ya se diferencian con una unidad de lo que buscas.

De paso, te cuento que usé estos códigos para la aproximación con R:

Código 2: Cálculo de los parámetros μ y s con R

```
t  <- c(qt(0.9,19,0), qt(0.08,19,0))
ta <- c(1.328, -1.488)

100*( 5*t[1] - 8*t[2] )/( t[1] - t[2] )
600*sqrt(5)/( t[1] - t[2] )
```

Como ya conoces los valores de los parámetros μ y s puedes calcular la probabilidad que buscas. En este caso, buscabas que el promedio sea mayor a 700; es decir, buscas

$$P\left(\overline{X} > 700\right) = 1 - P\left(\overline{X} \leq 700\right) = 1 - P\left(T_\nu \leq \frac{700 - \overline{x}}{s/\sqrt{n}}\right).$$

Para la aproximación decimal, tienes que

$$P\left(\overline{X} > 700\right) \approx 1 - P(T_\nu \leq 0.389) \approx 0.636.$$

La aproximación con R viene dada por

$$P\left(\overline{X} > 700\right) \approx 1 - P(T_\nu \leq 0.3977) \approx 0.638.$$

Para ello usas:

Código 3: Cálculo de la probabilidad $P\left(\overline{X} > 700\right)$ con R

```
1-dt( (700 - (100*(5*t[1] - 8*t[2])/(t[1] - t[2])) )/( (600*sqrt(5)/(t[1] - t[2]))/sqrt(20) ),19)
```

Esta vez el error de aproximación fue del 1 %, aunque si tomas el valor que la tabla te da para la probabilidad encontrarás un aproximado de 0.35067, lo cual representa un error de aproximación del 29 % y eso ya es grave.