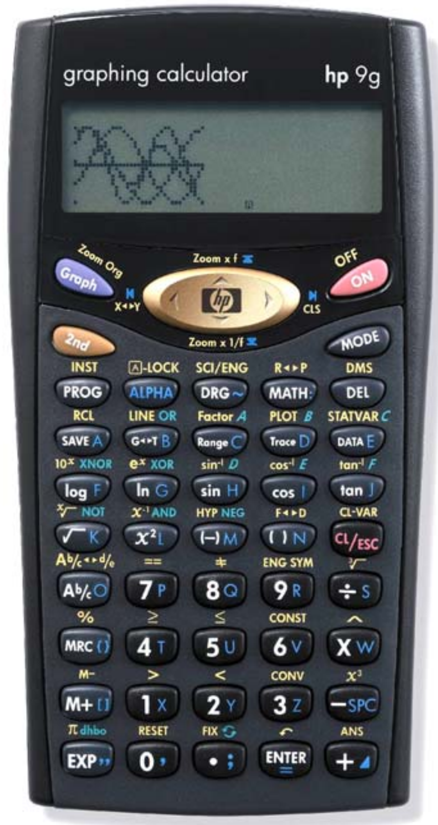
hp calculators



HP 9g Statistics – Normal Distribution

The Normal Distribution

Practice Solving Normal Distribution Problems

The normal distribution

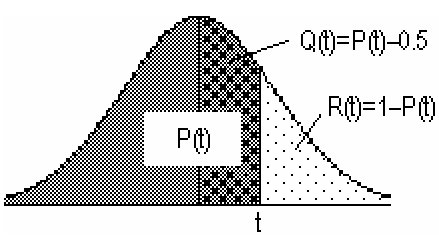
The most important continuous distribution of probability is the **normal** or **Gaussian** distribution, which arises in many practical cases. Its density function is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the arithmetic mean and σ^2 is the standard deviation (the square root of the variance). f(x) is a bell-shaped function, whose total area is 1. The **standard** normal distribution has a mean of 0 and a variance of 1, and is the form given in tables and used by some calculators (but you won't have to worry about this when using your HP 9g). This results in the familiar z value used in normal distribution problems to signify the number of standard deviations above or below the mean a particular observation falls. It is computed using this formula:

$$z = \frac{x-\mu}{\sigma}$$

where x is the observation. Z values are often called z-score or **standard score.** The HP 9g uses the letter t as the *non*-standardized random variable, and x is called a_x. The figure below shows the relationship between P, Q and R (as called by the HP 9g), i.e. the different areas of f(x). P(t) is the probability that the random variable is less than or equal to t, R(t) is the probability that it's greater than or equal to t, and Q is the probability that it's between t and the mean. Be warned that the notation used by the HP 9g may be different than that in some textbooks.



These functions are displayed in the STATVAR menu (2nd STATVAR ^) provided the current operating mode is 1-VAR (i.e. press MODE 1 X select 1-VAR and press ENTER). In addition to the value of t (which is calculated using the user-supplied a_x), P, Q and R also need to know the mean and standard deviation, which are calculated from the data keyed in by the user (at least two points with their frequencies have to be entered). The mean and standard deviation cannot be input by the user, they have to be calculated! This certainly seems a severe limitation, but we'll learn a way to work around it.

Practice solving normal distribution problems

<u>Example 1:</u> The scores on a final exam approximate a normal curve with a mean of 71 and standard deviation of 11. What percentage of the students scored between 70 and 89?

<u>Solution:</u> Let's first clear any previous statistical data in the calculator by pressing: MODE 1 X , then selecting D-CL and finally pressing ENTER . Normal distributions are calculated in the 1-VAR mode, so press: MODE 1 X , select 1-VAR and press enter ENTER .

No data is provided, other than μ and σ, which are the only parameters necessary to solve the problem anyway. On the HP 9g, we can't *enter* μ and σ directly, they have to be the mean and the standard deviation of the *entered data*. In other words, if there's no data, there's no μ or σ (nor P, Q, R !). Fortunately, there's a simple workaround:

Create a dataset consisting of only two points, namely $\mu + \sigma$ and $\mu - \sigma$, the frequencies of which are 1.

The mean and standard deviation that the HP 9g will calculate for these two points are exactly μ and σ !

Now, let's solve the example. Create the dataset as follows:

[DATA E] , select DATA-INPUT, [ENTER] [7 P] [1 X] [+] [1 X] [1 X] ▼ ▼ [7 P] [1 X] [– SPC] [1 X] [1 X] ▼

You can press [2nd] [STATVAR C] to verify that μ (i.e $\bar{x}$) = 71 and σ = 11 (they are displayed as $\bar{x}$ and σx, respectively). The percentage of the students who scored between 70 and 89 will be the difference between the probability that a student chosen at random obtained a score greater than 70 and the probability that s/he obtained a score greater than 89. That is to say: R(70) – R(89).

To enter the argument for the cumulative fraction functions P(t), Q(t) and R(t), press:

[DATA E] , select DISTR, [ENTER] and then key in the value (a_x) and press [ENTER] .

The P, Q and R functions, as well as the standard score t, are displayed in the STATVAR menu (press [2nd] [STATVAR C] ▲). To display their value, just select the desired function. The value appears in the result line (up to four decimal digits). The names of the functions can also be put into the entry line for further calculations, but keep in mind that there's no history stack in STAT mode!

Let's now find R(89):

[DATA E] ▲ [ENTER] [8 Q] [9 R] [ENTER] [2nd] [STATVAR C] ▲

R(89) = 0.0509. Let's store it in the variable A:

[SAVE A] [SAVE A] [ENTER]

To calculate R(70) – R(89) press:

[DATA E] ▲ [ENTER] [8 Q] [9 R] [ENTER] [2nd] [STATVAR C] ▲ [– SPC] [ALPHA] [SAVE A] [ENTER]

which returns 0.4853

Answer: About 49% of the students scored between 70 and 89. This can also be calculated using the P or the Q functions since: R(70) – R(89) = P(89) – P(70) = Q(70) + Q(89).

Example 2: Plot the normal curve corresponding to the previous example.

Solution: This is as easy as pressing the keys [2nd] [STATVAR C] [Graph] and [0 ,] . Ranges are automatically set to optimum values.

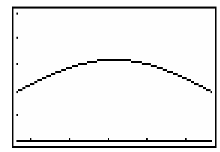


Figure 1

<u>Answer:</u> Figure 1 shows the resulting graph. Try changing the values of μ and σ (as described above) so as to see that they affect the position and the dispersion of the curve, respectively.

<u>Example 3:</u> If t ~ N(7,2) what is the probability that $5 < t \leq 10$?

<u>Solution:</u> The expression t ~N(μ,σ) means that t is a normally distributed random variable having a mean of μ and a standard deviation of σ. (μ=7 and σ=2 in this example). It can be proved that the variable: $z = \frac{t-\mu}{\sigma}$ ~ N(0,1). That is to say, z has a mean of 0 and a standard deviation of 1. Note that the probability requested can be written as:

$$P(5 < t \leq 10) = P\left(\frac{5-\mu}{\sigma} < \frac{t-\mu}{\sigma} \leq \frac{10-\mu}{\sigma}\right) = P\left(\frac{5-7}{2} < \frac{t-7}{2} \leq \frac{10-7}{2}\right) = P(-1 < z \leq 1.5)$$

It is P(z) that is given in tables, but we really don't need to standardize t before calculating the probability because the HP 9g can provide the P(t) values directly. Let's verify that $P(5 < t \leq 10) = P(-1 < z \leq 1.5)$. Remember that $P(a < x \leq b)$ is given by P(b) – P(a). First "enter" μ=7 and σ=2 by pressing:

DATA E ENTER 7 P + 2 Y ⌄ ⌄ 7 P –SPC 2 Y ⌄

To calculate P(10) – P(5) press:

DATA E DATA E ⌃ ENTER 5 U ENTER 2nd STATVAR C ⌃ ⌃ SAVE A SAVE A ENTER DATA E ⌃ ENTER 1 X 0 , ENTER 2nd STATVAR C ⌃ ⌃ –SPC ALPHA SAVE A ENTER

Jot down the result, and change μ and σ pressing:

DATA E ENTER 0 , + 1 X ⌄ ⌄ 0 , –SPC 1 X ⌄

We can now calculate $P(-1 < z \leq 1.5)$ as P(1.5) – P(–1)[1]

DATA E DATA E ⌃ ENTER (–) M 1 X ENTER 2nd STATVAR C ⌃ ⌃ SAVE A SAVE A ENTER DATA E ⌃ ENTER 1 X • ; 5 U ENTER 2nd STATVAR C ⌃ ⌃ –SPC ALPHA SAVE A ENTER

<u>Answer:</u> Both methods give the same result, 0.7745.

<u>Example 4:</u> Power lines can break down when voltage exceeds the capacity of the line. Calculate the probability of a power cut if voltage is N(100, 16) and capacity is N(140, 12).

<u>Solution:</u> If x and y are two *independent* random variables such as $x \sim N(\mu_1,\sigma_1)$ and $y \sim N(\mu_2,\sigma_2)$, then the random variables x+ y and x – y are distributed as follows:

[1] In fact, negative values of P are not tabulated because P(–z) = 1 – P(z).

$$x + y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$
$$x - y \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

Actually, x and y don't need to be normally distributed, the only condition is that they must be statistically independent of each other. Let's see how all this can be applied to our example:

We have two independent variables, namely the voltage, V ~ N(100, 16), and the capacity, C ~ N(140, 12). There will be a breakdown when V > C, that is to say: when V – C > 0. Therefore:

$$V - C \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}) = N(100 - 140, \sqrt{16^2 + 12^2}) = N(-40,20)$$

What we must find is the probability that the variable V – C is greater than 0, that is to say: R(0), which we can calculate on the HP 9g as follows. First, let's "enter" the new mean and standard deviation by pressing:

[DATA E] [ENTER] [(−) M] [4 T] [0 ,] [+] [2 Y] [0 ,] ▼ ▼ [(−) M] [4 T] [0 ,] [− SPC] [2 Y] [0 ,] ▼

Pr(V – C >0) can now be calculated easily:

[DATA E] [DATA E] ▲ [ENTER] [0 ,] [ENTER] [2nd] STATVAR C ▲

Answer: The probability of a blackout is 2.28%.