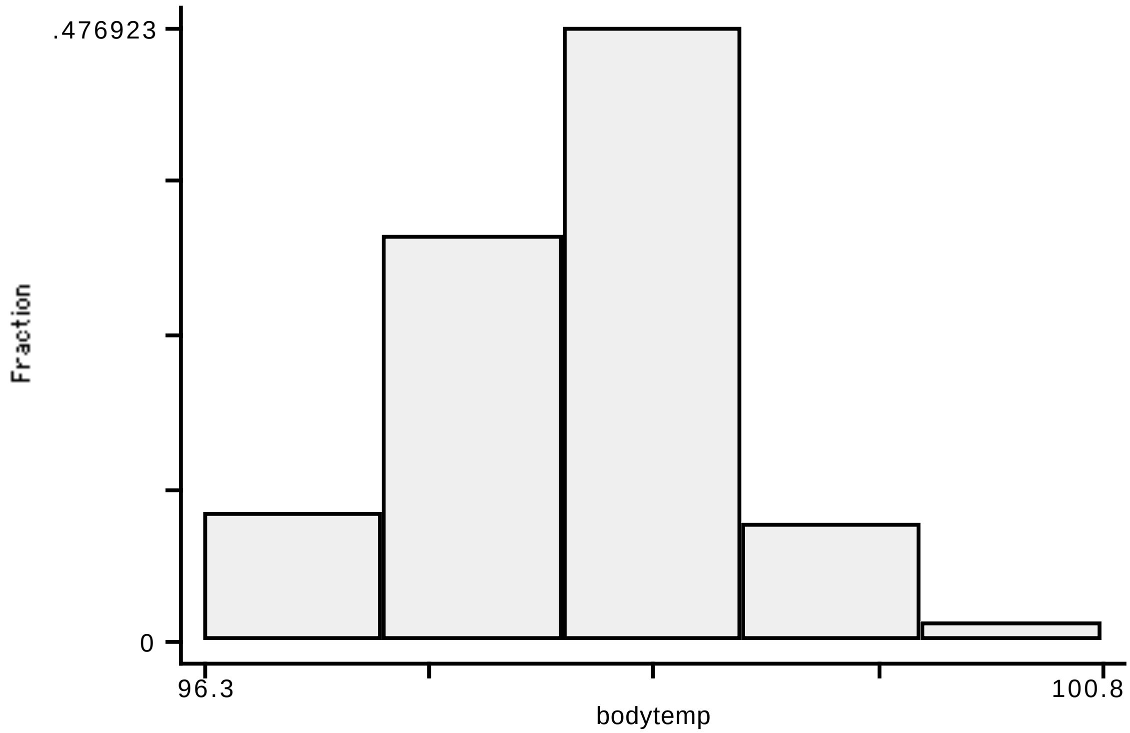
Body Temperature: What is "Normal"?

A study of 130 healthy men and women found these body temperatures:

bodytemp	
1.	96.3
2.	96.7
3.	96.9
4.	97
5.	97.1
6.	97.1
7.	97.1
8.	97.2
9.	97.3
10.	97.4
11.	97.4
12.	97.4
13.	97.4
14.	97.5
15.	97.5
16.	97.6
17.	97.6
18.	97.6
19.	97.7
20.	97.8
21.	97.8
22.	97.8
23.	97.8
24.	97.9
25.	97.9
26.	98
27.	98
28.	98
29.	98
30.	98
31.	98
32.	98.1
33.	98.1
34.	98.2
35.	98.2
36.	98.2
37.	98.2
38.	98.3
39.	98.3
40.	98.4
41.	98.4

42.	98.4
43.	98.4
44.	98.5
45.	98.5
46.	98.6
47.	98.6
48.	98.6
49.	98.6
50.	98.6
51.	98.6
52.	98.7
53.	98.7
54.	98.8
55.	98.8
56.	98.8
57.	98.9
58.	99
59.	99
60.	99
61.	99.1
62.	99.2
63.	99.3
64.	99.4
65.	99.5
66.	96.4
67.	96.7
68.	96.8
69.	97.2
70.	97.2
71.	97.4
72.	97.6
73.	97.7
74.	97.7
75.	97.8
76.	97.8
77.	97.8
78.	97.9
79.	97.9
80.	97.9
81.	98
82.	98
83.	98
84.	98
85.	98
86.	98.1
87.	98.2

88.	98.2
89.	98.2
90.	98.2
91.	98.2
92.	98.2
93.	98.3
94.	98.3
95.	98.3
96.	98.4
97.	98.4
98.	98.4
99.	98.4
100.	98.4
101.	98.5
102.	98.6
103.	98.6
104.	98.6
105.	98.6
106.	98.7
107.	98.7
108.	98.7
109.	98.7
110.	98.7
111.	98.7
112.	98.8
113.	98.8
114.	98.8
115.	98.8
116.	98.8
117.	98.8
118.	98.8
119.	98.9
120.	99
121.	99
122.	99.1
123.	99.1
124.	99.2
125.	99.2
126.	99.3
127.	99.4
128.	99.9
129.	100
130.	100.8



First, you'll note that there's a striking range of temperatures that are "normal". Second, the traditional gold-standard for "normal", 98.6 degrees fahrenheit turns out to be a little high for this group:

Variable	Obs	Mean	Std. Dev.	Min	Max
bodytemp	130	98.24923	.7331833	96.3	100.8

Is this average low because the "true" mean body temp. is lower than 98.6, or is this because of random fluctuations? (In other words, perhaps another group of 130 people would be closer to 98.6.) In fact, this data set was collected in part because the authors felt that 98.6 was too high for a healthy population.

We're setting ourselves up for an hypothesis test here:

Ho: Mean Body Temp = 98.6
Ha: Mean Body Temp < 98.6

How can we test it? The obvious choice seems to be to use the sample average as a test statistic. If the sample average is too small, we reject the null hypothesis. How small is too small?

Sample averages that come from a random sample are themselves random. So even if the null hypothesis is true, we'll see an average lower than 98.6 about half the time. But it should be rare that we see a sample average too far below 98.6. (If, indeed, 98.6 is the right mean.) For example, if body temperatures are normal, then about 68% of the body temperatures will be within 1 SD of the mean. And this means that the sample average, which we'll call Xbar, will be within 1 SD/sqrt(n) of the mean. (Because the SD of Xbar is SD/sqrt(n).)

So in this study, we observed xbar = 98.2. What we need is the p-value: Assuming that the null hypothesis is true and the mean really is 98.6, find

P(Xbar < observed value for Xbar.)

If this is a small number, then the outcome is unusually small and we should be suspicious of the null hypothesis. Otherwise, there is no grounds for suspicion. (Which is different from saying the null hypothesis is true. We simply have no reason to disbelieve it.)

To find this probability, we need to know the sampling distribution (the pdf) of Xbar. This is not too hard. If X1, X2, ..., Xn are independent observations from a normal distribution with mean mu and SD sigma, then Xbar is normal with mean mu and SD sigma/sqrt(n). And therefore (Xbar - mu)/(sigma/sqrt(n)) is N(0,1). So we can just do P(Xbar < 98.2) = P(Z < (98.2 - 98.6)/sigma(sqrt(130)).

Even if the X's aren't normal, this approach still works. The central limit theorem tells us that if n is sufficiently large, Xbar will be *approximately* normal. So the same approach gives us a p-value that is approximately correct. And for most problems, if n >= 30, this approximation is quite good.

But there's a problem here; we don't know the SD sigma. There's nothing in the theory that tells us what this should be. Our solution is the standard one. Let's estimate it with s, the sample standard deviation we learned in the second or third week of the course.

The output above gives the value for this. It tells us that for these 130 observations, s = 0.733. Of course, if you took another random sample from the same population, you would get a different value. In fact, s is NOT equal to sigma. Sigma represents the SD of the population, and s is the SD of a random sample from that population. So it is, in fact a random number because it varies with each random sample.

This means, unfortunately, that (Xbar - 98.2)/s/sqrt(n) is NOT a normal random variable. Why? Because it has two random variables in it. Xbar and s.

Remember the Guiness Beer story? This is where it comes into play. A statistician, named Gosset, was hired as a consultant to Guiness Beer,to worry over this very same problem. He discovered that the pdf for this statistic was something he called the t-distribution. He published this result under the pseudonym "Student" so that Guiness's competitors wouldn't catch on that they were on to something. And so it became known as "Student's t-test."

The t-distribution looks very much like the normal distribution except that

- it has "thicker" tails and
- it has just one parameter, called the degrees of freedom.
- Like the normal distribution , its symmetric.

To calculate the degrees of freedom, just take n-1, where n is the sample size.
The larger the degrees of freedom, the more the distribution looks like a normal distribution. In fact, for degrees of freedom >= 30, they are nearly indistinguishable.

Your book provides a table for the t-distribution up to 29 degrees of freedom. After that, you just use the last line of the table, labeled with the "infinity" symbol. (See Appendix F.)

Let's say, for example, you choose a significance level of alpha = 0.05. Our calculation for the p-value looks like this:
P(Xbar < 98.2) =
P(T < (98.2 - 98.6)/.733/(sqrt(130)) = P(T < - 6.22).
Note: the SD of Xbar is .733/sqrt(130) = .0643.

Now the table only gives us "right-hand" probabilities. But because the t-distribution is symmetric
P(T < -6.22) = P(T > 6.22).
From the table, we can see that we would reject the null hypothesis if our t-statistic was bigger than 1.64. Ours is much bigger (off the chart, in fact), and so we easily reject the null hypothesis.