How To ....
A Practical Guide to Psychometrics

Base Rates

Contents

If viewing on screen you can click on a contents item above to jump to the page the item is on

See also Some Problems of Using Tests for Diagnostic Classification

How to 1. What is a base rate?

A base rate tells you the proportion of people amongst those you test who have the disease or abnormality that you are interested in diagnosing. This base rate will vary from setting to setting, and different institutions and practitioners are quite likely to have different base rates amongst their clientele.

The base rate, which is most directly relevant to you, is the base rate for the target condition (schizophrenia, dementia, PTSD etc) amongst the people you assess for the target condition.

How to 2. How to use base rate information.

Without knowledge of the relevant base rates your conclusions after assessing somebody will be limited to statements such as; "This person obtained higher score than 'x' percent of the standardisation control group, and 'y' percent of the abnormal group obtained a more normal score than this person.

Let's make this concrete. Suppose we have a test intended to diagnose dementia. The normative data look like this.

Group	Above cut-off score	Below cut off score
Demented	60 (True Positive)	40 (False Negative)
Not demented	10 (False positive)	90 (True Negative)

The person we have assessed scores above the cut-off. Thus their score is higher than that of 90 percent of non-demented people, and 40 percent of people with dementia get a 'better' score than this person.

BUT, given this score, what is the probability that this patient is demented/

It will all depend on the base rate. The probability that the person assessed has the target condition is very heavily dependent on the base rate.

For example, lets us suppose that we were to use our dementia test on a group of lively, physically active elderly people who were enthusiastic member of a U3A group amongst whom the base rate for dementia was 5 percent.

Bayes Theorem can be used to work out the probability that somebody in this population who gets a dementia score is in fact demented.

This theorem states:

$$pD/T = (pD \times pT/D) / ((pD \times pT/D) + (pN \times pT/N))$$

Where:

> In general X/Y means the probability of someone having condition 'Y' given that we know that they have characteristic 'X'. More specifically in this case:

> D is the condition

> N is the absence of the condition

> T is a test result indicating the presence of the condition

> pD is the base-rate for the condition

> pN equals (1 - pD)

> pT/D is the true positive rate (the probability of getting the test score given that the target condition is present)

> pT/N is the false positive rate (the probability of getting the test score, given that the target condition is not present)

In our example, a proportion of 0.60 of those with dementia get a dementia score, so pT/D is .60.

And a proportion of .10 of those without dementia also get a dementia score, so. pT/N is .10

And in our example we have decided that 5 percent of this particular population actually have dementia – a proportion of .05 percent, which is the value of pD.

Placing these values into Bayes theorem we get:

> $pD/T = (.05 \times .60) / ((.05 \times .60) + (.95 \times .10)) = .03/.125 = 0.24$

So a dementia score amongst this group means that there is only about one chance in four that the person is demented.

Suppose instead that the group tested were patients referred specialist neurologists or psychiatrists who suspected the presence of dementia. And suppose that 70 percent of patients so referred did in fact suffer from dementia. What now would be the probability that an abnormal score on our test would indicate dementia?

Placing the appropriate values into Bayes theorem we get"

$$p\text{D/T} = (.70 \text{ x } .60) / ((.70 \text{ x } .60) + (.30 \text{ x } .10)) = .42/.45 = 0.93$$

This time there is a very high probability that the person suffers from dementia.

How to 3. Estimate base rates

A major problem has always been that there are very few situations in which we actually know the relevant base rate.

However base rates can be fairly easily estimated form the proportion of people to whom a test is administered who get abnormal scores.

You can estimate the number of people getting abnormal scores on a test that you use by either:

1. Going through your case files and taking all, or some random sample, of those to in which the particular test has been administered for the problem of detecting the target condition whose base rate you wish to estimate.

2. By administering the test to a new sample consisting of all those new referrals to whom the test is administered to detect the target problem.

In either case you will end up with an estimate of those who obtained an abnormal score on the test. From knowing what proportion of your clients/patients obtain an abnormal score, you can easily work out the base rate for the target condition amongst YOUR patients.

The formula for doing this is:

$$BR = (AS - FP) / (TP - FP)$$

Where:

BR = base rate

AS = proportion obtaining abnormal score

FP = false positive rate (as proportion)

TP - true positive rate (as proportion)

So, suppose we have a test for detecting schizophrenia, and, having searched through our case records we find that the proportion of those to whom we administered the test who obtained a score indicating schizophrenia was .60.

Further, suppose that the norms for the test were like this:

	Above test cut-off	**Below test cut-off**
Schizophrenia present	80 (True positives)	20 (False negatives)
Schizophrenia absent	25 (False positives)	75 (True negatives)

Filling in the appropriate values in the formula above we have:

$$BR = (.60 - .25) / (.80 - .25) = .35/.55 = .64$$

So, the proportion of people to whom we administer the test who suffer from schizophrenia is .64.

Given this information, and using Bayes theorem, we can estimate that the probability that someone obtaining an abnormal score is suffering from schizophrenia is going to be:

$$p\text{D/T} = (p\text{D x } p\text{T/D}) / ((p\text{D x } p\text{T/D}) + (p\text{N x } p\text{T/N}))$$

$$= (.64 \text{ x } .80)/((.64 \text{ x}.80)+(.36 \text{ x } .25)) = .51/.60 = .85$$

So if the test says that someone is suffering from schizophrenia, there is a probability of 0.85 that they early are.