1. *In the genetic screening example, calculate the (unconditional) probability of testing positive, and of having the genetic variant.*

We have: $P$[variant] = 0.01
$P$[no variant] = 0.99
$P$[test +ve| variant] = 0.8 (80% sensitivity)
$P$[test -ve| variant] = 0.2
$P$[test –ve|no variant] = 0.95 (95% specificity)
$P$[test +ve|no variant] = 0.05

So:

$P$[test +ve] = $P$[test+ve| variant]\*$ P$[variant] + $P$[test+ve|no variant]\*$ P$[no variant]
= 0.8\*0.01 + 0.05\*0.99 = 0.0575, i.e. 5.75%

$P$[have variant] = 0.01, i.e. 1%

*Compare the probability of having the genetic variant given a positive test to the unconditional probability of having the genetic variant: did getting a positive test result change much?*
$P$[have variant|+ve result] is 14%, up by a factor of 14 from $P$[have variant]=1%. So the amount you should be worried based on a positive test results is a *lot* more than you would be with no test at all. But it’s still a (fairly) small amount.

1. *In the ASE example, verify that when Y=0 (i.e. no expression observed in BY) the classical analysis give a point estimate =0 and estimated standard error=0. Why are both these values implausible?*

The point estimates is the sample mean, $\overbar{Y}$, which is zero. The standard error is $\frac{\sqrt{\overbar{Y}\left(1-\overbar{Y}\right)}}{√n}$, which is again zero. These are silly values! The first would imply that Brewers’ Yeast can *never* produce expression values at some particular location, i.e. ruling it out – but based on a small sample size and not using any other information source. The zero standard error indicates we would *never* expect this estimate to change, regardless of how many times we ran the experiment. Expressing complete confidence in a somewhat extreme situation, based on this data alone, would not be creditable.

1. *Human twins can be either monozygotic (from a single egg) or dizygotic (from two eggs). Among all sets of twins, approximately one third are monozygotic, and the ratio of males to females can be taken to be 50:50, regardless of monozygosity.

Suppose it is known that someone will have twins, e.g., from detection of two heartbeats, and a sonogram indicates there are twin girls. What is the probability that they are monozygotic?

Hint: make a Venn diagram for events G: the twins are both girls and M: the twins are monozygotic. We want Pr[M|G], which is related to Pr[G|M].*

Writing out all the algebra:

$P$[G|M] = 1/2
$P$[G] = $P$ [G|M]\*$P$[M] + $P$[G|not M]\*$P$[not M] = 1/2\*1/3 + 1/4\*2/3 =1/3
$P$[M] = 1/3
$P$[M|G] = $P$[G|M]\*$ P$[M]/$ P$[G] = 1/2 \* 1/3 / 1/3 = 1/2

So belief that there are identical twins should go up from 1/3 to 1/2 based on the sonogram. (Or use [cell-free DNA testing](https://www.natera.com/press-releases/publication-demonstrates-unique-ability-panorama-test-determine-zygosity-twin) and find out for sure!)

As a Venn diagram: (there are lots of ways to lay this out: the total area of the diagram is 1)



Or if you prefer to work with the sexes of both babies:



# Code

par(mar=c(0,0,0,0)+0.1)

plot(0,0,type="n", xlab="", ylab="", axes=FALSE, xlim=0:1, ylim=0:1, asp=1)

box()

rect(0.5-sqrt(1/3)/2, 0.5-3\*sqrt(1/3)/4, 0.5+sqrt(1/3)/2, 0.5+sqrt(1/3)/4,

col=adjustcolor("green",alpha=0.3), border="green", lwd=2)

rect(0.5-sqrt(1/3)/2, 0.5-sqrt(1/3)/4, 0.5+sqrt(1/3)/2,0.5+3\*sqrt(1/3)/4,

col=adjustcolor("goldenrod",alpha=0.3), border="goldenrod", lwd=2)

text( x=0.5, y=0.5+c(1,0,-1)\*sqrt(1/3)/2,

c("Pr[G and not M]=1/6","Pr[G and M]=1/6", "Pr[not G and M]=1/6"))

text(x=0.5 + sqrt(1/3)/2, y=0.5 + sqrt(1/3)/4, "Pr[G]=1/3", col="goldenrod", pos=4)

text(x=0.5 + sqrt(1/3)/2, y=0.5 - sqrt(1/3)/4, "Pr[M]=1/3", col="forestgreen", pos=4)

par(mar=c(0,0,0,0)+0.1)

plot(0,0,type="n", xlab="", ylab="", axes=FALSE, xlim=0:1, ylim=0:1, asp=1)

rect(rep(0,4), seq(0,0.75, 0.25), rep(2/3,4), seq(0.25, 1 ,0.25))

text(rep(1/3,4), seq(1/8,7/8,1/4), c("BB, not M: 1/4 \* 2/3 = 1/6", "BG, not M: 1/4 \* 2/3 = 1/6", "GB, not M: 1/4 \* 2/3 = 1/6", "GG, not M: 1/4 \* 2/3 = 1/6"))

rect(rep(2/3,2), c(0,0.5), rep(1,4), c(0.5, 1))

text(rep(5/6,4), c(0.25, 0.75), c("BB, M:\n1/2 \* 1/3 = 1/6", "GG, M:\n1/2 \* 1/3 = 1/6"))

rect(2/3, 0, 1, 1, , col=adjustcolor("green",alpha=0.3), border="green", lwd=2)

polygon(x=c(0,2/3,2/3,1,1,0), y=c(3/4,3/4,1/2,1/2,1,1),

col=adjustcolor("goldenrod",alpha=0.3), border="goldenrod", lwd=2)