

# In Class Assessment 3

Meetakshi Setiya

23/02/2022

*Note1 - Please enter the efficacy as a decimal fraction: like 85% efficacy becomes 0.85*

*Note2- I have assumed that once the antibiotic kills the bacteria, the patient is recovered. Thus these two statements are used interchangeably in the text below.*

*Note3- I have used seeds everywhere randomization was required to ensure consistent and replicable results everytime the code is run.*

**I have modeled the 100 patients in each sample as 100 randomly generated Bernoulli trials (0 or 1) with probability of 1 i.e. killing of bacteria taken as the efficacy of the antibiotic. Thus, 0 means that a patient has not recovered on administering the antibiotic while 1 means that a patient is recovered after taking the antibiotic (bacteria killed).**

The helper function below takes some probability p and seed s and returns 100 randomly generated 0s or 1s given by probability of occurrence of 1 (bacteria being killed) = p. The function `rbinom()` was used to model the Bernoulli distribution since a Binomial random variable with 2 outcomes is Bernoulli Random Variable.

```
Generate_Patients <- function(p, s)
{
  #generate 100 random bernoulli trials with the given probability(efficacy) and seed
  #0-patient i has not recovered, 1-patient i has recovered

  set.seed(s)
  smp <- rbinom(100, 1, p)
  sample <- data.frame(smp);
  return (sample)
}
```

The helper function below takes the number of samples n (here, 20) and the efficacy p as inputs, calls the function `Generate_Patients()` n times and returns a data frame with with separate n columns for each simulated sample of 100 patients.

```
Simulation_Study <- function(n, p)
{
  #return a data frame with the results of 100 (rows) simulated patients as a set of
  #binary values generated with probability p (efficacy), repeated n (columns) times

  records = data.frame()
  for (i in 1:n)
  {
    seedval = i+as.integer(p)
    samples <- Generate_Patients(p, seedval)
  }
}
```

```

nm <- paste(c("sample", i), collapse = "")
if(i==1)
  records <- rbind(records, samples)
else
  records <- cbind(records, samples)
colnames(records)[i] <- nm
}
return (records)
}

```

The function below takes the efficacy and number of samples `n_repeat` (here, 20) as inputs, calls the function `Simulation_Study()` to get randomly generated data samples that model if the patient recovered from the antibiotic as explained before. Now, each sample out of the `n_repeat` (here, 20) is checked to see the number of recovered patients. If they are at least 88, the new antibiotic is statistically better for that sample. The percentage of the number of such samples i.e. the statistical power is returned.

```

Statistical_Power <- function(efficacy, n_repeat)
{
  records <- Simulation_Study(n_repeat, efficacy)
  nm <- colnames(records)
  passed_samples = 0
  for (i in 1:n_repeat)
  {
    s <- sum(records[i])
    if(s>=88)
    {
      passed_samples <- passed_samples+1
    }
  }

  percentage = (passed_samples/n_repeat)*100
  return (percentage)
}

```

## Question 1

The code below uses the function `Statistical_Power` and prints the true efficacy and the calculated statistical power of the antibiotic.

```

Question_1 <- function()
{
  p = Statistical_Power(0.85, 20)
  print(sprintf("True efficacy of the antibiotic: %f", 0.85))
  print(sprintf("Statistical Power (in percent): %f", p))
}

Question_1()

```

```

## [1] "True efficacy of the antibiotic: 0.850000"
## [1] "Statistical Power (in percent): 25.000000"

```

## Question 2

The code below uses the function `Statistical_Power()` and prints the true efficacy and the calculated statistical power of the antibiotic for each (a), (b) and (c).

```
Question_2 <- function()
{
  p1 = Statistical_Power(0.80, 20)
  p2 = Statistical_Power(0.90, 20)
  p3 = Statistical_Power(0.95, 20)
  print("(a):")
  print(sprintf("True efficacy of the antibiotic: %f", 0.80))
  print(sprintf("Statistical Power (in percent): %f", p1))
  print("(b):")
  print(sprintf("True efficacy of the antibiotic: %f", 0.90))
  print(sprintf("Statistical Power (in percent): %f", p2))
  print("(c):")
  print(sprintf("True efficacy of the antibiotic: %f", 0.95))
  print(sprintf("Statistical Power (in percent): %f", p3))
}
```

```
Question_2()
```

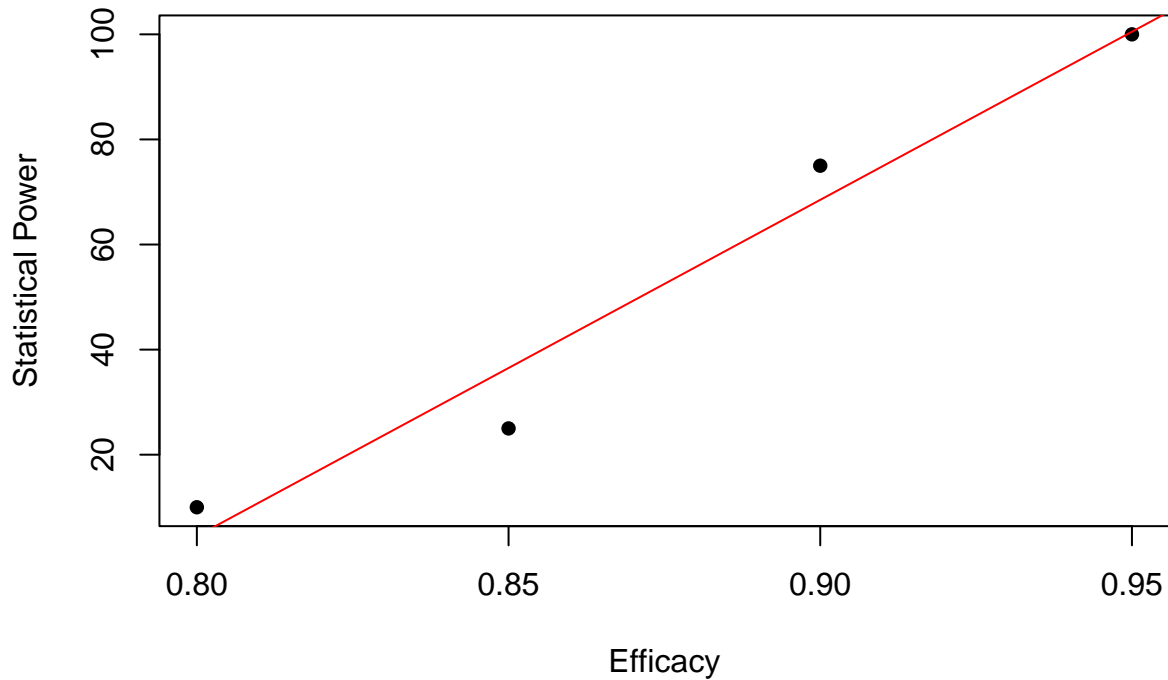
```
## [1] "(a):"
## [1] "True efficacy of the antibiotic: 0.800000"
## [1] "Statistical Power (in percent): 10.000000"
## [1] "(b):"
## [1] "True efficacy of the antibiotic: 0.900000"
## [1] "Statistical Power (in percent): 75.000000"
## [1] "(c):"
## [1] "True efficacy of the antibiotic: 0.950000"
## [1] "Statistical Power (in percent): 100.000000"
```

## Question 3

The code below plots the statistical power (y axis) vs true efficacy (x axis) for the antibiotic as a scatter plot at efficacies: 0.80, 0.85, 0.90 and 0.95. The output is a scatter plot along with a regression line.

```
Question_3 <- function()
{
  p1 = Statistical_Power(0.80, 20)
  p2 = Statistical_Power(0.85, 20)
  p3 = Statistical_Power(0.90, 20)
  p4 = Statistical_Power(0.95, 20)
  ind <- c(0.80, 0.85, 0.90, 0.95)
  dep <- c(p1, p2, p3, p4)
  plot(x=ind, y=dep, xlab="Efficacy", ylab="Statistical Power", pch=16)
  abline(lm(dep ~ ind), col="red")
}
```

```
Question_3()
```



**Do you think 100 patients is a sufficiently large sample to discover whether the new drug is “significantly better” if the true efficacy of the drug is 90%?**

No. In a real life scenario, the given sample size of 100 patients is not large enough. Smaller sample sizes are more prone to variability and have a risk of failing on a more generalised population. Thus, if the new drug’s significance over the standard antibiotic has to be measured, it better been done on a larger sample size. To find an optimal sample size, let us assume Type 1 error significance level to be 0.05 and power level to be 0.9, the standardised difference (S.D.) of the two drugs can be calculated from their efficacies as follows:

$$S.D. = \frac{p1-p2}{(P*(1-P))^{\frac{1}{2}}}$$

$$\text{where, } P = \frac{p1+p2}{2}$$

for  $p1 = 0.90$ ,  $p2=0.80$ , (efficacies of the new and standard antibiotics respectively)

$$P = \frac{0.90+0.80}{2} \quad P = 0.85$$

Now,

$$S.D. = \frac{0.90-0.80}{(0.85*(1-0.85))^{\frac{1}{2}}}$$

$$S.D. = \frac{0.10}{(0.85*0.15)^{\frac{1}{2}}}$$

$$S.D. = 0.28$$

Using the Altman nomogram, for  $SD=0.28$  on the left axis and drawing a line to the power value 0.90 on the right, we see that the line intersects the central axis at a point roughly between 500-600. This is the sample size that should be ideally used for trials to deem the new antibiotic “significantly better” if its efficacy holds.

In a real life scenario, a high efficacy as 90% should tail the drug being used on a relatively large sample size so that all potential outliers are captured. If the sample size is not appropriately large, it may be possible

that every single one of the patient ends up recovering and edge cases/outliers are never encountered which is not a correct estimate of how the drug would perform on the general population and thus would not be enough evidence to check if the new antibiotic is mathematically better than the standard antibiotic.

**Sources for answering this question:**

<https://emj.bmj.com/content/20/5/453>

[https://www.researchgate.net/figure/The-Altman-nomogram-The-left-hand-axis-represents-the-calculated-standardized-difference\\_fig4\\_241694867](https://www.researchgate.net/figure/The-Altman-nomogram-The-left-hand-axis-represents-the-calculated-standardized-difference_fig4_241694867)