Solutions:-

The article I reviewed is entitled as
"What Have We (Not) Learnt from Millions of Scientific Papers with ***P*** Values?"

I find this article interesting as it explains what we generally ignore about p value even if we read lots of articles for understanding it. This articles starts with the NHST (null hypothesis significance testing). Starting from empirical data across the biomedical literature in 1990 to 2015, p values 96% of them have P values of 0.05 or even less. The same value has been depicted in so many papers. So over 25 years use of p value has doubled and tripled for meta analysis. Then for couple of years abstracts tend to highlight further spin to unreliable conclusions. After that finding the biases like p-hacking has been used. Now the p value is shifted from $P < 0.05$ to $P < 0.005$ is kind of temporizing measure to contain death by significance.

In this article they has explained that majority of scientific disciplines are affected by epidemic p values. Strong selection of biases can make almost everything statistically significant. it seems these biases operate in many scientific fields which use P values, specially with lenient $P < 0.05$ thresholds for claiming the success. As 96% biomedical literature uses P values in their abstract or claims statistically significant results in 2016.

Its been argued that whichever fields with the highest level of proportion of significance claims might be least reliable. So this is basically that ecological relationship which serves the hierarchy of different scientific fields. NHST and P value, these both are most suitable inherently for minority of current research.

However, reducing the selection biases still require additional drastic measures rather than just a change in inferential method of a simple p value. Changing the choice of inferential methods do not address threat of selection bias. Direct protection against selection bias is to reproduce research practices which includes careful choices. Advance analysis should also be the highest priority when it comes to quality measure of statistical analysis.

Question 2:

2a

1 - Known σ

```r
#The change is actually 100. 75 is the standard deviation. So it will be
PowerZ <- function(diff = 100, sd = 75, n = 5, alpha = 0.05) {
    z.alpha <- qnorm(1 - (alpha/2))
    power <- pnorm( abs(diff) / (sd/sqrt(n) ) - z.alpha)
    return(power)
}
PowerZ( diff=100, sd=75, n=5, alpha=0.05)
```

0.84648172104949

Unknown sd

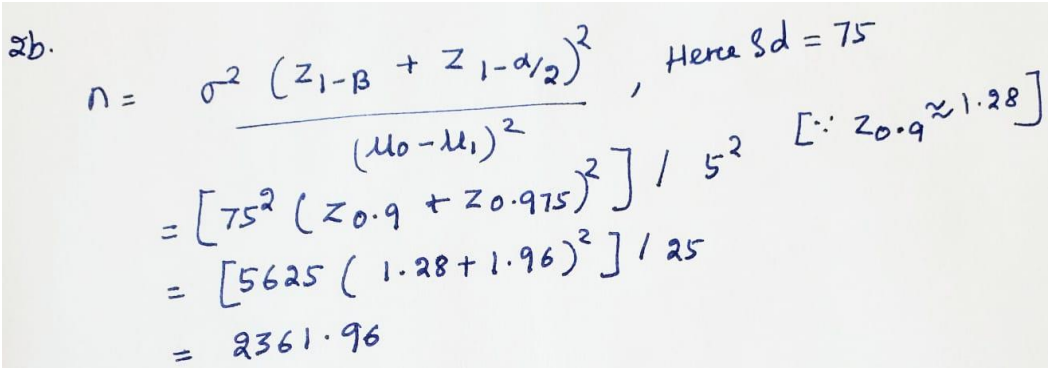
```r
pwrt_unknownsd <- power.t.test(n = 5, sd = 75, sig.level = 0.05, delta = 100, type = "one.sample", alternative = "two.sided")
pwrt_unknownsd
```

```
     One-sample t test power calculation

              n = 5
          delta = 100
             sd = 75
      sig.level = 0.05
          power = 0.6141832
    alternative = two.sided
```

```r
pwrt_unknownsd$power
```

0.614183212947759



Here the mean change in beta carotene that could 2361.96

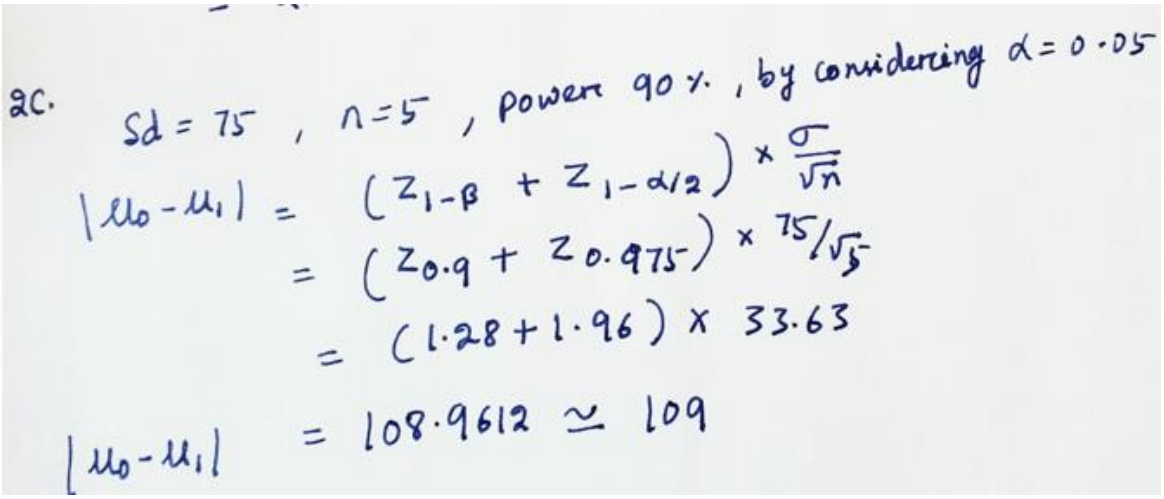
For alternative approach,

$|\mu_0 - \mu_1|$ ^2 = 100

So n = [5625(1.28+1.96)]/ 100

= 182.25

Here the mean change in beta carotene that could 182.25



With a sample size of 5, there is 90% power to detect a difference of at least 109, given that it exists.

```
power.t.test(n=5, sd=75, power=0.9, sig.level=0.05, delta=NULL,
```

```
 One-sample t test power calculation

          n = 5
      delta = 147.4417
         sd = 75
  sig.level = 0.05
      power = 0.9
alternative = two.sided
```

Our estimate here is rounded up to be 147

For validation that whether we are getting same power of 90% by using our resulted difference,

```
1 PowerZ( diff=109, sd=75, n=5, alpha=0.05)
```

```
0.901437887031338
```

2d.

sd = 75, n = 5, alpha = 0.05

$$|\mu_0 - \mu_1| = (Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}) \times \frac{\sigma}{\sqrt{n}}$$

$$= (Z_{0.8} + Z_{0.975}) \times 75/\sqrt{5}$$

$$= (0.84 + 1.96) \text{ X } 33.63$$

$$|\mu_0 - \mu_1| = 94.164$$

This result clearly states that if sample size decreases then Detectable difference increases as its inversely proportional.

sample size n **decreases:**

1. Detectable Difference |μ0−μ1|: *increases*

```
power.t.test(n=5, sd=75, power=0.8, sig.level=0.05, delta=NULL, type="one.sample", alternative="two.sided")
```

```
     One-sample t test power calculation

              n = 5
          delta = 126.1498
             sd = 75
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

Our estimate here is rounded up to be 126

With a sample size of 5, there is 80% power to detect a difference of at least 94.164, given that it exists.

For validation that whether we are getting same power of 90% by using our resulted difference,

```
PowerZ( diff=94.16, sd=75, n=5, alpha=0.05)
```

```
0.801598525862662
```

3a

```r
set.seed(2345)
n <- 5
mean <- 0
sd <- 75
numTrials <- 10000
alpha <- .05
count<- 0 # Set a counter to determine the number of rejected hypothesis tests
for( i in 1:numTrials){
  y <- rnorm(n=n, mean=0, sd=sd) # generate sample from desired distribution
  # Perform test
  t <- t.test(y, alternative = "two.sided")
  count <- count + (t$p.value < alpha)
}
# Power = proportion of rejections
power <- count/numTrials
power
```

0.0498

```
Here we got Power as 0.0498 that means nearly 4.9% of the result is False positive.
Only 4.9% is incorrectly rejected Null hypothesis
```

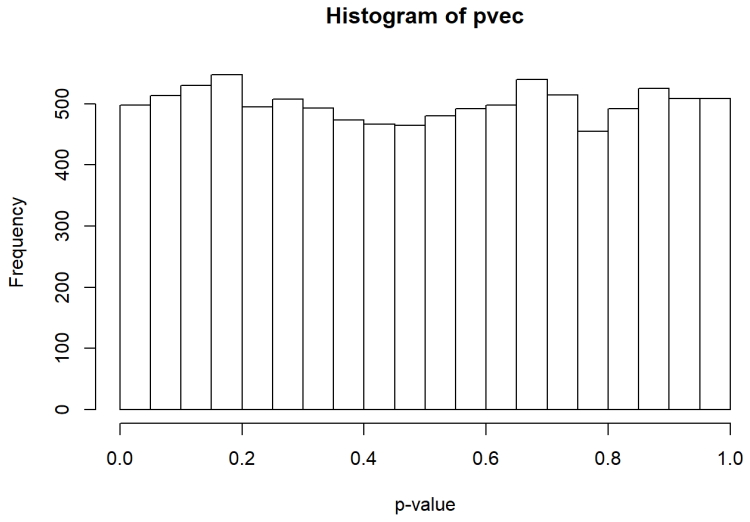
power is the probability of rejecting the false null hypothesis. We are calculating as the null is in fact false. Here sample size n is less so detectable difference might increase

3b

Null Hypothesis

```r
set.seed(2345)
n <- 5
mean <- 0
sd <- 75
numTrials <- 10000
alpha <- .05
count<- 0 # Set a counter to determine the num
for( i in 1:numTrials){
  y <- rnorm(n=n, mean=0, sd=sd) # generate sa
  # Perform test
  t <- t.test(y, alternative = "two.sided")
  count <- count + (t$p.value < alpha)
}
# Power = proportion of rejections
power <- count/numTrials
power
```

0.0498



it is almost uniform under the null hypothesis. Here the frequency is varying from 440 to 500.

3c

```r
set.seed(1796)
n <- 5
mean <- 100
sd <- 75
numTrials <- 10000
alpha <- .05
count<- 0 # Set a counter to determine the number
for( i in 1:numTrials){
  y <- rnorm(n=n, mean=100, sd=sd) # generate sam
  # Perform test
  t <- t.test(y, alternative = "two.sided")

  count <- count + (t$p.value < alpha)

}
# Power = proportion of rejections
power <- count/numTrials
power
```

0.613

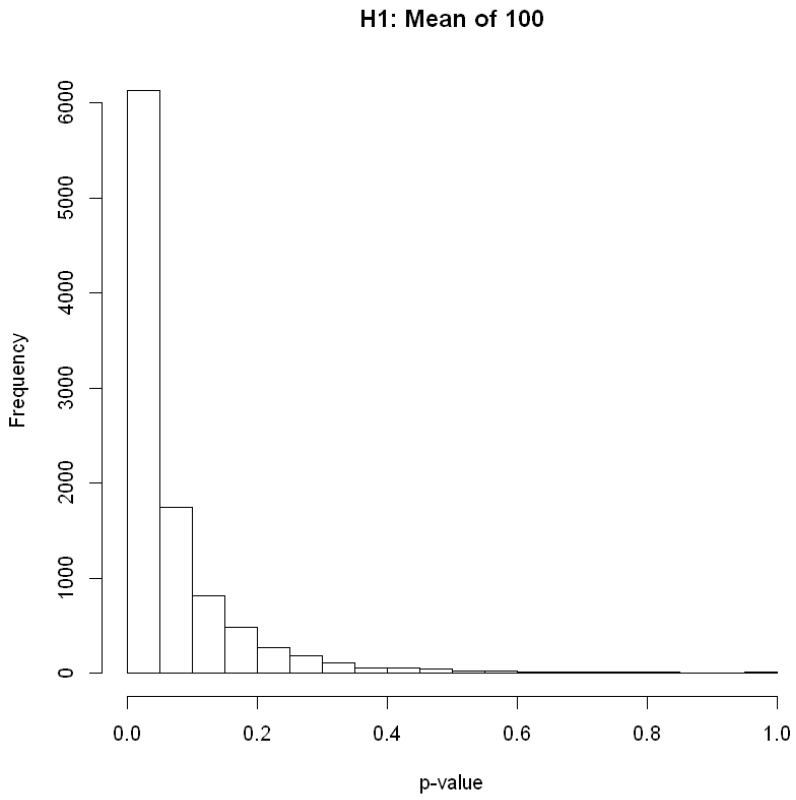
Here we got Power as 0.613 that means nearly 6.1% of the result is False positive.

Only 6.1% is incorrectly rejected Null hypothesis. This method can be useful while analysing different hypothesises. Even if we knew the underlying population distribution from the graph, we are still taking a sample, and therefore will have variability around our estimates.

```r
set.seed(1796)
n <- 5
mean <- 100
sd <- 75
numTrials <-10000
alpha <- 0.05

# For this for loop, we will save the p-value from e
pvec <- rep(NA, numTrials)
for(i in 1:numTrials){
  y <- rnorm(n,mean,sd) # simulate the data for the
  t <- t.test(y, mu = 0, alternative = "two.sided")
  pvec[i] <- t$p.value
}

hist( pvec, xlab='p-value', main='H1: Mean of 100' )
```



There is a minor skewness in the data as the frequency suddenly drops after 0.1. It might be due to lower or upper bounds on the data. The Lower bound are often skewed right while data that have an upper bound are often skewed left. Skewness can also result from start-up effects.

3d

Scenario 2:

Use simulation to obtain a power estimate when the alternative hypothesis assu
value 1796.

Alternative Hypotheses

```
set.seed(1796)
n <- 5
mean <- 100
sd <- 75
numTrials <- 10000
alpha <- .05
count<- 0 # Set a counter to determine the number of reject
for( i in 1:numTrials){
  y <- rnorm(n=n, mean=100, sd=sd) # generate sample from d
  # Perform test
  t <- t.test(y, alternative = "two.sided")

  count <- count + (t$p.value < alpha)

}
# Power = proportion of rejections
power <- count/numTrials
power
```

0.613

Here we got Power as 0.613 that means nearly 6.1% of the result is False positive.

Only 6.1% is incorrectly rejected Null hypothesis. Even if we knew the underlying population distribution from the graph, we are still taking a sample, and therefore will have variability around our estimates

3e

2a answer similar to scenario 2 as scenario 2 is having the mean change as 100 which same with the delta value in question 2a

As the power test here clearly explains delta as 100, n will b 5, standard deviation is also 75 as same as scenario 2.

So all the parameters along with the t test.

```
One-sample t test power calculation

              n = 5
          delta = 100
             sd = 75
      sig.level = 0.05
          power = 0.6141832
    alternative = two.sided
```