Stat 581 - Problem Set #3b Solutions

3. (a) $n = 125$ patients for each drug group

(b) see power curve graph

In either case, ($H_0: \delta = 0$ true or $H_A: \delta = \delta_A$ true), there is a low probability of committing an error. $\left(\begin{array}{l}\alpha = .05 \\ \beta = .20\end{array}\right)$

(c) The specific alternative δ_A is chosen to represent an effect size that is expected (past experience, related data), important (difference is non-negligible), and/or practical (cost considerations).

$\gamma = \frac{n \sum_i^a \tau_i^2}{\sigma^2} \geq \frac{n D^2}{2\sigma^2}$

4. (a) $D = \max\{\bar{Y}_i\} - \min\{\bar{Y}_i\} = 11.8$

$\sigma^2 = MSE = 8.06$

$n = 3$ fiber samples for each of $a = 5$ cotton blends

(b) (i) generate $\{Y_{ij}\}^{(\ell)} \sim$ indep $N(\mu_i, \sigma^2)$ — repeat for $\ell = 1, \dots, L$

(ii) compute $F_0^{(\ell)}$

(iii) power $= \frac{1}{L} \sum_{\ell=1}^{L} 1\{F_0^{(\ell)} > F_\alpha\}$

power = .977, see R code

(c) Specifying parameter values for a power analysis based on estimates from a pilot study, without accounting for estimation error, may lead to a hypothesis test that does not have adequate power.

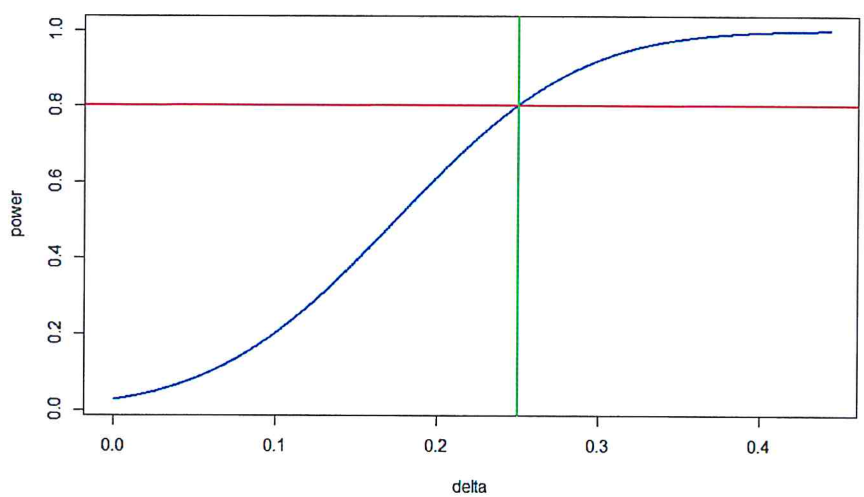
```
> power.t.test(n=NULL,delta=0.25,sd=0.7,sig.level = .05,power = .80,
type = "two.sample")

     Two-sample t test power calculation

              n = 124.0381
          delta = 0.25
             sd = 0.7
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group

>
>
> n = c(120,125,130)
> df = 2*(n-1)
> sd = 0.7
> delta = 0.25
> ncp = sqrt(n/2)*(delta/sd)
> alpha = .05
> power = .80
> 1 - pt(qt(1-alpha/2,df),df,ncp)
[1] 0.7867728 0.8030451 0.8182404
>
>
> n=125
> df = 2*(n-1)
> sd = 0.7
> alpha = .05
> delta = seq(from=0,to=5*sd/sqrt(n/2),length.out = 1000)
> power = 1 - pt(qt(1-alpha/2,df),df,ncp = sqrt(n/2)*(delta/sd))
>
> plot(delta,power,type = "l",lwd=2,col="blue")
> abline(h=.80,col="red,lwd=2)
> abline(v=0.25,col="green,lwd=2)
>
```



```
> hw3b.data = read_excel("handout2data.xlsx")
> str(hw3b.data)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':     25 obs. of  11 variables:
 $ strength: num  7 7 15 11 9 12 17 12 18 18 ...
 $ percent : num  15 15 15 15 15 20 20 20 20 20 ...
 $ 20g     : num  24 28 37 30 NA NA NA NA NA NA ...
 $ 30g     : num  37 44 31 35 NA NA NA NA NA NA ...
 $ 40g     : num  42 47 52 38 NA NA NA NA NA NA ...
 $ life    : num  17.6 18.9 16.3 17.4 20.1 21.6 16.9 15.3 18.6 17.1 ...
 $ fluid   : num  1 1 1 1 1 1 2 2 2 2 ...
 $ rate    : num  575 542 530 539 570 565 593 590 579 610 ...
 $ rf power: num  160 160 160 160 160 180 180 180 180 180 ...
 $ brand   : chr  "acme" "acme" "acme" "acme" ...
 $ wear    : num  2.1 2.4 2.5 2.3 2.2 2 1.9 2.1 2.2 2.4 ...
>
> percent = as.factor(na.omit(hw3b.data$percent))
> strength = na.omit(hw3b.data$strength)
>


> hw3.model = aov(strength~percent)

> summary(hw3.model)
            Df Sum Sq Mean Sq F value   Pr(>F)    
percent      4  475.8  118.94   14.76 9.13e-06 ***
Residuals   20  161.2    8.06                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> means = by(strength,percent,mean)
> means
percent: 15
[1] 9.8
-----------------------------------------------------------------------------
--------------
percent: 20
[1] 15.4
-----------------------------------------------------------------------------
--------------
percent: 25
[1] 17.6
-----------------------------------------------------------------------------
--------------
percent: 30
[1] 21.6
-----------------------------------------------------------------------------
--------------
percent: 35
[1] 10.8
```

```
> max.D = max(means) - min(means)
> max.D
[1] 11.8
> s2 = 8.06
> sd = sqrt(s2)
> a = 5
> power.anova.test(groups=a,between.var = max.D^2/2/(a-1),within.var = s2,pow
er = .8,sig.level = .05,n=NULL)

     Balanced one-way analysis of variance power calculation

         groups = 5
              n = 2.533845
    between.var = 17.405
     within.var = 8.06
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group

>
```

```
> n = 3
> decide.Ha = rep(NA,1000)
> for (k in 1:1000){
+   y1 = rnorm(n,means[1],sd)
+   y2 = rnorm(n,means[2],sd)
+   y3 = rnorm(n,means[3],sd)
+   y4 = rnorm(n,means[4],sd)
+   y5 = rnorm(n,means[5],sd)
+   ybar1 = mean(y1)
+   ybar2 = mean(y2)
+   ybar3 = mean(y3)
+   ybar4 = mean(y4)
+   ybar5 = mean(y5)
+   var1 = var(y1)
+   var2 = var(y2)
+   var3 = var(y3)
+   var4 = var(y4)
+   var5 = var(y5)
+   F.stat = n*var(c(ybar1,ybar2,ybar3,ybar4,ybar5)) / mean(c(var1,var2,var3,
var4,var5))
+   decide.Ha[k] = (F.stat>qf(.95,a-1,a*(n-1)))
+   k = k+1
+ }
> power = mean(decide.Ha)
> power
[1] 0.977
```