STAT 581 - HW #3

Alex Towell (atowell@siue.edu)

Problem 1

An experiment is conducted to study the effect of drilling method on drilling time. Each method (dry drilling, wet drilling) is used on $n = 12$ rocks. Drilling times are measured in 1/100 minutes.

Part (a)

Compute a 95% confidence interval for $\delta = \mu_1 - \mu_2$. Provide an interpretation, stated in the context of the problem.

A confidence interval for δ includes all parameter values compatible with the observed data $\hat{\delta}$,

$$\delta \in \left[\hat{\delta} + t_{\alpha/2, 2(n-1)} s_p \sqrt{2/n}, \hat{\delta} + t_{1-\alpha/2, 2(n-1)} s_p \sqrt{2/n}\right].$$

We compute this CI with:

```
library("readxl")
data = read_excel("./handout1data.xlsx")
data$method = as.factor(data$method)
dry = na.omit(data$time[data$method=='d'])
wet = na.omit(data$time[data$method!='d'])
alpha = .05

t.test(x=dry,
       y=wet,
       alternative=c("two.sided"),
       conf.level=1-alpha,
       var.equal=T)$conf.int[1:2]
```

```
## [1] 126.8757 276.4576
```

We estimate that the difference in drilling methods, $\delta = (\text{dry} - \text{wet})$, is between $[126.876, 276.458]$.

Part (b)

Explain how a confidence interval provides a complementary result to a hypothesis test.

A hypothesis test looks to determine if an effect exists. A CI looks to determine the size of the effect.

Problem 2

A prodcut developer is investigating the tensile strength of a new synthetic fiber. A completely randomized design with five levels of cotton content is performed, with $n = 5$ speciments per level.

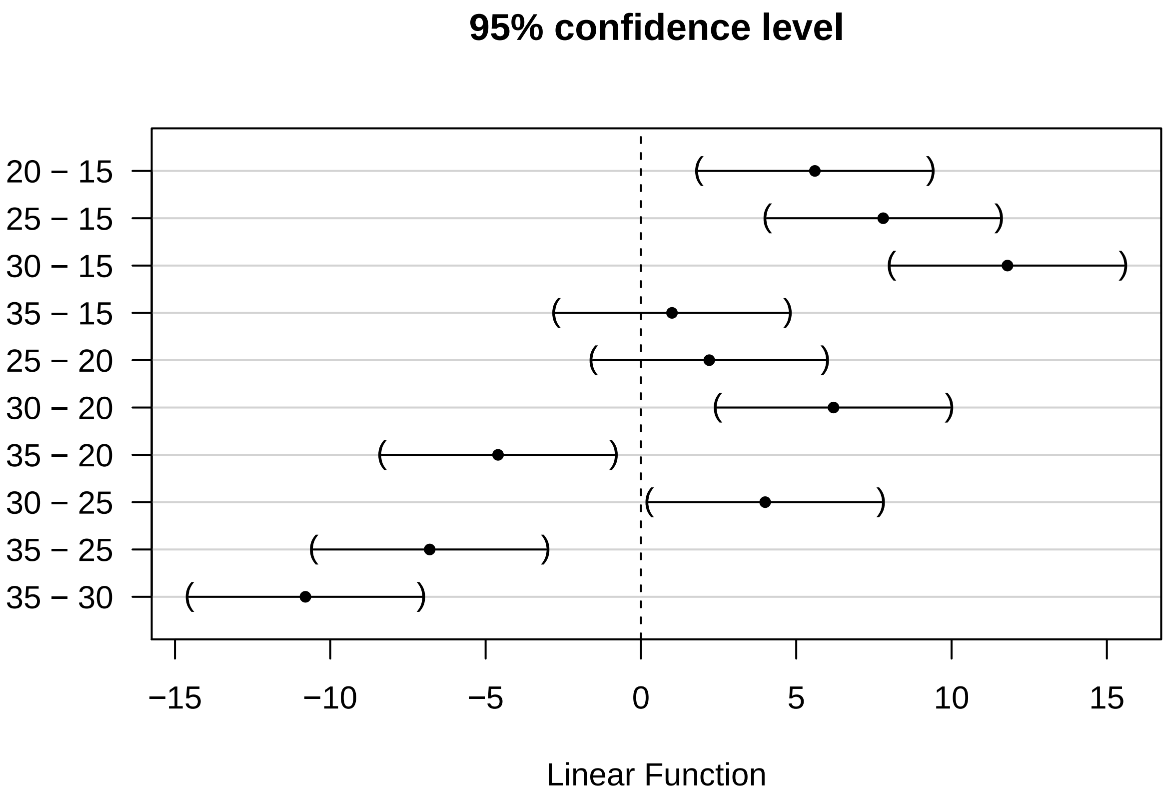
Part (a)

Compute and display 95% confidence intervals for all pairwise comparisons.

We show the confidence intervals with:

```
library("multcomp")
data.2 = read_excel("./handout2data.xlsx")
strength = na.omit(data.2$strength)
percent = na.omit(as.factor(data.2$percent))

m = aov(strength~percent)
m.lsd = glht(m,linfct=mcp(percent="Tukey"))
ci.lsd = confint(m.lsd,calpha=univariate_calpha())
plot(ci.lsd)
```




```
ci.lsd
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = strength ~ percent)
##
## Quantile = 2.086
## 95% confidence level
##
##
## Linear Hypotheses:
##              Estimate lwr     upr
## 20 - 15 == 0   5.6000   1.8545   9.3455
```

```
## 25 - 15 == 0   7.8000   4.0545  11.5455
## 30 - 15 == 0  11.8000   8.0545  15.5455
## 35 - 15 == 0   1.0000  -2.7455   4.7455
## 25 - 20 == 0   2.2000  -1.5455   5.9455
## 30 - 20 == 0   6.2000   2.4545   9.9455
## 35 - 20 == 0  -4.6000  -8.3455  -0.8545
## 30 - 25 == 0   4.0000   0.2545   7.7455
## 35 - 25 == 0  -6.8000 -10.5455  -3.0545
## 35 - 30 == 0 -10.8000 -14.5455  -7.0545
```

Part (b)

Explain how a confidence interval can be used in testing $H_0 : \mu_i = \mu_j$, for each pair of factor levels.

If the CI for $\mu_i - \mu_j$ contains 0, the decision is to decide $H_0^{(i,j)} : \mu_i - \mu_j$. Otherwise, the decision is to decide $H_A^{(i,j)} : \mu_i \neq \mu_j$.

Part (c)

Explain how computing multiple intervals impacts the probability of committing an error.

Suppose the hypothesis is given by

$$\begin{aligned} H_0 &: \mu_1 = \cdots = \mu_a \\ H_A &: \mu_i \neq \mu_j \text{ for at least one pair } (i,j),\ i \neq j. \end{aligned}$$

If we approximate this hypothesis test with $\binom{a}{2}$ pairwise tests of the form

$$\begin{aligned} H_0^{(i,j)} &: \mu_j = \mu_i \\ H_0^{(i,j)} &: \mu_j \neq \mu_i \end{aligned}$$

for $i = 1, \dots, a-1$ and $j = i+1, \dots, a$ where

$$\Pr\{\text{decide } H_A^{(i,j)} | H_0^{(i,j)} \text{ true}\} = \alpha,$$

then a type I error occurs if one or more of the tests is a false positive.

We denote this false positive rate by α_g which satisfies

$$\begin{aligned} \alpha_g &= \Pr\{\text{decide } H_A | H_0 \text{ true}\} \\ &= \Pr\{\text{decide } H_A^{(i,j)} \text{ for some } (i,j) | H_0 \text{ true}\} \\ &> \alpha. \end{aligned}$$

Proof. The equation for α_g may be rewritten as

$$\alpha_g = 1 - \Pr\{\text{decide } H_0^{(i,j)} \text{ for all } (i,j) | H_0 \text{ true}\}.$$

These pairwise tests are not necessarily independent, so assume k independent tests, $1 < k \leq \binom{a}{2}$, each with false positive rate α. The probability that a false positive does not occur on a pairwise test is $1 - \alpha$, so we rewrite the above as

$$\alpha_g = 1 - (1 - \alpha)^k,$$

which satisfies $\alpha_g > \alpha$ over its support $(0, 1)$. □

Observe that $\alpha_g \to 1$ as $k \to \infty$ and $\alpha_g = \alpha$ at $k = 1$.

Problem 3

An experiment to compare a new drug to a standard is in the planning stages. The response variable of interest is the clotting time (in minutes) of blood drawn from a subject. The experimenters want to perform a two sample t test at level $\alpha = .05$, having power $\pi = .8$ at $\delta_A = 0.25$, for standard deviation $\sigma = 0.7$.

Part (a)

Determine the sample size for each drug in order to achieve the stated test specifications.

We compute n with:

```
sd = .7
alpha = .05
h = .8  # power
v = .25 # alternative
power.t.test(n=NULL,delta=v,sd=sd,sig.level=alpha,power=h,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
```

We see that $n' = 124.0381$. If the power h is a lower-bound, then let $n = \lceil n' \rceil = 125$. If the power specification is not a lower-bound, but at approximate specification, it may be appropriate to round to the nearest integer, $n = [n'] = 124$.

Part (b)

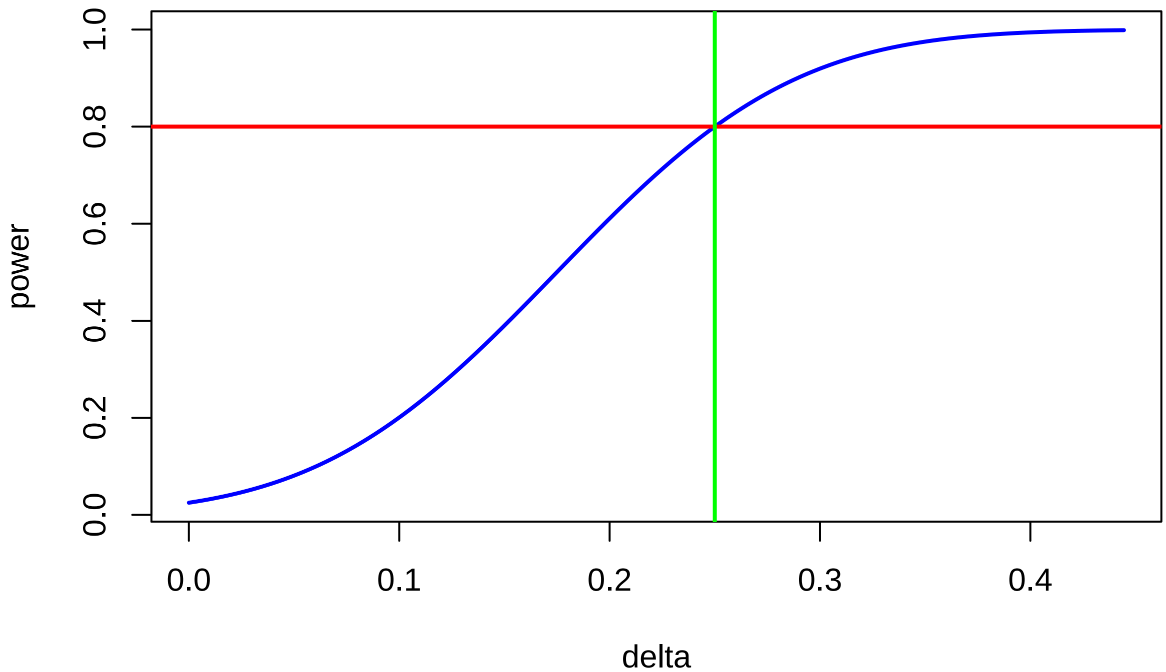
Graph the power curve for the chosen sample size. Explain how the power curve displays the desired properties of the test.

```
# power.curve
#
# create the power curve for the chosen sample size.
#
# arguments:
#   n: sample size,
#   sd: standard deviation
#   alpha: significance level
#   h: power (shows as a horizontal line)
#   v: specific alternative (shows as a vertical line)
#
# output:
#   graph of power curve
power.curve = function(n, sd, alpha, h, v)
{
  df = 2*(n-1)
  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=h,col="red",lwd=2)
  abline(v=v,col="green",lwd=2)
}

power.curve(n=124,sd=sd,alpha=alpha,h=h,v=v)
```



First, we see that the power at $\delta_A = 0.25$ obtains a power of $\pi(\delta_A) = 0.8$, as required.

Second, and this is not explicitly shown on the graph, whether $H_0 : \delta = 0$ is true or $H_A : \delta = \delta_A$ is true, there is a low probability of committing an error since the 95% confidence interval under the null model, approximately

$$\pm 1.96\sigma/\sqrt{n} = [-0.123, 0.123],$$

does not intersect with the 95% confidence interval under the alternative model,

$$\delta_A \pm 1.96\sigma/\sqrt{n} = [0.127, 0.373].$$

The small region between these confidence intervals may be classified as the "don't care" region.

Part (c)

Provide a general explanation of how δ_A can be determined.

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

Problem 4

Refer back to the tensile strength exampale of problem 2. Use the data from this study to perform a power analysis for a main study. The experimenters desire a level $\alpha = .05$ test with power $\pi = .8$.

Part (a)

Determine the sample size for each group based on specifying the maximum difference in means.

```r
pwr = .8
alpha = .05
a = 5
means = by(strength,percent,mean)
max.D = max(means) - min(means)
#summary(aov(strength~percent))
s2 = 8.06
power.anova.test(
  groups=a,
  between.var=max.D^2/2/(a-1),
  within.var=s2,
  power=pwr,
  sig.level=alpha,
  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
```

We see that $n' = 2.534$. Rounding to the nearest integer, we propose using a sample size of $n = 3$ for each group.

Part (b)

Use a simulation to compute power at $n = 3$ using the pilot study to specify the model parameters.

```r
sim.size = 10
decide.Ha = rep(NA,sim.size)
n = 3

for (k in 1:sim.size)
{
  y1 = rnorm(n,means[1],sqrt(s2))
  y2 = rnorm(n,means[2],sqrt(s2))
  y3 = rnorm(n,means[3],sqrt(s2))
  y4 = rnorm(n,means[4],sqrt(s2))
  y5 = rnorm(n,means[5],sqrt(s2))
  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(1-alpha,a-1,a*(n-1)))
}
power = mean(decide.Ha)
power
```

```
## [1] 0.9
```

Part (c)

Comment on the use of pilot study data in a power analysis.

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.