Homework 10

ANOVA

Langdon Holmes

1.1 Complete the table (with sphericity assumption). Show your work. Obtain the p-value of the F test.

	SS	df	MS	F
Treatment	150	3	50	5
Subject	900	9	100	
Interaction	270	27	10	
Total	1320			

Sample Descriptives:
$k = df_\tau + 1 = 4$
$n = df_\pi + 1 = 10$

Treatment:
$MS_\tau = \frac{SS_\tau}{df_\tau} = \frac{150}{3} = 50$

Subject:
$MS_\pi = \frac{SS_\pi}{df_\pi} = \frac{900}{9} = 100$

Interaction:
$SS_{\pi\tau} = SS_T - SS_\pi - SS_\tau = 1,320 - 900 - 150 = 270$
$df_{\pi\tau} = (k-1)(n-1) = 27$
$MS_{\pi\tau} = \frac{SS_{\pi\tau}}{df_{\pi\tau}} = \frac{270}{27} = 10$

$F = \frac{MS_\tau}{MS_{\pi\tau}} = \frac{50}{10} = 5$

```r
# pf is for p-values
# qf is for critical values
cat(pf(q = 5, df1 = 3, df2 = 27, lower.tail = FALSE),
    qf(p=.05, df1 = 3, df2 = 27, lower.tail = FALSE))
```

```
0.006922033 2.960351
```

$F_{3,27,.05} = 2.96, p = 0.007$

There is a significant effect of treatment. I don't believe we can say in which direction with the given summary statistics.

1.2 How many subjects are there? How many treatments are there?

There are 10 subjects (n) and 4 treatments (k). I showed "calculations" in 1.1.

1.3 Calculate $\hat{\omega}^2$.

$\hat{\omega}^2 = \frac{SS_{\tau} - (k-1)MS_{\pi\tau}}{SS_T + MS_{\pi}} = \frac{150-3(10)}{1{,}320+100} = \frac{120}{1{,}420} = .085$

1.4 Suppose the correction factor for sphericity is $\hat{\epsilon} = 0.6$, what is the F statistic? What is its sampling distribution? Find p-value.

```r
# pf is for p-values
# qf is for critical values
cat(pf(q = 5, df1 = .06*3, df2 = .06*27, lower.tail = FALSE),
    qf(p=.05, df1 = .06*3, df2 = .06*27, lower.tail = FALSE))
```

```
0.1103312 19.03088
```

$F_{\epsilon*3,\epsilon*27,.05} = F_{.18,1.62,.05} = 19.03, p = 0.11$

With the sphericity adjustment, the main effect of treatment is not significant.

2. Consider a within-subject design with 30 subjects repeated measured under 3 conditions.

2.1 Suppose a computer program calculated two estimates of the correction factor $\epsilon : 0.6$ and 0.75. Which is Huynh-Feld estimate and which is Greenhouse-Geisser estimate? Calculate the lower bound estimate.

Greenhouse-Geisser tends to underestimate ϵ, so that is likely to result in the lower value of 0.6.

The lower bound estimate is $\epsilon = 1/(k-1) = \frac{1}{3} = 0.\overline{33}$. This is even more conservative than Greenhouse-Geisser.

2.2 What is the power of this design to detect $\omega^2 = 0.1$? Assume $\epsilon = 0.6$ and $\rho = 0.5$. Present screenshots of WebPower.

```r
# d = 0.5
omega <- 0.1
rho <- 0.5
f <- sqrt( 3 * omega/(1-omega) * 1/(1-rho) )
print(f)
```

```
[1] 0.8164966
```

```r
wp.rmanova(ng=1, # num groups
           n=30, # subjects
           nm=3, # number of measurements
           nscor=0.6, # nonsphericity correction coefficient
           f=f,
           type=1 # 1 = "Within Effect"
           )
```

```
Repeated-measures ANOVA analysis

     n         f ng nm nscor alpha     power
    30 0.8164966  1  3   0.6  0.05 0.9059011

NOTE: Power analysis for within-effect test
URL: http://psychstat.org/rmanova
```

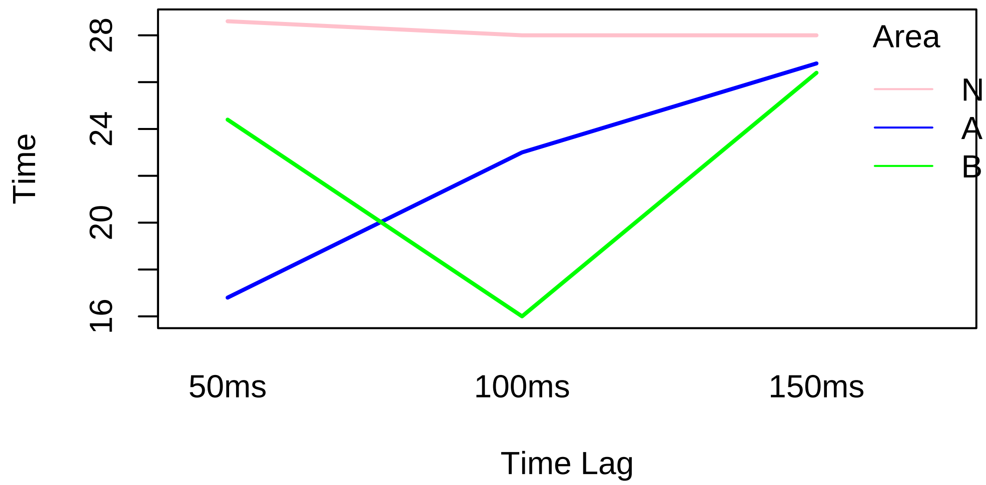
3. In the attachment you find the data from the Brain Area experiment described in HW9. The three levels 1, 2 and 3 of Area are N, A and B brain areas. The three levels 1, 2 and 3 of Lag are 50, 100 and 150ms. Do the following using SPSS or R.

```
df <- read.table("Avoidance.dat", header = TRUE) %>%
  mutate(
    Area = factor(Area, labels = c("N", "A", "B")),
    Lag = factor(Lag, labels = c("50ms", "100ms", "150ms"), ordered = TRUE)
    )
summary(df)
```

```
 Area      Lag          Time
 N:15   50ms :15   Min.   : 9.00
 A:15   100ms:15   1st Qu.:20.00
 B:15   150ms:15   Median :24.00
                   Mean   :24.22
                   3rd Qu.:28.00
                   Max.   :40.00
```

3a obtain an interaction plot with brain region as separate lines and Lag on the x axis.

```
interaction.plot(
  x.factor = df$Lag,
  trace.factor = df$Area,
  response = df$Time,
  fun = mean,
  ylab = "Time",
  xlab = "Time Lag",
  col = c("pink", "blue", "green"),
  lty = 1, #line type
  lwd = 2, #line width
  trace.label = "Area"
  )
```



3b obtain ANOVA table.

```
fit <- aov(Time ~ Area*Lag, data=df)
summary(fit)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)   
Area         2  356.0  178.02   6.074 0.00534 **
Lag          2  188.6   94.29   3.217 0.05184 . 
Area:Lag     4  372.0   92.99   3.172 0.02482 * 
Residuals   36 1055.2   29.31                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3c obtain the unadjusted p-value for the interaction contrast in Problem 3 of HW9

	Area N	Area A	Area B	Row Means
50ms	28.6	16.8	24.4	23.27
100ms	28	23	16	22.33
150ms	28	26.8	26.4	27.07

	Area N	Area A	Area B	Row Means
Column Means	28.2	22.2	22.27	24.22

$\hat{\psi} = (16.8 - 24.4) - (23.0 - 16.0) = -14.6$
$df = (5 \times 9) - 9 = 36$
$SE = \sqrt{MS_{error} \times 4/n} = \sqrt{29.311 \times (1+1+1+1)/5} = 4.84$
$t = \hat{\psi}/SE = -14.6/4.84 = -3.016529$

```
p <- pt(q = -3.016529, df = 36, lower.tail=TRUE)
print(p)
```

```
[1] 0.002335165
```

```
# multiply by 2 since this is not a directional hypothesis.
print(p*2)
```

```
[1] 0.004670329
```

The two-sided, unadjusted p-value is 0.005.

3d obtain unadjusted two-sided p-values for the family of 6 pairwise comparisons in the simple effects of Area (those in Problem 2 of HW9).

I am borrowing the t-statistics from HW9. Hopefully that is allowed. And the p-values are calculated below.

```
# All p-values are multiplied by 2 because these are two-sided tests.
p <- pt(q = 3.45, df = 36, lower.tail=FALSE)*2
cat(sprintf("$t_{50A} = 3.45, p = %1.3f$  \n", round(p, 3)))
```

$t_{50A} = 3.45, p = 0.001$

```
p <- pt(q = 1.23, df = 36, lower.tail=FALSE)*2
cat(sprintf("$t_{50B} = 1.23, p = %1.3f$  \n", round(p, 3)))
```

$t_{50B} = 1.23, p = 0.227$

```r
p <- pt(q = 1.46, df = 36, lower.tail=FALSE)*2
cat(sprintf("$t_{100A} = 1.46, p = %1.3f$  \n", round(p, 3)))
```

$t_{100A} = 1.46, p = 0.153$

```r
p <- pt(q = 3.50, df = 36, lower.tail=FALSE)*2
cat(sprintf("$t_{100B} = 3.50, p = %1.3f$  \n", round(p, 3)))
```

$t_{100B} = 3.50, p = 0.001$

```r
p <- pt(q = 0.35, df = 36, lower.tail=FALSE)*2
cat(sprintf("$t_{150A} = 0.35, p = %1.3f$  \n", round(p, 3)))
```

$t_{150A} = 0.35, p = 0.728$

```r
p <- pt(q = 0.47, df = 36, lower.tail=FALSE)*2
cat(sprintf("$t_{150B} = 0.47, p = %1.3f$  \n", round(p, 3)))
```

$t_{150B} = 0.47, p = 0.641$