Statistics 641, Fall 2014
Take Home Final Exam
Solutions

1. Suppose that we have a binomial response with expected failure probabilities $p_0 = 0.2$ and $p_1 = 0.1$ for treatments 0 (control) and 1 (experimental) respectively. Calculate the required sample sizes to achieve 90% power (type II error rate of 10%) at level $\alpha = .01$ assuming

   (a) Equal allocation to groups 0 and 1

   $$N = \frac{\bar{p}(1-\bar{p})(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\xi_0\xi_1(p_0 - p_1)^2}$$

   and with equal allocation $\bar{p} = (p_0 + p_1)/2 = 0.15$.

   $$\frac{4 \times .15(1 - .15)(2.58 + 1.28)^2}{(0.2 - 0.10)^2} = 759.9$$

   so $N = 760$.

   (b) 2:1 randomization to groups 0, 1 respectively

   $\bar{p} = (2p_0 + p_1)/3 = 0.167$ and

   $$\frac{9 \times .167(1 - .167)(2.58 + 1.28)^2}{2(0.2 - 0.10)^2} = 931.2$$

   so $N = 932$.

   (c) 1:2 randomization to groups 0, 1 respectively

   $\bar{p} = (p_0 + 2p_1)/3 = 0.133$ and

   $$\frac{9 \times .133(1 - .133)(2.58 + 1.28)^2}{2(0.2 - 0.10)^2} = 774.8$$

   so $N = 775$.

   (d) Why are these different?

   N is sensitive to $\bar{p}$, which in turn depends on ξ_j. When more subjects are allocated to the treatment with the lower rate, fewer subjects are required to detect a fixed difference in probabilities.

   Note that the odds ratio is outside the interval $[.5, 2]$, so the approximation to the sample size based on the log odds ratio will not be good, so that formula involving the log odds ratio should not be used.

2. Suppose we have the following observations for response X:

$$4,\ 16,\ 24,\ 27,\ 28,\ 29,\ 32,\ 37,\ 43,\ 48$$

and observations are randomly assigned either $Z = 0$ or $Z = 1$. Assume we use the model $X = \alpha + \beta Z + \varepsilon$.

(a) Calculate the size of the reference set for complete randomization and calculate the variance of $\hat{\beta}$ under the randomization distribution. (It is possible to enumerate all possible allocations, but you may use simulation if necessary.) Use the convention that when all observations are allocated to the same group, $\hat{\beta} = 0$.

The reference set has size

$$2^{10} = 1024$$

Using simulation (10,000 replicates)

```
> var(replicate(1e4, {z <- sample(0:1,10,repl=T)
+     if(all(z==0)|all(z==1)) 0 else lm(x~z)$coef[2]}))
[1] 76.06725
```

Using exact randomization distribution, use recursion to build matrix of all possible treatment assignments:

```
> Z <- t(0:1)  ## for one subject.  Columns are treatment assignments
> Z
     [,1] [,2]
[1,]    0    1
> Z <- rbind(cbind(Z,Z),rep(0:1,each=ncol(Z)))  #for 2 subjects:
> Z
     [,1] [,2] [,3] [,4]
[1,]    0    1    0    1
[2,]    0    0    1    1
> Z <- rbind(cbind(Z,Z),rep(0:1,each=ncol(Z)))  #for 3 subjects:
> Z
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    1    0    1    0    1    0    1
[2,]    0    0    1    1    0    0    1    1
[3,]    0    0    0    0    1    1    1    1
> for(i in 1:7) Z <- rbind(cbind(Z,Z),rep(0:1,each=ncol(Z)))  ## for 10 subjects
> dim(Z)
[1]   10 1024
> var(apply(Z,2,function(z) if(all(z==0)|all(z==1)) 0 else lm(x~z)$coef[2]))
[1] 73.97587 ## exact variance for randomization distribution
```

(b) Calculate the size of the reference set for the random allocation rule and calculate the variance of $\hat{\beta}$ under the randomization distribution. (The `R` function `combn` can let you easily enumerate all possible allocations.)

The reference set has size

$$\binom{10}{5} = 252$$

Using exact randomization distribution, use `combn` to build matrix of all possible treatment assignments:

```
> M <- combn(10,5)
> dim(M)
[1]   5 252
> M[,1:10]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    1    1    1    1    1    1    1    1    1     1
[2,]    2    2    2    2    2    2    2    2    2     2
[3,]    3    3    3    3    3    3    3    3    3     3
[4,]    4    4    4    4    4    4    5    5    5     5
[5,]    5    6    7    8    9   10    6    7    8     9
> Z <- matrix(0,10,252)
> Z[cbind(c(M),c(col(M)))] <- 1        ## clever way
# or
> for(i in 1:ncol(Z)) Z[M[,i],i] <- 1  ## clunkier way
> var(apply(Z,2,function(z) lm(x~z)$coef[2]))
[1] 64.86183                           ## exact variance for random allocation rule
```

(c) If a proportion ξ_1 is allocated to group $Z = 1$, then efficiency relative to perfect balance is $4\xi_1\xi_0$. Calculate the expected relative efficiency of complete randomization relative to the random allocation rule for a total sample size of 10. Compare this to the ratio of the variance from the previous two parts.

If n_1 is the number assigned to $Z = 1$, under complete randomization, n_1 is binomial with size 10 and probability 1/2. $4\xi_1\xi 0 = 4n_1n_0/100$, and we can find its expectation using

```
> sum(4*0:10*10:0/100*dbinom(0:10,10,.5))
[1] 0.9
```

Furthermore, from the first two parts, the relative efficiency of complete randomization to random allocation is

$$\frac{64.86183}{73.97587} = 0.8767971$$

which is close to 0.9, the theoretical value.

3. Suppose we conduct a randomized trial and observe

   - binary treatment `z` (0,1)
   - continuous baseline covariate `w`
   - continuous response `y`

   We perform the following analyses (some output deleted):

```
> summary(lm(y~z,data=data))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.4295     0.2242   10.84   <2e-16
z             0.7545     0.3170    2.38   0.0199

Residual standard error: 1.382 on 74 degrees of freedom
F-statistic: 5.665 on 1 and 74 DF,  p-value: 0.01989
```

```
> summary(lm(w~z,data=data))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.9332     0.1678  29.402   <2e-16
z             0.1574     0.2373   0.663    0.509

Residual standard error: 1.034 on 74 degrees of freedom
F-statistic: 0.4399 on 1 and 74 DF,  p-value: 0.5092

> summary(lm(y~w,data=data))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.5474     0.7769   0.705  0.48329
w             0.4508     0.1519   2.968  0.00403

Residual standard error: 1.355 on 74 degrees of freedom
F-statistic:  8.81 on 1 and 74 DF,  p-value: 0.004034
```

Finally, we fit the model $y = \alpha + \beta z + \gamma w + \varepsilon$ below. Using the previous analyses and the partial information in the output below, find $\hat{\beta}$ from this model (the ***** in the first column). Is `w` a confounder?

```
> summary(lm(y~z+w,data=data))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.3330     0.7620   0.437  0.66341
z             ******     ******   *****  *******
w             0.4250     0.1482   2.867  0.00542

Residual standard error: 1.319 on 73 degrees of freedom
F-statistic: 7.218 on 2 and 73 DF,  p-value: 0.001379
```

Letting $\hat{\beta}_U$ and $\overline{\beta}_A$ be the unadjusted and adjusted estimates of β respectively, and $\overline{W}_0$ and $\overline{W}_1$ be the means of W in the groups $Z = 0$ and $Z = 1$ respectively, geometry (see figure on page 16, lecture 4) says that

$$\begin{aligned} \hat{\beta}_A &= \hat{\beta}_U - \hat{\gamma}(\overline{W}_1 - \overline{W}_0) \\ &= 0.7545 - 0.4250 \times 0.1574 \\ &= 0.6876 \end{aligned}$$

For extra credit, find the standard error of $\hat{\beta}$, the corresponding t-statistic and the p-value.

The t-test for $H_0 \colon \beta = 0$ is the square root of the F-statistic for the same hypothesis, and F is the model MSE divided by the residual MSE. From the model `y~w`, F is 8.81 with 1 and 74 df, and Residual SE 1.355, so the model sum of squares is

$$8.81 \times 1.355^2 = 16.17$$

From model `y~z+w`, F is 7.218 with 2 and 73 df, and RSE 1.319, so the model sum of squares is

$$2 \times 7.218 \times 1.319^2 = 25.12$$

This model sum of squares is the sum of squares for both W and Z, so the SS for Z adjusted for W is the difference

$$25.12 - 16.17 = 8.94$$

and the t is

$$\frac{\sqrt{8.94}}{1.319} = 2.26$$

and the SE for $\hat{\beta}_U$ is

$$\frac{0.6876}{2.26} = 0.3042$$

with 83 df, the p-value is 0.026.

4. Suppose we conduct a trial with 3 dose groups (low, medium, high) and placebo (none), and a binary outcome (dead/alive). At the conclusion of the trial we observed the following table:

Treatment:	placebo ($i = 0$)	low ($i = 1$)	medium ($i = 2$)	high ($i = 3$)
Alive	10	18	22	25
Dead	90	82	78	75
	100	100	100	100

   (a) Letting $p_i, i = 0, 1, 2, 3$ be the probability of death in group i (where i is defined in the table), the global null hypothesis is

   $$H^{0123}\colon p_0 = p_1 = p_2 = p_3.$$

   By reversing the roles of outcome and treatment, we may test this null hypothesis using the Wilcoxon rank-sum test (uncorrected). (This test is sensitive to the alternative hypothesis in which subjects who die are, for example, more likely to have received a lower dose than subjects who survive.)

   Conduct this test of H^{0123} at level $\alpha = 0.05$.

   (For convenience, the dataset `dataF.csv` has one record per subject with a treatment variable, `z`, and death variable `dead`. You can perform the Wilcoxon test directly using this dataset)

```
> dataF <- read.csv("dataF.csv")
> wilcox.test(z~dead, data=dataF, correct=F)
Wilcoxon rank sum test
data:  z by dead
W = 14637.5, p-value = 0.005053
alternative hypothesis: true location shift is not equal to 0
```

(b) Note that there are 6 pairwise hypotheses, denoted $H^{ij}: p_i = p_j$. Pairwise intersections of these take two forms. E.g.,

- $H^{01} \cap H^{02} = H^{012}: p_0 = p_1 = p_2$
- $H^{01} \cap H^{23} = H^{01,23}: p_0 = p_1$ and $p_2 = p_3$.

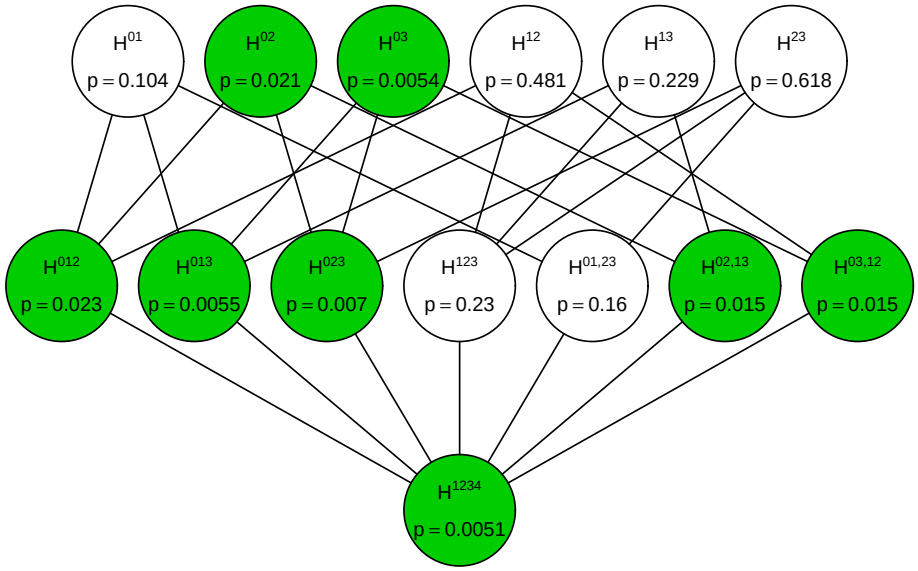
We can implement a closed testing procedure as follows. First test the global null hypothesis H^{0123} at level $\alpha = .05$. If we reject H^{0123}, then test each of the pairwise intersections (H^{012}, $H^{01,23}$, etc.) at level $\alpha = .05$.

Intersection hypotheses of the form H^{ijk} can be tested using the Wilcoxon rank-sum test restricted to the treatment groups $\{0, 1, 2\}$, $\{0, 1, 3\}$, etc. (Note that the function `Wilcox.test` takes a `subset` argument, e.g., `subset=z%in%c(0,1,2)` to use only groups 0,1,2.)

Tests for intersection hypotheses of the form $H^{ij,kl}$ can be constructed using stratified 2×2 tables as in homework 2, problem 2, where within each stratum, the lower dose acts as "treatment A" and the higher dose acts as "treatment B".

Depending upon which hypotheses H^{jkl} and $H^{ij,kl}$ are rejected, we can proceed to test pairwise hypotheses $H^{jk}: p_j = p_k$ at level $\alpha = 0.05$. (Note that for 2 groups and 2 outcomes, the Wilcoxon test (uncorrected) and the Pearson chi-square test (uncorrected) give identical results, so either can be used.)

Draw the diagram (similar to the diagrams in the Dec 9 and 11 lectures) showing the relationships between all 14 hypotheses and using this procedure, determine which of the pairwise hypotheses, H^{jk}, can be rejected at level $\alpha = 0.05$.



Tests with extraneous output deleted:

```
## Hypotheses of the form H^ijk
> wilcox.test(z~dead, data=dataF, correct=F,subset=z!=3)
W = 7450, p-value = 0.02302
> wilcox.test(z~dead, data=dataF, correct=F,subset=z!=2)
W = 8045.5, p-value = 0.005496
```

```
> wilcox.test(z~dead, data=dataF, correct=F,subset=z!=1)
W = 8425.5, p-value = 0.006951
> wilcox.test(z~dead, data=dataF, correct=F,subset=z!=0)
W = 8337.5, p-value = 0.2303
## Hypotheses of the form H^{ij,kl}
> Tab <- table(dd$dead, dd$z)
> Tab01.23 <- Tab               ## Maybe there's a slick way to do this in one step
> dim(Tab01.23) <- c(2,2,2)
> mantelhaen.test(Tab01.23,correct=F)
Mantel-Haenszel X-squared = 2.0054, df = 1, p-value = 0.1567
> Tab02.13 <- Tab[,c(1,3,2,4)]
> dim(Tab02.13) <- c(2,2,2)
> mantelhaen.test(Tab02.13,correct=F)
Mantel-Haenszel X-squared = 5.9239, df = 1, p-value = 0.01494
> Tab03.12 <- Tab[,c(1,4,2,3)]
> dim(Tab03.12) <- c(2,2,2)
> mantelhaen.test(Tab03.12,correct=F)
Mantel-Haenszel X-squared = 5.9005, df = 1, p-value = 0.01514
## Hypotheses of the form H^{ij}
> wilcox.test(z~dead, data=dataF, correct=F,subset=z%in%0:1)
W = 2808, p-value = 0.1039
> wilcox.test(z~dead, data=dataF, correct=F,subset=z%in%c(0,2))
W = 3288, p-value = 0.02096
> wilcox.test(z~dead, data=dataF, correct=F,subset=z%in%c(0,3))
W = 3637.5, p-value = 0.005362
> wilcox.test(z~dead, data=dataF, correct=F,subset=z%in%c(1,2))
W = 3400, p-value = 0.4806
> wilcox.test(z~dead, data=dataF, correct=F,subset=z%in%c(1,3))
W = 3725.5, p-value = 0.2294
> wilcox.test(z~dead, data=dataF, correct=F,subset=z%in%c(2,3))
W = 3745.5, p-value = 0.6177
```

The hypotheses in green have p-values less than 0.05. For hypotheses H^{02} and H^{03}, all hypotheses below are also green, and therefore, the pairwise hypotheses H^{02} and H^{03} can be rejected at 0.05. No other pairwise hypotheses can be rejected.

(c) Using the Bonferroni procedure for all 6 pairwise hypotheses, which can be rejected at overall level $\alpha = 0.05$?

Using the Bonferroni procedure for 6 hypotheses, we reject any particular hypothesis if the corresponding p-value is less than 0.05/6=0.00833. The only one of the 6 pairwise hypotheses meeting this criterion is H^{03}.

(d) Using the Bonferroni procedure for the three pairwise hypotheses that compare an active dose to placebo, which can be rejected at overall level $\alpha = 0.05$?

Using the Bonferroni procedure for 3 hypotheses, we reject any particular hypothesis if the corresponding p-value is less than 0.05/3=0.01667. The only one of the 3 pairwise hypotheses meeting this criterion is H^{03}.

(e) Using the Holm procedure for all 6 pairwise hypotheses, which can be rejected at overall level $\alpha = 0.05$?

Using Holm, we order the 6 pairwise hypotheses and compare them to 0.05/6=0.00833, 0.05/5=0.01, 0.05/4=0.0125, 0.05/3=0.0167, 0.05/2=0.025, and 0.05 respectively. The ordered p-values are 0.00536, 0.0210, 0.1039, 0.229, 0.481, and 0.618. The first (for H^{03}) is less then its critical value, 0.00833, however, the second, (for H^{02}) is not less than its critical value, 0.01, so we cannot reject it, and testing stops.

(f) Using the Holm procedure for the three pairwise hypotheses that compare an active dose to placebo, which can be rejected at overall level $\alpha = 0.05$?

Using Holm, we order the 3 pairwise hypotheses and compare them to 0.05/3=0.0167, 0.05/2=0.025, and 0.05 respectively. The ordered p-values are 0.00536, 0.0210, and 0.1039. The first (for H^{03}) is less then its critical value, 0.0167, however, the second, (for H^{02}) is also less than its critical value, 0.025, however $0.104 > 0.05$, so we can reject H^{03} and H^{02} but not H^{01}.