Statistics 641, Fall 2011
Homework #3
Answers

1. Suppose that we have a phase II, single arm trial using a two stage design. The hypotheses of interest are H_0: $\pi \leq 0.15$ versus H_1: $\pi \geq 0.4$ where π is the true success rate. Let y_k be the total number of successes through stage k, $k = 1, 2$. Note: you can use the functions `dbinom` and `pbinom` in R to calculate binomial probabilities.

   (a) We enroll 16 subjects in stage 1 and stop and accept H_0 if we observe $a_1 = 3$ or fewer responses ($y_1 \leq 3$), otherwise we continue to stage 2. Find the stopping probabilities under both H_0 and H_1.

   Under H_0, the stopping probability is $\Pr\{y_1 \leq 3\} = .7899$, and under H_1 it is .0651.

   In `R`:

   ```
   > pbinom(3, 16, .15)
   [1] 0.7898907
   > pbinom(3, 16, .40)
   [1] 0.06514674
   ```

   (b) At stage 2 we enroll an additional 16 subjects and reject H_0 if $y_2 > 8$. Compute the overall probabilities of rejection under both H_0 and H_1 for the two-stage trial.

   We accept H_0 if $y_1 \leq 3$ and $y_2 \leq 8$. Under H_0 this probability is

   $$\Pr\{y_1 \leq 3\} \quad + \quad \sum_{i=4}^{8} \Pr\{y_1 = i\} \Pr\{x_1 \leq 8 - i\} = .9659$$

   In `R`:

   ```
   > pbinom(3, 16, .15) + sum(dbinom(4:8,16,.15)*pbinom(4:0, 16, .15))
   [1] 0.9658661
   ```

   Under H_1, this probability is 0.0969.

   ```
   > pbinom(3, 16, .40) + sum(dbinom(4:8,16,.40)*pbinom(4:0, 16, .40))
   [1] 0.09691022
   ```

   Therefore, the rejection probabilities are 1-0.9659=0.0341 and 1-0.0969=0.9031 under H_0 and H_1 respectively.

   Alternatively,

   ```
   > 1-pbinom(8,16,.15) + sum(dbinom(4:8,16,.15)*(1-pbinom(4:0, 16, .15)))
   [1] 0.03413386
   > 1-pbinom(8,16,.40) + sum(dbinom(4:8,16,.40)*(1-pbinom(4:0, 16, .40)))
   [1] 0.9030898
   ```

(c) Compute the expected sample sizes for $\pi = 0.15$ and $\pi = 0.4$.

N is either 16 or 32, depending on whether we stop at stage 1. Under H_0,
$E[N] = 16 \times 0.7899 + 32 \times (1 - 0.7899) = 19.36$ and under H_1,
$E[N] = 16 \times 0.0651 + 32 \times (1 - 0.0651) = 30.96$.

(d) Suppose, instead, we perform a single stage trial with $N = 32$ subjects and we reject H_0 if we observe more than 9 successes. Find the type I and type II error rates. What is the advantage of the two-stage trial?

Under H_0, probability of rejection (type I error) is $\Pr\{y > 9\} = 1 - 0.9844 = .0156$, and under H_1 the acceptance probability (type II error) is $\Pr\{y \leq 9\} = 0.1155$. The advantage of the 2 stage trial is that we can have the potential to stop earlier and possibly discard ineffective treatments sooner.

2. The dataset `data3.csv` contains data collected from a crossover study with 40 subjects per sequence. The variables in the dataset are:

`seq`	Assigned treatment sequence
`y`	Response
`id`	Subject id
`period`	Period
`z`	Treatment ("A" or "B")

(a) Calculate the means within each treatment group separately for periods 1 and 2. Using these means, calculate the estimate of the treatment difference assuming no carryover.

```
> m <- with(data, tapply(y, list(z, period), mean))
> m
        1       2
A  9.6625 22.7025
B 10.1000 22.9200
# difference in "AB" group:
> m[2,2]-m[1,1]
[1] 13.2575
# difference in "BA" group:
> m[2,1]-m[1,2]
[1] -12.6025
# mean difference:
> (m[2,2]-m[1,1] + m[2,1]-m[1,2])/2
[1] 0.3275
```

Note that the estimate of period effect is:

```
> (m[2,2]-m[1,1] - (m[2,1]-m[1,2]))/2
[1] 12.93
```

(b) Fit a regression model that estimates the treatment difference and its standard error.

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

Coefficients:
              Estimate Std. Error   t value Pr(>|t|)
(Intercept) -3.409e+00  6.695e-01    -5.091 2.41e-06 ***
zB           3.275e-01  1.412e-01     2.320 0.022947 *
period       1.293e+01  1.412e-01    91.602  < 2e-16 ***
ids02        1.750e+00  8.927e-01     1.960 0.053535 .
...
ids80       -1.400e+00  8.927e-01    -1.568 0.120878
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
```

The estimate of treatment effect is 0.3275 with SE 0.1412. The p-value is 0.0229.

(c) Fit a regression model to estimate the effect of treatment using only period 1. (This is equivalent to a parallel group trial in which subjects are assigned only one of "A" or "B.")

```
> summary(lm(y ~ z, data=data, subset=period==1))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.6625     0.4734  20.412   <2e-16 ***
zB            0.4375     0.6695   0.654    0.515
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(d) Comment on the differences between the analyses in parts (b) and (c).

The standard errors of the two estimates are quite different

- Cross-over analysis: SE = 0.1412
- Parallel group (period 1 only): SE = 0.6695

This suggests that there is high correlation between the period 1 and period 2 observations from each subject. Because the cross-over model is based on within-subject differences, the subject-level effects are accounted for and the variability is significantly reduced providing greater power.

Even though the point estimate of the difference is larger in period 1 analysis, the increased variance results in a statistically in-significant difference.

3. Suppose we conduct a no-inferiority trial with a binary outcome (failure/success) and probabilities of failure π_C and π_E in the control and experimental arms, respectively. The investigators believe that if the control failure rate were $\pi_C = .13$, a non-inferiority margin of .1 is acceptable. Letting $\Delta = \pi_E - \pi_C$, this corresponds to null and alternative hypotheses of:

$$H_0^A: \Delta \geq 0.1 \quad \text{and} \quad H_1^A: \Delta < 0.1$$

Furthermore, the rates above correspond approximately to null and alternative hypotheses in terms of the odds ratio $\psi = \pi_E(1-\pi_C)/(1-\pi_E)\pi_C$ of:

$$H_0^B: \psi \geq 2 \quad \text{and} \quad H_1^B: \psi < 2.$$

After enrolling 240 subjects per group, we observe the following table:

	Failure	Success	Total
C	20	220	240
E	27	213	240

(a) Verify that when $\pi_C = .13$, the two hypotheses H_0^A and H_0^B are equivalent.

If $\pi_C = 0.13$, then under H_0^A, $\pi_E \geq \pi_C + 0.1 = 0.23$. Under H_0^B, $\pi_E/(1-\pi_E) \geq 2\pi_C/(1-\pi_C) = 0.2989$. Solving for π_E we have $\pi_E \geq 0.2989/(1-.2989) = 0.23$.

(b) Calculate the observed difference $\widehat{\Delta}$ and a 95% confidence interval. Can we reject H_0^A?

- $\widehat{\pi}_C = 20/240 = 0.0833$, $\text{Var}(\widehat{\pi}_C) = \pi_C(1-\pi_C)/240$ which we estimate as $20 \times 220/240^3 = 0.0003183$.
- $\widehat{\pi}_E = 27/240 = 0.1125$, $\text{Var}(\widehat{\pi}_C) = \pi_C(1-\pi_C)/240$ which we estimate as $27 \times 213/240^3 = 0.0004160$.

The 95% CI for Δ is $0.1125 - 0.08333 \pm 1.96\sqrt{0.0003183 + 0.0004160} = (-0.0239, 0.0823)$. This confidence interval excludes $\Delta \geq 0.1$, so we can reject H_0^A.

(c) Calculate the observed odds ratio $\widehat{\psi}$ and a 95% confidence interval. Can we reject H_0^B?

The observed odds ratio is $\widehat{\psi} = 27 \times 220/20 \times 213 = 1.394$. The variance of $\log \widehat{\psi}$ is

$$\frac{1}{20} + \frac{1}{27} + \frac{1}{220} + \frac{1}{223} = 0.09628$$

The 95% CI for $\log \widehat{\psi}$ is $\log \widehat{\psi} \pm 1.96\sqrt{\text{Var}(\log \widehat{\psi})} = \log 1.394 \pm 1.96\sqrt{0.09628} = (-0.2757, 0.9406)$. The CI for ψ is $(0.759, 2.56)$ which overlaps $\psi \geq 2$. Therefore we cannot reject H_0^B.

Notes: This problem illustrates the importance of scale in non-inferiority trials. The observed control group rate is quite different than the rate for which the two null hypotheses were "equivalent." When rates are low, the variance of the observed difference in rates goes down ($\pi(1-\pi)$ decreases), so the confidence intervals are narrower and it's easier to show non-inferiority. Conversely, when rates are low, the variance of $\widehat{\psi}$ increases (the reciprocals of the cell entries goes up more in the small cells than they go down in the large cells), so the CI for $\widehat{\psi}$ is wider, making it more difficult to show non-inferiority. For these and other reasons, hypothesis testing in non-inferiority trials if fraught with difficulties not present in superiority trials.