Math 3070 § 1. Treibergs | **Heart Attack Example: Statistic for the Ratio of Two Proportions** | Name: Example July 24, 2011

Today's example was motivated by problem 9.53 of Devore, *Probability and Statistics for Engineering and the Sciences, 5nd ed.*, Brooks Cole 2000, that discusses a statistic for the ratio of two proportions. In medical research, it is sometimes more interesting to know $\theta = p_1/p_2$, the ratio of two proportions, rather than the difference $p_1 - p_2$. We ask, how much larger is the incidence of heart attack with no treatment than the incidence of those given aspirin?

Devore describes the data which comes from a study reported in the *New York Times,* 1987. The study compared the number of heart attacks in a randomized control group given a placebo vs. the number of heart attacks in the group given aspirin. Of the $m = 11,034$ subjects in the placebo group, $x = 189$ developed heart attacks. But of the $n = 11,037$ subjects in the aspirin group, $y = 104$ developed heart attacks. By what factor is the incidence of heart attacks reduced? Find a .005 confidence interval for θ. An estimator for θ is

$$\hat{\theta} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{x/m}{y/n} = \frac{xn}{my}$$

which works out to be 1.818.

Devore describes the statistic that works for large samples. $\log(\hat{\theta})$ has approximately a normal distribution, with approximate mean $\log(\theta)$ and standard deviation

$$\hat{s} = \sqrt{\frac{\hat{q}_1}{\hat{p}_1 m} + \frac{\hat{q}_2}{\hat{p}_2 n}} = \sqrt{\frac{m-x}{mx} + \frac{n-y}{ny}}$$

where $\hat{q}_i = 1 - \hat{p}_i$. Thus, assuming that $\theta = \theta_0$, the normalized statistic is

$$zQ = \frac{\log\left(\frac{xn}{ym}\right) - \log\theta_0}{\sqrt{\frac{m-x}{mx} + \frac{n-y}{ny}}}.$$

To see how this behaves, we take $B = 10,000$ samples $x \sim \text{Binomial}(m, p_1)$ and $y \sim \text{Binomial}(n, p_2)$, compute zQ and tabulate. The histogram agrees with the standard normal curve. In our simulation $p_1 = .6$, $p_2 = .7$, $m = 150$ and $n = 100$.

Now, using the asymptotic normality, we obtain α confidence intervals for $\log(\theta)$. If z_α is chosen so that $\Phi(z_\alpha) = 1 - \alpha$, then the lower and two-sided CI for $\log(\theta)$ is

$$(\log(\hat{\theta}) - z_\alpha \hat{s}, \infty);$$
$$(\log(\hat{\theta}) - z_{\alpha/2}\hat{s}, \log(\hat{\theta}) + z_{\alpha/2}\hat{s})$$

In case $\alpha = .001$, this works out to be $(0.2226368, \infty)$. Taking exponentials, we are $1-\alpha$ confident that θ satisfies the CI, $1.219 < \theta$. The $\alpha = .05$ CI is $1.489 < \theta$. Taking aspirin reduces incidence of heart attack by 30% $(= 1 - 1/1.489)$.

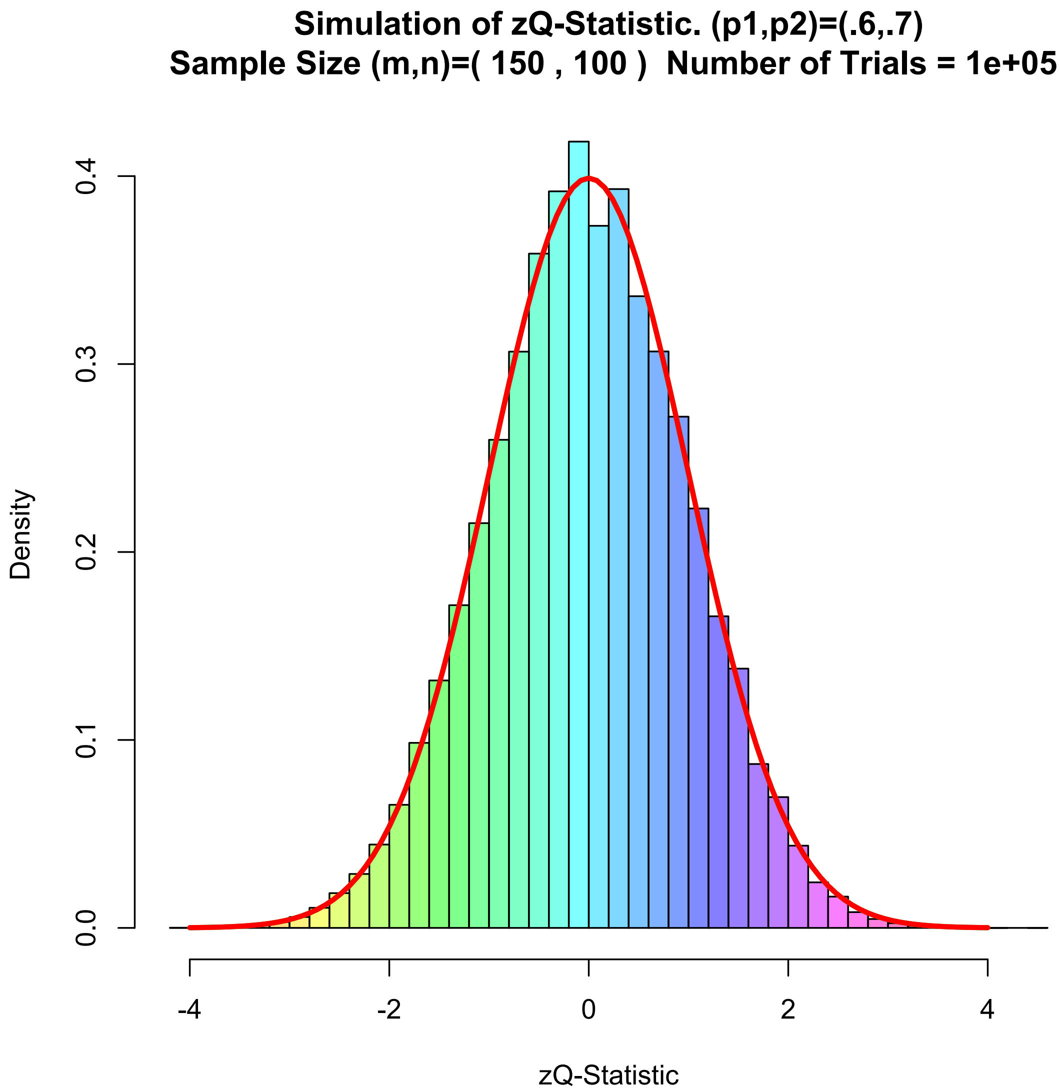
R Session:

```
R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
```

```
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.41 (5874) i386-apple-darwin9.8.0]

[History restored from /Users/andrejstreibergs/.Rapp.history]

> ############### SIMULATE zQ STATISTIC ########################
> n <- 100
> m <- 150
> st <- function(x,y){sqrt((m-x)/(m*x)+(n-y)/(n*y))}
> mu <- log(.6/.7)
> mu
[1] -0.1541507

> s <- sqrt((1-.6)/(.6*m)+(1-.7)/(.7*m))
> s
[1] 0.08544933

> # Function to take two random binomial variables and compute zQ
> zl <- function(z){
+                   x <- rbinom(1,m,.6)
+                   y <- rbinom(1,n,.7)
+                   (log(x*n/(m*y))-mu)/st(x,y)
+                  }
>
> B <- 100000
> hist(replicate(B, zl(1)), breaks = 40, freq = F,
+ xlab = "zQ-Statistic", main =
+ paste("Simulation of zQ-Statistic. (p1,p2)=(.6,.7)\n
+ Sample Size (m,n)=(", m, ",", n, ")  Number of Trials =", B),
+ col = rainbow(40, alpha = .5))
> curve(dnorm(x), -4, 4, add = T, lwd = 3, col = 2)
> # M3074HeartAttack1.pdf
```

```
> ############### HEART ATTACK DATA #############################
> # heart Attack data: Placebo
> x <- 189
> m <- 11034
> # after aspirin:
> y <- 104
> n <- 11037
> # zQ statistic
> mu=0
> zQ <- (log(x*n/(m*y))-mu)/st(x,y)
> ZQ
[1] 4.92494

> # P-value for two sided test with theta0=1
> pvalue=pnorm(zQ, 0, 1, lower.tail = F)*2
> pvalue
[1] 8.438621e-07

> s <- st(x,y)
> alpha <- .001
> za2 <- qnorm(alpha/2,0,1,lower.tail=F); za2
[1] 3.290527
> za <- qnorm(alpha,0,1,lower.tail=F); za
[1] 3.090232
>
> lthetahat <- log(x*n/(m*y)); lthetahat
[1] 0.597628
>
> # Lower CI
> c(-Inf, lthetahat+za*s)
[1]      -Inf 0.9726192
> # 2-sided .005 CI
> c(lthetahat-za2*s, lthetahat+za2*s)
[1] 0.1983316 0.9969244
>
> # CI on p1/p2
> # Lower CI
> c(thetahat-za*s,Inf)
[1] 0.2226368       Inf
> # 2-sided .005 CI
> c(thetahat-za2*s,thetahat+za2*s)
[1] 0.1983316 0.9969244
>
> # CI on p1/p2
> # Lower CI
> c(exp(thetahat-za*s),Inf)
[1] 1.249367      Inf
> # 2-sided .005 CI
> c(exp(thetahat-za2*s),exp(thetahat+za2*s))
[1] 1.219367 2.709934
```

```
> ############### alpha=.05 CI's ################################
> alpha <- .05
> za <- qnorm(alpha,0,1,lower.tail=F); za
[1] 1.644854

> za2 <- qnorm(alpha/2,0,1,lower.tail=F); za2
[1] 1.959964

> # Lower CI
> c(thetahat-za*s,Inf)
[1] 0.3980295       Inf

> # 2-sided .05 CI
> c(thetahat-za2*s,thetahat+za2*s)
[1] 0.3597917 0.8354642

> # CI on p1/p2
> # Lower CI
> c(exp(thetahat-za*s),Inf)
[1] 1.488888      Inf

> # 2-sided .05 CI
> c(exp(thetahat-za2*s),exp(thetahat+za2*s))
[1] 1.433031 2.305884
```


Simulation of zQ-Statistic. (p1,p2)=(.6,.7)
Sample Size (m,n)=(150 , 100) Number of Trials = 1e+05
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
zQ-Statistic