Aspirin and Myocardial Infarction

Greg Johnson

8/9/2017

Introduction

We have the following data as the result of a simple clinical trial on the efficacy of an Aspirin regimen on likelihood of having a heart attack (myocardial infarction).

```
tab <- matrix(c(189, 104, 10845, 10933), dimnames = list(Group = c("Placebo",
    "Aspirin"), `Myocardial Infarction` = c("Yes", "No")), 2)
n <- sum(tab)
tab
```

```
##          Myocardial Infarction
## Group     Yes    No
##   Placebo 189 10845
##   Aspirin 104 10933
```

There are a variety of tests that may be used to test the independence of our two variables. One measure worth exploring is Yule's Q (which may be generalized to Kruskal's γ in ordinal variables). It is a measure of independence that is monotonically related to the odds ratio. The statistical translation of our research hypothesis using these two statistics is: $H_0 : Q = 0$ or equivalently, $H_0 : \theta = 1$.

Asymptotic Tests of Estimators

We can appeal to the asymptotic null distributions of these two estimators or we can resample the data to create an empirical null sampling distribution for each statistic. First let's use the theoretical sampling distributions.

$$\hat{Q} \sim N(Q, (\sum_i \sum_j \pi_{ij}^{-1})(1 - Q^2)^2/4) \text{ as n} \to \infty$$

$$log\hat{\theta} \sim N(log\theta, \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}) \text{ as n} \to \infty$$

```
Q.observed <- (tab[1, 1] * tab[2, 2] - tab[1, 2] * tab[2, 1])/(tab[1, 1] * tab[2,
    2] + tab[1, 2] * tab[2, 1])
OR.observed <- -(Q.observed + 1)/(Q.observed - 1)

# p-value for Q
tab_pi <- tab/n  #table of estimated joint probabilities
Q.SE <- sqrt(sum(1/tab_pi) * (1 - Q.observed^2)^2/4)
Q.null <- 0
z.approx.Q <- (sqrt(n) * (Q.observed - Q.null))/Q.SE
(p.Q <- 2 * pnorm(z.approx.Q, lower.tail = FALSE))
```

```
## [1] 1.647426e-07
```

```
# p-value for odds ratio
logOR.SE <- sqrt(sum(1/tab))
logOR.null <- 0
```

```r
z.approx.OR <- (log(OR.observed) - logOR.null)/logOR.SE
(p.OR <- 2 * pnorm(z.approx.OR, lower.tail = FALSE))
```

```
## [1] 8.281927e-07
```

Using the theoretical sampling distributions, our null hypothesis is rejected! Assuming there is no association between Aspirin and Myocardial Infarction, the probability of getting a Q and Odds ratio as large as the ones we observed in our sample size is so small ($p_Q < 0.05, p_{OR} < 0.05$), that we can confidently claim that there is indeed an association.

Resampling Tests of Estimators

Now let's compare these p-values to those obtained from resampling the data and building an empirical null sampling distribution.

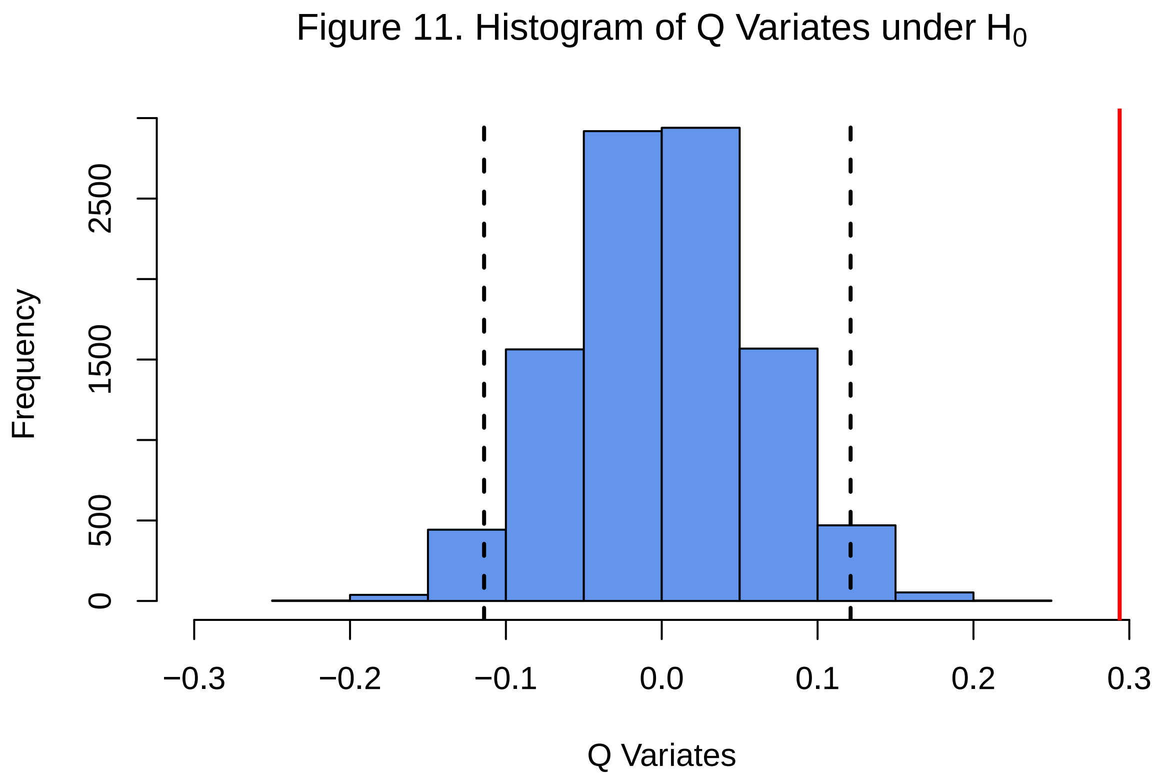
```r
# create ungrouped data from grouped data
taby <- apply(tab, 1, sum)
tabx <- apply(tab, 2, sum)
y <- NULL  #group
for (i in 1:length(taby)) {
    y <- c(y, rep(names(taby)[i], taby[i]))
}
x <- NULL  #heart attack
for (i in 1:length(tabx)) {
    x <- c(x, rep(names(tabx)[i], tabx[i]))
}

Q.resampled <- numeric(10000)
OR.resampled <- numeric(10000)
for (i in 1:10000) {
    y.s <- sample(y)
    tab.s <- table(x, y.s)
    Q.resampled[i] <- (tab.s[1, 1] * tab.s[2, 2] - tab.s[1, 2] * tab.s[2, 1])/(tab.s[1,
        1] * tab.s[2, 2] + tab.s[1, 2] * tab.s[2, 1])
    # directly transform Q into Odds Ratio
    OR.resampled[i] <- -(Q.resampled[i] + 1)/(Q.resampled[i] - 1)
}

# Q hist, p, ci
(p.Q.resampled <- sum(Q.observed <= Q.resampled)/10000)
```

```
## [1] 0
```

```r
(Q.ci <- quantile(Q.resampled, prob = c(0.025, 0.975)))
```

```
##       2.5%      97.5%
## -0.1139878  0.1211735
```

```r
hist(Q.resampled, col = "cornflowerblue", xlim = c(-0.3, 0.3), xlab = "Q Variates",
    main = expression("Figure 11. Histogram of Q Variates under" ~ H[0]))
abline(v = Q.observed, col = "red", lwd = 2)
abline(v = Q.ci, col = "black", lwd = 2, lty = 2)
```



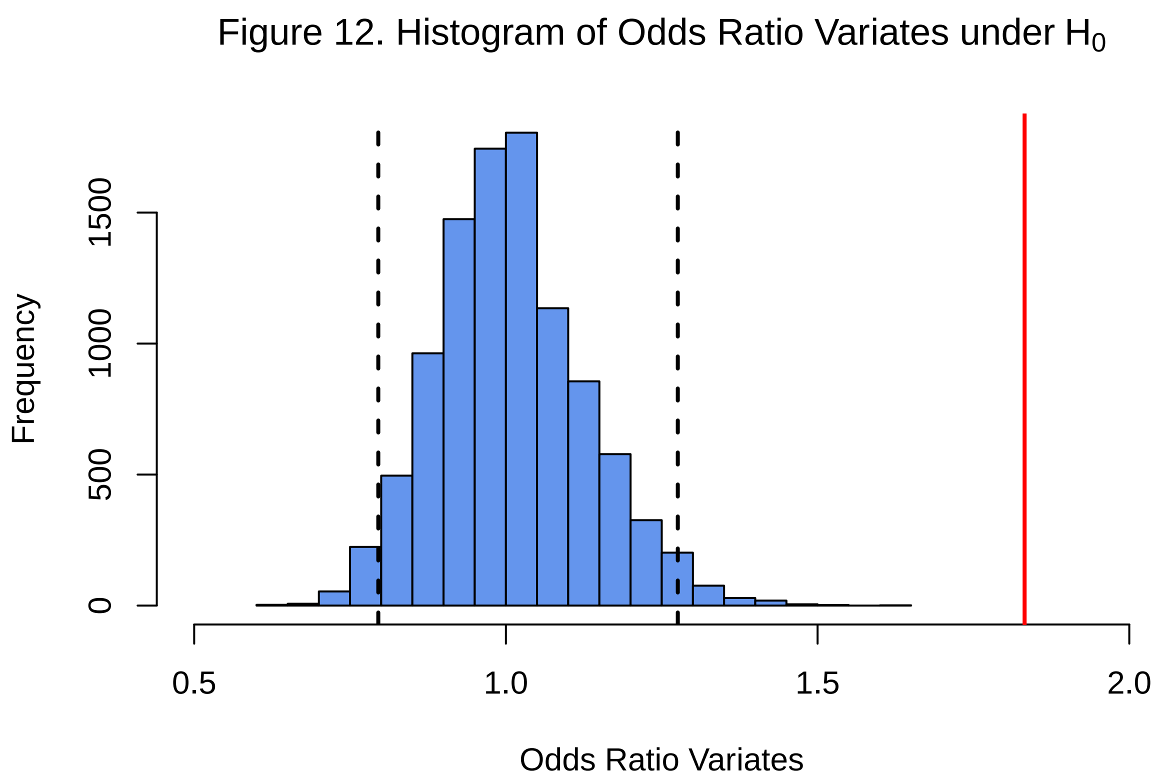

```r
# Odds ratio hist, p ,ci
(p.OR.resampled <- sum(OR.observed <= OR.resampled)/10000)
```

```
## [1] 0
```

```r
(OR.ci <- quantile(OR.resampled, prob = c(0.025, 0.975)))
```

```
##      2.5%     97.5%
## 0.7953517 1.2757622
```

```r
hist(OR.resampled, col = "cornflowerblue", xlim = c(0.5, 2), xlab = "Odds Ratio Variates",
    main = expression("Figure 12. Histogram of Odds Ratio Variates under" ~
        H[0]))
abline(v = OR.observed, col = "red", lwd = 2)
abline(v = OR.ci, col = "black", lwd = 2, lty = 2)
```



Our resampled p-values are in agreement ($p_Q < 0.05, p_{OR} < 0.05$). Our 95% confidence intervals tell us the same story in a different way. They tell us where we would expect our Q's and Odds Ratios would fall 95% of the time if the null hypothesis was true. Since our observed Q and Odds Ratio fall so far out of the 95% confidence interval, we can say that they're sufficiently rare events under the null hypothesis and thus the null hypothesis probably isn't true and there actually is an association between aspirin and myocardial infarction.

Conclusion

The odds ratio gives us a nice interpretable effect size. Odds of a myocardial infarction increase a whole 83% when a placebo is used instead of aspirin.

$$\frac{\text{Odds}_{\text{placebo}}}{\text{Odds}_{\text{aspirin}}} = 1.83 \;\rightarrow\; \text{Odds}_{\text{placebo}} = 1.83 \cdot \text{Odds}_{\text{aspirin}}$$