

Chapter 5: Cochran's Q

Giulia Bertoldo

4/9/2022

Based on the book “Doing Meta-analysis with R”

Random effect model, one equation:

- $\hat{\theta}_j = \hat{\theta} + u_j + r_j$

Random effect model, two equations:

- $\hat{\theta}_j = \theta_j + r_j$

- $\theta_j = \hat{\theta} + u_j$

Cochrane's Q:

- $Q = \sum_{k=1}^K w_k (\hat{\theta}_k - \hat{\theta})^2$

Simulation

We want to inspect how Q behaves under two different scenarios:

1. When there is no between-study heterogeneity
2. When heterogeneity exists

Scenario 1: No between-study heterogeneity

In this case $u_j = 0$ and that the residuals $(\hat{\theta}_k - \hat{\theta})$ are only a product of the sampling error r_j .

We are going to draw a number of studies, for example $K = 40$, from a distribution. Because we assume no heterogeneity between studies (only sampling variance), the studies are drawn from the very same distribution. Here we model the deviations between a study's effect size and the overall effect size as a Standard Normal distribution (mean 0 and standard deviation 1): $(\hat{\theta}_k - \hat{\theta}) \sim N(0, 1)$.

```
set.seed(123)
rnorm(n = 40, mean = 0, sd = 1)
```

```
## [1] -0.56047565 -0.23017749 1.55870831 0.07050839 0.12928774 1.71506499
## [7] 0.46091621 -1.26506123 -0.68685285 -0.44566197 1.22408180 0.35981383
## [13] 0.40077145 0.11068272 -0.55584113 1.78691314 0.49785048 -1.96661716
## [19] 0.70135590 -0.47279141 -1.06782371 -0.21797491 -1.02600445 -0.72889123
## [25] -0.62503927 -1.68669331 0.83778704 0.15337312 -1.13813694 1.25381492
## [31] 0.42646422 -0.29507148 0.89512566 0.87813349 0.82158108 0.68864025
## [37] 0.55391765 -0.06191171 -0.30596266 -0.38047100
```

Now let us repeat this process of drawing n=40 samples many, many times:

```
set.seed(123)
error_fixed <- replicate(n=10000, rnorm(n = 40)) # rnorm is by default standard normal
```

Scenario 2: There is between-study heterogeneity

We want to simulate the scenario where between-study heterogeneity (u_j errors) exists in addition to the sampling error (r_j).

Let us assume that the true effect sizes of the studies have been sampled from a standard normal distribution. Because there is this added component of variability, we add an `rnorm` call to the previous code.

```
set.seed(123)
error_random <- replicate(n=10000, rnorm(n = 40) + rnorm(n = 40))
```

Cochrane's Q

Here we are interested in evaluating the behavior of Q in the two scenarios. Then, we need to simulate many Q statistics for the scenario with heterogeneity and without heterogeneity.

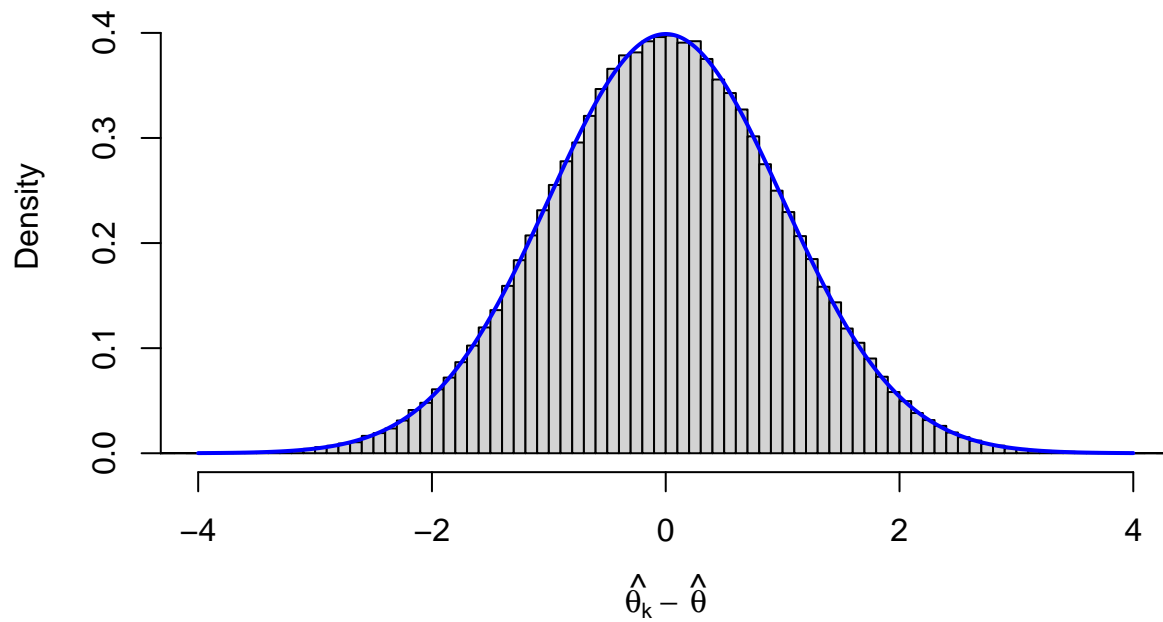
Assumption: to simplify, we assume that all studies have the same weight with value 1: $w_k = 1$.

```
set.seed(123)
Q_fixed <- replicate(n=10000, sum(rnorm(n = 40)^2))
Q_random <- replicate(n=10000, sum((rnorm(n = 40) + rnorm(n = 40))^2))
```

Important: Q is assumed to follow a χ^2 distribution with k-1 degrees of freedom **if effect size differences are only caused by sampling error**. So, the mean of the χ^2 distribution with k-1 degrees of freedom tells us the value of Q that we expect through sampling error alone.

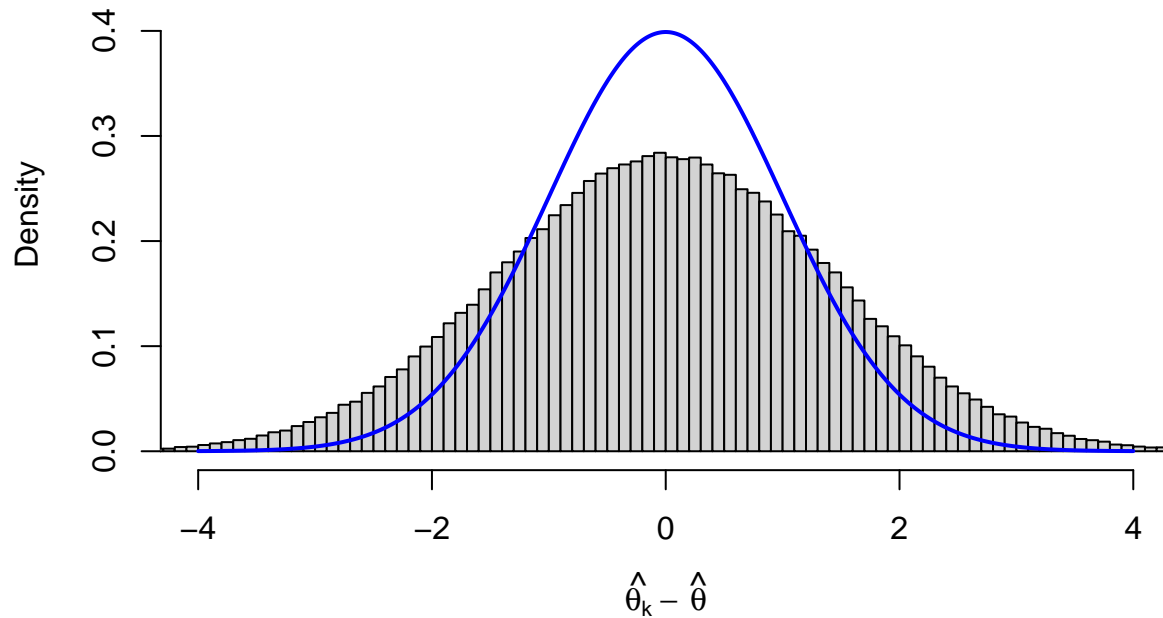
```
# Histogram of the residuals (theta_k - theta)
# - We produce a histogram for both the simulated values in
#   error_fixed and error_random
# - `lines` is used to add a normal distribution in blue.
hist(error_fixed,
     xlab = expression(hat(theta[k]) ~~~ hat(theta)), prob = TRUE,
     breaks = 100, ylim = c(0, .45), xlim = c(-4,4),
     main = "No Heterogeneity")
lines(seq(-4, 4, 0.01), dnorm(seq(-4, 4, 0.01)),
     col = "blue", lwd = 2)
```

No Heterogeneity



```
hist(error_random,
      xlab = expression(hat(theta[k]) - hat(theta)), prob = TRUE,
      breaks = 100, ylim = c(0, .45), xlim = c(-4, 4),
      main = "Heterogeneity")
lines(seq(-4, 4, 0.01), dnorm(seq(-4, 4, 0.01)),
      col = "blue", lwd = 2)
```

Heterogeneity

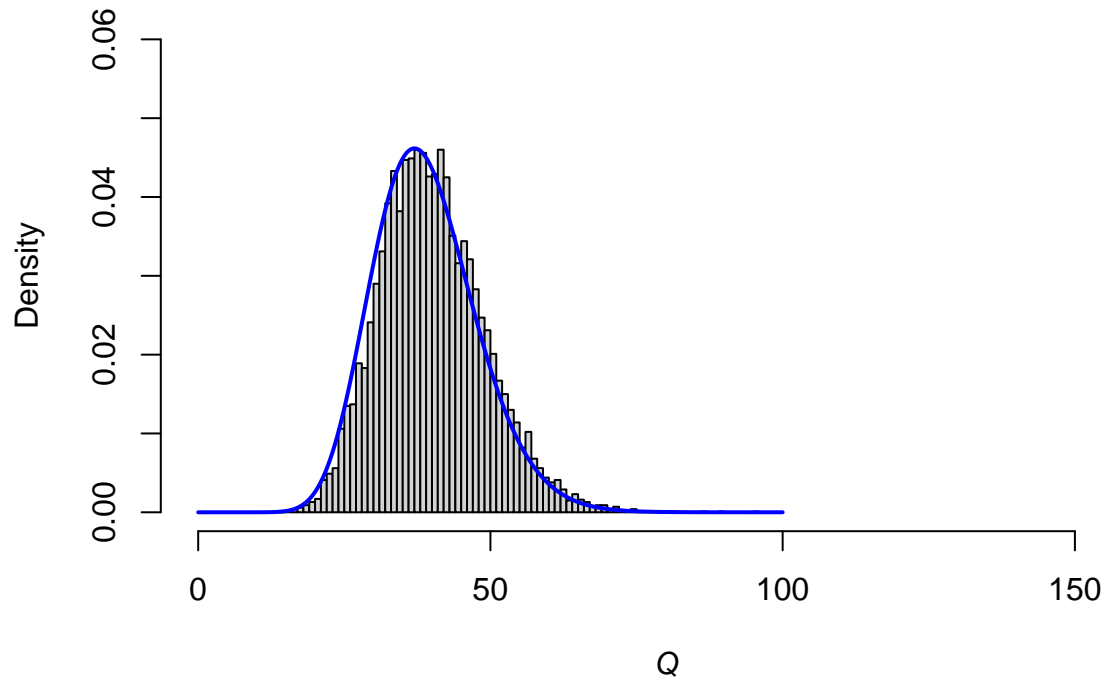


When there is no heterogeneity the ‘residuals’ ($\hat{\theta}_k - \hat{\theta}$) follow a standard normal distribution (because that is the distribution we used to generate the data).

When there is heterogeneity the distribution departs from the standard normal.

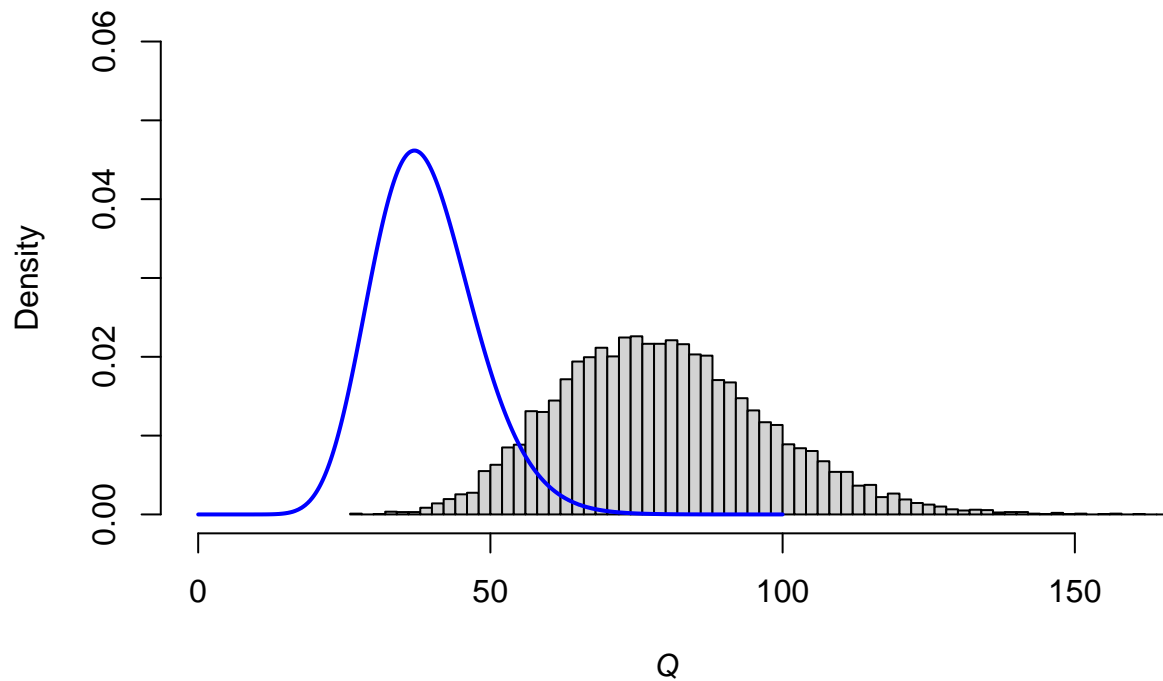
```
# Histogram of simulated Q-values
# - We produce a histogram for both the simulated values in
#   Q_fixed and Q_random
# - `lines` is used to add a chi-squared distribution in blue.
# First, we calculate the degrees of freedom (k-1)
# remember: k=40 studies were used for each simulation
df <- 40-1
hist(Q_fixed, xlab = expression(italic("Q")), prob = TRUE,
     breaks = 100, ylim = c(0, .06), xlim = c(0, 160),
     main = "No Heterogeneity")
lines(seq(0, 100, 0.01), dchisq(seq(0, 100, 0.01), df = df),
     col = "blue", lwd = 2)
```

No Heterogeneity



```
hist(Q_random, xlab = expression(italic("Q")), prob = TRUE,
     breaks = 100, ylim = c(0, .06), xlim = c(0,160),
     main = "Heterogeneity")
lines(seq(0, 100, 0.01), dchisq(seq(0, 100, 0.01), df = df),
     col = "blue", lwd = 2)
```

Heterogeneity



If we have no heterogeneity, the distribution of Q values follows a Chi-Square distribution with $k-1$ degrees of freedom. In this case variation in the observed effect sizes exists only because of sampling error.

If we do have heterogeneity, Q departs from the Chi-Square distribution with $k-1$ degrees of freedom. The variation that we observed is not due to sampling error alone.