STA 5364, Report 3.2

Carson Slater *Baylor University*

Problem

When the large-sample approximation is applied in CML data example above, the approximate p-value is 0.07. Confirm this.

We have that the large sample approximation is,

$$R^* = \frac{R - \mathbb{E}_0(R)}{\sqrt{\mathbb{V}_0(R)}},$$

where

$$\mathbb{E}_0(R) = \frac{n(n-1)(n-2)}{8},$$

and

$$\mathbb{V}_0(R) = \frac{3}{2}n(n-1)(n-2)\left[\frac{5}{2591}(n-3)(n-4) + (n-3)\frac{7}{432} + \frac{1}{48}\right].$$

Then, we have that applying this to the CML data yields the following result:

```
cml <- c(
  7, 47, 58, 74, 177, 232, 273, 285, 317, 429, 440, 445,
  455, 468, 495, 497, 532, 571, 579, 581, 650, 702, 715,
  779, 881, 900, 930, 968, 1077, 1109, 1314, 1334, 1367,
  1534, 1712, 1784, 1877, 1886, 2045, 2056, 2260, 2429,
  2509
)

R <- 8324
n <- length(cml)
ER <- n*(n-1)*(n-2)/8
VR <- (3/2)*n*(n-1)*(n-2)*((5/2592)*(n-3)*(n-4) + (n-3)*(7/432) + 1/48)

(R_star <- (R - ER)/sqrt(VR))
```

```
## [1] -1.457749
```

```
pnorm(R_star)
```

```
## [1] 0.0724548
```

Hence the large sample approximation yields a p-value of approximately 0.07.