

**Exercise 1:** Conduct the quadrat test for complete spatial randomness on the cells data set. Carry out the analysis using a 2x2 grid, a 3x3 grid, and a 4x4 grid. Is there evidence of regularity or clustering on one scale but not the others? Use  $\alpha = .05$ , and carry out just 2-sided test.

Please summarize your results, including a table of counts for each for each of the grid sizes and data sets.

the cell data is included in the spatstat package.

```
library(spatstat)
data(cells)
```

Sample code, which you'll need to modify:

```
tmp <- quadratcount(cells, nx=2, ny=2 )
tmp
plot(tmp)
quadrat.test(cells, nx=2, ny=2, alternative="two.sided")
```

### Solution:

Let's recall a few ideas behind conducting a formal test for CSR using quadrat data. First note that we obtain counts  $y_i$  for each of the  $m$  quadrats, and remember that as a rule of thumb it is suggested that we have  $m > 6$  and  $y_i > 5$ . Our test statistic, the index of dispersion (or  $\chi^2$  goodness of fit) is computed by,

$$I = \frac{(m-1)s^2}{\bar{y}}.$$

Note that  $I$  is approximately distributed by a  $\chi^2$  with  $(m-1)$  degrees of freedom. Computing the p-values for the various tests we get, two-tailed (2x the smaller direction, this was the procedure B statistic which applies distance data. I suppose it is similar for I),

$H_0$  : CSR

$H_a$  : Not CSR

right-tailed,

$H_0$  : CSR

$H_a$  : Clustered

and left-tailed,

$H_0$  : CSR

$H_a$  : Regular

Conducting the two tailed tests in R we get the following counts, p-values, and test statistics,

grid length	2	3	4
p-value	0.1518	0.3391	0.0007283
test statistic	0.47619	4.2857	2.9524
count	9	3	2
	11	6	3
	12	3	3
	8	4	2
		7	2
		6	3
		3	4
		6	3
		4	2
			2
			2
			2
			3
			4
			2
			3

**Code:**

```
-----
> tmp <- quadratcount(cells , nx=2, ny=2 )
> tmp
      x
y      [0,0.5) [0.5,1]
  [0.5,1]      9      11
  [0,0.5)     12      10
> plot(tmp)
> quadrat.test(cells , nx=2, ny=2, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: cells
X2 = 0.47619, df = 3, p-value = 0.1518
alternative hypothesis: two.sided
Quadrats: 2 by 2 grid of tiles
```

```
-----
> tmp1 <- quadratcount(cells , nx=3, ny=3 )
> tmp1
      x
y      [0,0.333) [0.333,0.667) [0.667,1]
  [0.667,1]      3      6      3
  [0.333,0.667)  4      7      6
  [0,0.333)     3      6      4
> plot(tmp1)
> quadrat.test(cells , nx=3, ny=3, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: cells
X2 = 4.2857, df = 8, p-value = 0.3391
alternative hypothesis: two.sided
```

Quadrats: 3 by 3 grid of tiles

Warning message:

Some expected counts are small; chi^2 approximation may be inaccurate

```
-----
> tmp2 <- quadratcount(cells, nx=4, ny=4)
> tmp2
```

	x			
y	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
[0.75,1]	2	2	2	2
[0.5,0.75)	2	3	4	3
[0.25,0.5)	3	4	2	3
[0,0.25)	2	3	3	2

```
> plot(tmp2)
> quadrat.test(cells, nx=4, ny=4, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: cells
X2 = 2.9524, df = 15, p-value = 0.0007283
alternative hypothesis: two.sided
```

Quadrats: 4 by 4 grid of tiles

Warning message:

Some expected counts are small; chi^2 approximation may be inaccurate

From the various quadrat tests we find that when using a 4x4 grid we conclude that the on an  $\alpha = .05$  significance level the data does not exhibit complete spatial randomness. This is somewhat expected given the low value of the counts. Considering the ratio between mean and variance for the poisson distribution, with low counts in each quadrant we will find that there is little to no variance across quadrants which suggests regularity (performing the one-sided tests corroborates this).

**Exercise 2:** Repeat problem 1 for the redwood seedlings / saplings data set. This data set is included in spatstat:

```
library(spatstat)
data(redwood)
```

**Solution:**

Conducting the two tailed tests in R we get the following counts, p-values, and test statistics,

grid length	2	3	4
p-value	0.1781	0.007333	0.0004102
test statistic	6.5161	22.774	42.258
count	10	0	0
	17	6	5
	23	13	1
	12	13	9
		8	5
		2	0
		5	7
		9	0
		6	8
			3
			4
			2
			2
			10
			1
			5

### Code:

```
-----
> tmp <- quadratcount(redwood, nx=2, ny=2 )
> tmp
      x
y      [0,0.5) [0.5,1]
[-0.5,0]      10      17
[-1,-0.5)      23      12
> plot(tmp)
> quadrat.test(redwood, nx=2, ny=2, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: redwood
X2 = 6.5161, df = 3, p-value = 0.1781
alternative hypothesis: two.sided
```

Quadrats: 2 by 2 grid of tiles

```
-----
> tmp <- quadratcount(redwood, nx=3, ny=3 )
> tmp
      x
y      [0,0.333) [0.333,0.667) [0.667,1]
[-0.333,0]       0           6       13
[-0.667,-0.333)  13           8        2
[-1,-0.667)      5           9        6
```

```
> plot(tmp)
> quadrat.test(redwood, nx=3, ny=3, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: redwood
X2 = 22.774, df = 8, p-value = 0.007333
alternative hypothesis: two.sided
```

Quadrats: 3 by 3 grid of tiles

```
-----
> tmp <- quadratcount(redwood, nx=4, ny=4 )
> tmp
```

	x			
y	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
[-0.25,0]	0	5	1	9
[-0.5,-0.25)	5	0	7	0
[-0.75,-0.5)	8	3	4	2
[-1,-0.75)	2	10	1	5

```
> plot(tmp)
> quadrat.test(redwood, nx=4, ny=4, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: redwood
X2 = 42.258, df = 15, p-value = 0.0004102
alternative hypothesis: two.sided
```

Quadrats: 4 by 4 grid of tiles

Warning message:

Some expected counts are small; chi^2 approximation may be inaccurate

Here we can see that with 4 quadrats the csr test concludes that the data exhibit complete spatial randomness, which is questionable given the small number of quadrats. When we consider the 9 and 16 csr tests we find that at an  $\alpha = .05$  level we get that the data do not exhibit complete spatial randomness. Given the variance (variance is likely larger than the mean) among the counts for the 9 and 16 quadrat csr tests we would likely conclude that the data show significant clustering.

**Exercise 3:** Repeat problem 1 for the black pine seedlings data set. This data set is posted on Canvas as the file black\_pine\_saplings.txt.

### Solution:

Importing the data, creating a ppp object, then conducting the quadrat test for csr we get the following p-values, test statistics, and counts,

grid length	2	3	4
p-value	0.6762	0.1119	0.9028
test statistic	3.3692	15.169	15
count	22	6	4
	17	15	8
	13	7	4
	13	10	4
		4	5
		3	5
		4	5
		8	4
		8	4
			1
			2
			0
			2
			6
			6
			5

**Code:**

```
> blackpine = read.table('black_pine_saplings-1.txt', header = TRUE)
> blackpine <- ppp(blackpine[,1], blackpine[,2])
```

```
-----
> tmp <- quadratcount(blackpine, nx=2, ny=2)
> tmp
```

```
      x
y      [0,0.5) [0.5,1]
[0.5,1]      22      17
[0,0.5)      13      13
```

```
> plot(tmp)
> quadrat.test(blackpine, nx=2, ny=2, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: blackpine
X2 = 3.3692, df = 3, p-value = 0.6762
alternative hypothesis: two.sided
```

Quadrats: 2 by 2 grid of tiles

```
-----
> tmp <- quadratcount(blackpine, nx=3, ny=3)
> tmp
```

```
      x
y      [0,0.333) [0.333,0.667) [0.667,1]
[0.667,1]        6          15          7
[0.333,0.667)    10          4          3
[0,0.333)        4          8          8
> plot(tmp)
```

```
> quadrat.test(blackpine, nx=3, ny=3, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: blackpine
X2 = 15.169, df = 8, p-value = 0.1119
alternative hypothesis: two.sided
```

Quadrats: 3 by 3 grid of tiles

```
-----
> tmp <- quadratcount(blackpine, nx=4, ny=4)
```

```
> tmp
```

	x			
y	[0,0.25)	[0.25,0.5)	[0.5,0.75)	[0.75,1]
[0.75,1]	4	8	4	4
[0.5,0.75)	5	5	5	4
[0.25,0.5)	4	1	2	0
[0,0.25)	2	6	6	5

```
> plot(tmp)
```

```
> quadrat.test(blackpine, nx=4, ny=4, alternative="two.sided")
```

Chi-squared test of CSR using quadrat counts

```
data: blackpine
X2 = 15, df = 15, p-value = 0.9028
alternative hypothesis: two.sided
```

Quadrats: 4 by 4 grid of tiles

Warning message:

Some expected counts are small; chi^2 approximation may be inaccurate

-----

With every grid we find that at a  $\alpha = .05$  significance level we fail to reject the null hypothesis and conclude that the data exhibit complete spatial randomness. Of the different number of quadrats, I trust the 9 quadrat test the most, since it has more than 6 quadrats with the counts, generally being greater than 5.

**Exercise 4:** Carry out a quadrat test for spatial randomness for each of the three data sets above, using a hexagonal grid with an appropriate grid size. Note the recommendations on page 172 of the lecture notes. Summarize your results including a table of counts for each data set.

You should exclude partial cells, which means you'll have to do some of the calculations yourself.

Sample code, which you'll need to modify:

```
plot(redwood)
H <- hextess(redwood, s=2) # s = side length of hexagons
plot(H)
```





Data	Redwood	Cells	Blackpine
p-value	0.249304	0.0002143529	0.1822028
test statistic	10	0.1764706	10.91304
count	5	4	11
	5	5	8
	11	5	2
	6	5	8
	11	5	6
	9	5	9
	2	5	2

**Code:**

```

-----
Hredwood <- hextess(redwood, s=.2)
plot(quadratcount(redwood, tess=Hredwood))

## Counts
Hcounts <- c(5,5,11,6,11,9,2)

m <- length(Hcounts)
ybar <- mean(Hcounts)
ssquared <- var(Hcounts)
I_Test_Statistic <- ((m-1)*ssquared)/ybar

## Computing test statistic, and two.tailed chi-squared csr test
regularpvalue <- pchisq(I_Test_Statistic, (m-1), lower.tail = TRUE)
clusterpvalue <- pchisq(I_Test_Statistic, (m-1), lower.tail = FALSE)

if (regularpvalue > clusterpvalue){
  twosidedpvalue <- 2*clusterpvalue
} else {
  twosidedpvalue <- 2*regularpvalue
}

## regularpvalue = 0.875348
## clusterpvalue = 0.124652
## twosidedpvalue = 0.249304
## I_Test_Statistic = 10

-----

Hcells <- hextess(cells, s=.2)
plot(quadratcount(cells, tess=Hcells))

## Counts
Hcounts <- c(4,5,5,5,5,5,5)

m <- length(Hcounts)
ybar <- mean(Hcounts)

```

```

ssquared <- var(Hcounts)
I_Test_Statistic <- ((m-1)*ssquared)/ybar

## Computing test statistic , and two.tailed chi-squared csr test
regularpvalue <- pchisq(I_Test_Statistic , (m-1), lower.tail = TRUE)
clusterpvalue <- pchisq(I_Test_Statistic , (m-1), lower.tail = FALSE)

if (regularpvalue > clusterpvalue){
  twosidedpvalue <- 2*clusterpvalue
} else {
  twosidedpvalue <- 2*regularpvalue
}

## regularpvalue = 0.0001071765
## clusterpvalue = 0.9998928
## twosidedpvalue = 0.0002143529
## I_Test_Statistic = 0.1764706

-----

Hblackpine <- hextess(blackpine , s=.2)
plot(quadratcount(blackpine , tess=Hblackpine))

## Counts
Hcounts <- c(11,8,2,8,6,9,2)

m <- length(Hcounts)
ybar <- mean(Hcounts)
ssquared <- var(Hcounts)
I_Test_Statistic <- ((m-1)*ssquared)/ybar

## Computing test statistic , and two.tailed chi-squared csr test
regularpvalue <- pchisq(I_Test_Statistic , (m-1), lower.tail = TRUE)
clusterpvalue <- pchisq(I_Test_Statistic , (m-1), lower.tail = FALSE)

if (regularpvalue > clusterpvalue){
  twosidedpvalue <- 2*clusterpvalue
} else {
  twosidedpvalue <- 2*regularpvalue
}

## regularpvalue = 0.9088986
## clusterpvalue = 0.0911014
## twosidedpvalue = 0.1822028
## I_Test_Statistic = 10.91304
-----

```

From the tests we found that the Redwood and Blackpine data failed to reject the null hypothesis at an  $\alpha = .05$  significance level and therefore exhibited complete spatial randomness. The tests on the Cells data showed that the at a  $\alpha = .05$  significance level the data did not exhibit complete spatial randomness. The one-sided tests on the Cells data

showed that the data was regularly spaced. interestingly the Redwood data had a higher p-value than the Blackpine data, even though when plotted the Redwood data looks to be more clustered.

**Exercise 5:** How are the p-values related to the three alternative of the quadrat test, two.sided, regular, and clustered? Answer this question by carrying out test with each of the alternatives for the cells data, using a 3x3 grid. Also, explain what the results of these three tests tell us about the cells data; When we look at the data, they look regular is this reflected in the test results? If not, why not?

**Solution:**

As stated before, our test statistic approximates a  $\chi^2$  distribution on  $n - 1$  degrees of freedom, where  $n$  is the number of quadrats. The three alternative quadrat tests come the upper, lower, and two.sided p-values of this null distribution. Carrying out the three alternative tests on the Cells data using a 3x3 grid we get the following summary outputs.

**Code:**

```
-----
> tmp <- quadratcount(cells , nx=3, ny=3 )
> tmp
      x
y      [0,0.333) [0.333,0.667) [0.667,1]
[0.667,1]      3      6      3
[0.333,0.667)  4      7      6
[0,0.333)     3      6      4
-----
> quadrat.test(cells , nx=3, ny=3, alternative="two.sided")

Chi-squared test of CSR using quadrat counts

data:  cells
X2 = 4.2857, df = 8, p-value = 0.3391
alternative hypothesis: two.sided

Quadrats: 3 by 3 grid of tiles
Warning message:
Some expected counts are small; chi^2 approximation may be inaccurate

-----
> quadrat.test(cells , nx=3, ny=3, alternative="regular")

Chi-squared test of CSR using quadrat counts

data:  cells
X2 = 4.2857, df = 8, p-value = 0.1695
alternative hypothesis: regular
```

Quadrats: 3 by 3 grid of tiles

Warning message:

Some expected counts are small; chi^2 approximation may be inaccurate

```
-----
> quadrat.test(cells, nx=3, ny=3, alternative="cluster")
```

Chi-squared test of CSR using quadrat counts

data: cells

X2 = 4.2857, df = 8, p-value = 0.8305

alternative hypothesis: clustered

Quadrats: 3 by 3 grid of tiles

Warning message:

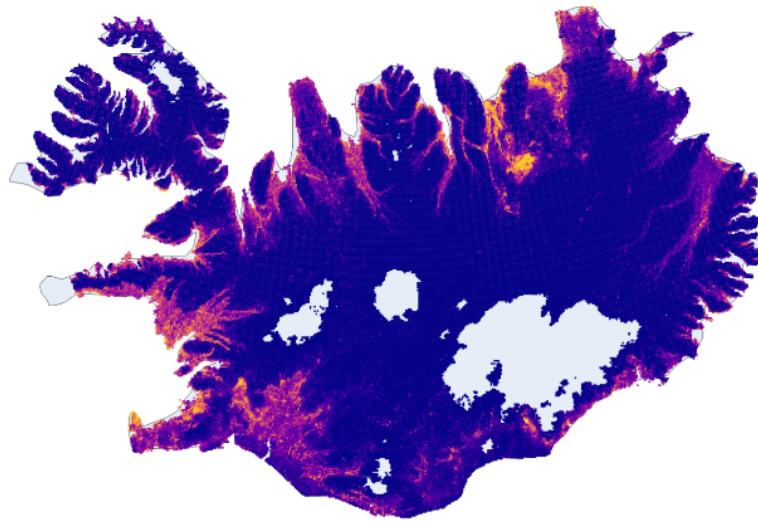
Some expected counts are small; chi^2 approximation may be inaccurate

-----

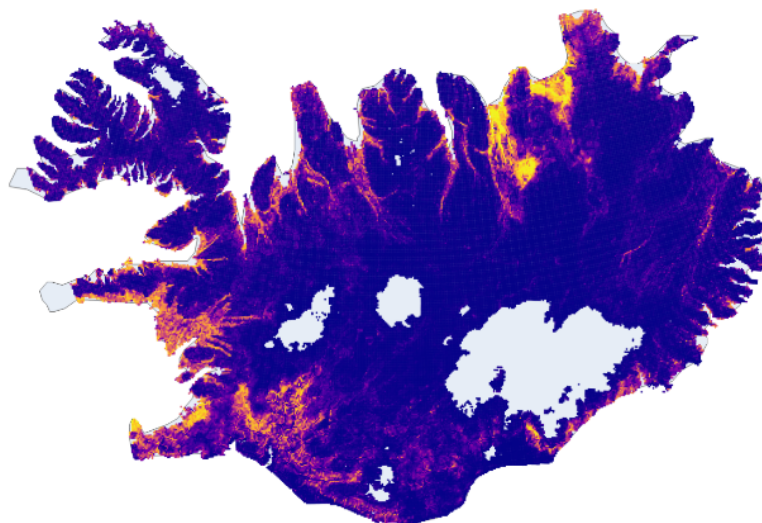
Note that the two.sided test produced a p-value of .3391 which is not low enough to reject the null of not exhibiting complete spatial randomness, despite the data looking clearly regular when plotted normally. Considering the two alternative tests for regular and clustered data we found that they got p-values of 0.1695 and 0.8305 respectively. Clearly the test statistic on the  $\chi^2$  distribution on 8 degrees of freedom is more towards the left, making the p-value for the regular test smaller, and the p-value for the clustered larger. In any case, the quadrat test is doing a poor job, because the expected counts in each quadrat is lower than what is suggested for performing this test (greater than 5). Looking at the two alternative test we might be able to surmise the the data exhibits more regularity than clustering, but is still not significantly regular to break from the assumption of complete spatial randomness. Also we can see that to compute the two sided p-value, we double the lower p-value of the other to alternative tests ( $0.3391 = 2 * 0.1695$ ).

**Project Proposal:** For my project I was thinking about mainly working with an existing set of data in the realm of ecological niche modeling. The data that I have to work with uses presence/absence of the Icelandic Rock Ptarmagin as a response, and several environmental layers like elevation, temperature, soil quality, and grass cover (as is the case with most habitat modeling endeavour). I have already experimented with making models to predict the RIO (relative index of occurrence, sometimes relative occurrence rate), this work resulted in the creation of the following maps,

Random Forest RIO Map



Gradient Boost RIO Map



The main problem that I am hoping to explore is in regard to model validation. The maps were created using Random Forest and Gradient Boosting tree models, for validation I have separate test data in the form of [presence, long, lat], but the rest of my environmental layers are not smoothed maps so using that data to test any model isn't really possible unless we perform some kriging analysis to generate the data at those long, lat pairs.