GESC 258- Lab 2 - Discrete Probability Distributions

Marking Key

Part I Assignment

1. Calculate the probability of observing a cell with 15 pits based on the sample intensity (λ computed from all cells in the image). Write a sentence interpreting what this means. Include commands used to generate the answer. (out of 2)

- 1 mark for the code and correct solution
- 1 mark for the description

```
answer <- c(4,7,9,4,5,0,4,15,9,9,10,3,7,14,16,7,12,7,11,7,15,2,9,3,1,6,14,11
  ,3,1,2,7,8,6,1,6)
dpois(x = 15, lambda = mean(answer))
```

```
## [1] 0.003637584
```

2. What is the probability of observing a cell with between 3 and 5 pits (λ computed from all cells in the image)? Write a sentence interpreting what this means. Include commands used to generate the answer. (out of 3)

- 2 mark for the code and correct solution

- 1 mark for the description

```r
p3 <- dpois(x = 3, lambda = mean(answer))
p4 <- dpois(x = 4, lambda = mean(answer))
p5 <- dpois(x = 5, lambda = mean(answer))
p3+p4+p5
```

```
## [1] 0.2623633
```

3. There are two important assumptions to using the Poisson distribution we should consider.

- The probability that an event will occur within a given unit must be the same for all units (i.e. the underlying process governing the phenomenon must be invariant)

- The number of events occurring per unit must be independent of the number of events occurring in other units (no interactions/dependencies).

Write short paragraph (200-300 words) explaining why or why not these assumptions are met in this analysis of the weathering pits dataset. (out of 5)

```r
# Students should be aware that the effect of location of each cell can play
  a role in the first assumption. The process of formation of these landscap
  es are complex and there can be underlying processing there. However since
  our sample is from one stone some of the processes can be ignored.


# The main porpuse of this question is to ask students to think about the as
  sumptions of probability models before using them.
```

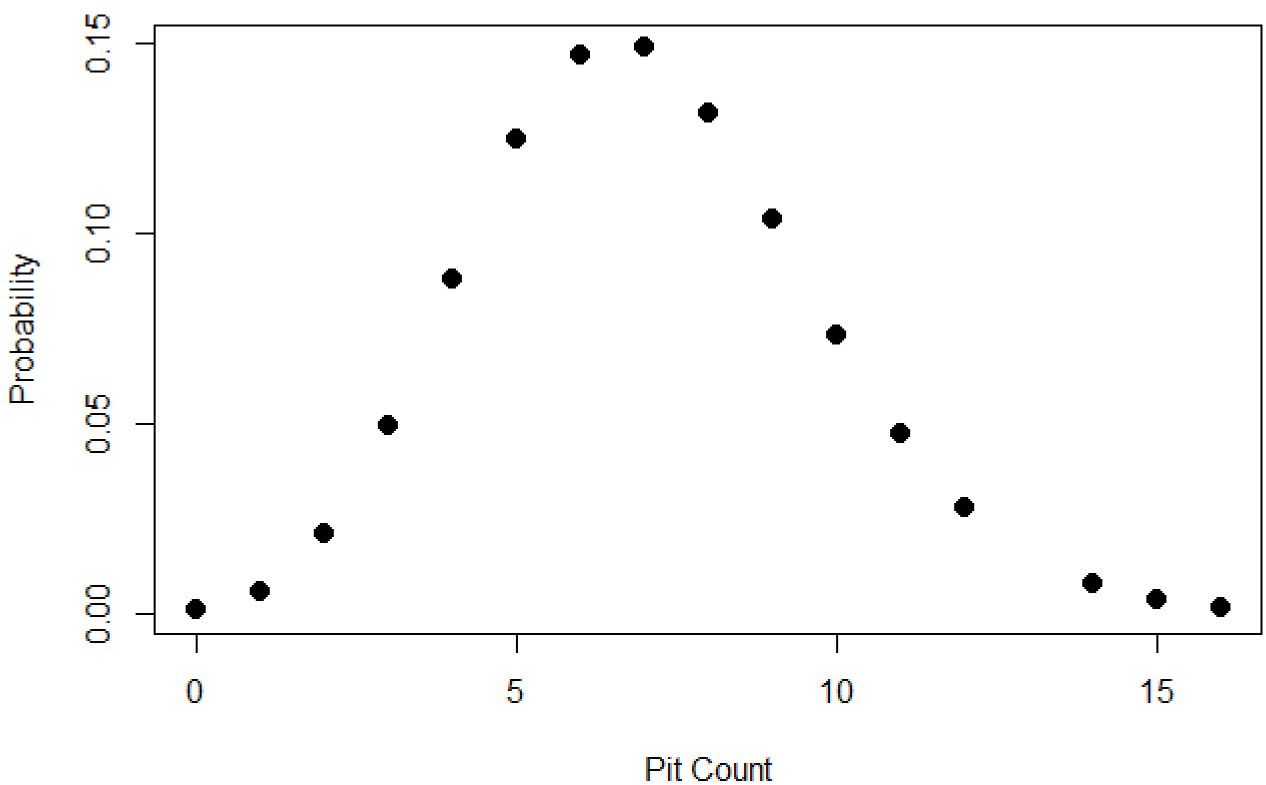
4. Explore the assumption of independence by calculating the probability of each observed count and noting where on the image any counts with a probability less than 0.10 occur. Comment on whether these are distributed randomly over the image or clumped in specific parts of the image and what this means for the independence assumption. Include an image showing which cells have an unusual (i.e., $p < 0.10$) number of pits. (out of 5)

> *Hint: to answer 4 you can take a screenshot from the lab and use MSPaint (Windows), Photos (Mac) or another graphics program to identify which cells have unusual counts.*

- 1 mark for calculating all the probabilities for all cells
- 1 mark for making image
- 3 marks on the discussion on spatial distribution of the pits

```
prop <- dpois(x = answer, lambda = mean(answer)) # This line is the answer f
  or calculating all the probabilities for all cells

# we can also plor them and show them as a table.
plot(answer, prop, pch=20, cex=2, xlab = "Pit Count", ylab = "Probability")
```




```
cbind(prop,answer)
```

```
##                prop answer
##  [1,] 0.0880008560      4
##  [2,] 0.1489294690      7
##  [3,] 0.1037823508      9
##  [4,] 0.0880008560      4
##  [5,] 0.1246678794      5
##  [6,] 0.0008389719      0
##  [7,] 0.0880008560      4
##  [8,] 0.0036375838     15
##  [9,] 0.1037823508      9
## [10,] 0.1037823508      9
## [11,] 0.0735124985     10
## [12,] 0.0496946011      3
## [13,] 0.1489294690      7
## [14,] 0.0077031186     14
## [15,] 0.0016103886     16
## [16,] 0.1489294690      7
## [17,] 0.0279423296     12
## [18,] 0.1489294690      7
## [19,] 0.0473375937     11
## [20,] 0.1489294690      7
## [21,] 0.0036375838     15
## [22,] 0.0210471252      2
## [23,] 0.1037823508      9
## [24,] 0.0496946011      3
## [25,] 0.0059427177      1
## [26,] 0.1471773576      6
## [27,] 0.0077031186     14
## [28,] 0.0473375937     11
## [29,] 0.0496946011      3
## [30,] 0.0059427177      1
## [31,] 0.0210471252      2
## [32,] 0.1489294690      7
## [33,] 0.1318646340      8
## [34,] 0.1471773576      6
## [35,] 0.0059427177      1
## [36,] 0.1471773576      6
```

Part II Assignment

1. Calculate the probability of finding a tree with a dbh of 90 or greater based on the sample mean and sample standard deviation above. What is the *z-score* associated with a dbh of 90 cm? Include commands used to generate the answer. (out of 3)

- 1 mark for the code
- 1 marks for z-score and mentioning it
- 1 mark for the final probability

```
x <- c(272,272,236,154,256,156,143,269,205,175)
dbh <- x/pi

pnorm(q=90, mean = mean(dbh), sd=sd(dbh), lower.tail = FALSE) # method 1
```

```
## [1] 0.09809276
```

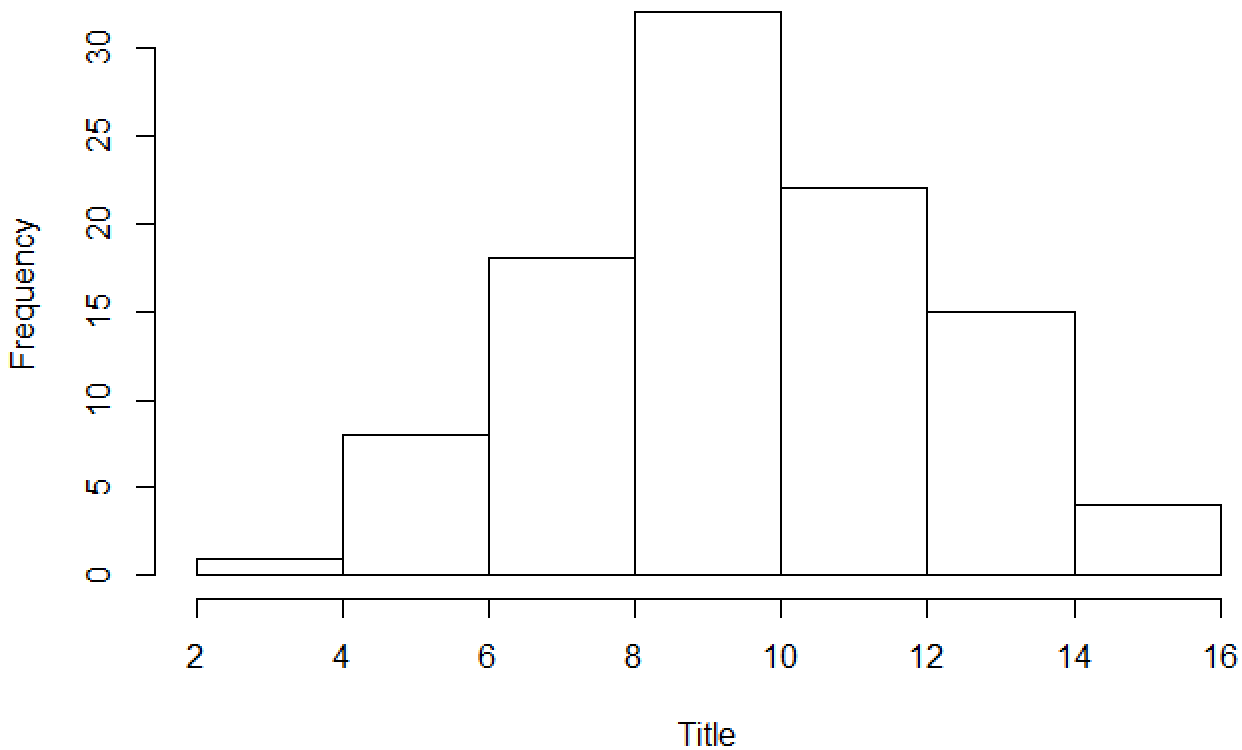
```
# using the second method
z_score <- (90 - mean(dbh)) / sd(dbh)
pnorm(z_score, lower.tail=FALSE)
```

```
## [1] 0.09809276
```

2. Create a new dataset with a mean and standard deviation of your choosing and a sample size of 100. Plot the histogram being sure to label axes appropriately. (out of 3)

- 1 mark for the code
- 2 mark for the final histogram

```
sample_mean <- 10
sample_sd <- 2.5
sim_x <- rnorm(n=100, mean=sample_mean, sd=sample_sd)
hist(sim_x, xlab = "Title", main="")
```



3. Do you think the sample of trees we collected was representative of the wider forest? What would be some potential issues with inferring characteristics of the forest from this dataset? (out of 4)