ENV 710

Poisson distribution

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

roadmap

- quick summary of 'where we are'
- review of discrete answer questions
- pod work

descriptive statistics

types of data
measures of central tendency and spread
kurtosis, skewness, modality, "outliers"

↓

discrete probability/distributions

Bernoulli random variable
binomial probability/distribution
Poisson probability/distribution

⇣

continuous probability/distributions

PDF, CDF
uniform distribution
Normal distribution

SUMMARY

discrete distributions

1. use the probability mass function to determine the probability of a specific outcome

2. $P(X = k) = \frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k}$ binomial probability mass function for binary data

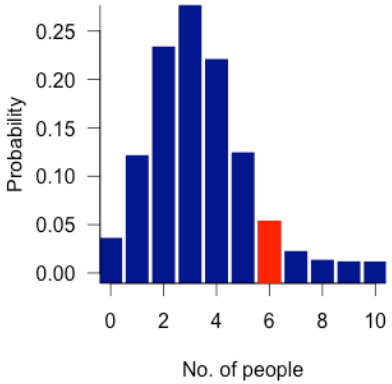
3. $\mu = np$

   $\sigma = \sqrt{np(1-p)}$

   mean and standard deviation of binomial distribution

4. `dbinom(x, size, prob)`
   `pbinom(q, size, prob)`

   `dbinom()` calculates probability and `pbinom()` calculates cumulative probability



5. $P(X) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$

   Poisson probability mass function for count data

   as λ increases the Poisson distribution looks more normally distributed (bell-shaped)

6. $\mu = \lambda$
   $\sigma^2 = \lambda$

   mean and variance of Poisson distribution

7. $E(X) = \sum_{i=1}^{n} x_i p_i$ expected value (average) of counts

8. `dpois(x, lambda)`
   `ppois(q, lambda)`

   `dpois()` calculates probability and `ppois()` calculates cumulative probability

SUMMARY

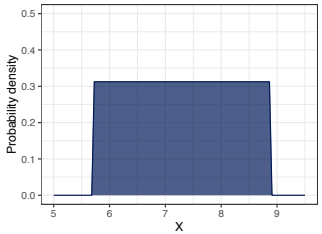
continuous distributions

linear models
linear regression

1. use the probability density function to determine the area/integral within an interval of *x* values to calculate probability

2. $$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b. \\ 0 & \text{for } x < a \text{ or } x > b. \end{cases}$$

uniform probability density function for continuous data



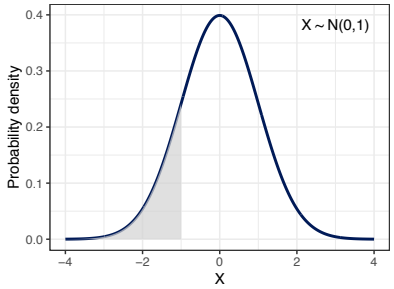
3. $$E(X) = (b+a)/2$$
$$\sigma^2(X) = \frac{(b-a)^2}{12}$$

mean and variance of uniform distribution

4. `dunif(x, min, max)` `dunif()` calculates density and
`punif(q, min, max)` `punif()` calculates cumulative probability

5. $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

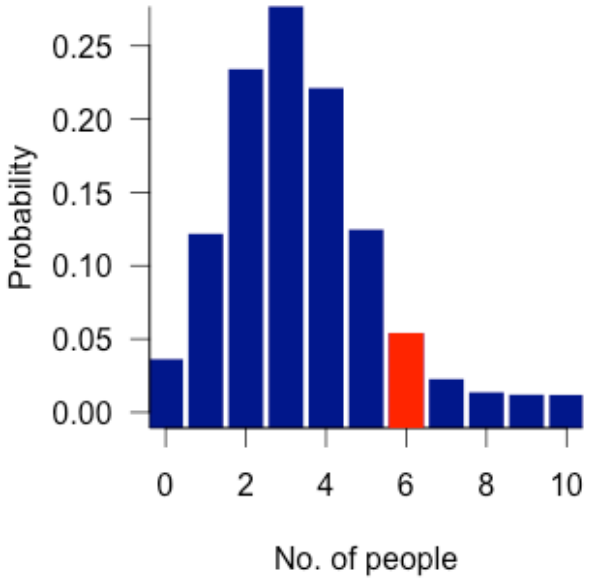
normal probability density function for continuous data



6. `dnorm(x, mean, sd)`
`pnorm(q, mean, sd)`

`dnorm()` calculates density and
`pnorm()` calculates cumulative probability

"Could you please explain the quantile part of the pbinom function and how you go from the sum(dbinom) to pbinom?"



According to a 2014 Gallup poll, 31% of Americans indicate that they worry "a great deal" about the quality of the environment, marking the lowest level of worry about the environment since Gallup began measuring it in 2001. What is the probability that in a random sample of 10 people, exactly 6 are worried a great deal about the environment?

```
> dbinom(x = 6, size = 10, prob = 0.31)
[1] 0.04224602
```

P(X<3)?

P(X>3)?

2 – Poisson random variable

Discuss the following and solve the problems together (10 min).

your team is conducting research on dolphin abundance using aerial flights to count the number of pod sightings in 2 hours of flight through a region

after 30 flights, you have collected the following data: 8 1 4 0 4 4 6 2 2 4 5 1 4 3 4 3 3 4 6 3 3 8 6 5 2 3 0 8 6 6

- what is the expected value for the number of counts of dolphin pods?
- what is the probability of seeing 5 dolphin pods on an aerial flight?
- what is the probability of seeing 6 or more dolphin pods on a flight?

2 – Poisson random variable

Discuss the following and solve the problems together (10 min).

your team is conducting research on dolphin abundance using aerial flights to count the number of pod sightings in 2 hours of flight through a region

- what is the expected value for the number of counts of dolphins? ~ 3.93
- what is the probability of seeing 5 dolphin pods on an aerial flight? 0.153

```
dpois(x = 5, lambda = 3.93)
```

- what is the probability of seeing 6 or more dolphin pods on a flight? 0.204

```
sum(dpois(x = 6:20, lambda = 3.93))
1-sum(dpois(x = 0:5, lambda = 3.93))
1-(ppois(q = 5, lambda = 3.93))
ppois(q = 5, lambda = 3.93,lower.tail = F)
```

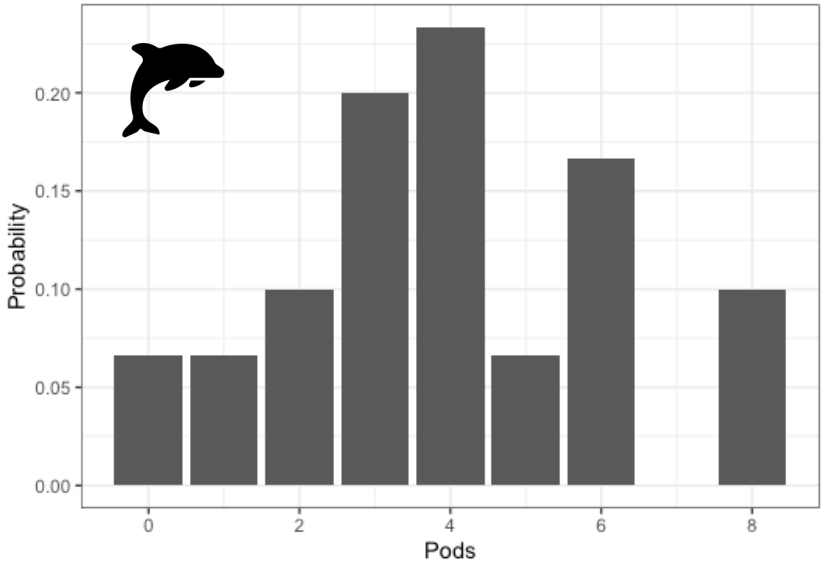
2 – Poisson random variable

Discuss the following and solve the problems together (10 min).

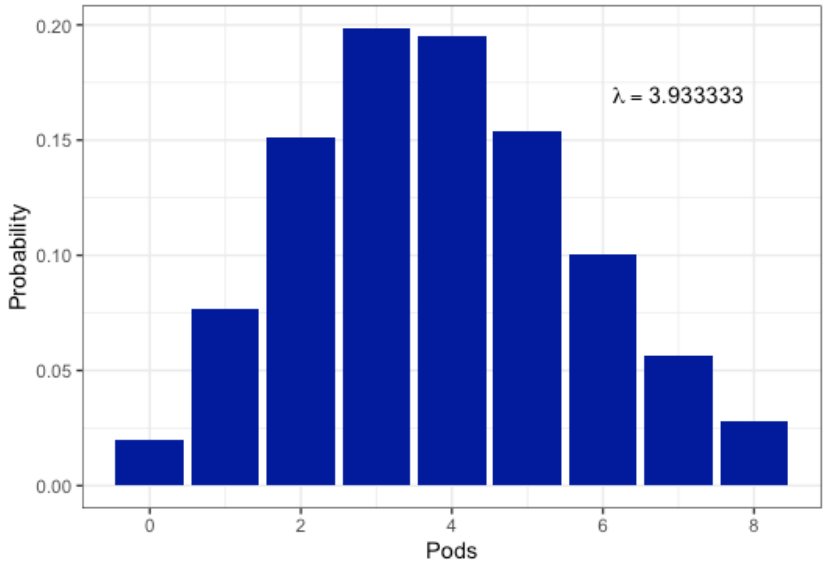
your team is conducting research on dolphin abundance using aerial flights to count the number of pod sightings in 2 hours of flight through a region

- create bar plots representing the raw data and the theoretical distribution (with mean 3.93).

```
pods <- c(8,1, 4, 0, 4, 4, 6, 2, 2, 4,
5, 1, 4, 3, 4, 3, 3, 4, 6, 3, 3, 8, 6,
5, 2, 3, 0, 8, 6, 6)

pods <- data.frame(cbind(p = pods, n =
c(1:length(pods))))

ggplot(pods, aes(x = p)) +
       geom_bar(stat="count") +
        ylab("Counts") + xlab("Pods") +
       theme_bw()
```

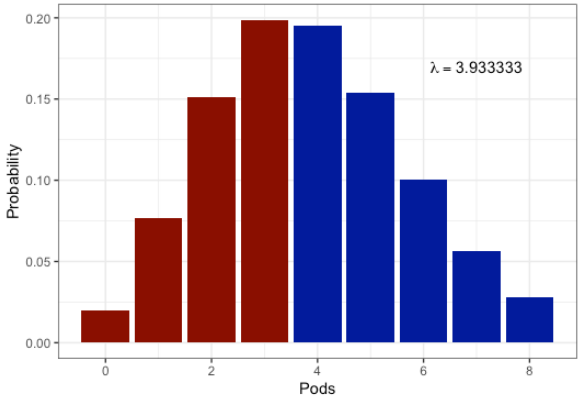


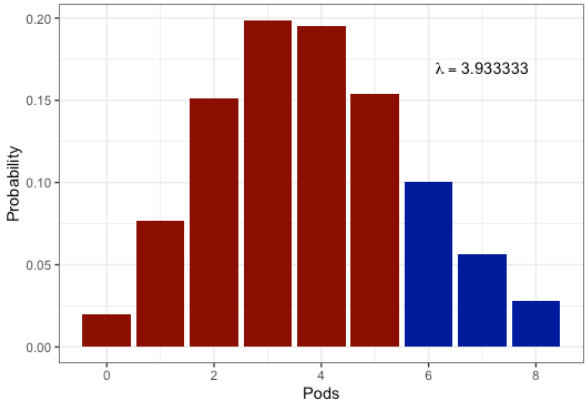

```
pods2 <- data.frame(p = dpois(0:8,
         lambda = 3.93), n = c(1:9))

ggplot(pods2, aes(x = n, y=p)) +
       geom_bar(stat="identity",
                fill = "magenta") +
        ylab("Probability") + xlab("Pods") +
       theme_bw()
```

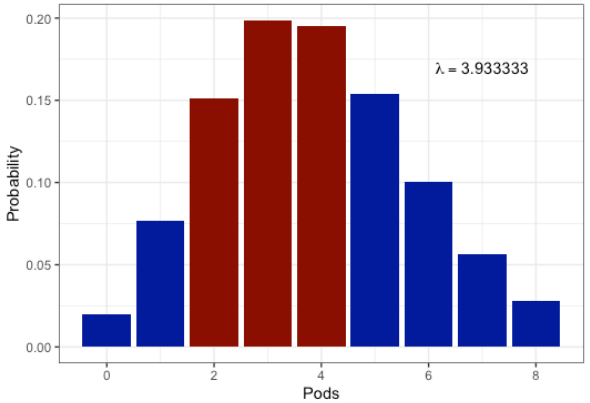


3 – CDF & PDF

Find the cumulative probability for the dark red area of each graph using ppois(). (5 mins)



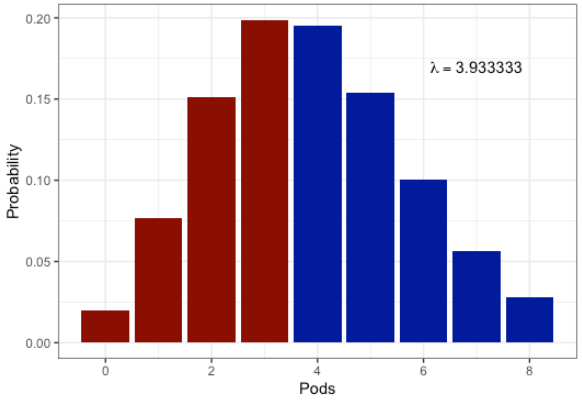




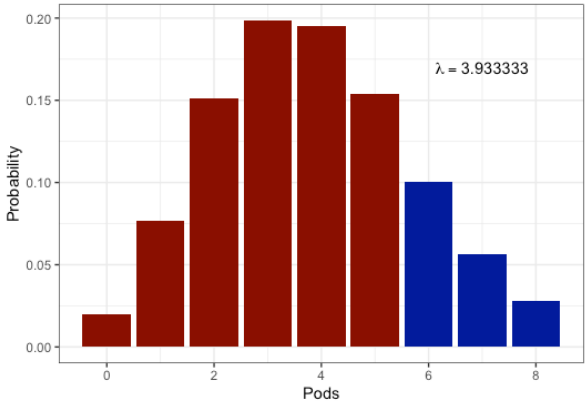
3 – CDF & PDF

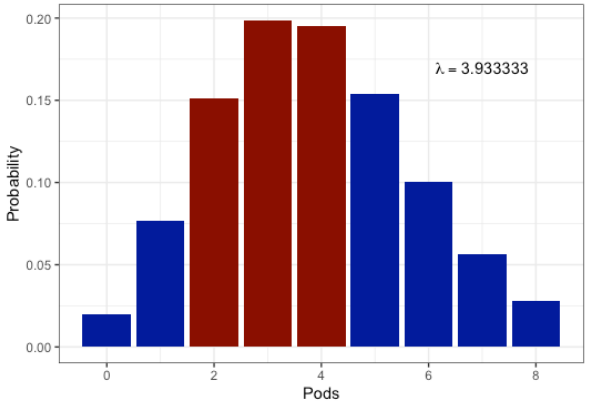
Find the cumulative probability for the dark red area of each graph using ppois(). (5 mins)

```
ppois(q = 3, lambda = 3.93)
```




```
ppois(q = 5, lambda = 3.93)
```

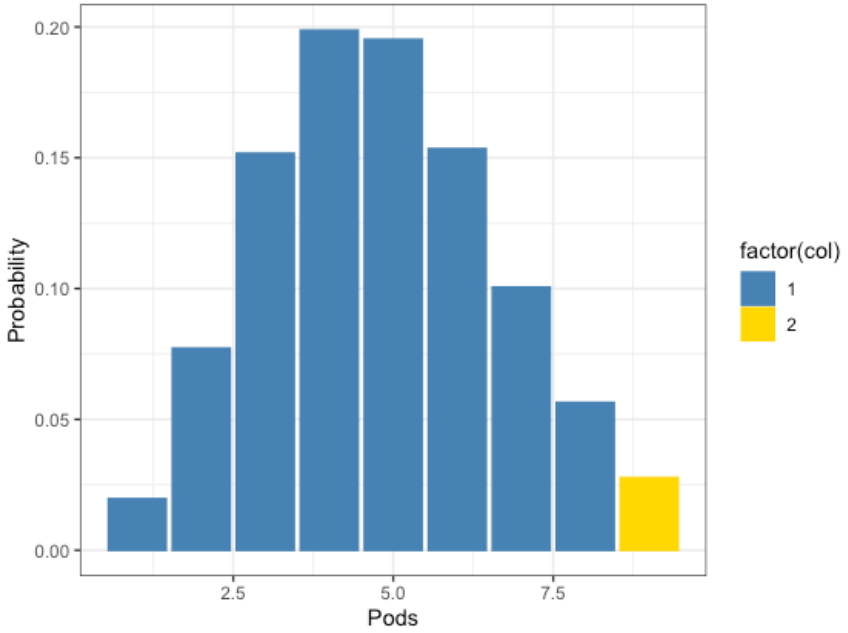




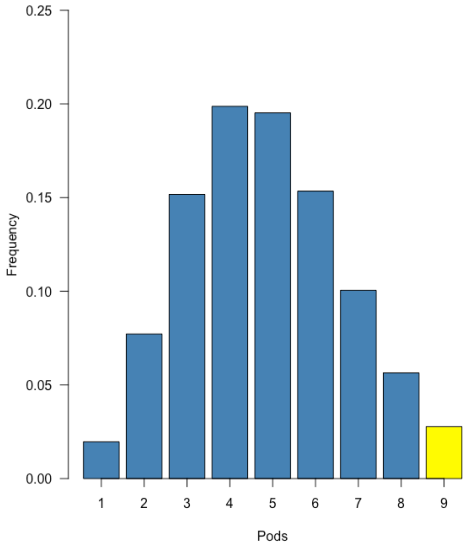

```
> ppois(q = 4, lambda = 3.93) - ppois(q = 1, lambda = 3.93)
[1] 0.5456665
```

3 – CDF & PDF

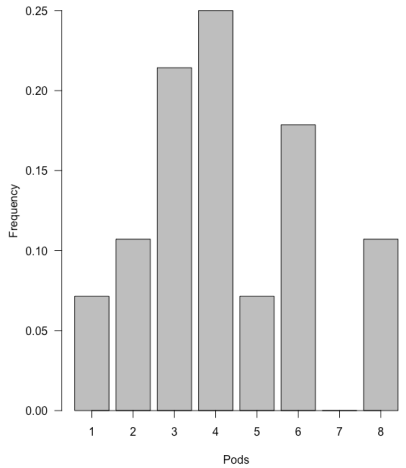
From the previous example, use either `ggplot()` or `barplot()` to create this graph. (10 mins)



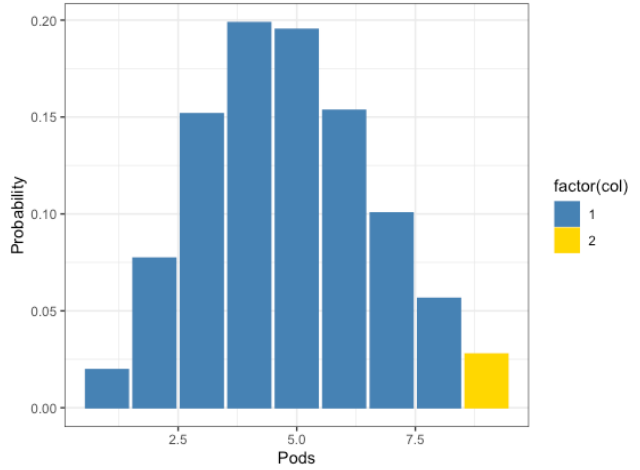
```
my_cols <- c("steelblue", "yellow")

with(pods2, barplot(p,
     col = my_cols[factor(col)], axis.lty = 1,
     las = 1, names.arg = n, ylim = c(0, .25)))
```



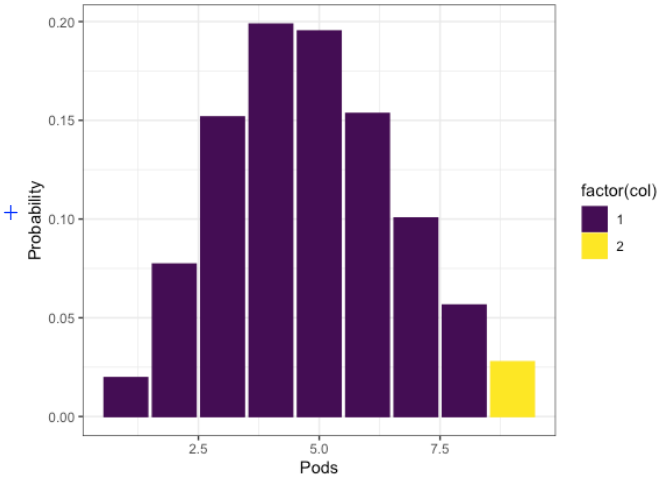

```
pods_bp <- data.frame(p = tabulate(pods$p))
pods_bp <- with(pods_bp, data.frame(p,
           freq = p/sum(p),
           n = c(1:nrow(pods_bp))))

with(pods_bp, barplot(freq, axis.lty = 1,
     las = 1, names.arg = n,
     ylim = c(0, .25)))
```



```
col <- c(rep(1,8), 2)
clr <- c("1" = "steelblue", "2" = "yellow")

ggplot(pods2, aes(x = n, y=p, colour = factor(col),
                  fill = factor(col))) +
       geom_bar(stat="identity") +
        labs(y = "Probability", x = "Pods") +
        scale_colour_manual(values = clr,
                            aesthetics = c("colour", "fill")) +
       theme_bw()
```




```
ggplot(pods2, aes(x = n, y=p, colour = factor(col),
                  fill = factor(col))) +
       geom_bar(stat="identity") +
        labs(y = "Probability", x = "Pods") +
       scale_color_viridis_d(aesthetics = c("colour", "fill")) +
      theme_bw()
```



4 – CDF & PDF

If 5% of the population is left-handed, what is the probability that 2 people out of a sample of 50 are left-handed?

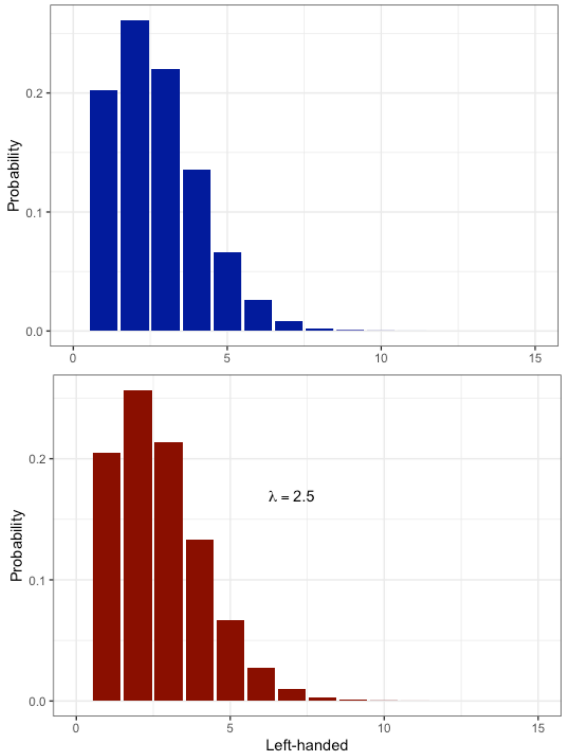
4 – CDF & PDF

If 5% of the population is left-handed, what is the probability that 2 people out of a sample of 50 are left-handed?

```
> dbinom(x = 2, size = 50, prob = 0.05)
[1] 0.2611014
```

```
> exp <- 0.05 * 50
[1] 2.5
```

```
> dpois(x = 2, lambda = exp)
[1] 0.2565156
```



Questions?