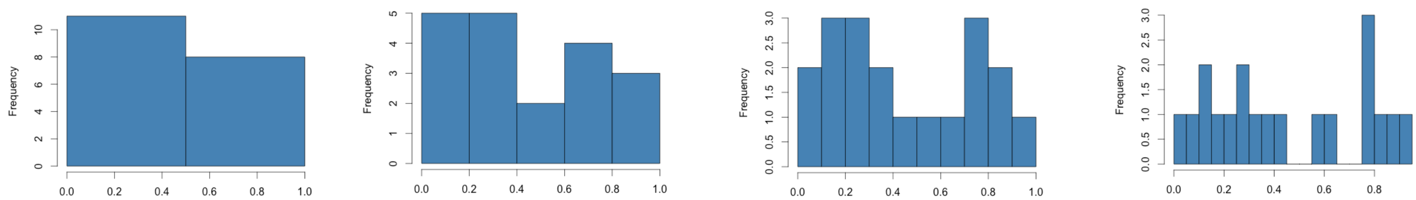
Activity #10: Continuous Distributions (Uniform, Exponential, Normal)

1. I had a computer generate the following 19 numbers between 0-1. Were these numbers randomly selected?

   0.12374454, 0.19609266, 0.44248450, 0.78311352, 0.86777448 , 0.61630168, 0.31731518, 0.77020822, 0.25598677, 0.01568532, 0.37556068, 0.26352634, 0.08893067, 0.12402286, 0.84960023, 0.59903364, 0.21330093, 0.76167311, 0.93632686

   If these numbers are random, we might expect them to be distributed uniformly over the interval from 0-1. We could plot a histogram to see if the distribution looks uniform (flat)…



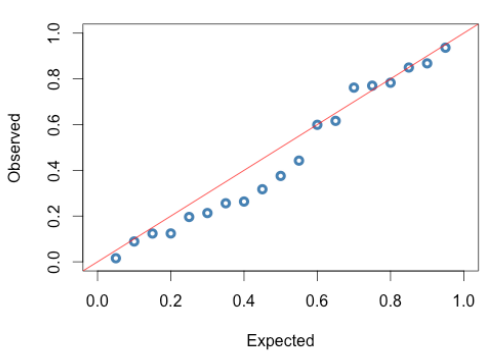
   … but our perception of "flatness" changes depending on how many *bins* we group our data into.

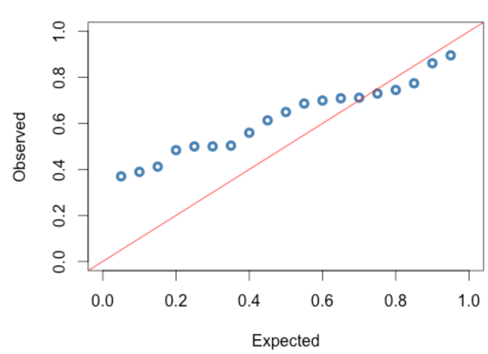
   Suppose we arrange our numbers from smallest to largest. If they follow a perfect uniform distribution, we'd expect (on average) each observation to represent an equal portion of the interval from 0-1. In other words, the first number should represent the first 5% of the interval; the next number should represent the next 5%, and so on:

Observed value	Expected Percentile	Observed value	Expected Percentile
0.01568532	0.05	0.44248450	0.55
0.08893067	0.10	0.59903364	0.60
0.12374454	0.15	0.61630168	0.65
0.12402286	0.20	0.76167311	0.70
0.19609266	0.25	0.77020822	0.75
0.21330093	0.30	0.78311352	0.80
0.25598677	0.35	0.84960023	0.85
0.26352634	0.40	0.86777448	0.90
0.31731518	0.45	0.93632686	0.95
0.37556068	0.50		

   We can then create a Q-Q plot showing the level of agreement between the numbers and expected percentiles:

   Below, I've plotted two Q-Q plots. One is for the 19 numbers listed above; the other is for 19 non-uniformly distributed numbers. How can we use Q-Q plots to evaluate whether a set of values follows a uniform distribution?





2. This semester, I noticed students yawn in class at an average rate of 6 students per hour. In other words, it takes an average of 10 minutes before I see the next student yawning. I want to calculate the probability that I must wait 15 minutes before the next student yawns.

If we let X represent the time I spend waiting before the next yawn, is X a discrete or continuous random variable?

X is: ________________________

Suppose we treat X as a discrete random variable (and arbitrarily set each minute as a discrete trial). We want to calculate the probability that we must wait at least 15 minutes (trials) before the first success (yawn). Which probability distribution would we use to model this scenario?

Probability distribution: __________________________________

Suppose we also define λ to represent the average number of yawns we get per minute. In this case, λ = 1/10.

Our decision to split time into discrete minutes is arbitrary (we could have chosen seconds, milliseconds, nanoseconds, etc.), so let's go ahead and divide our unit of time (a minute) into *n* very small subintervals. With this notation:

P(student yawning in any particular subinterval) = λ/n

With these subintervals, we're interested in P(X ≤ b) = the probability a student yawns within the first *b* subintervals.

If we let Y represent the number of subintervals we must wait for the next yawn, we know Y follows a geometric distribution with parameter *p* = P(a student yawns in any trial) = λ/n. In other words:

P(X ≤ b) = P(wait b time for next yawn) = P(Y ≤ bn) = P(it takes bn subintervals until we find the next yawn)

We can now use the geometric distribution to calculate P(Y ≤ bn):

$$P(Y \leq bn) = \sum_{k=1}^{bn} p(1-p)^{k-1} = \sum_{k=1}^{bn} \frac{\lambda}{n}\left[1-\frac{\lambda}{n}\right]^{k-1} = \frac{\lambda}{n}\sum_{k=1}^{bn}\left[1-\frac{\lambda}{n}\right]^{k-1} = \frac{\lambda}{n}\sum_{k=0}^{bn-1}\left[1-\frac{\lambda}{n}\right]^{k}$$

The expression on the right is a finite geometric series. Depending on how recently you took Calc II, you may recall:

$$\sum_{k=0}^{m} a^k = \frac{1-a^{m+1}}{1-a}$$

In our scenario: m = bn-1 and a = 1-λ/n. Substituting those values, we find:

$$P(Y \leq bn) = \frac{\lambda}{n}\left[\frac{1-\left(1-\lambda/n\right)^{bn}}{1-\left(1-\lambda/n\right)}\right] = \frac{\lambda}{n}\left[\frac{1-\left(1-\lambda/n\right)^{bn}}{\lambda/n}\right] = 1-\left(1-\lambda/n\right)^{bn} = 1-\left[\left(1-\lambda/n\right)^{n}\right]^{b}$$

Again, if you remember your Calculus II class well, you'll recall that as *n* approaches infinity:

$$P(X \leq b) = \lim_{n\to\infty}\left(1-\left[\left(1-\lambda/n\right)^{n}\right]^{b}\right) = \lim_{n\to\infty}\left(1-\left[e^{-\lambda}\right]^{b}\right) = 1-e^{-\lambda b}$$

That represents the cumulative distribution function to calculate waiting time.

3. If you didn't like that derivation, try this one. Let's start with the Poisson distribution, which models the number of times an event occurs within a given unit of time, distance, weight, etc.

We know that if λ = the average rate of occurrences per unit of time, the Poisson distribution is represented as:

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

We can define an **exponential random variable**, X, as a model for the time we must wait until the next occurrence.

To calculate the CDF –– P(X ≤ x) –– let x represent a positive amount of time:

$$F(x) = P(X \le x) = 1 - P(X > x) = 1 - P(\text{time to next occurrence} > \text{x}) =$$
$$= 1 - P(\text{no occurrences in the interval from 0 to x}) =$$
$$= 1 - \left[\frac{\lambda^0}{0!} e^{-\lambda x}\right] =$$
$$= 1 - e^{-\lambda x}$$

Exponential Distribution.

Conditions: • Can be used to model waiting times

Expected Value = 1 / λ, so λ = 1 / E[x]

Calculating:

Hand calculator: $P(X \le \#) = F(\#) = 1 - e^{-\lambda \#}$

R: CDF: P(X ≤ #) = `pexp(#, λ, lower.tail=TRUE, log.p=FALSE)`
CDF: P(X > #) = `pexp(#, λ, lower.tail=FALSE, log.p=FALSE)`
Simulating # random values: `rexp(#, λ)`

Online: http://www.stat.berkeley.edu/~stark/Java/Html/ProbCalc.htm

4. Let's calculate our probability of interest. If it takes an average of 10 minutes before I see the next student yawning in class, what's the probability that I must wait 15 minutes before the next student yawns?

5. An engineer assumes the life span of a particular type of lightbulb is exponentially distributed. Based on previous testing, a lightbulb lasts, on average, 1000 hours.

   a) Calculate the probability a randomly selected lightbulb lasts more than 200 hours.

   b) Given a lightbulb has already lasted 900 hours, calculate the probability it lasts another 200 hours.

   c) Calculate the probability a lightbulb lasts between 200 and 900 hours.

   d) Calculate the probability a lightbulb lasts exactly 200 hours.

6. According to the National Transportation Safety Board, there were 14 commercial airline crashes (not all fatal) in 2007. Using this, we can assume a plane crashes, on average, every 26 days.

   a) What's the median number of days we must wait until the next plane crash?

   b) Given a plane hasn't crashed for 90 days, what's the probability the next crash will occur within 30 days?

7. The manager of a local supermarket decides to offer a \$0.50 coupon to anyone who waits in line longer than 2.5 minutes. Before offering the deal, the manager wants to find the probability that a randomly selected customer will receive a coupon. In other words, this manager wants to know: $P(X > 2.5)$.

This manager assumes that since the exponential distribution is used to model waiting times, it should be used to calculate this probability.

To test this assumption, the manager observes the waiting times of 30 customers:

```
.03   .21   .33   .66   .91   1.33  1.56  1.84  2.49  3.46
.04   .21   .36   .87   .92   1.40  1.60  1.87  2.79  3.80
.10   .22   .47   .90   .96   1.51  1.76  2.36  3.24  4.10
```

a) Let's group these observations into minute-long bins:

Interval	**0-1**	**1-2**	**2-3**	**3-4**	**4-5**
Frequency	15	8	3	3	1
Relative freq.	0.50	0.27	0.10	0.10	0.03
Theoretical prob.	__________	__________	__________	__________	__________

We'll fill-in those blanks in just a bit. For now, note that the average of these 30 wait times is 1.41.

Given this average, find lambda: __________________________

If we have an exponential distribution with this parameter, then we can calculate the probability of waiting 0-1 minutes.

$P(X \leq 1) =$ __________________________________. This is the theoretical probability for our first bin.

Calculate the other theoretical probabilities and determine the reasonableness of assuming an exponential distribution in this scenario.

b) What's the probability a customer will receive the \$0.50 coupon?

8. In a previous activity, we learned both the geometric and negative binomial distributions. A geometric distribution was used to model the probability of the first success occurring on the kth trial. The negative binomial distribution was used to model the probability of the rth success occurring on the kth trial.

In this activity, we modeled the likelihood of waiting a certain amount of time for the next student to yawn. We can extend that to model the time until the rth student yawns.

If we let W = the time until the rth yawn, we could show that: $f(w) = \dfrac{\lambda^r}{(r-1)!} w^{r-1} e^{-\lambda w}, \ \text{for } w > 0$

This is the **Gamma Distribution**. Notice that if r=1, this reduces to our exponential distribution.

a) A telephone wire experiences an average of 1 flaw per 1000 feet. Suppose a customer receives a shipment of wire, randomly selects a 150-foot section, inspects it, and rejects shipment if there are 2 or more flaws. Using the gamma distribution, calculate the probability this shipment is rejected.