Suppose there are only 2 observations, \( x_i, \ i=1,2 \).
Show that sample variance of \( x \) is \( \frac{1}{2} ( x_2 - x_1 )^2 \)

\[
S^2 = \frac{1}{2-1} \sum_{i=1}^{2} (x_i - \overline{x})^2 = \left[ x_1 - \frac{1}{2} (x_1+x_2) \right]^2 + \left[ x_2 - \frac{1}{2} (x_1+x_2) \right]^2
= \left[ \frac{1}{2} (x_1-x_2) \right]^2 + \left[ \frac{1}{2} (x_2-x_1) \right]^2 = \frac{1}{2} (x_2-x_1)^2 = \frac{1}{2} \text{ (difference)}^2
\]

Note that \( S^2 \) is a generalization of the concept of difference.

Consider the adjacent histogram.

a) Compute/find the sample mean \( \overline{x} \). Show work.

\[
\overline{x} = \frac{1}{9} \left( 0 + 1 + 1 + 2 + 2 + 2 + 3 + 3 + 4 \right) = \frac{18}{9} = 2
\]

b) Find the sample std. dev. \( s \), using the defining formula.

\[
s^2 = \frac{1}{9-1} \sum_{i=1}^{9} (x_i - \overline{x})^2
= \frac{1}{8} \left[ (0-2)^2 + (1-2)^2 + (2-2)^2 + (2-2)^2 + (2-2)^2 + (2-2)^2 + (3-2)^2 + (3-2)^2 + (4-2)^2 \right]
= \frac{1}{8} \left[ 4 + 1 + 1 + 0 + 0 + 0 + 1 + 1 + 4 \right] = \frac{12}{8} = \frac{3}{2} = 1.5
\]

\[
s = \sqrt{1.5} \approx 1.22
\]

c) Draw the mean and the sample std. dev. on the histogram.

Note that \( \overline{x} \) and \( s \) agree with the eye-ball estimates I told you about on Day 1.
(a) For the bottom/left distribution, find the distr. mean \( \mu_x = E[x] \).

\[
E[x] = \sum_{x=1}^{3} x \cdot P(x) = 1(1) + 2(3) + 3(6)
\]

(b) For the bottom/right distribution, find the distr. mean \( \mu_y = E[y] \). Note that \( \mu_y = \mu_x + 1 \).

\[
E[y] = E[x] + 1
\]

\[\text{hw.lect7-4 \hspace{1cm}}\text{Find The mean of The exponential (} \lambda \text{). }\]

Hint: You may use \( \int_{0}^{\infty} ye^{-y} dy = 1 \).

\[
\mu_x = \int_{0}^{\infty} x f(x) dx = \int_{0}^{\infty} x e^{-\lambda x} dx \quad \text{change of Var.} \, \lambda x = y \Rightarrow \lambda dx = dy
\]

\[
= \frac{1}{\lambda} \int_{0}^{\infty} xe^{-y} dy = \frac{1}{\lambda} \int_{0}^{\infty} ye^{-y} dy = \frac{1}{\lambda}
\]

So mean of \( \text{Exp}(\lambda) \) is \( \frac{1}{\lambda} \).

\[\text{2.4 \hspace{1cm}}\text{Let } \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i:
\]

\[
\bar{X}_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} X_i = \frac{1}{n+1} \left[ \frac{n}{n+1} \bar{X}_n + X_{n+1} \right] = \frac{n}{n+1} \bar{X}_n + \frac{X_{n+1}}{n+1}
\]

This way, in a large data, when a new obs. is made, you don’t have to compute the mean of the whole data.

There is an analogous formula for variance, which is hard to prove:

\[
S_{n+1}^2 = \frac{n-1}{n} S_n^2 + \frac{(X_{n+1} - \bar{X}_n)^2}{n+1}
\]
a) Consider the binomial dist. with params n, pi. Draw four figures that show qualitatively how its mean (μ_x) and variance (σ_x^2) vary with n and pi.

Suppose we toss n=100 unfair coins, with an unknown pi.

b) What is the expected number of heads out of n? (The answer depends on pi).

c) What is the typical deviation in the number of heads out of n? (The answer depends on pi).

d) What is the largest typical deviation of the number of heads out of n? (The answer is a number!)

Hint: Consult your graph of variance vs. pi, in part a).

\[ \mu_x = np \]

\[ \sigma_x^2 = np(1-p) \]

b) \( 100 \pi \)

c) \( \sqrt{100(\pi)(1-\pi)} \)

d) The largest \( \sigma_x^2 \) (i.e., peak of the curve in \( \sigma_x^2 \) vs. \( \pi \)) occurs at \( \pi = \frac{1}{2} \), and it is \( 100 \cdot \frac{1}{2} \left(1-\frac{1}{2}\right) = 25 \). So, the max dev. is 5.

This is an important property of binomial dist.

Remember it!
For the uniform distribution between $a$, $b$, show that the mean (or expected value) and the variance are $(b+a)/2$, and $(b-a)^2/12$, respectively.

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{else} \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = \int_{a}^{b} x \cdot \frac{1}{b-a} \, dx$$

$$= \frac{1}{b-a} \int_{a}^{b} x \, dx = \frac{1}{b-a} \left[ \frac{1}{2} x^2 \right]_{a}^{b}$$

$$= \frac{1}{b-a} \frac{1}{2} (b^2-a^2) = \frac{1}{b-a} \frac{1}{2} (b-a)(b+a) = \frac{1}{2} (b+a)$$

$$V[X] = \int (x - E[X])^2 \cdot f(x) \, dx = \int_{a}^{b} (x - \frac{1}{2} (b+a))^2 \cdot \frac{1}{b-a} \, dx$$

$$= \frac{1}{b-a} \int_{a-rac{1}{2} (b+a)}^{b-rac{1}{2} (b+a)} y^2 \, dy$$

$$y = x - \frac{1}{2} (b+a) \Rightarrow dy = dx$$

$$= \frac{1}{3} (b-a) \left. \frac{1}{3} y^3 \right|_{a-rac{1}{2} (b-a)}^{b-rac{1}{2} (b+a)} = \frac{1}{3} (b-a) \left[ \frac{1}{8} (b-a)^3 + \frac{1}{8} (b-a)^3 \right]$$

$$= \frac{1}{12} (b-a)^3 = \frac{(b-a)^2}{12}$$
The mean of the Poisson distribution $p(x)$ with parameter $\lambda$ is $\lambda$ itself. Show that

$$\sum_{x=0}^{\infty} x^2 p(x) = \lambda (\lambda + 1)$$

$$\sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{(x-1)!} \uparrow x=1 \quad \sum_{y=0}^{\infty} (y+1) e^{-\lambda} \frac{\lambda^y}{y!}$$

$$\lambda \cdot \sum_{y=0}^{\infty} (y+1) \frac{\lambda^y}{y!} = \lambda \left[ \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} + \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \right] = \lambda (\lambda + 1)$$

$$\mu_x = \lambda$$

This result is useful when finding the variance of Poisson($\lambda$).
Find the prob. that $x$ is within 1 std-dev. of the mean, for

a) Binomial($n=20, \pi=\frac{1}{4}$)

$$
\begin{align*}
\mu_x &= n\pi = 20 \left( \frac{1}{4} \right) = 5, \\
\sigma_x &= \sqrt{n\pi(1-\pi)} = \sqrt{20 \left( \frac{1}{4} \right) \left( \frac{3}{4} \right)} = 1.9
\end{align*}
$$

$$
\mu_x - \sigma_x < x < \mu_x + \sigma_x \Rightarrow 5-1.9 < x < 5+1.9 \Rightarrow 3.1 < x < 6.9
$$

Integer $x$ values in this range are $x = 4, 5, 6,$

\[ \text{area} = p(x=4) + p(x=5) + p(x=6) = 0.19 + 0.202 + 0.169 = 0.561 \]

\text{Table II.}

b) Poisson ($\lambda=5$)

$$
\begin{align*}
\mu_x &= \lambda = 5, \\
\sigma_x &= \sqrt{\lambda} = \sqrt{5} = 2.24
\end{align*}
$$

$$
\mu_x - \sigma_x < x < \mu_x + \sigma_x \Rightarrow 5-2.24 < x < 5+2.24 \Rightarrow 2.76 < x < 7.24
$$

Integer $x$ values in this range are $x = 3, 4, 5, 6, 7,$

\[ \text{area} = p(x=3) + \cdots + p(x=7) = 0.14 + 0.175 + 0.175 + 0.146 + 0.046 \]

\[ = 0.742 \]

\text{Table III.}

c) Normal($\mu=5, \sigma=1$)

\[ z = \frac{x-5}{1} \]

\[ \frac{5-1}{5+1} \rightarrow x = \frac{4}{6} \]

\[ \frac{-1}{0} \rightarrow z = \frac{x-5}{1} \]

\[ = 0.8413 - 0.1587 = 0.6826 \]
It can be shown that the $p^{th}$ percentile of $\text{Unif}(a, b)$ is given by

$$
\eta_p(a,b) = a + (b-a) \frac{p}{100}
$$

a) What is the $p^{th}$ percentile of $\text{Unif}(0,1)$, i.e., $\eta_p(0,1)$?

$$
\eta_p(0,1) = 0 + \frac{p}{100}
$$

b) Write $\eta_p(a,b)$ in terms of $\eta_p(0,1)$.

$$
\eta_p(a,b) = a + (b-a) \eta_p(0,1)
$$

c) What will the plot of $\eta_p(a,b)$ vs. $\eta_p(0,1)$ look like? What are the slope & y-intercept?

A straight line with y-intercept $a$, and slope $(b-a)$.

Moral: If you make a qq plot of your data, but with $\eta_p(0,1)$ on the x-axis, then a straight line would imply that your data came from a $\text{Unif}(a,b)$ with $a,b$ estimated from y-intercept & slope $(b-a)$. 
Do a qq-plot of each of the 2 cont. vars. in the data you collected. By R. Describe/Interpret the result. Note: If you find out that there is not much you can say about the qq-plot, it may be that your data is not appropriate. This is another chance to correct the error, because later you will be doing more hw problems using your data. So, see me, if you are not sure.

The results will vary across students, but if you see a "straight" qq plot, then that's evidence that your data came from a Normal distr. with $\mu$ equal to the y-intercept, and $\sigma = \text{slope of the qq plot}$. If your qq plot looks curved, then try doing the qq plot of the log or $\sqrt{}$ of the data. If that makes the qq plot "straight" then use the log or $\sqrt{}$ of the data for all future work.