Lecture 9 (Ch. 2)

Keep the "big picture" in mind: We are looking for

<table>
<thead>
<tr>
<th>Measure of</th>
<th>Sample</th>
<th>pop./distr.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>location</strong></td>
<td>Sample mean</td>
<td>Below</td>
</tr>
<tr>
<td>$\bar{x} = \frac{1}{n} \sum_{i} x_i$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{x} \sim &quot;typical \ x&quot;$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>spread</strong></td>
<td>Sample variance</td>
<td>Tomorrow.</td>
</tr>
<tr>
<td>$s^2 = \frac{1}{n-1} \sum_{i} (x_i - \bar{x})^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s \sim &quot;typical deviation \ in \ x&quot;$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As a warm-up for a long calculation do this:

**Q2:** For $n=2$, which of the following is WRONG?

A) $\sum_{i} (x_i - \bar{x}) = \sum_{i} x_i - \bar{x} \cdot \bar{x} = (x_1 + x_2) - 2 \bar{x}$

B) $\sum_{i} (x_i - \bar{x}) = \sum_{i} x_i - \frac{1}{2} \sum_{j} x_j = (x_1 \cdot \frac{1}{2} (x_1+x_2)) + (x_2 \cdot \frac{1}{2} (x_1+x_2))$

C) $\sum_{i} (x_i - \bar{x}) = \sum_{i} (x_i - \frac{1}{2} \sum_{i} x_i) \times (x_1 - \frac{1}{2} x_1) + (x_2 - \frac{1}{2} x_2)$

D) None of the above.
Another way of computing $s^2$. — Sometimes more useful, always faster (1 loop vs. 2 loops)

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

\[ = \frac{1}{n-1} \sum_{i=1}^{n} \left( x_i^2 - 2 \bar{x} x_i + \bar{x}^2 \right) \]

\[ = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - 2 \bar{x} \sum_{i=1}^{n} x_i + n \bar{x}^2 \right] \]

\[ = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - 2 n (\bar{x})^2 + n (\bar{x})^2 \right] \]

\[ = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - n (\bar{x})^2 \right] \]

\[ = \frac{1}{n-1} \left[ n (\frac{1}{n} \sum_{i=1}^{n} x_i^2) - n (\bar{x})^2 \right] \]

\[ = \frac{n-1}{n-1} \left[ \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\bar{x})^2 \right] = \frac{n-1}{n-1} \left[ \bar{x^2} - (\bar{x})^2 \right] \]

In summary:

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad "\text{Defining formula}" \]

\[ s^2 = \frac{n}{n-1} \left[ \bar{x^2} - (\bar{x})^2 \right] \quad "\text{Computational formula}" \]

Example

\[ x = \{1, 3, 8\} \rightarrow x^2 = \{1, 9, 64\} \rightarrow \bar{x^2} = \frac{74}{3} \]

\[ s^2 = \frac{3}{2} \left[ \frac{74}{3} - 16 \right] = \frac{3}{2} \cdot \frac{74-48}{3} = \frac{26}{2} = 13 \]
Now, we need to come up with corresponding things in the pop

So, switch to distributions \((p(x), f(x))\), no data!

1) Distribution mean = \(E[x] = \mu_x = \cdot \frac{\infty}{\int_{-\infty}^{\infty} x \cdot p(x) \, dx}

Motivation: Consider a "pop" of size 10:
\[
\{3, 2, 2, 1, 3, 2, 3, 1, 2, 2\}
\]

mean = \(\frac{1}{10} \left[ 3 \cdot 3 + 2 \cdot 2 + \cdots \right] = \frac{1}{10} \left[ 3 \cdot 3 + 5 \cdot 2 + 2 \cdot 1 \right] = \frac{3}{10} \cdot 3 + \frac{5}{10} \cdot 2 + \frac{2}{10} \cdot 1 \]

Compare:

Sample mean: \(\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i\),

\(E[x] \equiv \mu_x = \sum_{x} x \cdot p(x) \), \(\int_{-\infty}^{\infty} x \cdot f(x) \, dx\)

The book drops the \(x\) on \(\mu_x\), but then \(\mu\) can be confused with the parameter of the normal distr.

\(E[x]\) does not mean that \(E\) is a function of \(x\). In fact, \(E\) is a \(\Sigma\) or an \(\int_{-\infty}^{\infty}\), and so it is not a function of \(x\).

\(E[x]\) simply means that you need \(p(x)\) or \(f(x)\) to find it.

See binomial example, below.
**Example** Binomial \( (n, p) \)

\[
E(x) = \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \cdot x
\]

\[
= \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \cdot \frac{x}{1} \\
= \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-y-1)!} \cdot \frac{y+1}{1} \cdot (1-p)^{n-y-1} \\
= n \cdot \frac{(n-1)!}{y!(n-y-1)!} \cdot (1-p)^{n-y-1} \\
= (m+1) \cdot \sum_{y=0}^{m} \frac{m!}{y!(m-y)!} \cdot (1-p)^{m-y} \\
= \sum_{y=0}^{m} \frac{m!}{y!(m-y)!} \cdot (1-p)^{m-y} = 1 = \sum_{y=0}^{\infty} p(y)
\]

\[E[x] = np\]

Note \( E[x] \) is not a function of \( x \).

**E.g. 1.23:**

<table>
<thead>
<tr>
<th># of Bads</th>
<th>( p(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.6058</td>
</tr>
<tr>
<td>1</td>
<td>0.3044</td>
</tr>
<tr>
<td>2</td>
<td>0.0757</td>
</tr>
<tr>
<td>3</td>
<td>0.0124</td>
</tr>
</tbody>
</table>

\[E[x] = \sum x \cdot p(x) = 0(0.6058) + 1(0.3044) + \ldots \]

\[= np = 100(0.005) = 0.5\]

On avg. 0.5 out of 100 (i.e. 1 out of 200) computers are defective.
For the other distributions, same tricks:

\[ \text{Poisson}(\lambda) : \mu_x = E[x] = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \ldots = \lambda \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = \lambda \]

Now you can see why \( \lambda \) is called mean.

\[ 1 = \sum_{x=0}^{\infty} p(x) \]

\[ \text{Normal}(\mu, \sigma) : \]

\[ \mu_x = E[x] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma}\right)^2} \, dx = \ldots (t = \frac{x-\mu}{\sigma}) \ldots = \mu \]

Now you can see why \( \mu \) (the param of Normal) is a mean.

Etc. We can find the mean of any distribution in terms of parameters of that distv. See \( \text{(a)} \).

**Warning:** Don't confuse \( \bar{x}, \mu_x, \mu \)

\[ \text{Sample mean} : \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i ; \]

\[ \text{distv. mean} : \quad E[x] = \mu_x \]

- binomial \((n, p)\)
- Poisson \((\lambda)\)
- Normal \((\mu, \sigma)\)

Note about \( \sum_{x} x \cdot p(x) \):

If \( x \) = Qualitative (Categorical)

- E.g. \( x \) = Computer brand = \{ Apple, Dell, Lenovo \}

Then \( \sum_{x} x \cdot p(x) \) makes no sense!
Let $X$ denote the number of heads out of 10 fair coins. Five observations of $X$ are: 3, 3, 5, 6, 3. The expected number of heads out of 10 is

a) 3  b) 4  c) 5  d) none of the above.

Summary

Histogram location: Sample mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

~ typical $X/obs.$

Histogram spread: Sample variance:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

~ Typical deviation/spread

distribution/population location: dist. (pop mean, or $E[X]$)

$$\mu_X = E[X] = \sum_{x} x \cdot p(x) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$

distr./pop. spread: Next time.
B) Consider the binomial distribution \( p(x) \) with parameters \( N = 4, \pi = \frac{1}{4} \).

a) Compute specific values of \( p(x) \) for all possible values of \( x \). (By hand or by R).

b) Compute \( E[x] = \sum_{x} x \cdot p(x) \), and compare the answer with the value of \( (N\pi) \). (By hand or by R).

c) Take a sample of size 100 from \( p(x) \), compute the sample mean of the 100 numbers, and compare the answer with the answer in part b. (By R)

B) For the uniform distribution (see 1.19) between \( a, b \), show that the expected value is \( \frac{1}{2} (a+b) \).

B) For the exponential distribution with parameter \( \lambda \), find \( \mu_x \).

Hints: \( \int_0^\infty y e^{-y} \, dy = 1 \) \( \int_0^\infty (y-1) e^{-y} \, dy = 1 \)

B) Find the \( \mu_x \) for

a) The \( p(x) \) given in exercise 1.27, with the two "?" given as 0.1 and zero, respectively.

b) The \( f(x) \) given in exercise 1.21