Lecture 15 (CH3, 5.5, 5.6, 7)

This page FYI

Why \((n-1), (n-2), \ldots, n-(k+1), \ldots\) ?

Q \[ \bar{y} = \frac{1}{n} \sum_{i} y_i \]  
Why \(n\) ?

A \( \{y_1, y_2, \ldots, y_n\} \) are all independent \(\implies\) \(\text{df}(\text{of } \sum_{i} y_i) = n\)

Q \[ s^2 = \frac{1}{n-1} \sum_{i} (y_i - \bar{y})^2 \]  
Why \((n-1)\) ?

A \( \{y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y}\} \) are not all independent.
There is 1 constraint on them \(\sum_{i} (y_i - \bar{y}) = 0\)
I.e., there are \((n-1)\) independent terms \(\implies\) \(\text{df}(\text{of } s^2) = n-1\)

[There are other reasons for \((n-1)\), too].

Q \[ s_0^2 = \frac{1}{n-2} \sum_{i} (y_i - \hat{\gamma}_i)^2 \]  
Why \((n-2)\) ?

A \( \{y_1 - \hat{\gamma}_1, y_2 - \hat{\gamma}_2, \ldots, y_n - \hat{\gamma}_n\} \) satisfy 2 constraints (below) \(\implies\) \(\text{df}(\text{of } \text{SSE}) = n-2\)

1st constraint: \(\frac{1}{n} \sum_{i} (y_i - \hat{\gamma}_i) = \frac{1}{n} \sum_{i} (y_i - \hat{\alpha} - \hat{\beta} x_i) = \bar{y} - \hat{\beta} \bar{x} = 0\)  
\(\bar{y} - \hat{\beta} \bar{x}\) (see \(\hat{\alpha}\) eqn.)

2nd constraint: \(\frac{1}{n} \sum_{i} (y_i - \hat{\gamma}_i) x_i = \frac{1}{n} \sum_{i} (y_i x_i - (\hat{\alpha} + \hat{\beta} x_i) x_i) = \bar{y} \bar{x} - \hat{\alpha} \bar{x} - \hat{\beta} \bar{x}^2\)

\[= \frac{xy - (\bar{y} - \hat{\beta} \bar{x}) \bar{x} + \hat{\beta} \bar{x}^2}{\bar{x}^2 - \bar{x}}\]
\[= (\bar{xy} - \bar{y} \bar{x}) - \hat{\beta} (\bar{x}^2 - \bar{x}) = 0\]
\[\frac{\bar{xy} - \bar{y} \bar{x}}{\bar{x}^2 - \bar{x}}\] (see \(\hat{\beta}\) eqn.)
1) That's the sampling dist. of the sample mean. But one can talk about the sampling distribution of the sample median, the sample standard deviation, the sample proportion, the sample anything.... Later (in ch 11), we will even talk about the sampling distribution of the sample fit. They all say something about the typical value and the typical fluctuation of the respective quantity.

2) Like the name suggests, it is a distribution, i.e. $p(x)$ or $f(x)$, that can be derived mathematically, or simply assumed as a description of the population of all $x$'s. In fact, you have already seen some sampling distributions, e.g. distribution of minimum, maximum, ... of sample of size 2 or 3 taken from Bernoulli. The only reason I talk about a histogram is to make the concept of the sampling dist. more intuitive; the talk of taking a zillion samples etc. is just thought experiment; in practice, we take only one sample of size $n$. The histogram is sometimes called the "empirical sampling dist."

2 important quantities:

One estimates the pop. parameter (e.g. $\mu$), the other tells us how certain that estimate is. Precise

So $\bar{X}$ and $\sigma_{\bar{X}}$ of the sampling dist. tells us something about $\mu$ (the pop. mean!) and our uncertainty in it.
ntrial = 64
xbar = numeric(ntrial)
par(mfrow=c(8,8))
for( trial in 1:ntrial ){
x = rnorm(50, 0, 1)
hist(x, breaks=10)
xbar[trial] = mean(x)
}
hist(xbar, main="")

Q: What's $\bar{x}$ in each hist above?
Q: What's the mean of the $\bar{x}$'s?
Q: What's $s$ in each hist above?
Q: What's $s$ of the $\bar{x}$'s?

Try `rexp(50, 1)`
What is the sampling distr. of $\bar{x}$? Normal, Poisson, ... later!

But even without knowing the distr., we can still find its mean ($E[\bar{x}]$ or $\mu_{\bar{x}}$) and variance ($V[\bar{x}]$ or $\sigma_{\bar{x}}^2$):

**Theorem:** If the pop. (dist.) has mean & std. dev. $\mu_x$, $\sigma_x$, then

$$
\mu_{\bar{x}} = E[\bar{x}] = \mu_x \quad \text{pop. mean} \\
\sigma_{\bar{x}} = \sqrt{V[\bar{x}]} = \frac{\sigma_x}{\sqrt{n}} \quad \text{pop. std. dev.}
$$

Sometimes called "standard error of mean."

where $\mu_{\bar{x}} = \text{Mean of the Sampling distr. of sample mean}$

$$
\sigma_{\bar{x}} = \text{Std. dev. of sampling dist. of sample mean}
$$

---

**General notation/jargon:**

<table>
<thead>
<tr>
<th><strong>statistics</strong></th>
<th><strong>pop. parameters</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$ (sample mean) is a point estimate of $\mu_x$ (pop. mean)</td>
<td>$\mu_x$ (pop. mean)</td>
</tr>
<tr>
<td>$s$ (s std. dev.)</td>
<td>$\sigma_x$ (s std. dev.)</td>
</tr>
<tr>
<td>$p$ (s prop.)</td>
<td>$\pi_x$ (s prop.)</td>
</tr>
</tbody>
</table>

$n$ (s size) is NOT related to pop. size. $n = \infty$ for us.

$\bar{x}$ = mean of The sampling distr. of $\bar{x}$

$\sigma_{\bar{x}}$ = Std. dev. of sampling distr. of $\bar{x}$

---

We skipped Sec. 3.6, but it has one result that we need:

**constant**

$$
E[a \cdot x] = a \cdot E[x] \quad V[a \cdot x] = a^2 V[x]
$$

$$
E[x \pm y] = E[x] \pm E[y]
$$

always + partial proof below

$$
V[x \pm y] = V[x] \cup V[y] + \bigcirc
$$

$x, y = \text{independent}$
Proof: Suppose we do not know the distrib. of the population \((x, f(x))\), but we do know its \(\mu_x\) and \(\sigma_x\).

Of course, if you do know the pop. distr., then you can compute \(\mu_x\), \(\sigma_x\) as before:

\[
\begin{align*}
\mu_x &= \mathbb{E}[x] = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[x_i] = \frac{1}{n} \sum_{i=1}^{n} \mu_x = \mu_x, \\
\sigma_x^2 &= \mathbb{V}[x] = \mathbb{V}[\frac{1}{n} \sum_{i=1}^{n} x_i] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}[x_i] = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_x^2 = \frac{\sigma_x^2}{n}.
\end{align*}
\]

Now, start!

The \(i\)th obs. is a random value, there is nothing special about the \(i\)th obs. So, just drop the \(i\)th obs.

Then \(\mathbb{E}[x_i] = \mathbb{E}[x] = \mu_x p(x) = \mu_x\).

Alternatively, work out \(\mathbb{E}[x_i]\) for each \(i\), e.g. \(i=1\)

\[
\mathbb{E}[x_i] = \frac{1}{\mu_x} x_i p(x_i) = \mu_x, \quad \mathbb{E}[x_i] = \mu_x, \quad \mathbb{E}[x] = \mu_x.
\]

The var. of each element in the sample is the variance of the pop.
Pf. of $E[aX] = aE[X]$:

$E[aX] = \int (ax)f(x) \, dx = a \int x \, f(x) \, dx = a \cdot E[X]$.

Pf. of $E[X+Y] = E[X] + E[Y]$:

With 2 variables (i.e., $X,Y$), $E[\cdot]$ is defined this way:

$E[X] = \iiint x \cdot f(x,y) \, dx \, dy$.

Then

$E[X+Y] = \iiint (x+y) \cdot f(x,y) \, dx \, dy$

$= \iiint x \cdot f(x,y) \, dx \, dy + \iiint y \cdot f(x,y) \, dx \, dy$

$= \int x \left( \int f(x,y) \, dy \right) \, dx + \int y \left( \int f(x,y) \, dx \right) \, dy$

$= \int x f(x) \, dx + \int y f(y) \, dy$


\[ \bar{x} = \frac{1}{n} \sum x \]

Tells us that the typical deviation in \( \bar{x} \) is \( \frac{\sigma_x}{\sqrt{n}} \), and so it tells us how precise is our estimate of \( \mu_x \).

Note that \( \mu_x, \sigma_x, \bar{x}, \sigma_x \) are means and std. dev. of distributions, not of data, even though the thought-experiment led to a histogram. That’s why we use \( \mu, \sigma, \bar{x}, \sigma_x \) notation.

\[ \bar{x} \text{ and } s_x \text{ are measures of Accuracy & Precision:} \]

<table>
<thead>
<tr>
<th>Accuracy ( \bar{x} - \mu_x )</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

True/pop mean

\[ \mu_x = \mathbb{E}[x] = \bar{x} \]

Tells us that we can use the sample mean (from the one sample of size \( n \)) to estimate the pop. mean \( \mu_x \) with accuracy. (see box, below)
Important
Distinguish between random things (like \( \bar{x} \)) and non-random things (like \( \bar{x}_{obs} \), \( \mu_x \)). In lower-level stat classes this is not an important distinction; but at the 390 level, it is. And this important distinction will stay with us until the end of the quarter.
And, yes, in all of the above calculations of prob, we need to know the mu_x and sigma_x of population. So, this whole lecture does not seem to deliver on the promise of being able to determine mu_x and sigma_x of the population from a sample. For the delivery of that promise, wait for next lecture.
a) write R code to produce the sampling distribution of the sample maximum, for samples of size 50 taken from a standard Normal. Use 5000 trials, 
b) Repeat for the sample minimum.

Turn-in your code, and the resulting 2 histograms.
FYI, these distributions arise naturally when one tries to model extreme events, e.g. the biggest storms, the strongest earthquakes, the brightest stars, the smallest forms of life, etc.

a) write R code to take 5000 samples of size n=100 from an exponential distr. with parameter lambda=2, and make a qq-plot of the 5000 means. Recall that if the qq-plot is a straight line, then the histogram of the sample means is Normal. This will show that the sampling distr. of sample means is Normal, even when the pop. is not!

b) using the qq-plot, estimate the mean and std. dev. of the sampling dist. of sample means. Are they consistent with what you would expect from our formulas for the mean and standard deviation of the sampling distribution? show work.

A sample of size 36 from a Normal pop. yields the observed values xbar= 3.5 and s=1.
a) Under the assumption that mu_x = 2.5, and sigma_x = 2, what's the prob of a sample mean larger than the one observed?
b) Under the assumption that mu_x = 2.5, and sigma_x = 2, what's the prob of a sample mean smaller than the one observed?
c) Under the assumption that mu_x = 3.5, and sigma_x = 2, what's the prob of a sample mean larger than the one observed?
d) Under the assumption that mu_x = 3.5, and sigma_x = 2, what's the prob of a sample mean smaller than the one observed?
e) Now, suppose we know that sigma_x = 2, but we don't know mu_x. What is the observed 95% Confidence Interval for mu_x. Interpret it, in TWO ways.

Students are often suspicious of my claim that the E (i.e., distribution mean) of the ith element of a sample of size n is equal to the population mean, i.e. E[x_i] = E[x]. To convince yourself, write code to
- take 10^7 samples of size 50 from a normal distribution with mu = 2 and sigma = 3,
- select the 3rd element in each of the 10^7 samples, and store them in an array called x3.
- compute the mean of x3.
Convinced?!