Sections 5.1 and 5.2

Timothy Hanson

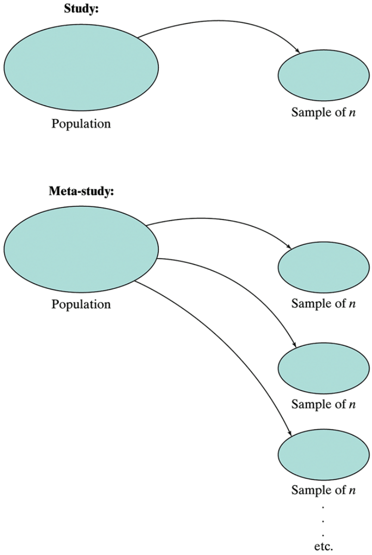
Department of Statistics, University of South Carolina

Stat 205: Elementary Statistics for the Biological and Life Sciences

Sampling variability

- A random sample is exactly that: *random*.
- You can collect a sample of n observations and compute the mean $\bar{Y}$. Before you do it, $\bar{Y}$ is random.
- If you you randomly sample a population two different times, taking, e.g. $n = 5$ each time, the two sample means $\bar{Y}_1$ and $\bar{Y}_2$ will be different.
- Example: sampling $n = 5$ ages from Stat 205.
- Variability among random samples is called **sampling variability**.
- Variability is assessed through a hypothetical "mind experiment" called a **meta-study**.

Study and meta-study



Example 5.1.1 Rat blood pressure

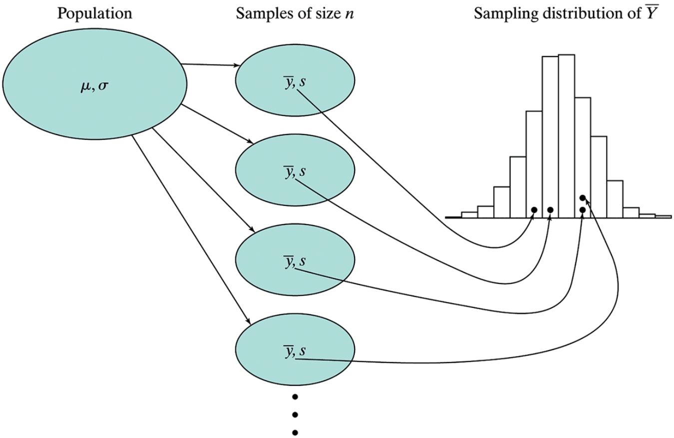
- Study is measuring change in blood pressure in $n = 10$ rats after giving them a drug, and computing a mean change $\bar{Y}$ from $Y_1, \ldots, Y_{10}$.
- Meta study (which takes place in our mind) is simply repeating this study over and over again on different samples of $n = 10$ rats and computing a mean each time $\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \ldots$
- Because the sample is random each time, the means will be different.
- A (hypothetical) histogram of the $\bar{Y}_1, \bar{Y}_2, \bar{Y}_3, \ldots$ would give the **sampling distribution** of $\bar{Y}$, and smoothed version would give the density of $\bar{Y}$.
- Restated: the sample mean *from one randomly drawn sample of size* $n = 10$ has a density.

The density of $\bar{Y}$

- $\bar{Y}$ estimates $\mu_Y = E(Y_i)$, the mean of all the observations in the population.
- We'll first look at a picture of where the **sampling distribution of $\bar{Y}$** comes from.
- Then we'll discuss a Theorem that tells us about the mean $\mu_{\bar{Y}}$, standard deviation $\sigma_{\bar{Y}}$, and shape of the density for $\bar{Y}$.

Sampling distribution of $\bar{Y}$

"Meta-experiment..."



Sampling distribution of $\bar{Y}$

Theorem 5.2.1: The Sampling Distribution of $\overline{Y}$

1. Mean The mean of the sampling distribution of $\overline{Y}$ is equal to the population mean. In symbols,

$$\mu_{\overline{Y}} = \mu$$

2. Standard deviation The standard deviation of the sampling distribution of $\overline{Y}$ is equal to the population standard deviation divided by the square root of the sample size. In symbols,

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}$$

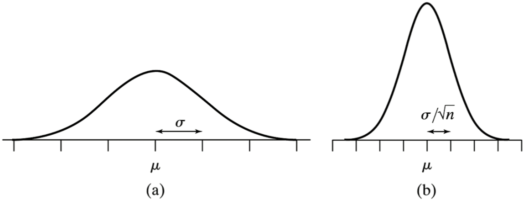
3. Shape

   (a) If the population distribution of Y is normal, then the sampling distribution of $\overline{Y}$ is normal, regardless of the sample size n.

   (b) *Central Limit Theorem* If n is large, then the sampling distribution of $\overline{Y}$ is approximately normal, even if the population distribution of Y is not normal.

Sampling distribution of $\bar{Y}$ from normal data

If data $Y_1, Y_2, \ldots, Y_n$ are normal, then $\bar{Y}$ is *also normal*, centered at the same place as the data, but with smaller spread.



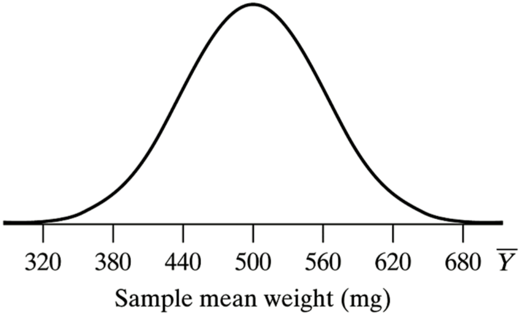
(a) population distribution of normal data $Y_1, \ldots, Y_n$, and (b) sampling distribution of $\bar{Y}$.

Example 5.2.2 Seed weights

- The population of weights of the princess bean is *normal* with $\mu = 500$ mg and $\sigma = 120$ mg. We intend to take a samplle of $n = 4$ seeds and compute the (random!) sample mean $\bar{Y}$.
- $E(\bar{Y}) = \mu_{\bar{Y}} = \mu = 500$ mg. *On average, the sample mean gets it right.*
- $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{4}} = 60$ mg. 68% of the time, $\bar{Y}$ will be within 60 mg of $\mu = 120$ mg.

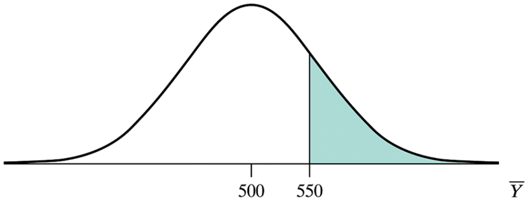
Sampling distribution for $\bar{Y}$ for Example 5.2.2

$\mu_{\bar{Y}} = 500$ mg and $\sigma_{\bar{Y}} = 60$ mg.



$\Pr\{\bar{Y} > 550\}$ for $n = 4$

Recall for $n = 4$ that $\mu_{\bar{Y}} = 500$ mg and $\sigma_{\bar{Y}} = 60$ mg.




```
> 1-pnorm(550,500,60)
[1] 0.2023284
```

What happens when n is increased?

- As n gets bigger, $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$ gets smaller. The density of $\bar{Y}$ gets more focused around μ .
- If $Y_1, \ldots, Y_n$ come from a normal density, then so does $\bar{Y}$, *regardless of the sample size*.
- Even if $Y_1, \ldots, Y_n$ *do not* come from a normal density, the *Central Limit Theorem* guarantees that the density of $\bar{Y}$ will look more and more like a normal distribution as n gets bigger.
- This is in Section 5.3; have a look if you're interested.

Sampling dist'n for $\bar{Y}$ from different sample sizes n