Econometrics (Econ 452) – Fall 2022 – Instructor: Daniele Girardi

2 – REVIEW OF PROBABILITY

University of Massachusetts Amherst BE REVOLUTIONARY™

THE PLAN

1. Random Variables and Probability Distributions
2. Expected Value, Mean and Variance
3. Two Random Variables
4. The Normal Distribution
5. Random Samples and the Distribution of Sample Averages

WHY PROBABILITY THEORY?

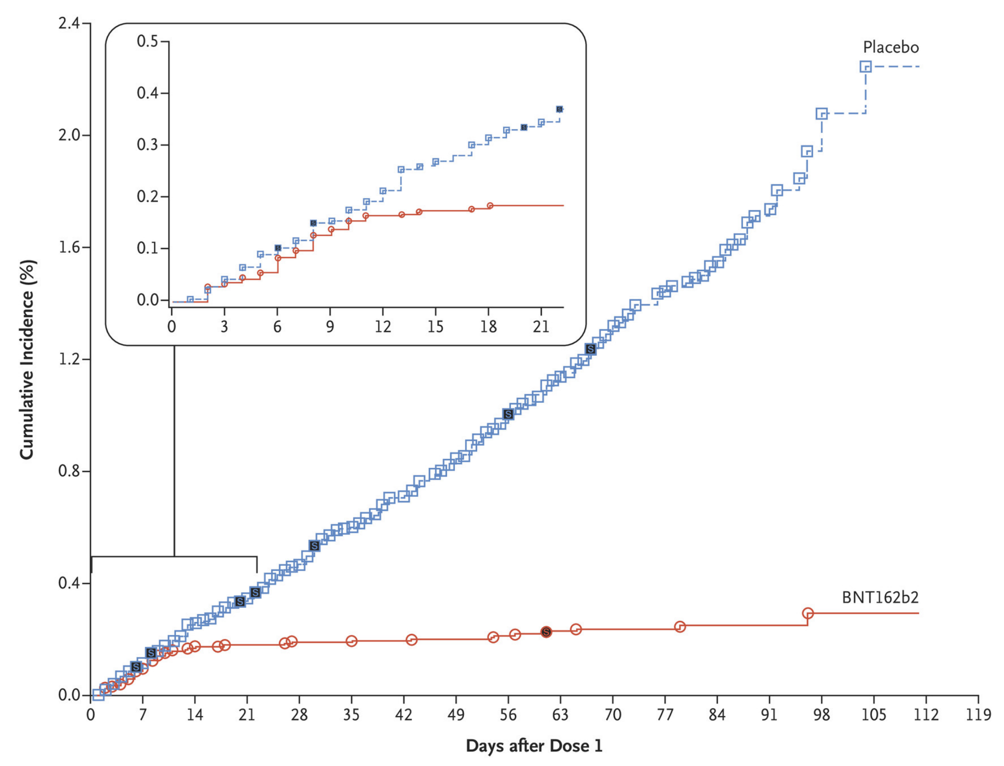
- **We live in an uncertain & risky world.**

- **Application to econometrics: *statistical inference***
  - *Ability to discern random fluctuations from changes caused by some intervention.*

EXAMPLE: EVALUATING COVID VACCINES

Clinical trial for Pfizer COVID vaccine

- 43,448 participants worldwide.
- ½ received vaccine, ½ a placebo.
- 8 infections among vaccinated vs. 162 in placebo group.
- *How likely is this difference to have arisen randomly, rather than reflecting an immunization effect of the vaccine?*
- ***Probability theory*** provides the tools to answer this question!



Efficacy End-Point Subgroup	BNT162b2, 30 μg (N=21,669)		Placebo (N=21,686)		VE (95% CI)
	No. of participants	Surveillance time *person-yr (no. at risk)*	No. of participants	Surveillance time *person-yr (no. at risk)*	*percent*
Covid-19 occurrence					
After dose 1	50	4.015 (21,314)	275	3.982 (21,258)	82.0 (75.6–86.9)
After dose 1 to before dose 2	39		82		52.4 (29.5–68.4)
Dose 2 to 7 days after dose 2	2		21		90.5 (61.0–98.9)
≥7 Days after dose 2	9		172		94.8 (89.8–97.6)

Source: [Polack et al (2020) "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine"]

2.1 RANDOM VARIABLES AND PROBILITY DISTRIBUTIONS

KEY CONCEPTS

- **Random processes**
  - Number of times your computer crashes while writing an essay.
  - Time it takes you to get on campus tomorrow.
  - Number of times campus closes due to snowstorm.
  - *Influenced by something not known in advance, that will eventually be revealed.*
- **Outcomes** of random processes
- **Probabilities** of different outcomes
- **Sample space**
- **Event**
- **Random variables (RV)**
  - o discrete vs. continuous

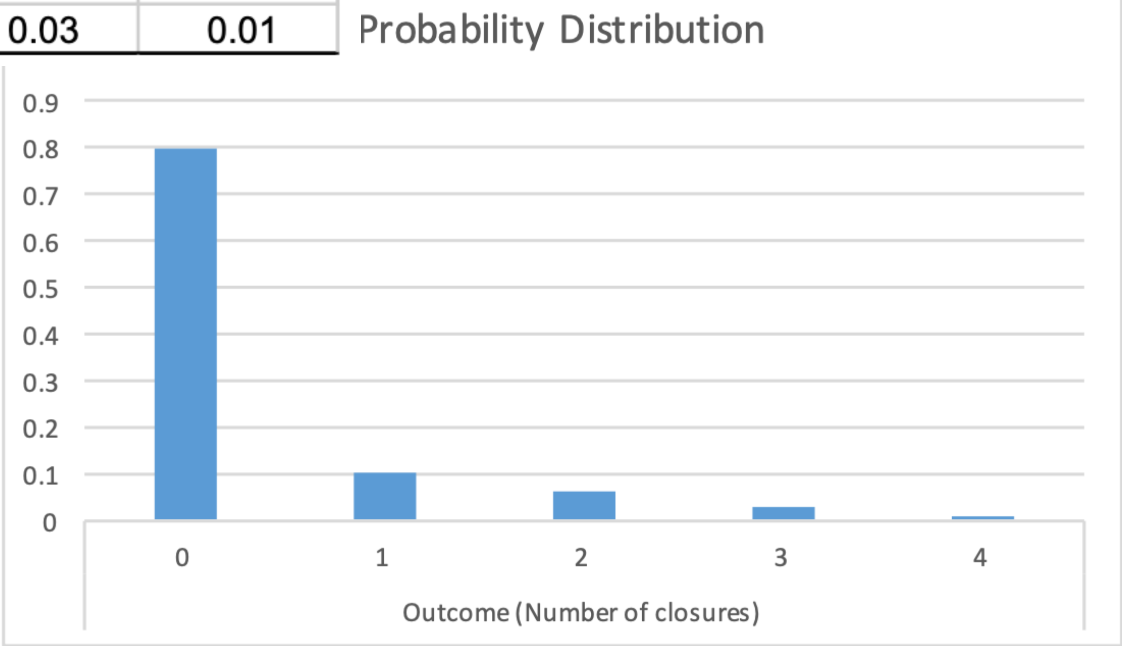
PROBABILITY DISTRIBUTION

- **Probability distribution of a discrete RV:** a list all possible outcomes and the probability that each will occur.
- The probabilities must sum to one.
- **Example:** number of snow closures at UMass in the next Spring semester
  - M = number of snow closures.
  - $\Pr(\mathrm{M} = 0)$ is the probability of no closures.
  - $\Pr(\mathrm{M} = 1)$ is the probability of one closure
  - $\Pr(\mathrm{M} = 2)$ is the probability of two closures
  - ….

A (HYPOTETICAL) PROBABILITY DISTRIBUTION FOR SNOW CLOSURES AT UMASS

	Outcome (Number of closures)				
	0	1	2	3	4
Probability Distribution	0.8	0.1	0.06	0.03	0.01

- *How to interpret probabilities?*
- *What does it mean for the probability of one closure to be 10%?*
- *Thought experiment: repeat the same semester many times under the same conditions.*



A (HYPOTETICAL) PROBABILITY DISTRIBUTION FOR SNOW CLOSURES AT UMASS

	Outcome (Number of closures)				
	0	1	2	3	4
Probability Distribution	0.8	0.1	0.06	0.03	0.01

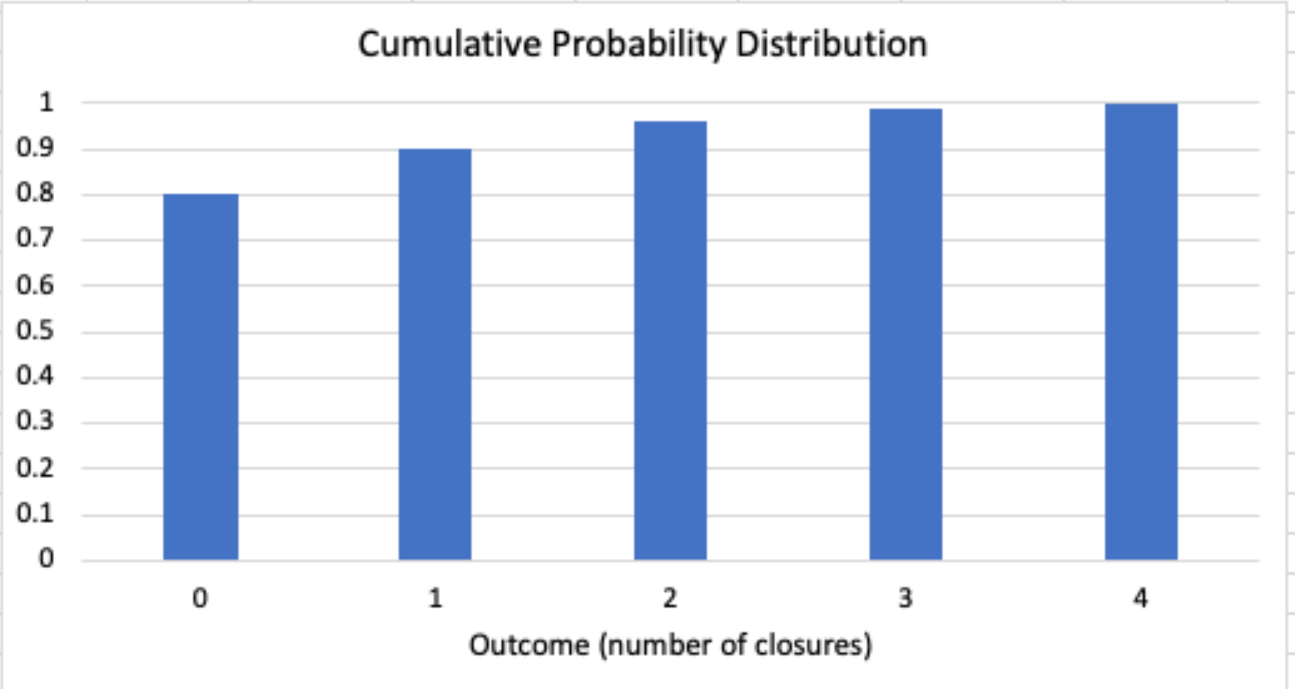
Examples of *events*

- "not more than one closure"
  - $\Pr(\text{M} = 0 \text{ or } \text{M} = 1) = \Pr(\text{M} = 0\ ) + \Pr(\text{M} = 1) = 0.8 + 0.1 = 0.9$
- "either zero or two closures"
  - $\Pr(\text{M} = 0 \text{ or } \text{M} = 2) = \Pr(\text{M} = 0\ ) + \Pr(\text{M} = 2) = 0.8 + 0.06 = 0.86$
- "not more than two closures"
  - $\Pr(\text{M} = 0 \text{ or } \text{M}=1 \text{ or } \text{M}=2) = \Pr(\text{M} = 0\ ) + \Pr(\text{M} = 1) + \Pr(\text{M} = 2) = 0.8+0.1+0.06$

THE CUMULATIVE PROBABILITY DISTRIBUTION

- Gives the probability that the RV is less than or equal to a particular value.
- In our “UMass snow closures” example:

	Outcome (Number of closures)				
	0	1	2	3	4
Probability Distribution	0.8	0.1	0.06	0.03	0.01
Cumulative probability distribution	0.8	0.9	0.96	0.99	1

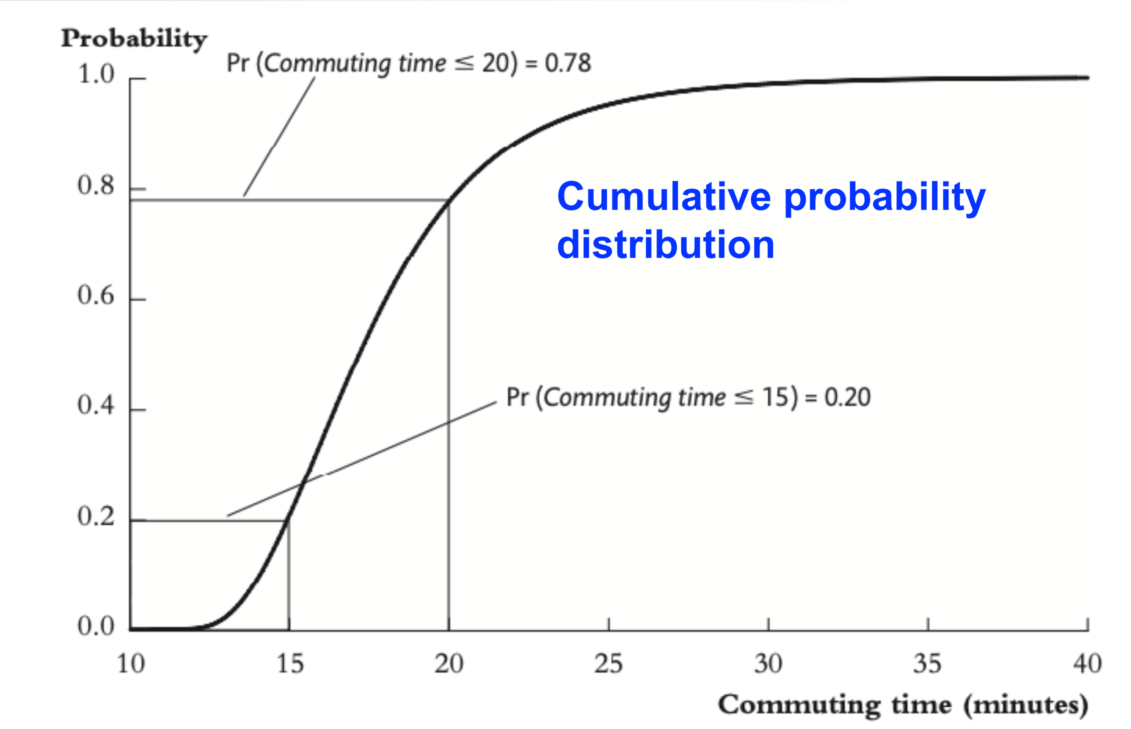


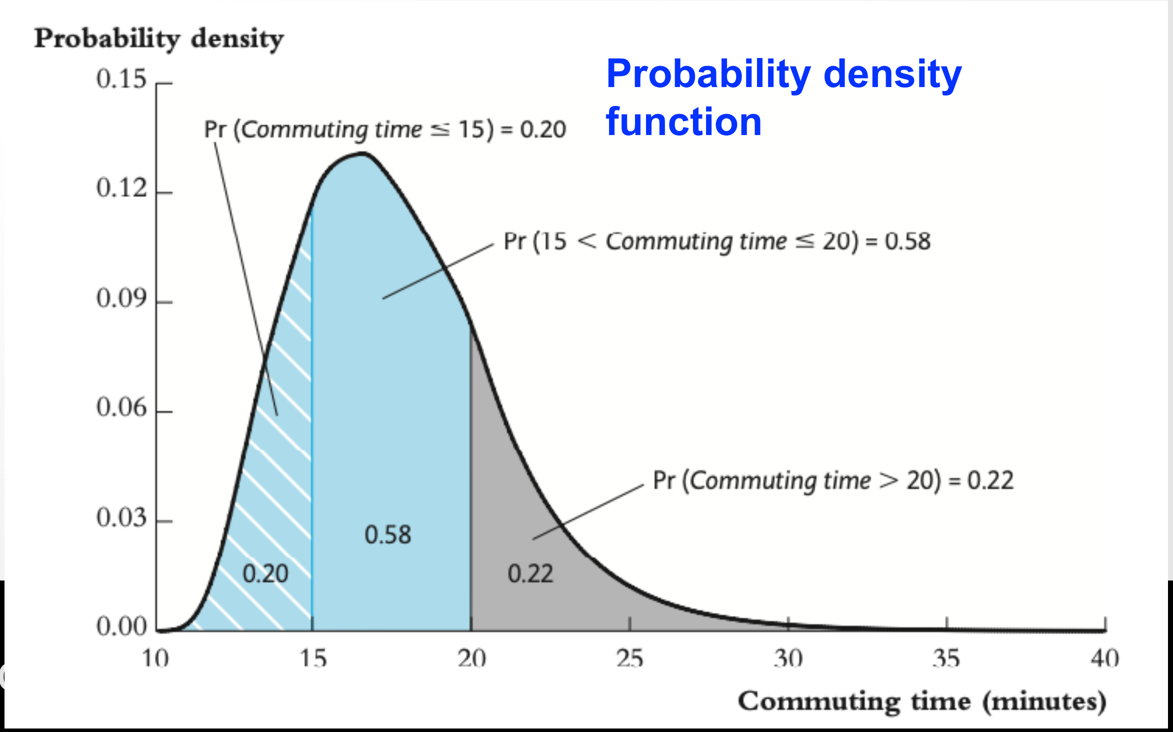
THE BERNOULLI DISTRIBUTION

- Binary (or Bernoulli) RV
  - possible outcomes are either 0 or 1.
  - Examples:
    - Result of a coin toss (1=Heads; 0=Tails).
    - Pass/Fail an Econometrics Exam (1=Pass; 0=Fail)
    - US citizenship status (1= citizen; 0=non-citizen)
- Its distribution is called ‘Bernoulli distribution’.
  - $\Pr(G = 1) = p$
  - $\Pr(G = 0) = 1 - p$

PROBABILITY DISTRIBUTION OF A CONTINUOUS RV

- Similar concepts, more complicated math.
- **Cumulative probability distribution:** $\boldsymbol{Pr(X \leq x)}$
- **Probability density function (p.d.f.):** $\boldsymbol{Pr(a \leq X \leq b)}$
- **Example:** time it gets me to drive from Noho to UMass





2.2 EXPECTED VALUE, MEAN AND VARIANCE

EXPECTED VALUE

- Denoted as $\boldsymbol{E(Y)}$ or $\boldsymbol{\mu_Y}$.
- Also called *expectation* or *mean*.
- The long-run average value of Y over many repeated occurrences.
- Weighted average of all possible outcomes, with weights given by probabilities.
- In our UMass snow closures example:

	Outcome (Number of closures)				
	0	1	2	3	4
Probability Distribution	0.8	0.1	0.06	0.03	0.01

$$\boldsymbol{E(Y)} = (0 \, x \, 0.80) + (1 \, x \, 0.1) + (2 \, x \, 0.06) + (3 \, x \, 0.03) + (4 \, x \, 0.01) = \mathbf{0.35}$$

EXPECTED VALUE

- If Y is a discrete RV with *k* possible outcomes:

$$E(Y) = y_1 p_1 + y_2 p_2 + \cdots + \ y_k p_k = \sum_{i=1}^{k} y_i p_i$$

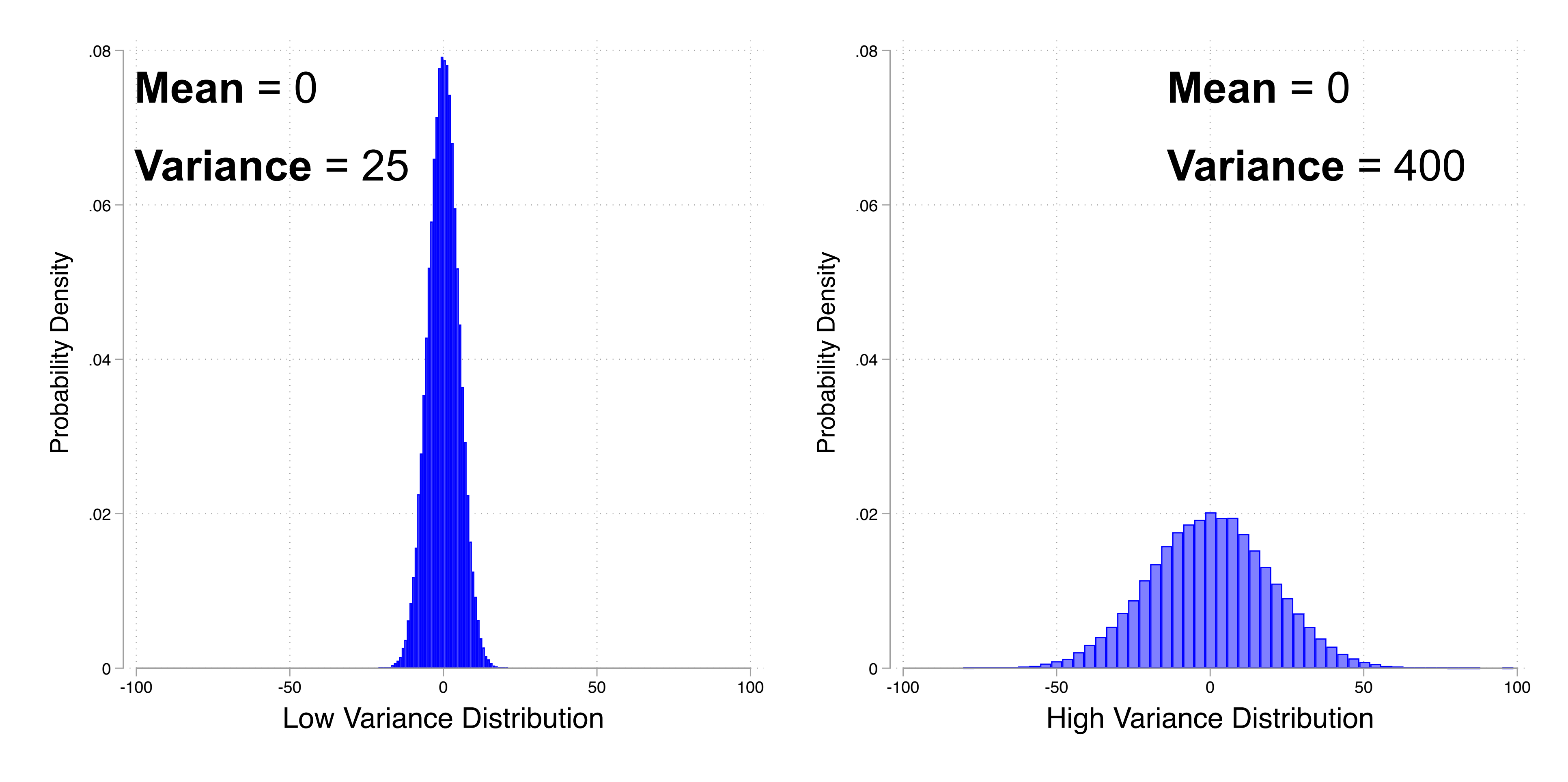
- If G is a Bernoulli (aka binary) RV:

$$E(G) = [0\ x(1-p)] + [1\ x\ p] = p$$

- If X is a continuous RV:

$$E(X) = \int_{-\infty}^{+\infty} x \ \cdot f(x) dx$$

STANDARD DEVIATION AND VARIANCE



STANDARD DEVIATION AND VARIANCE

- Measure the dispersion ("spread") of a probability distribution.
- **Variance** = Mean *squared deviation* of Y from E(Y)
- ***Var(Y)*** $= E(Y-\mu_Y)^2 = \underbrace{\sum_{i=1}^{k}(y_i-\mu_Y)^2 p_i}$

  $= \sigma_Y^2$

(Assuming Y is a discrete RV with *k* possible outcomes)

- **Standard deviation** = square root of the variance
- $\sigma_Y = \sqrt{Var(Y)}$

STANDARD DEVIATION AND VARIANCE

- **Your turn:** calculate the variance and standard deviation of the number of campus closures from our example distribution

	Outcome (Number of closures)				
	0	1	2	3	4
Probability Distribution	0.8	0.1	0.06	0.03	0.01

- Remember:
  - $\text{Var(y)} = \sum_{i=1}^{k} (y_i - \mu_Y)^2 p_i$
  - $\sigma_Y = \sqrt{Var(Y)}$
- From our previous calculation: μ_Y=0.35

STANDARD DEVIATION AND VARIANCE

	Outcome (Number of closures)				
	0	1	2	3	4
Probability Distribution	0.8	0.1	0.06	0.03	0.01

- Solution:
  - Var(y) = $\sum_{i=1}^{k}(y_i - \mu_Y)^2 p_i$ = 0.6475
  - $\sigma_Y = \sqrt{Var(Y)}$ = 0.80

VARIANCE OF A BERNOULLI RV

- Applying the variance formula to a Bernoulli RV:

$$Var(G) = \sigma_G^2 = p(1-p)$$

$$\sigma_G = \sqrt{p(1-p)}$$

MEAN AND VARIANCE OF LINEAR FUNCTIONS

- A linear function of X: $Y = 2000 + 0.8\,X$
- In general: $Y = a + b\,X$
- Mean: $E(Y) = E(a + b\,X) = a + b\,E(X)$
- Variance: $Var(Y) = Var(a + b\,X) = b^2 Var(X)$

STANDARDIZED RANDOM VARIABLES

- Consider random variable Y, with mean μ_Y and variance σ_Y.
- *Standardized* version of Y:

$$Z = \frac{(Y - \mu_Y)}{\sigma_Y}$$

- By design, if Z is a standardized RV, we always have

$$E(Z) = 0$$

$$\mathrm{V}ar(Z) = 1$$

2.3 TWO RANDOM VARIABLES

University of Massachusetts Amherst

TWO RANDOM VARIABLES

- Are Democrats more likely to get vaccinated against Covid than Republicans?

- Are graduates more likely to find a job than non-graduates?

- How do women and men's average earnings differ?

- All these Qs involve the relationship between **two RVs**.

JOINT PROBABILITY DISTRIBUTION

- *Joint probability distribution* of X and Y:

$$Pr(X = x, Y = y)$$

- *Conditional distribution* of Y given X:

$$Pr(Y = y|X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)}$$

- *Conditional expectation* (or conditional mean) of Y given X:

$$E(Y|X = x) = \sum_{i=1}^{k} y_i \, Pr(Y = y_i|X = x)$$

EXAMPLE: *JOINT DISTRIBUTION OF COVID VAX STATUS AND PARTISANSHIP IN THE US:*

	Dem (X=1)	Rep (X=0)	**Total**
Vaccinated (Y=1)	0.37	0.21	0.58
Not Vaccinated (Y=0)	0.18	0.24	0.42
Total	0.55	0.45	1.00

Your Turn: Figure out the following

1. Pr(Y=1, X=0)
2. Pr(Y=1 | X=1)
3. E(Y|X=1)

Source: calculated and adapted from data in [New York Times "The Vaccine Class Gap", 5-24-2021]

Reminder:

- $Pr(Y = y|X = x) = \frac{Pr(X=x, Y=y)}{Pr(X=x)}$
- $E(Y|X = x) = \sum_{i=1}^{k} y_i Pr(Y = y_i|X = x)$

	Dem (X=1)	Rep (X=0)	**Total**
Vaccinated (Y=1)	0.37	0.21	0.58
Not Vaccinated (Y=0)	0.18	0.24	0.42
Total	0.55	0.45	1.00

1. Pr(Y=1, X=0) = 0.21
2. Pr(Y=1 | X=1) = 0.67
3. E(Y|X=1) = 0.67

Reminder:

- $Pr(Y = y|X = x) = \frac{Pr(X=x,Y=y)}{Pr(X=x)}$
- $E(Y|X = x) = \sum_{i=1}^{k} y_i Pr(Y = y_i|X = x)$

CONDITIONAL VARIANCE

- Variance of Y conditional on X.
- It is the variance of the conditional distribution of Y given X.
- Conditional variance of Y given X:

$$var(Y|X = x) = \sum_{i=1}^{k} [y_i - E(Y|X = x)]^2 \, Pr(Y = y_i|X = x)$$

VARIANCE OF VAX STATUS CONDITIONAL ON PARTISANSHIP

	Dem (X=1)	Rep (X=0)	**Total**
Vaccinated (Y=1)	0.37	0.21	0.58
Not Vaccinated (Y=0)	0.18	0.24	0.42
Total	0.55	0.45	1.00

$$Var(Y|X=1) = [1 - E(Y|X=1)]^2 \Pr(Y=1|X=1) + [0 - E(Y|X=1)]^2 \Pr(Y=0|X=1) =$$

$$[1 - 0.67]^2\, 0.67 + [0 - 0.67]^2\, 0.33 =$$

$$= 0.22$$

$$Var(Y|X=0) = 0.25$$

INDEPENDENCE

- X and Y are *independently distributed* (or *independent*) if

$$\Pr(Y = y|X = x) = \Pr(Y = y) \qquad \text{for all possible X and Y}$$

- If X and Y are independent, then

$$\Pr(X = x, Y = y) = \Pr(X = x) \Pr(Y = y)$$

COVARIANCE

- How much do X and Y move together?
- Covariance:

$$cov(X,Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{l} (x_j - \mu_x)(y_i - \mu_Y)\Pr(X = x_y, Y = y_i)$$

CORRELATION

- The units of covariance are awkward (units of X * units of Y).
- Correlation:

$$corr(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- Correlation is unit free and always between -1 and +1.

SUMS OF RANDOM VARIABLES

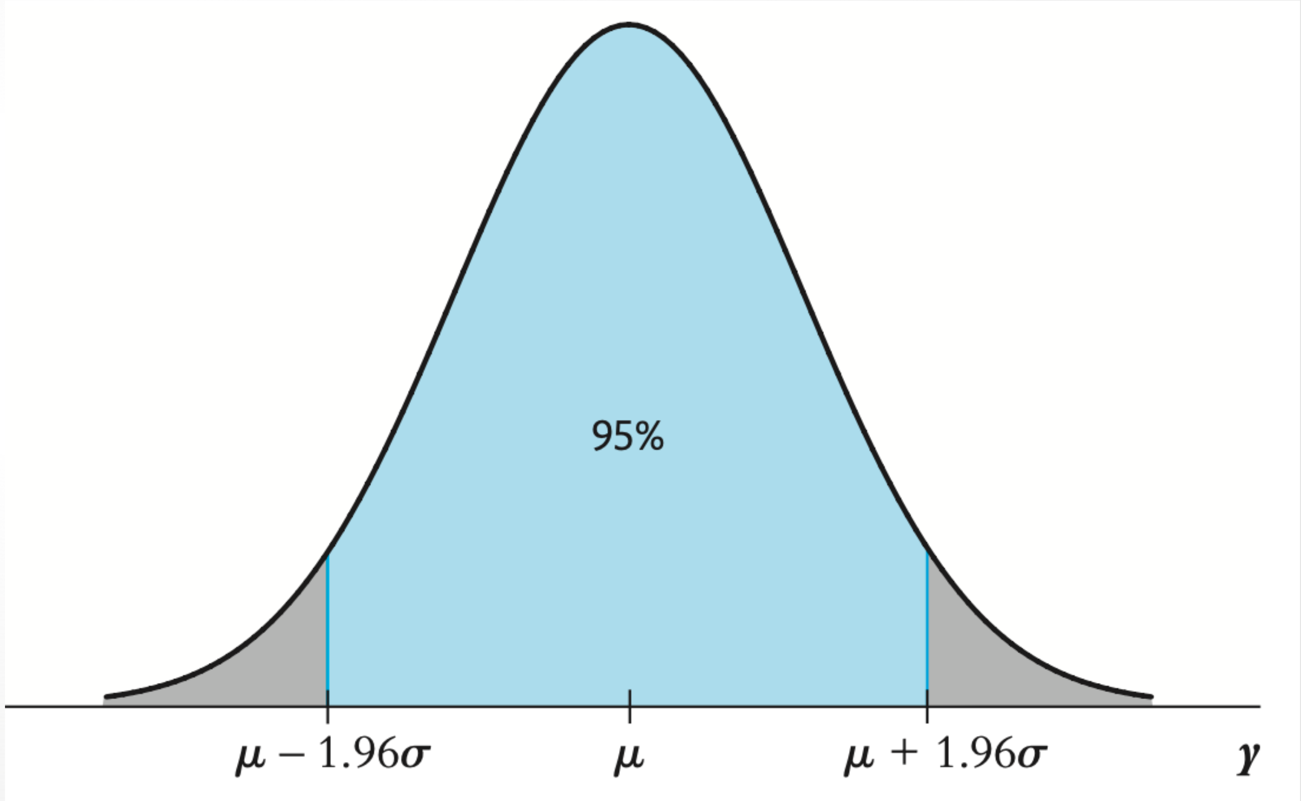
- $E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$
- $Var(X+Y) = \mathrm{var(X)} + \mathrm{var(Y)} + 2\mathrm{cov(X, Y)}$
- See textbook "Key Concept" 2.3 for useful equations related to means, variances and covariances of sums of RVs

2.4 THE NORMAL DISTRIBUTION

University *of* Massachusetts Amherst

THE NORMAL DISTRIBUTION

- Bell-shaped probability density.
- Symmetric around its mean.
- 95% of probability mass between $\mu - 1.96\sigma$ / $\mu + 1.96\sigma$
- Written as $N(\mu, \sigma^2)$
- Some random variables are distributed normally.

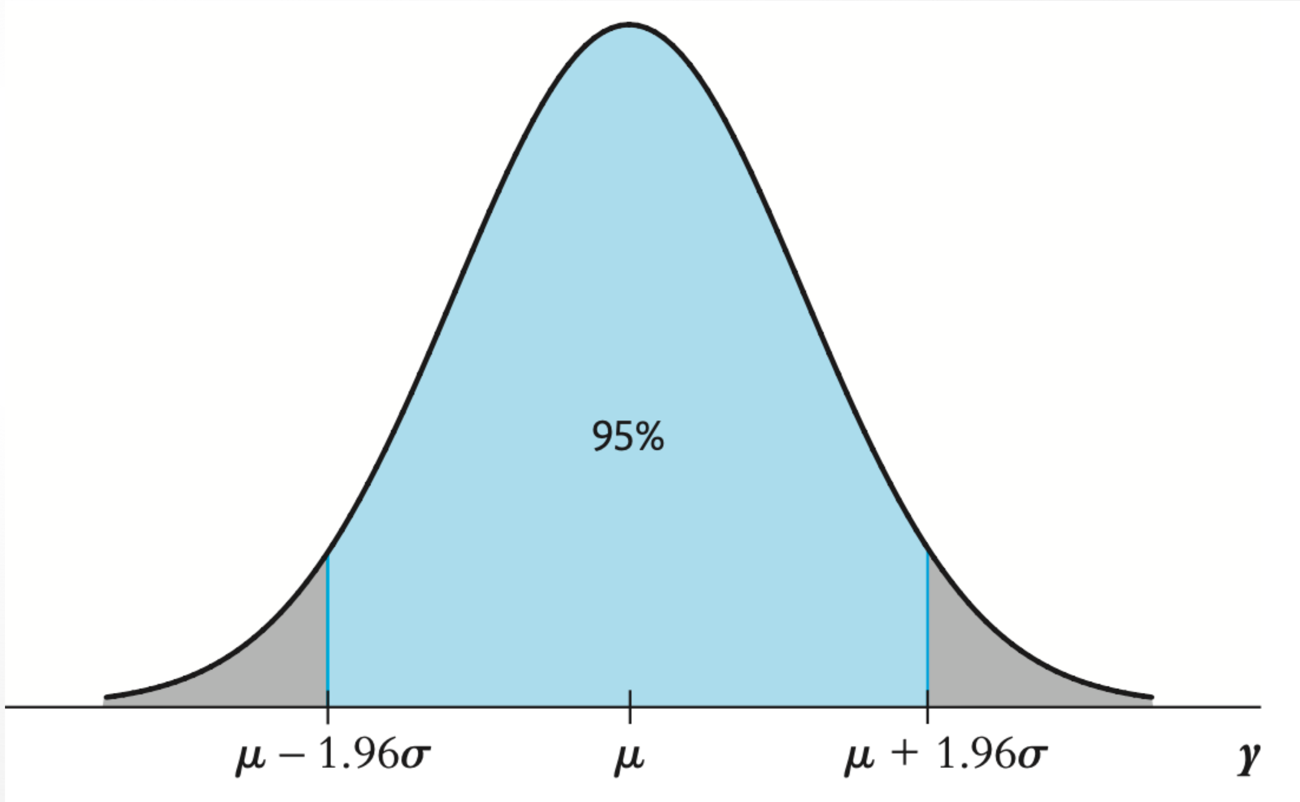


THE *STANDARDIZED* NORMAL DISTRIBUTION

1. Take a variable Y distributed $N(\mu, \sigma^2)$.
2. Standardize it:

$$Z = \frac{(Y - \mu_Y)}{\sigma_Y}$$

3. Z is distributed $N(0,1)$.



THE *STANDARDIZED* NORMAL DISTRIBUTION

- Z is distributed $N(0,1)$
- Then $\Pr(Z \leq z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$
- To get $\Phi(z)$ for any z:
  - Excel: NORMSDIST(z)
  - STATA: display normal (z)
  - Appendix Table 1: look up the probability of the desired value.

THE NORMAL DISTRIBUTION: EXAMPLE

- Say $X \sim N(1,4)$
- What is the probability that $X \leq 2$?
- $\Pr(X \leq 2) = \Pr\left(Z \leq \frac{2-1}{\sqrt{4}}\right) = \Pr(Z \leq 0.5) = \Phi(0.5)$
- Look up 0.5 in the table or "display normal(0.5)" in STATA.

```
. display normal(0.5)
.69146246
```

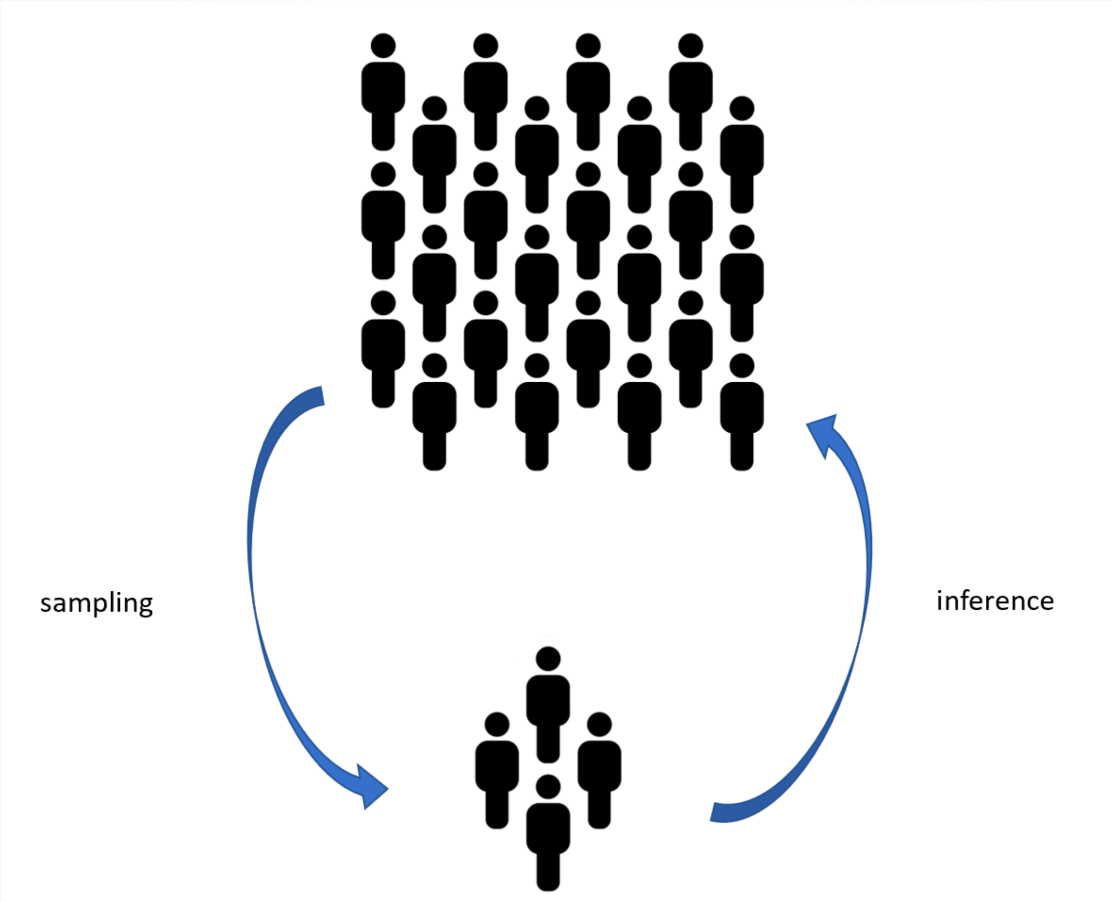
OTHER IMPORTANT DISTRIBUTIONS

- Chi-Squared distribution
- Student-t distribution
- F distribution
- …
- We don't really need to study them, at least for now.

2.5 RANDOM SAMPLES AND THE DISTRIBUTION OF SAMPLE AVERAGES

University of Massachusetts Amherst

POPULATION, SAMPLE, INFERENCE



RANDOM SAMPLING

- *n* randomly drawn observations of Y:

$$Y_1, Y_2, Y_3, \dots, Y_n$$

- $Y_1, \dots, Y_n$ are *random variables*: different from one random sample to the next.
- $Y_1, \dots, Y_n$ are i.i.d.
- The sample mean

$$\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{1}{n}\sum_{i=0}^{n} Y_i$$

is also a random variable.

- *Sampling distribution*: the probability distribution of $\bar{Y}$.

THE SAMPLING DISTRIBUTION OF $\bar{Y}$

If sample observations $Y_1, \ldots, Y_n$ are i.i.d.,

- $E(\bar{Y}) = \mu_Y$
- $var(\bar{Y}) = \sigma_{\bar{Y}}^2 = \frac{1}{n}\sigma_Y^2$
- $std.dev(\bar{Y}) = \sigma_{\bar{Y}} = \frac{1}{\sqrt{n}}\sigma_Y$

THE LAW OF LARGE NUMBERS

- **Law of large numbers**:
  - If n is larger, $\bar{Y}$ is more likely to be close to μ_Y.
- → $\bar{Y}$ is **consistent** estimator of μ_Y

SAMPLING DISTRIBUTIONS IN LARGE SAMPLES

- What does the probability distribution function of $\bar{Y}$ look like?
- If $Y \sim N(\mu_Y, \sigma_Y^2)$ then $\bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$ *irrespective of sample size*.
- **Central Limit Theorem**:
  - When n is large, $\bar{Y}$ is (approximately) normally distributed *even if Y is not*.
  - The larger n, the closer the distribution of $\bar{Y}$ to a normal.
  - → $\bar{Y}$ is asymptotically normally distributed.

THE CENTRAL LIMIT THEOREM