Independent Trials Model (and the Binomial Distribution)

- n independent trials conducted
- Success or Failure (*S* or *F*) on each trial
- P(Success) = p is the same for each trial

Example: **Z** = # of heads in n=4 tosses of a coin with P(*Heads*)=p

Consider **3** Heads in **4** tosses:

Outcome	Probability	z
HHHT	p*p*p*(1-p)	3
HHTH	p*p*(1-p)*p	3
HTHH	p*(1-p)*p*p	3
THHH	(1-p)*p*p*p	3

⇨ **$P(Z=3) = 4 p^3 (1-p)$**

***4* is the number of ways of ordering the Heads and Tails**

Binomial Random Variable:

- **X** = # of Successes in n independent trials
- $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ gives the number of ways to order the Successes & Failures
  - $n! = n \bullet (n-1) \cdots \bullet 2 \bullet 1$ → e.g., $4! = 4 \bullet 3 \bullet 2 \bullet 1 = 24$

Is it a binomial random variable?

1. A coin is weighted in such a way so that there is a 70% chance of getting a head on any particular toss. Toss the coin, in exactly the same way, 100 times. Let X equal the number of heads tossed.

2. A college administrator randomly samples students until he finds four that have volunteered to work for a local organization. Let X equal the number of students sampled.

3. A Quality Control Inspector investigates a lot containing 15 skeins of yarn. The Inspector randomly samples (without replacement) 5 skeins of yarn from the lot. Let X equal the number of skeins with acceptable color.

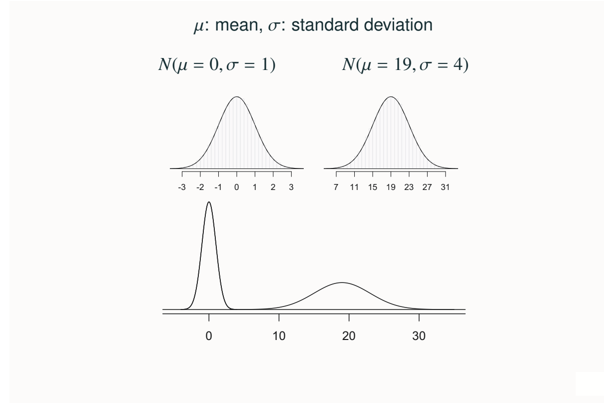
It is estimated that 20% of US citizens have no health insurance. A SRS of n = 15 is taken and let X denote the number in the sample with no health insurance.

What is the probability that exactly 3 of the 15 sampled have no health insurance?

What is the probability that at most one of those sampled has no health insurance?

A **density curve** is an idealized description of the distribution of a set of data.

- $\bar{x}$ and s denote the mean and SD of the sample observations
- μ and σ denote the mean and SD of the density curve



Normal Density Curves $N(\mu,\sigma)$

- Represent normal probability distributions: $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$

- Curvature changes at μ-σ and μ+σ

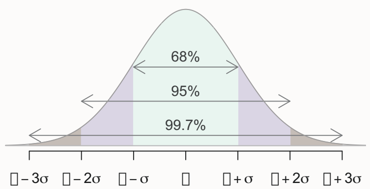
Standard Normal Distribution $N(0,1) \rightarrow$ mean=0 SD=1

Table 1:

$$P(Z \leq 1.3) = \int_{-\infty}^{1.3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.9032 \quad \text{or} \quad 90.32\%$$

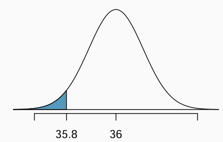
$$= \text{Proportion of obs. } \leq 1.3$$

- For nearly normally distributed data,
  - about 68% falls within 1 SD of the mean,
  - about 95% falls within 2 SD of the mean,
  - about 99.7% falls within 3 SD of the mean.



At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection. What percent of bottles have less than 35.8 ounces of ketchup?

Let X *= amount of ketchup in a bottle:* $X \sim N(\mu = 36, \sigma = 0.11)$



Standardized values ("z-scores")

- If x is an observation from a distribution with mean=μ and SD=σ

  Then $z = \frac{x-\mu}{\sigma}$ is the *standardized value* for x (or "z-score")
- If X is $N(\mu,\sigma)$ then Z is $N(0,1)$

Example:
Suppose that a small company produces "one-pound" chocolate bars whose weights are actually approximately normally distributed with μ=450g and σ=5g

1. What proportion of chocolate bars are more than 454g?
2. What proportion are between 440g and 454g?
3. How heavy is a bar that is at the 95th percentile of weights?

Blood Glucose levels for women (40 years of age) are roughly normal with

μ = 90 mg/dl and σ = 20 mg/dl

1) What proportion have levels less than 80 mg/dl ?

2) What proportion have levels greater than 130 mg/dl ?

3) What proportion have levels between 90 and 130 mg/dl ?

4) How high does one's glucose level have to be in order to be at the 90th percentile?