Biostat 200 – Lab 3
Lecture Review

October 3, 2013

Probability Distributions

- Describe the possible values of your random variable
- A theoretical distribution of an infinite number of experiments
- Discrete distributions
  - Bernoulli
  - Binomial
  - (Poisson)
- Continuous distributions
  - Normal

What probability distributions tell you

- For discrete variables the probability distribution describes the probability of each possible value
  - $P(X=x)$
- For continuous variables, the distribution describes the probability of a *range* of values
  - $P(Z>x)$, $P(Z<x)$
- All are characterized (or summarized) by certain parameters

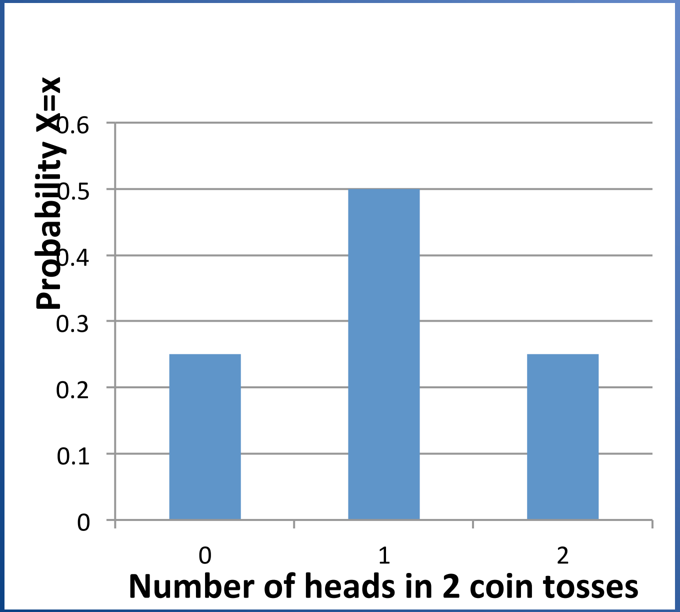
Discrete distributions

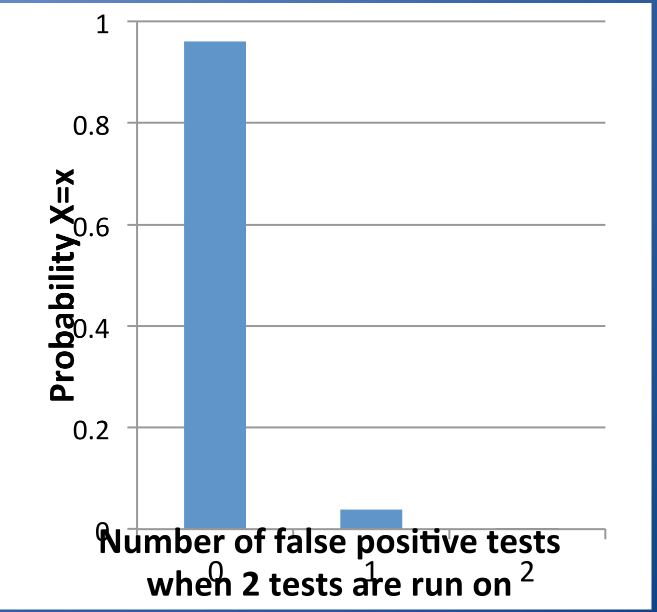
Bernoulli

Binomial

Discrete distributions

- Probabilities of all possible values add up to 1
- Examples
  - flipping a coin twice
  - 2 independent diagnostic tests





Bernoulli distribution

- Bernoulli random variable: can take 1 of 2 values with a constant probability *p*
  - $P(Y=1) = p$
  - $P(Y=0) = 1-p$
- Parameter: *p*
- Outcome: 0 or 1 only
- Describes only one trial (theoretical, not practical)

Binomial distribution

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- An extension of Bernoulli
- Example: probability of 5 people having a disease, knowing the prevalence of that disease
- Parameters: *n,p*
  - *n*: is the number of "trials" (e.g. people sampled)
  - *p*: probability of "success" in each "trial" (e.g. disease prevalence)
- Outcome: x, the number of "successes"
- Stata and Table A.1 use the symbol k for x

Binomial distribution - assumptions

- Fixed number of trials n
- Two mutually exclusive outcomes for each trial
- Outcomes of the n trials are independent (allowing multiplicative rule)
- Probability of success p is constant for each trial

n choose *x*

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{where } \mathrm{n}! = \mathrm{n} * \ldots * 3 * 2 * 1$$

- number of different ways to get *x* successes in *n* trials
- Examples:
  - 5 ways for 1 success in 5 trials
  - 10 ways for 2 successes in 5 trials
- Stata: `display comb (n,k)`

Using binomial distribution - formula

- What is the probability of exactly 2 cases of disease in a sample of n=5 with p=0.15?
- Use binomial formula

```
. display comb (5,2)
10
```

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

- Calculate (or use Stata):
  - $(10)(0.15)^2 (1\text{-}0.15)^{5-2}$
  - (10)(0.0225) (0.614) = 0.1382

Using binomial distribution – Table A.1

- What is the probability of exactly 2 cases of disease in a sample of n=5 with p=0.15?
- Use Table A.1
  - Gives you P(X=k)
  - Look up p=0.15, n=5, k=2
  - Answer = 0.1382

Using binomial distribution – Stata

- What is the probability of exactly 2 cases of disease in a sample of $n=5$ with $p=0.15$?
- Use Stata
  - binomialp (n,k,p)
  - $P(X=k)$ in *n* trials with probability of success *p*

```
display binomialp(5,2,.15)
.13817813
```

Using binomial distribution - formula

- What is the probability of 1 or more cases of disease in a sample of n=5 with p=0.15?
- Use binomial formula
- $P(X \geq 1) = 1 - P(X=0)$

$$P(X=0) = \binom{5}{0} .15^0 (.85)^5 = 1 * .85^5$$

```
. display comb (5,0)*0.85^5
.44370531
```

- $P(X \geq 1) = 1 - 0.4437 = 0.5563$

Using binomial distribution – Table A.1

- What is the probability of <u>1 or more cases</u> of disease in a sample of $n=5$ with $p=0.15$?
- Use Table A.1
  - $P(X \geq 1) = 1 - P(X=0)$
  - Look up $P(X=0)$, $p=0.15$, $n=5$
  - Get 0.4437
  - $P(X \geq 1) = 1 - 0.4437 = 0.5563$

Using binomial distribution – Stata

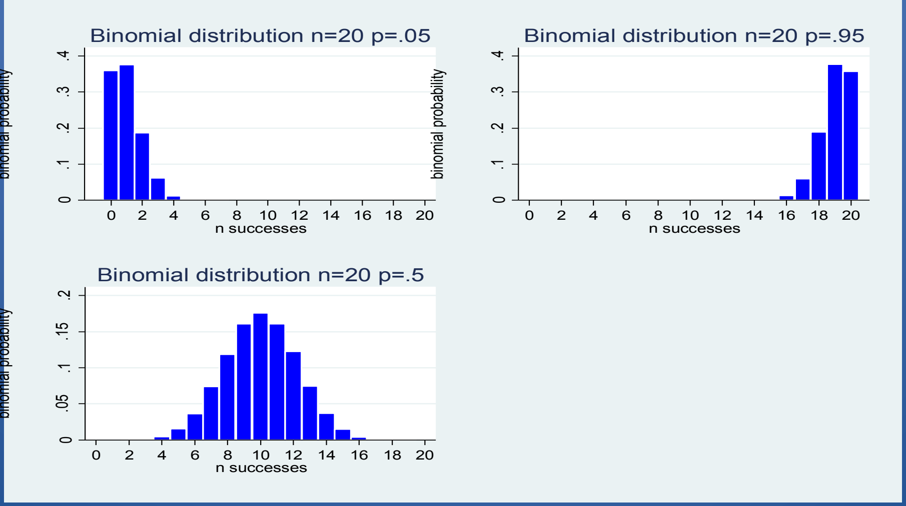
- What is the probability of <u>1 or more cases</u> of disease in a sample of n=5 with p=0.15?
- Use Stata
  - $P(X \geq 1) = 1 - P(X=0)$
  - binomialtail(n,k,p) gives $P(X \geq k)$

```
display binomialtail(5,1,.15)
 .55629469
```

  - (could use 1 – binomialp, keeping in mind your k)

Binomial distribution

- Binomial mean = np
- Binomial variance = np(1-p)
  - When p=0.5, variance is the largest and the distribution is symmetric
  - Variance is smaller when p is closer to 0 or 1
  - The distributions for p (0.05) and 1-p (0.95) are mirror images



Continuous distributions

Normal

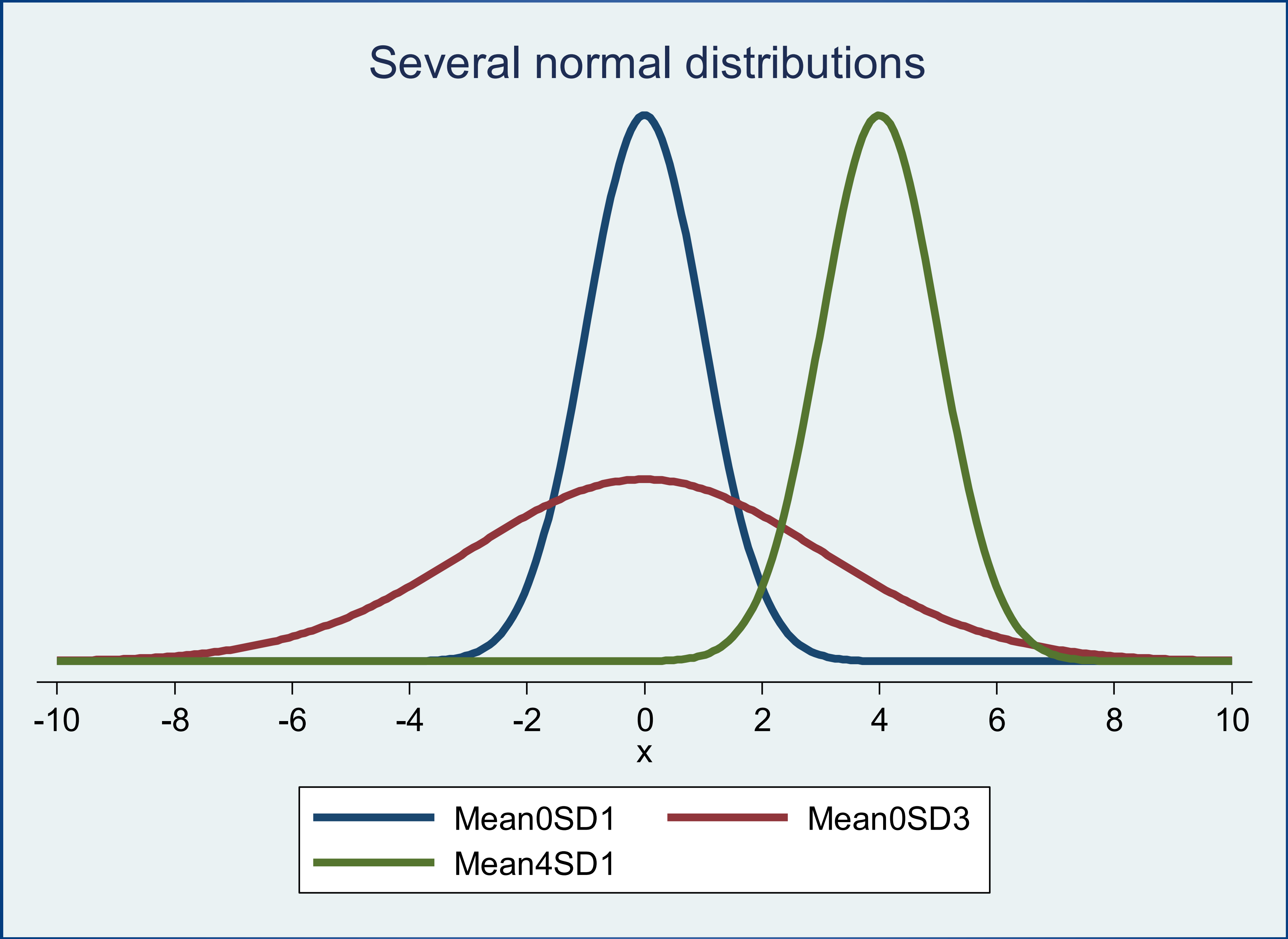
For continuous variables, the distribution describes the probability of a range of values

Normal distribution

- Probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{where } -\infty < x < \infty$$

- Parameters: μ, σ
  - μ: mean
  - σ: standard deviation
- The area under the curve = 1
- The continuous variable X can take on an infinite number of values


Several normal distributions
-10
-8
-6
-4
-2
0
2
4
6
8
10
x
Mean0SD1
Mean0SD3
Mean4SD1

Standard normal distribution

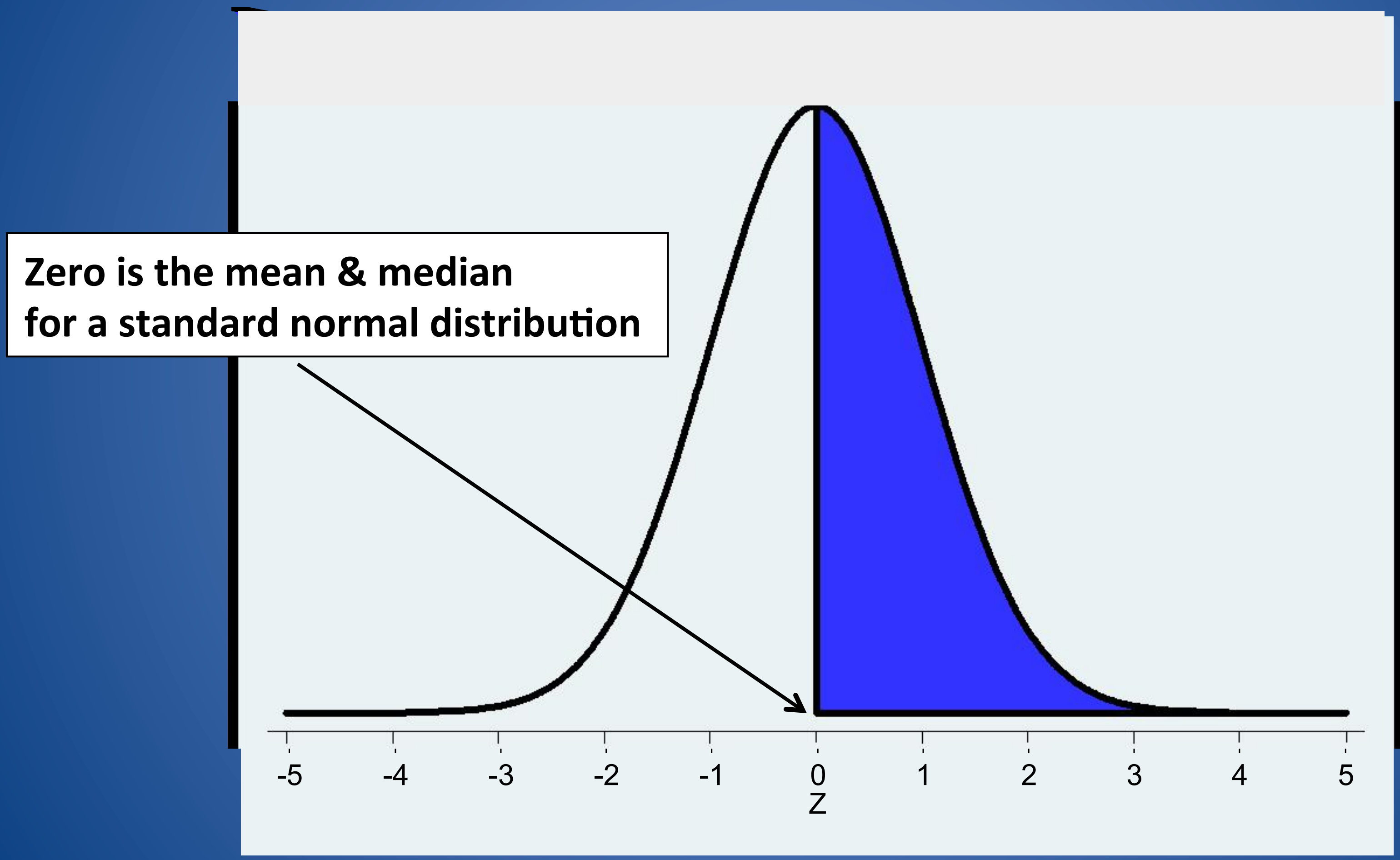
- μ and σ can take on an infinite number of values
- The standard normal (reference) curve has
  - Mean $\mu = 0$
  - Standard deviation $\sigma = 1$ (and variance $\sigma^2 = 1$)
- N(0,1)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{where } -\infty < x < \infty$$

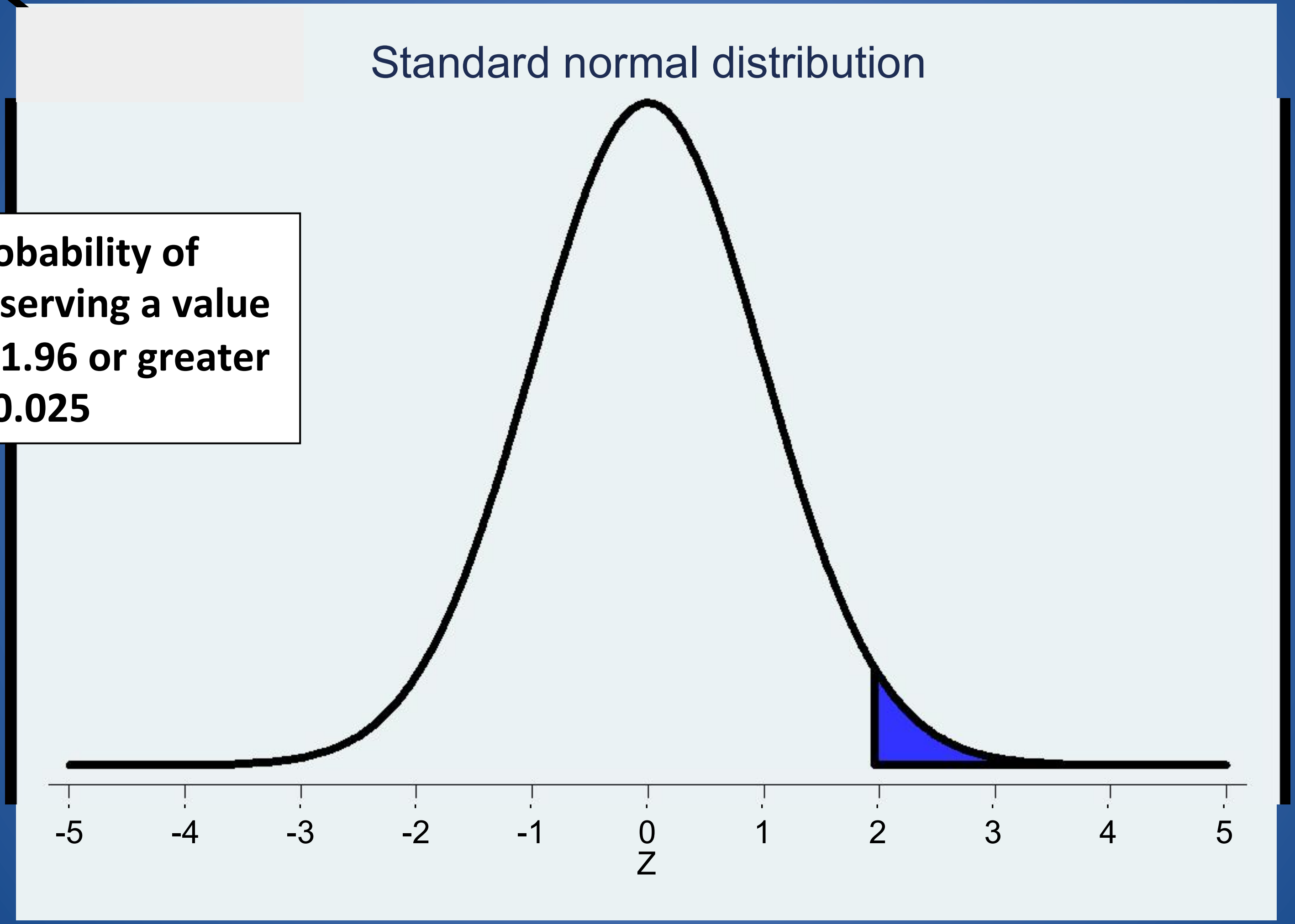
Standard normal random variable

- X: normally distributed random variable with mean μ and standard deviation σ
- Then $Z = (X - \mu)/\sigma$
- Z is a standard normal random variable with mean 0 and standard deviation 1

For Z ~ N(0,1) and P(Z≥0) = 0.50



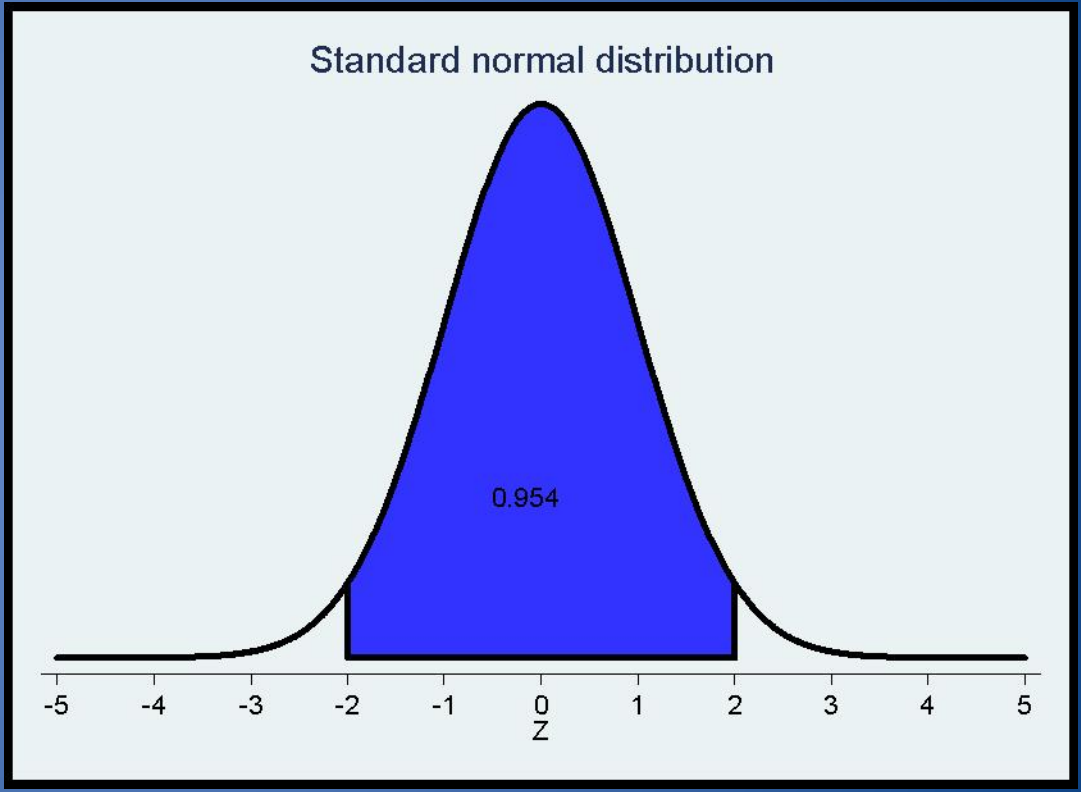
For Z ~ N(0,1) and P(Z≥1.96) = 0.025



Probability of observing a value of 1.96 or greater is 0.025

Area within 2 standard deviations of the mean

- $P(\mu-2\sigma \leq Z \leq \mu+2\sigma)$
- $\mu=0$ and $\sigma=1$, so:
- $P(-2<Z<2) = 0.954$
- Approximately 95.4% of the area of the standard normal is within 2 SD of the mean



Calculating standard normal probabilities (Stata)

- Left portion of the curve P(Z<z)
  - use the `normal` command

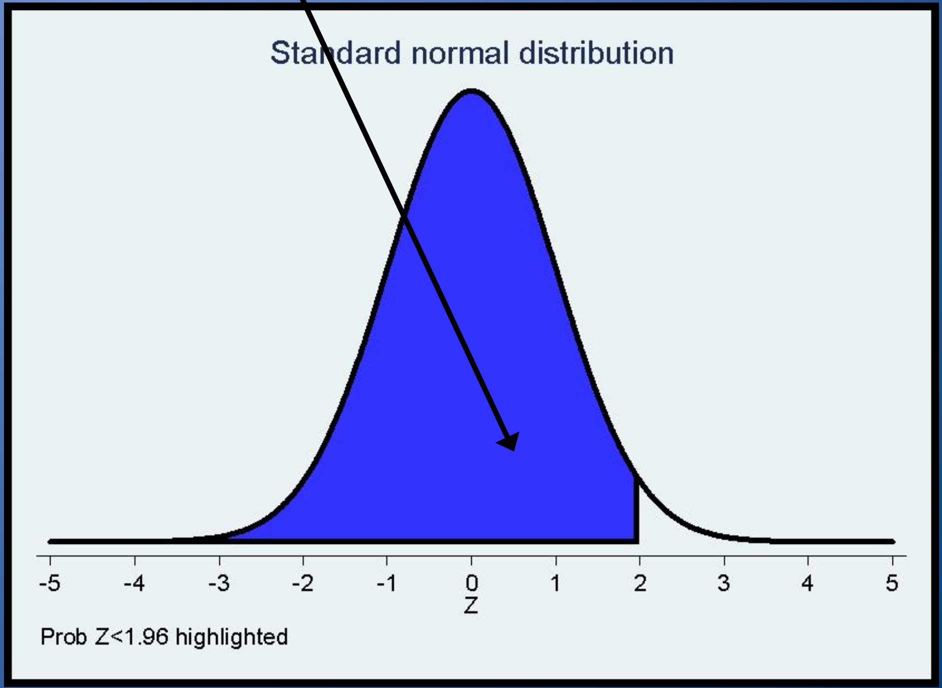
```
display normal(1.96)
.9750021
```

- Right portion of the curve P(Z≥z)
  - use `1-normal`

```
display 1-normal(1.96)
.0249979
```

- Middle
  - Subtract the right tail from the left portion

```
display normal(1.96)-
normal(-1.96)
.95000421
```



Calculating z-values for probabilities - Stata

- To get the z value for $P(Z \leq z) = p$ use

```
display invnormal(p)
```

- E.g. what is the z value for $P(Z \leq z) = 0.025$

```
display invnormal(0.025)
-1.959964
```

- To get the z value for $P(Z > z) = p$ use

```
display invnormal(1-p)
```

- E.g. what is the z value for $P(Z > z) = 0.025$

```
display invnormal(1-.025)
1.959964
```

Calculating z-values for probabilities – Table A.3

- To get the z value for $P(Z>z) = p$
  – Find p in the table, read corresponding z
- E.g. what is the z value for $P(Z>z) = 0.025$
  – For $p=0.025$ the table value is 1.96

- To get the z value for $P(Z\leq z) = p$
  – Find p in the table, and use (-1)*corresponding p
- E.g. what is the z value for $P(Z\leq z) = 0.025$
  – For $p=0.025$ the table value is 1.96, so the answer is -1.96