Chương 6
Phân phối Chuẩn

6-1 Phân phối chuẩn

6-2 Chuẩn hóa phân phối chuẩn

6-3 Phân phối mẫu và công cụ ước lượng

6-4 Định lý Giới hạn Trung tâm

6-5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Phân phối liên tục

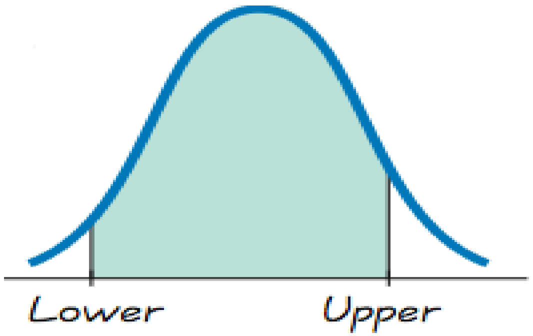
- Phân phối được gọi là liên tục nếu
  - biến ngẫu nhiên nhận giá trị trong một miền vô hạn không đếm được
  - hàm phân phối tích lũy tạo thành một đường cong liên tục
- Nếu X là một biến ngẫu nhiên liên tục
  - không thể sử dụng hàm độ lớn xác suất (pmf) cho X
  - ta có thể tính xác suất cho một khoảng giá trị của X
  - xác suất X = a với a là bất kỳ giá trị cụ thể nào đó đều bằng 0

Phân phối liên tục

- Được đặc trưng bởi hàm mật độ xác suất (pdf) *f(x)* thỏa:

$$\Pr(a \le X < b) = \int_a^b f(x)dx \qquad \text{với } a \le b$$

- Để tìm xác suất của một biến ngẫu nhiên liên tục, ta tính diện tích phần dưới đường cong nằm giữa 2 điểm cần tính xác suất



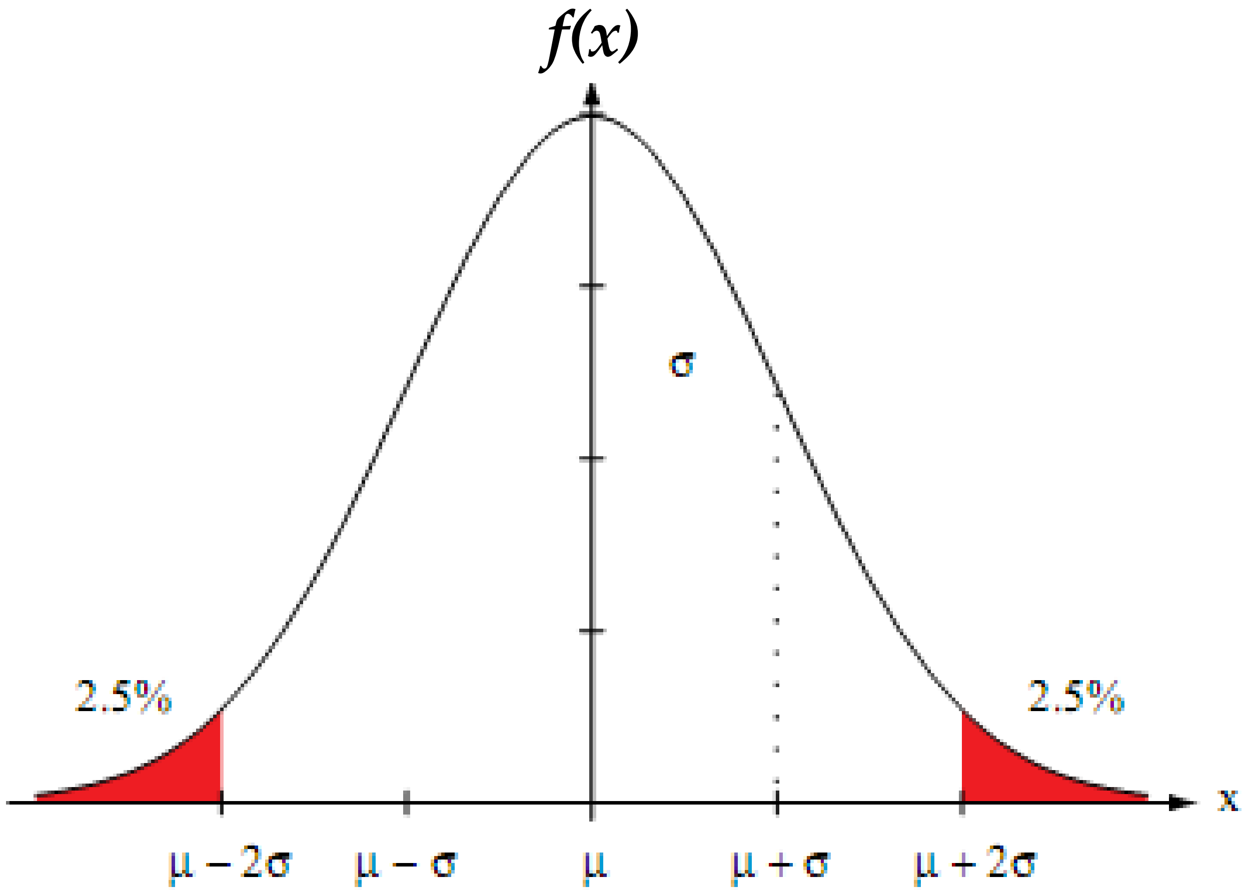
Phân phối chuẩn

- Phân phối chuẩn là mô hình xác suất được đặc trưng bởi hai đại lượng
  - – trung bình μ
  - – phương sai σ^2
- Ký hiệu $N(\mu,\sigma^2)$ hay $N(\mu,\sigma)$
- Hàm mật độ xác suất

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$$

  - – Hình chuông
  - – Khác nhau ở tâm và độ rộng

Phân phối chuẩn



Tổng phần diện tích dưới đường mật độ bằng 1.

Phân phối Z

- Là phân phối chuẩn có:
  - Trung bình $\mu = 0$
  - Phương sai $\sigma^2 = 1$
- Phân phối Z còn được gọi là phân phối chuẩn chính tắc
- Hàm mật độ:

$$f(z) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{(z-0)^2}{2\times 1}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Ký hiệu

$$P(a < z < b)$$

biểu thị xác suất của z nằm giữa a và b.

$$P(z > a)$$

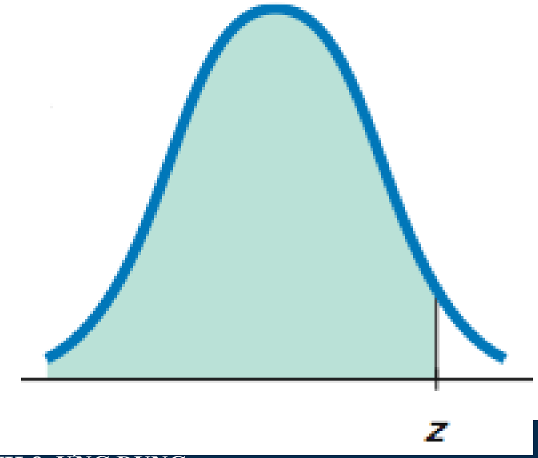
biểu thị xác suất của z lớn hơn a.

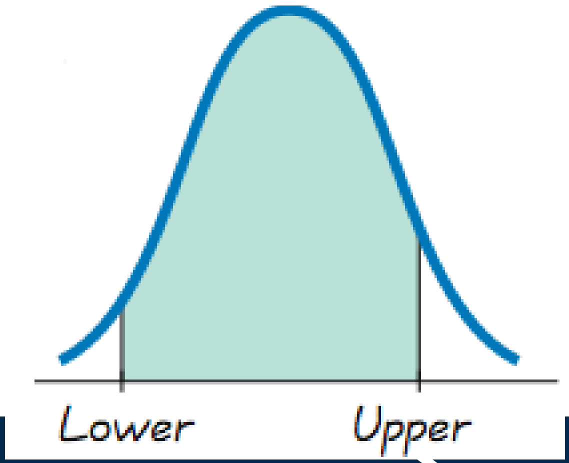
$$P(z < a)$$

biểu thị xác suất của z nhỏ hơn a.

Tìm giá trị xác suất

- Có thể tính diện tích (xác suất) của các vùng khác nhau của phân phối chuẩn chính tắc sử dụng các công cụ có sẵn hoặc bằng cách tra bảng Z (A-2).
- Tra bảng Z trong trường hợp ta cần tính xác suất tích lũy của phân phối chuẩn chính tắc (P(z<a)).
- Nếu không phải xác suất tích lũy của phân phối chuẩn chính tắc, ta cần chuyển về xác suất tích lũy của phân phối chuẩn chính tắc mới tra được bảng Z.
  - P(z>a) = 1- P(z<a)
  - P(a<z<b) = P(z<b) - P(z<a)





Bảng Z (Bảng A-2)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Ví dụ - Kiểm tra mật độ xương

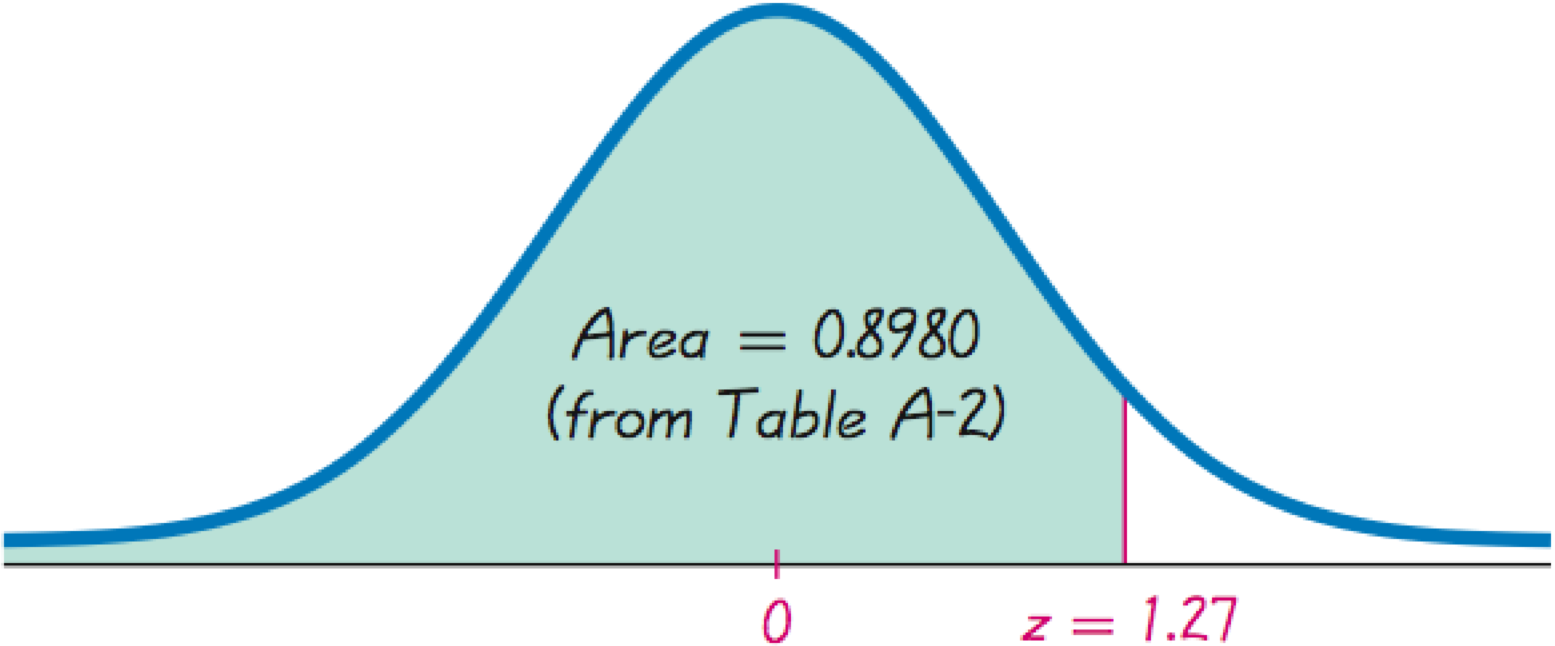
Đo mật độ Canxi trong xương để xác định một người lớn có bị loãng xương hay không.

Giả sử z là kết quả mật độ Canxi trong xương của một người lớn được chọn ngẫu nhiên, **z có phân phối chuẩn** với **giá trị trung bình là 0** và **độ lệch chuẩn là 1.**

Xác suất của một người lớn được chọn ngẫu nhiên có mật độ Canxi trong xương dưới 1,27 là bao nhiêu?

Ví dụ (tt)

$$P(z < 1.27) =$$



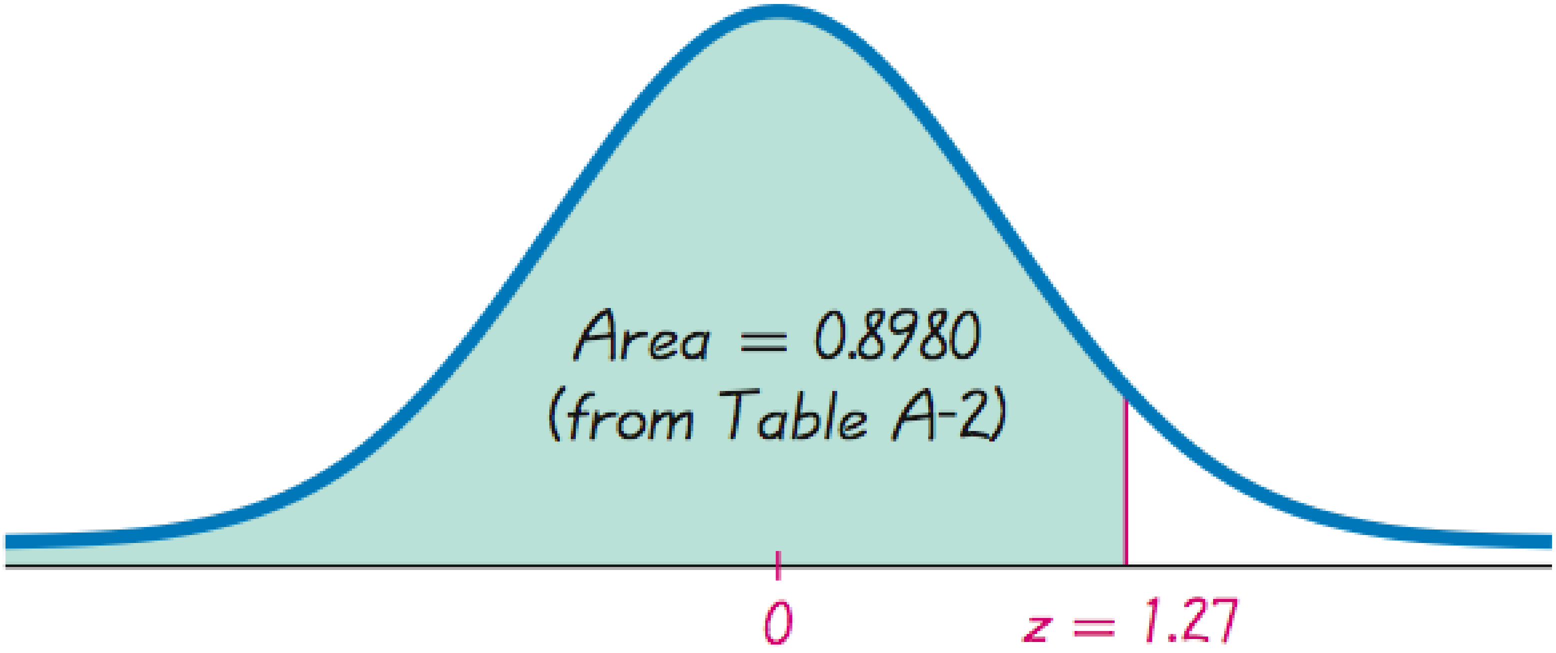
Tra bảng A-2

TABLE A-2 (*continued*) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292

Ví dụ (tt)

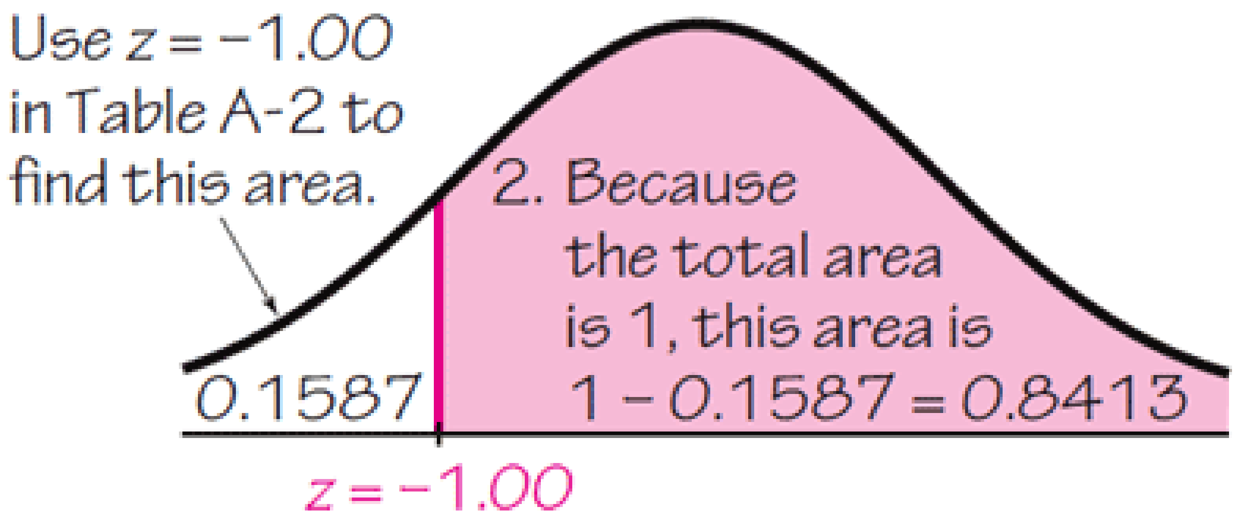
$$P(z < 1.27) = 0.8980$$



Xác suất của một người lớn được chọn ngẫu nhiên có mật độ Canxi trong xương dưới 1,27 là 0,8980.

Ví dụ:

Sử dụng cùng một thử nghiệm đo mật độ Canxi trong xương, tìm xác suất mà một người lớn được chọn ngẫu nhiên có mật độ Canxi trong xương trên -1,00 (được coi là trong phạm vi "bình thường" của các chỉ số mật độ Canxi trong xương)?

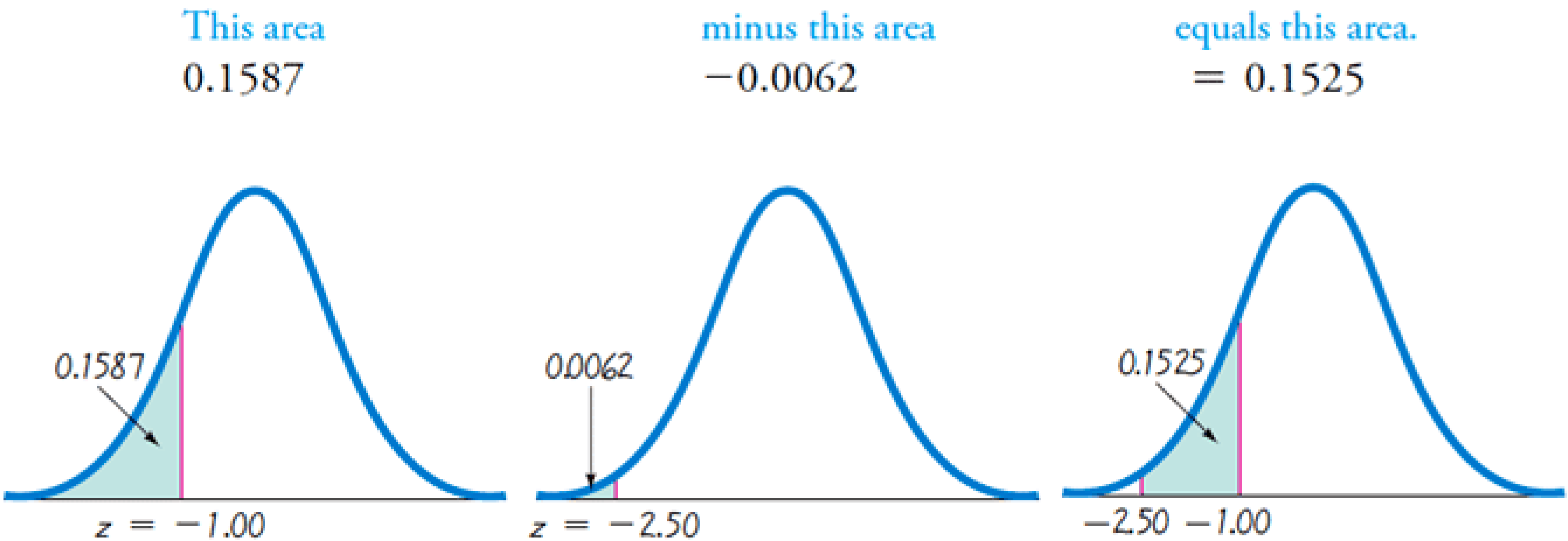


Xác suất của một người lớn được chọn ngẫu nhiên có mật độ Canxi trong xương trên –1 là 0,8413.

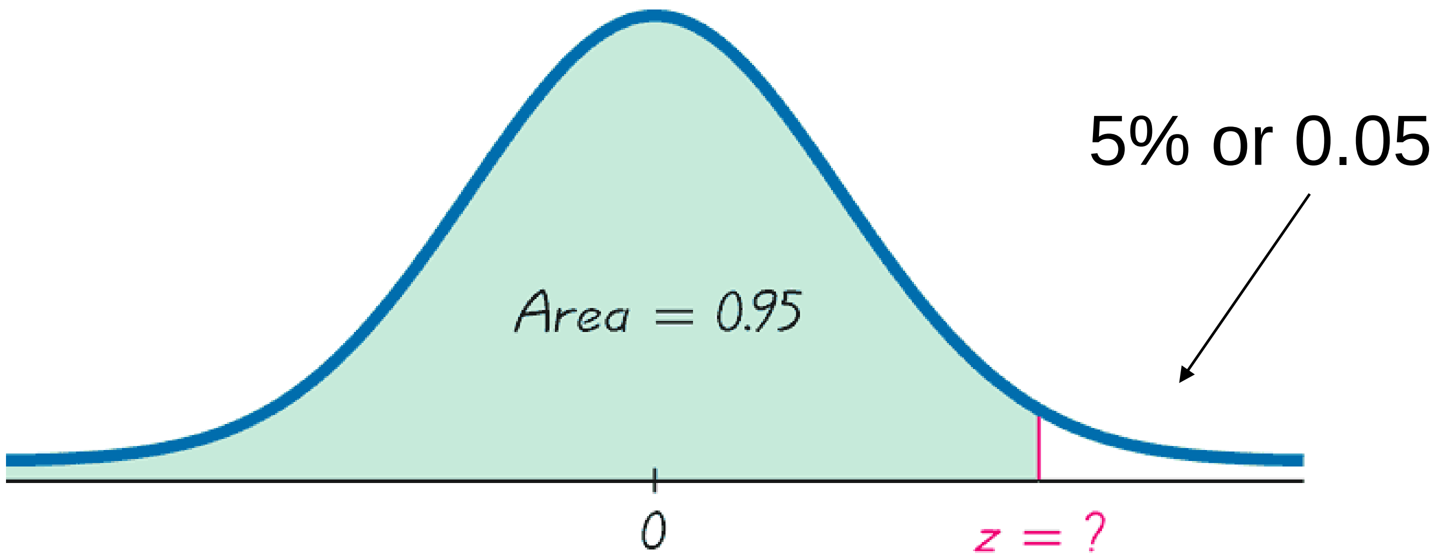
Ví dụ (tt)

Tính xác suất của một người lớn được chọn ngẫu nhiên có mật độ Canxi trong xương từ –2,50 đến -1,00 ?

1. Diện tích bên trái của z = –2.50 là 0,00262.
2. Diện tích bên phải của z = –1.00 là 0.1587.
3. Diện tích vùng giữa z = –2.50 và z = –1.00 khác so với hai vùng trên.



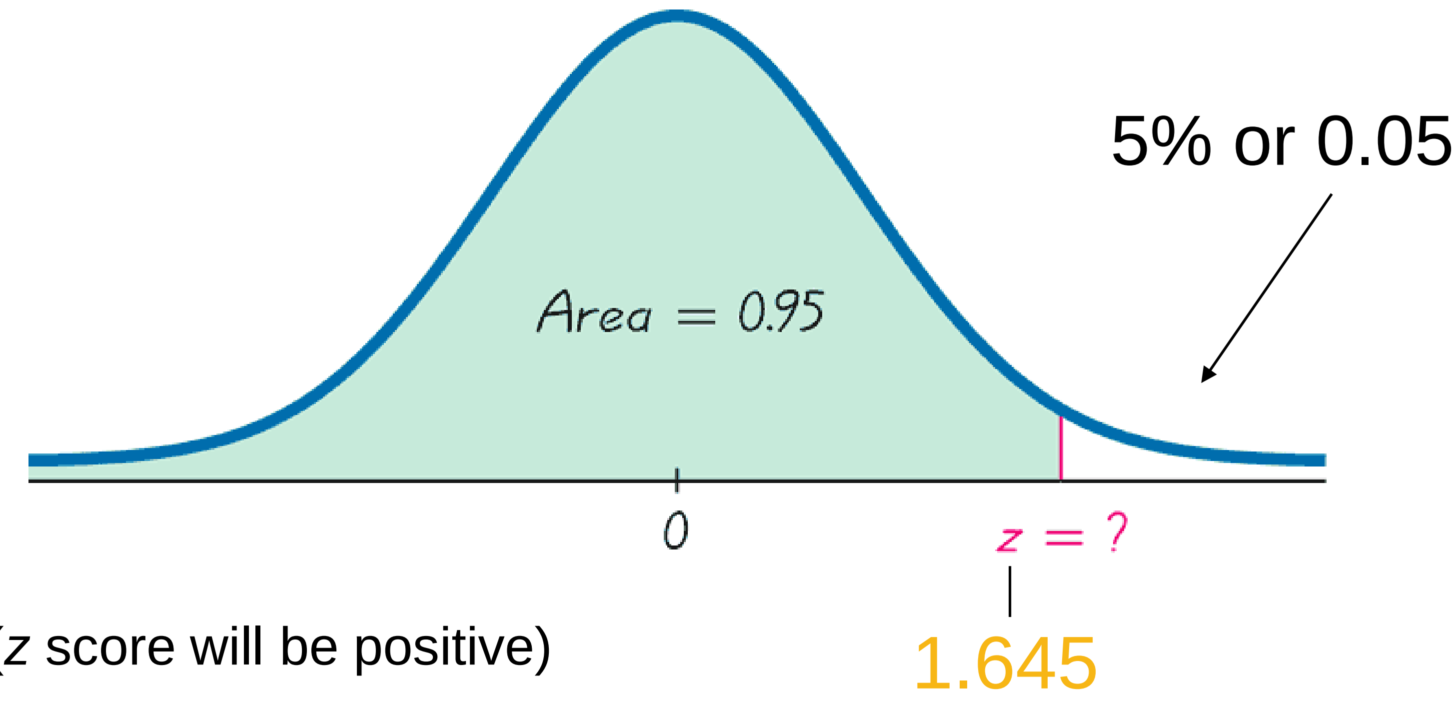
Tìm z khi biết xác suất



(*z* score will be positive)

Finding the 95th Percentile

Tìm z khi biết xác suất

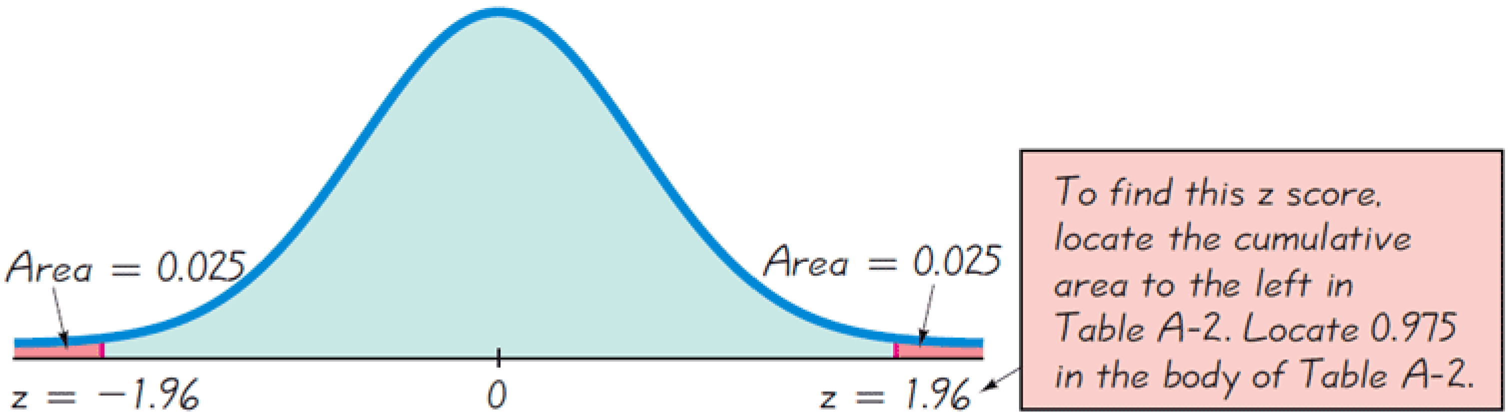


(*z* score will be positive)

Finding the 95th Percentile

Ví dụ

Sử dụng cùng một thử nghiệm mật độ Canxi trong xương, hãy tìm tỉ số z để xác suất một người lớn được chọn ngẫu nhiên có mật độ xương=z là lớn hơn 25% và nhỏ hơn 95%



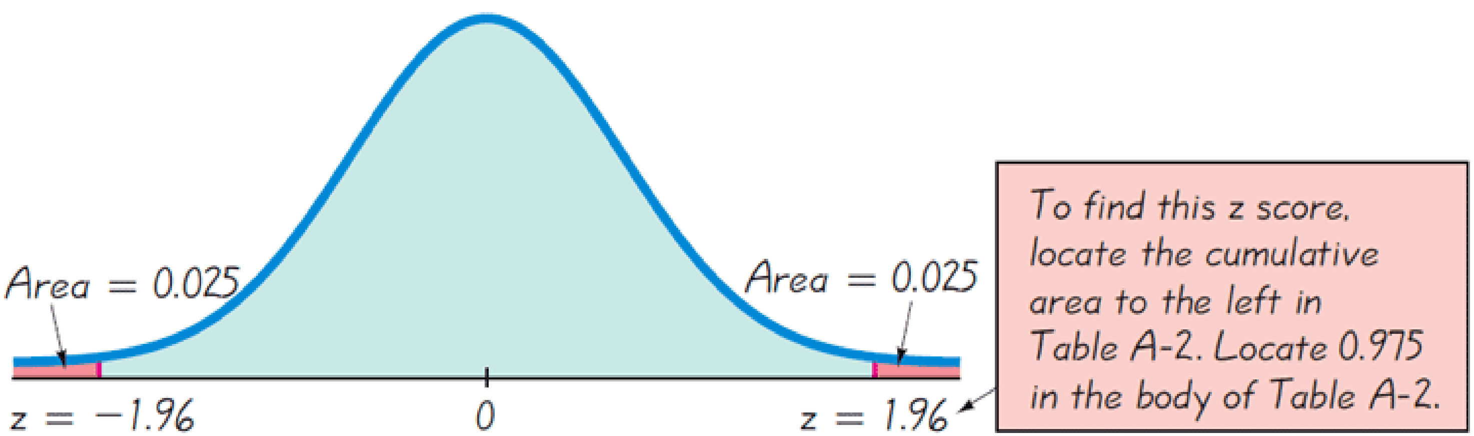
Định nghĩa

Đối với phân phối chuẩn, critical value là giá trị z mà tại đó phần diện tích bên phải z bằng một giá trị α

Ví dụ: Nếu $\alpha = 0.025$ thì $z_{0.025} = 1.96$.

Nghĩa là, critical value $z_{0.025} = 1.96$ có diện tích phần bên phải là 0.025

Quay lại ví dụ về mật độ xương, $z_{0.025} = 1.96$



Chương 6
Phân phối Chuẩn

6-1 Phân phối chuẩn

6-2 Chuẩn hóa phân phối chuẩn

6-3 Phân phối mẫu và công cụ ước lượng

6-4 Định lý Giới Hạn Trung Tâm

6-5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

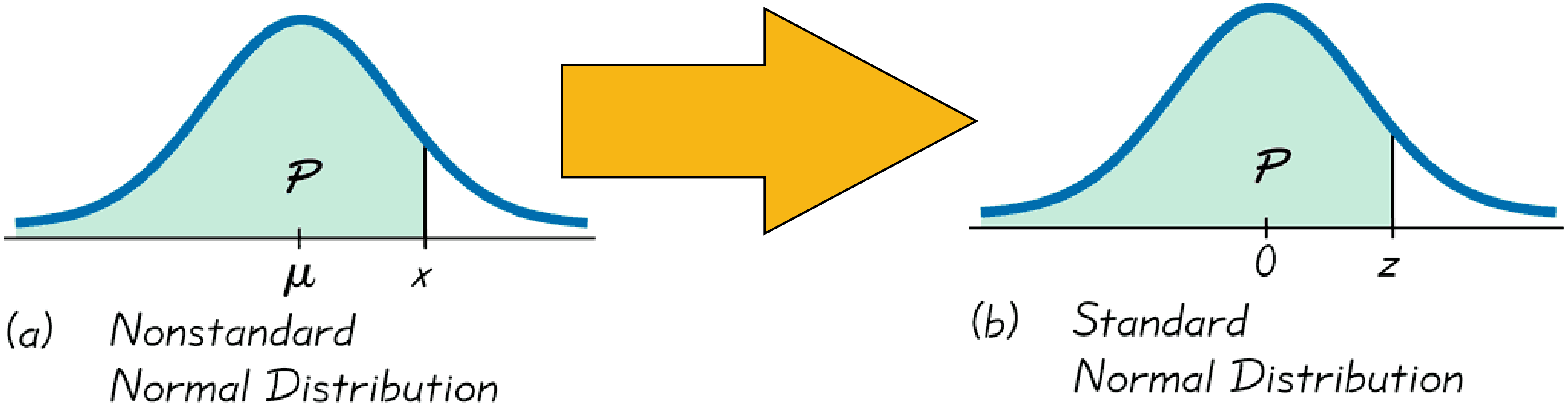
Chuẩn hóa phân phối chuẩn

- Chuẩn hóa phân phối chuẩn N(μ,σ) là biến đổi phân phối chuẩn đã cho sang phân phối Z với N(μ = 0,σ = 1) hay N(0,1).

$$Z = \frac{X - \mu}{\sigma}.$$

- Việc chuẩn hóa phân phối chuẩn cho trước để có thể sử dụng được bảng phân phối Z không làm ảnh hưởng gì đến các xác suất cần tính và như vậy, không ảnh hưởng đến kết quả bài toán gốc

Chuyển đổi sang phân phối chuẩn



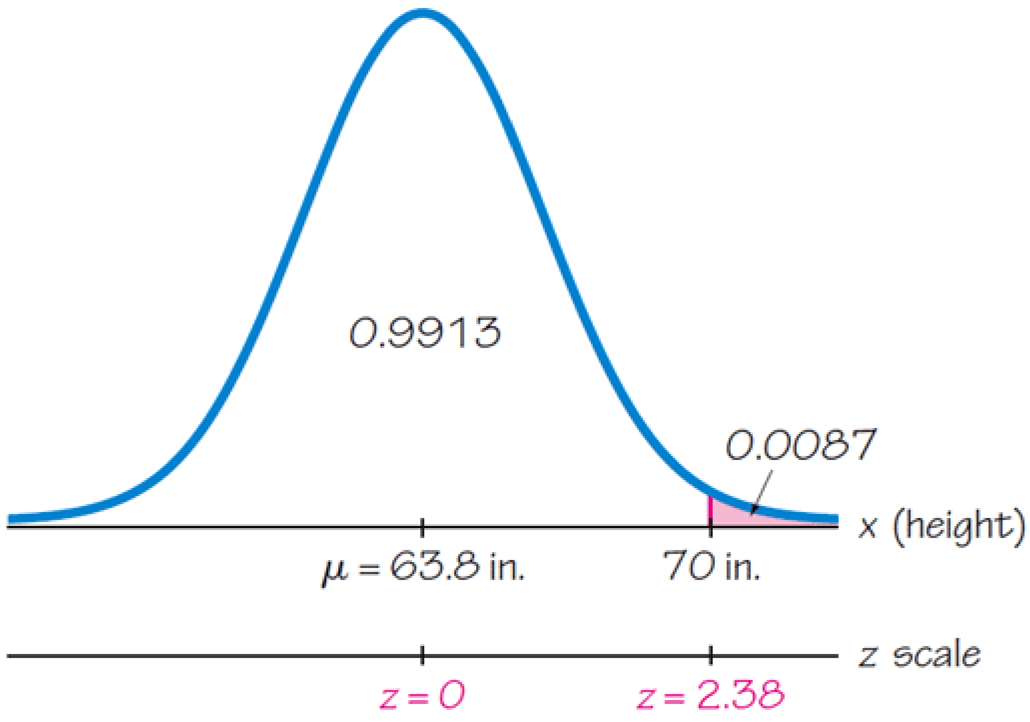
Ví dụ

Câu lạc bộ về chiều cao yêu cầu các phụ nữ phải cao tối thiểu 70 inch.

Giả sử rằng chiều cao của phụ nữ tuân theo phân phối chuẩn với giá trị trung bình là 63.8 inch và độ lệch chuẩn là 2.6 inch. Hãy tìm phần trăm phụ nữ đáp ứng yêu cầu về chiều cao đó.

Ví dụ (tt)

Vẽ phân phối chuẩn và hình dạng của vùng cần tính xác suất



Ví dụ (tt)

Chuyển đổi biến z và tra bảng Z để tìm diện tích vùng tô đậm.

$$z = \frac{x-\mu}{\sigma} = \frac{70-63.8}{2.6} = 2.38$$

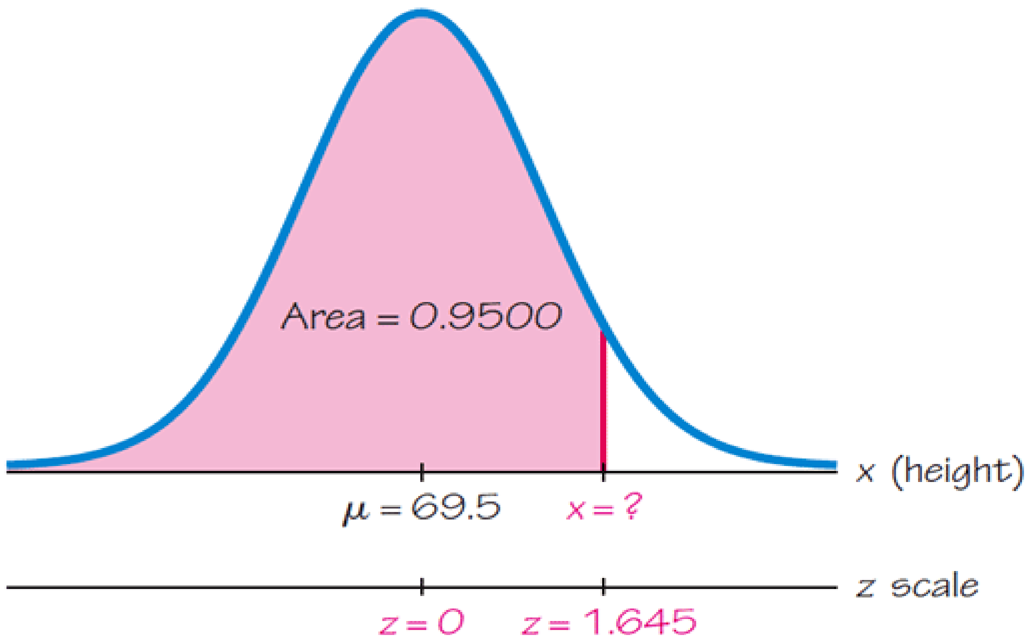
Diện tích vùng bên phải của z=2.38 là 0.008656, vì vậy khoảng 0.87% nữ có chiều cao tối thiểu 70 inch.

.

Ví dụ - Buồng máy bay

Khi thiết kế cabin máy bay, chiều cao trần như thế nào sẽ cho phép 95% nam đứng lên mà không đụng đầu họ? Chiều cao của nam giới thường được phân phối với giá trung bình 69,5 inch và độ lệch chuẩn là 2,4 inch.

Trước tiên, vẽ phân phối chuẩn.



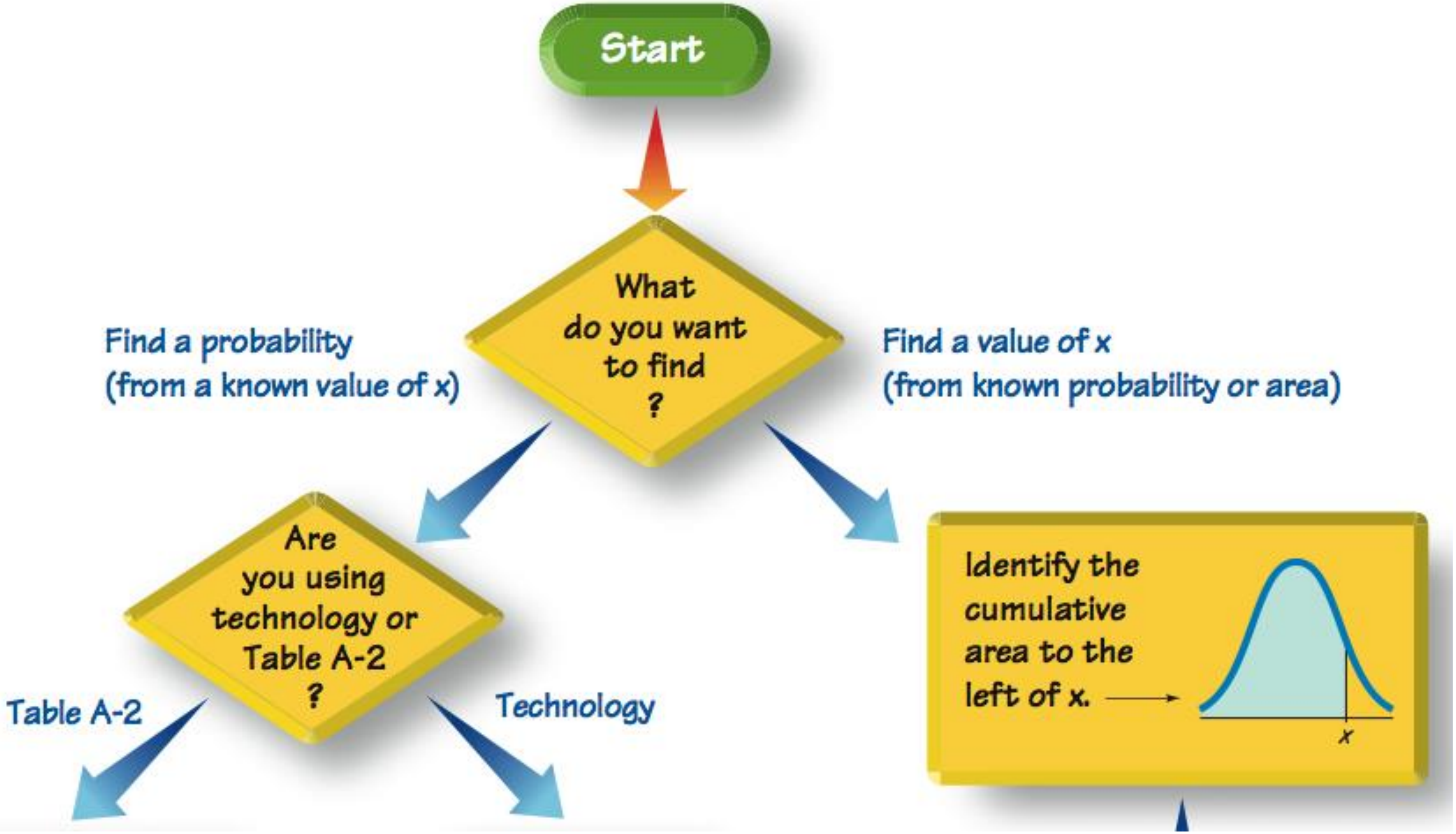
Ví dụ - Buồng máy bay

Khi thiết kế cabin máy bay, giá trị chiều cao của trần là bao nhiêu thì cho phép 95% nam đứng lên mà không đụng đầu họ? Chiều cao của nam giới thường được phân phối chuẩn với giá trung bình 69,5 inch và độ lệch chuẩn là 2,4 inch.

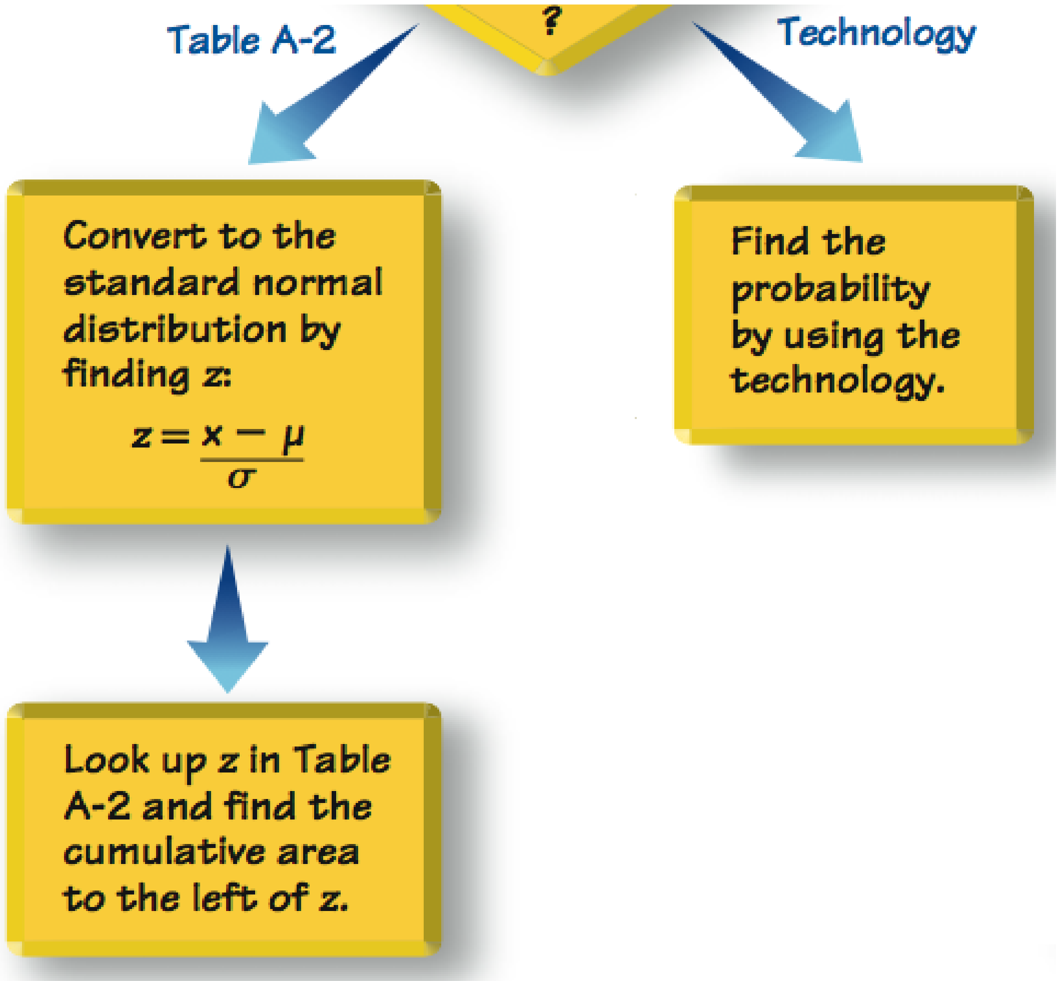
Với $z = 1.645$, $\mu = 69.5$, và $\sigma = 2.4$. chúng ta có thể tính x.

$$x = \mu + (z * \sigma) = 69.5 + 1.645 * 2.4 = 73.448 \text{ (inch)}$$

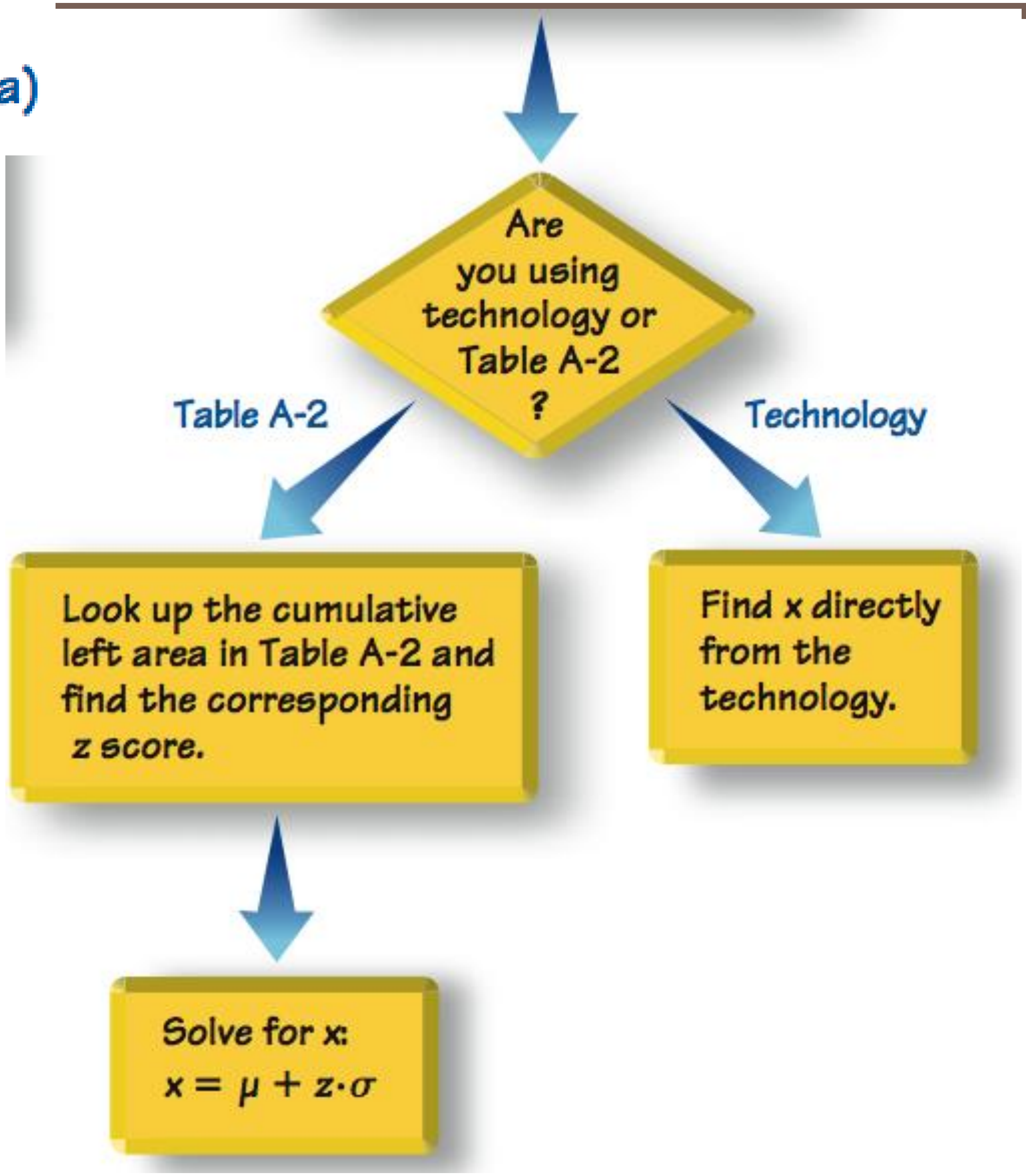
Applications with Normal Distributions



Find a probability (from a known value of x)



Find a value of x
(from known probability or area)



Bài tập

- Chiều dài cá được mô hình hóa bằng phân phối chuẩn N(μ=16 (cm), σ=4 (cm)). Ta cần trả lời các câu hỏi sau:
  - Câu hỏi 1: Xác suất bắt được con cá nhỏ (nhỏ hơn 8 (cm))?
  - Câu hỏi 2: Giả sử, ai bắt được con cá lớn (lớn hơn 24(cm)) sẽ được thưởng. Hỏi xác suất được thưởng là bao nhiêu?
  - Câu hỏi 3: Xác suất bắt được con cá vừa (trong khoảng 16-24(cm))?

Chương 6
Phân phối Chuẩn

6-1 Phân phối chuẩn

6-2 Chuẩn hóa phân phối chuẩn

6-3 Phân phối mẫu và công cụ ước lượng

6-4 Định lý Giới Hạn Trung Tâm

6-5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Phân phối mẫu

Phân phối mẫu (sampling distribution): với quần thể và cỡ mẫu n, phân phối mẫu của 1 giá trị thống kê là phân phối tất cả của các giá trị thống kê cho tất cả các mẫu có thể có với kích thước n.

Phân phối mẫu của trung bình

Phân phối mẫu (sampling distribution) của trung bình là phân phối của tất cả các giá trị trung bình mẫu (sample mean) nếu như ta liệt kê được mọi tập mẫu có thể lấy với kích thước cố định n.

Giá trị trung bình của trung bình mẫu: $\mu_{\bar{x}}$

Độ lệch chuẩn của trung bình mẫu: $\sigma_{\bar{x}}$

Phân phối mẫu của trung bình

2,3	2,4	2,5
$\bar{x} = 2.5$	$\bar{x} = 3$	$\bar{x} = 3.5$
3,4	3,5	4,5
$\bar{x} = 3.5$	$\bar{x} = 4$	$\bar{x} = 4.5$

- $\mu = 3.5$
- $\sigma \approx 1.12$
- $\mu_{\bar{x}} = 3.5$
- $\sigma_{\bar{x}} \approx 0.79$

Chương 6
Phân phối Chuẩn

6-1 Phân phối chuẩn

6-2 Chuẩn hóa phân phối chuẩn

6-3 Phân phối mẫu và công cụ ước lượng

6-4 Định lý Giới Hạn Trung Tâm

6-5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Phân phối mẫu

Định lý giới hạn trung tâm: Khi ta lấy mẫu ngẫu nhiên, kích thước tập mẫu càng lớn thì phân phối xác suất của đặc trưng trung bình của tập mẫu càng gần với phân phối chuẩn với giá trị trung bình của phân phối mẫu và phương sai của phân phối mẫu là:

$$\mu_{\overline{x}} = \mu \quad \text{and} \quad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Phân phối mẫu

Khi chọn một mẫu ngẫu nhiên đơn giản gồm n đối tượng trong một quần thể có giá trị trung bình μ và độ lệch chuẩn là σ, cần áp dụng các nguyên tắc sau:

- Đối với 1 quần thể có phân phối bất kỳ, **nếu n>30**, trung bình mẫu có thể xấp xỉ với phân phối chuẩn với giá trị trung bình là

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Nếu $n \leq 30$, quần thể ban đầu có phân phối chuẩn thì trung bình mẫu có phân phối chuẩn với

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Nếu $n \leq 30$, quần thể ban đầu không theo phân phối chuẩn thì không áp dụng nguyên tắc này.

Phân phối mẫu của trung bình

2,3	2,4	2,5
$\bar{x} = 2.5$	$\bar{x} = 3$	$\bar{x} = 3.5$
3,4	3,5	4,5
$\bar{x} = 3.5$	$\bar{x} = 4$	$\bar{x} = 4.5$

- $\mu = 3.5$
- $\mu_{\bar{x}} = 3.5$

$\bar{x}$: bộ ước lượng không lệch

- Range: 5 – 2 = 3
- Sample Ranges: 1,2,3,1,2,1
- $\mu_{Range} \approx 1.67$
- Sample Range: bộ ước lượng lệch

Bộ ước lượng không lệch và lệch

Bộ ước lượng không lệch: một giá trị thống kê mà có phân phối mẫu có giá trị trung bình bằng với tham số của quần thể.

- Trung bình mẫu là bộ ước lượng không lệch với trung bình của quần thể.
- Tỉ lệ mẫu là bộ ước lượng không lệch với tỉ lệ của quần thể.
- Phương sai mẫu là bộ ước lượng không lệch với phương sai của quần thể.
- Sample range là bộ ước lượng lệch với range của quần thể

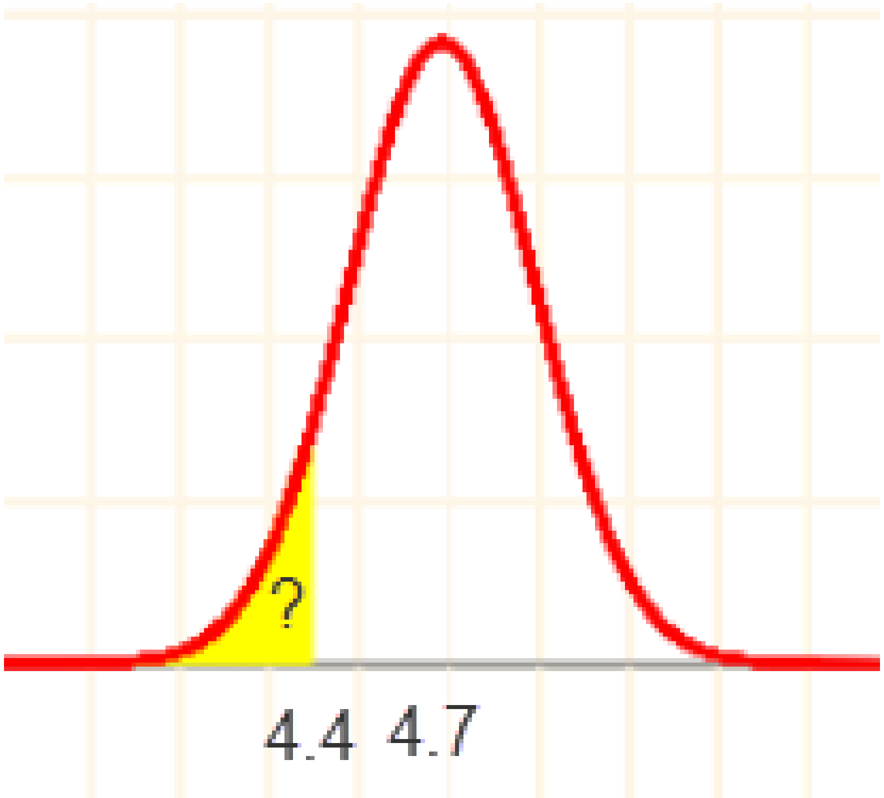
Ví dụ

Giả sử thời gian trung bình mà sinh viên đại học cần để hoàn thành bằng cấp là 4,7 năm. Độ lệch chuẩn là 0,3. Xác suất mà 40 sinh viên đại học được chọn ngẫu nhiên sẽ có thời gian hoàn thành trung bình dưới 4,4 năm là bao nhiêu?

Ví dụ

$\mu = 4.7$, $\sigma = 0.3$, n = 40, $P(\overline{x} < 4.4) = ?$

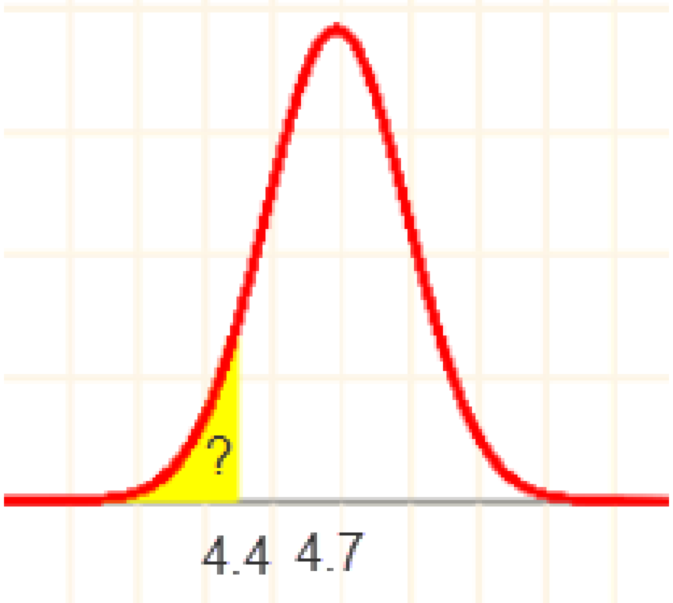
$$\overline{x} \sim N(4.7, \frac{0.3}{\sqrt{40}})$$



Ví dụ

P($\overline{x}$ < 4.4) = ?

$$\overline{x} \sim N(4.7, \frac{0.3}{\sqrt{40}})$$



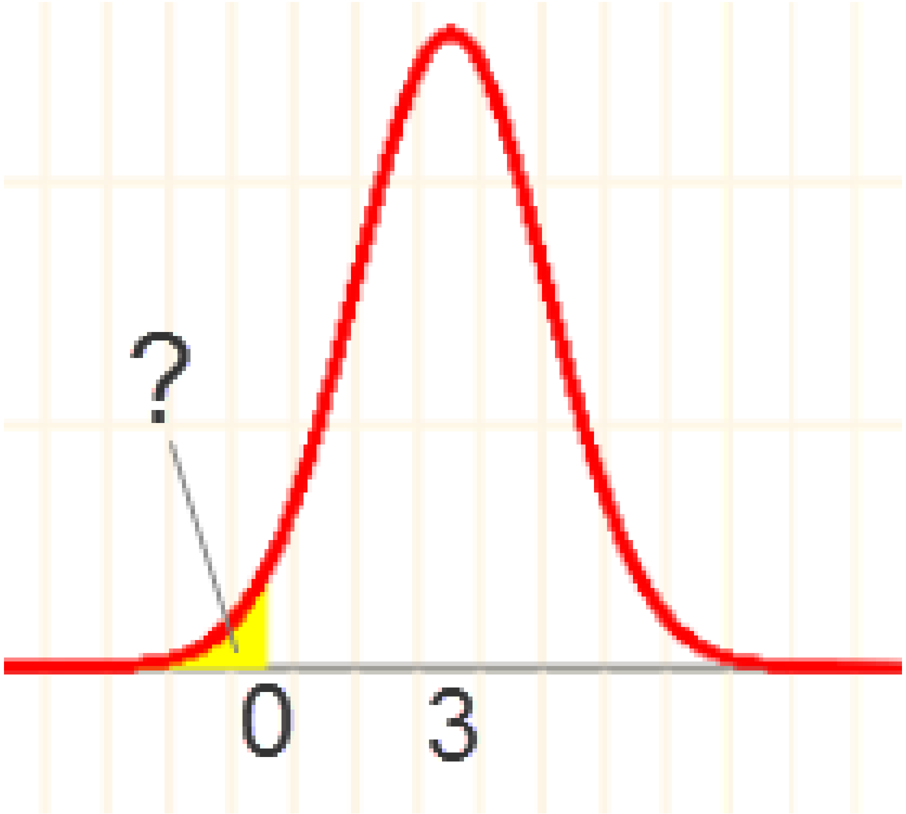
Ví dụ

Giả sử việc tăng cổ phiếu tuân theo phân phối chuẩn với giá trị trung bình là 3 phần trăm và độ lệch chuẩn là 5 phần trăm. Nếu danh mục đầu tư (được chọn ngẫu nhiên) của bạn bao gồm 20 cổ phiếu, xác suất danh mục đầu tư của bạn sẽ mất tiền là bao nhiêu?

Ví dụ

$\mu = 3, \sigma = 5, n = 20, P(\overline{x} < 0) = ?$

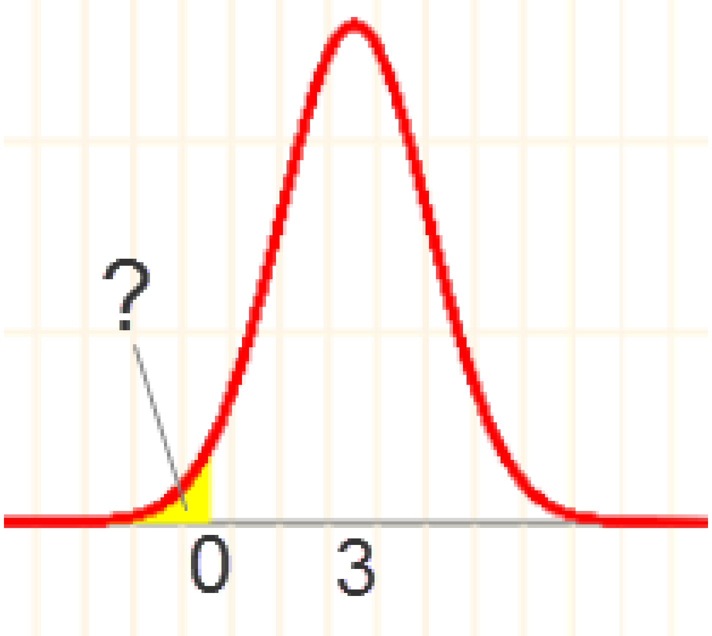
$$\overline{x} \sim N(3, \frac{5}{\sqrt{20}})$$



Ví dụ

$P(\overline{x} < 0) = ?$

$$\overline{x} \sim N(3, \frac{5}{\sqrt{20}})$$



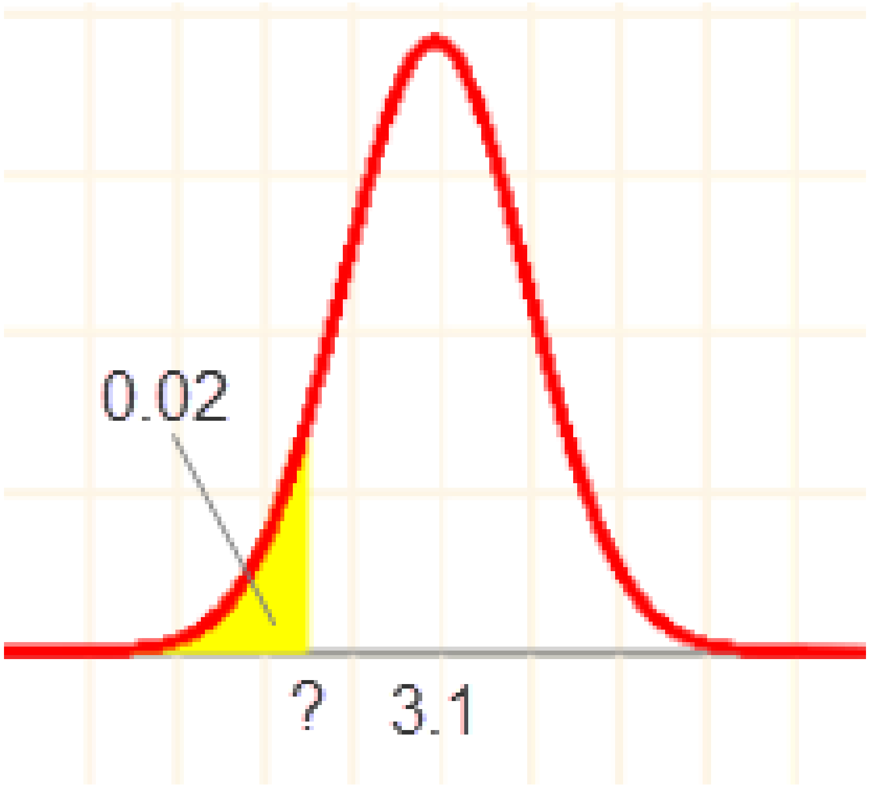
Ví dụ

Giả sử điểm trung bình chung của sinh viên đại học là 3,1 và độ lệch chuẩn là 0,7. Một lớp gồm 35 sinh viên được chọn ngẫu nhiên sẽ được coi là có rủi ro cao nếu điểm trung bình trung bình của họ nhỏ hơn 2%. Điểm trung bình lớn nhất sẽ được coi là rủi ro cao là bao nhiêu?

Ví dụ

$\mu = 3.1$, $\sigma = 0.7$, n = 35, P($\overline{x} < ?$) = 0.02

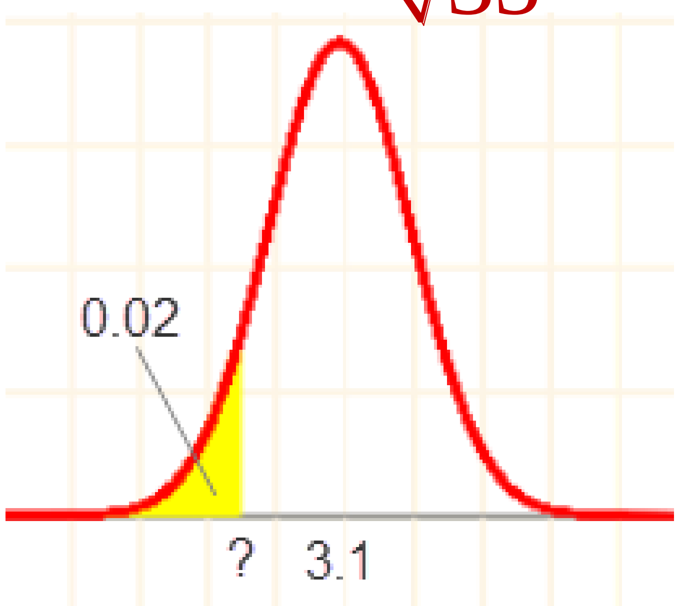
$$\overline{x} \sim N\left(3.1, \frac{0.7}{\sqrt{35}}\right)$$



Ví dụ

P($\bar{x}$ < ?) = 0.02

$$\bar{x} \sim N(3.1, \frac{0.7}{\sqrt{35}})$$



Hệ số hiệu chỉnh đối với quần thể hữu hạn

Khi áp dụng định lý giới hạn trung tâm, sử dụng $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, chúng ta giả định rằng quần thể có vô số thành viên. Khi chúng ta lấy mẫu theo phương pháp thay thế, quần thể là vô hạn.

Tuy nhiên, nhiều ứng dụng thực tế liên quan đến lấy mẫu mà không cần thay thế, do đó các mẫu liên tiếp phụ thuộc vào kết quả trước đó.

Hệ số hiệu chỉnh đối với quần thể hữu hạn

Trong sản xuất, kiểm soát chất lượng thanh tra thường lấy mẫu các mặt hàng từ một hoạt động sản xuất hữu hạn mà không thay thế. Đối với quần thể hữu hạn như vậy, chúng ta có thể cần điều chỉnh. Dưới đây là quy tắc chung:

- Khi lấy mẫu không thay thế và cỡ mẫu n lớn hơn 5% kích thước quần thể hữu hạn N, cầu điều chỉnh độ lệch chuẩn của trung bình mẫu như sau:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$$

Phân phối mẫu của tỉ lệ

Phân phối mẫu (sampling distribution) của tỉ lệ là phân phối của tất cả các giá trị tỉ lệ mẫu (sample proportion) nếu như ta liệt kê được mọi tập mẫu có thể lấy với kích thước cố định n.

Giá trị trung bình của tỉ lệ mẫu: $\mu_{\hat{p}}$

Độ lệch chuẩn của tỉ lệ mẫu: $\sigma_{\hat{p}}$

Phân phối mẫu của tỉ lệ

Giả sử rằng chúng ta có quần thể gồm 6 người:

Alice Ben Charles Denise Edward Frank






Tỉ lệ nữ trong quần thể là bao nhiêu?

p=1/3

Tham số nào cần quan tâm trong quần thể này?

Tỉ lệ nữ

Liệt kê các mẫu gồm 2 người trong quần thể này

Có bao nhiêu mẫu khác nhau có thể có?

$_6C_2 = 15$

Phân phối mẫu của tỉ lệ

Tìm 15 mẫu khác nhau có thể và tìm tỷ lệ mẫu của số lượng nữ trong mỗi mẫu.

Alice & Ben	0.5
Alice & Charles	0.5
Alice & Denise	1.0
Alice & Edward	0.5
Alice & Frank	0.5
Ben & Charles	0.0
Ben & Denise	0.5
Ben & Edward	0.0
Ben & Frank	0.0
Charles & Denise	0.5
Charles & Edward	0.0
Charles & Frank	0.0
Denise & Edward	0.5
Denise & Frank	0.5
Edward & Frank	0.0

Phân phối mẫu của tỉ lệ

Alice & Ben	0.5
Alice & Charles	0.5
Alice & Denise	1.0
Alice & Edward	0.5
Alice & Frank	0.5
Ben & Charles	0.0
Ben & Denise	0.5
Ben & Edward	0.0
Ben & Frank	0.0
Charles & Denise	0.5
Charles & Edward	0.0
Charles & Frank	0.0
Denise & Edward	0.5
Denise & Frank	0.5
Edward & Frank	0.0

Tính giá trị trung bình của tất cả các tỉ lệ mẫu và độ lệch chuẩn của tất cả các tỉ lệ mẫu?

Phân phối mẫu của tỉ lệ

$\mu_{\hat{p}} = \frac{1}{3}$

$\sigma_{\hat{p}} = 0.29814$

Phân phối mẫu của tỉ lệ

Ta có tỉ lệ mẫu: $\widehat{p} = \frac{x}{n}$

Trong đó: x là số phần tử trong mẫu mà ta quan tâm.

n là kích thước mẫu.

Đối với một mẫu ngẫu nhiên đơn giản từ một quần thể lớn, giá trị của x là một biến ngẫu nhiên tuân thep phân phối nhị thức.

Vì n là hằng số nên $\frac{x}{n}$ cũng tuân theo phân phối nhị thức.

Phân phối mẫu của tỉ lệ

Trong chương 5, phân phối nhị thức có thể xấp xỉ phân phối chuẩn khi kích thước mẫu lớn thỏa các điều kiện sau:

$$np \geq 5 \text{ và } n(1-p) \geq 5$$

Khi đó, trung bình và phương sai của tỉ lệ mẫu như sau:

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Phân phối mẫu của tỉ lệ

Khi lấy mẫu không thay thế và cỡ mẫu n lớn hơn 5% kích thước quần thể hữu hạn N, cầu điều chỉnh độ lệch chuẩn của tỉ mẫu như sau:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}}$$

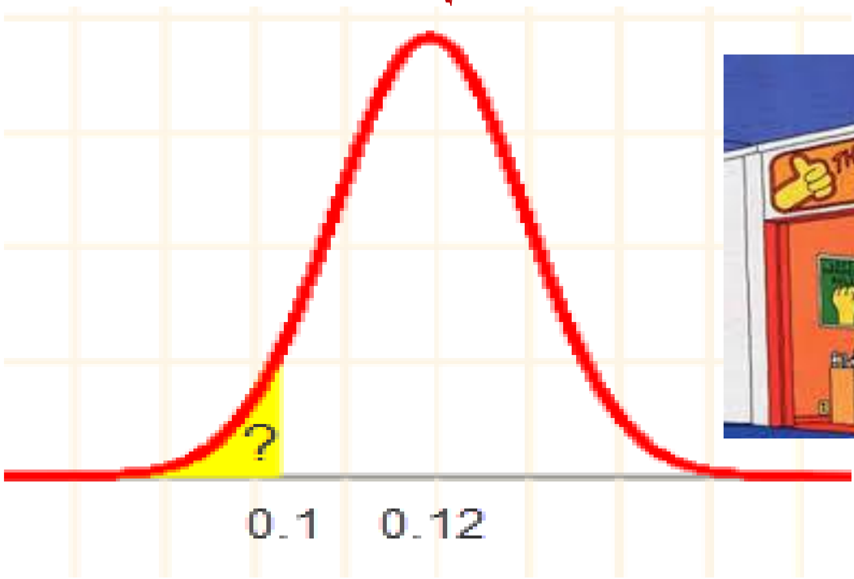
Ví dụ

12% dân số Mỹ là người thuận tay trái. Nếu 200 người Mỹ được chọn ngẫu nhiên để khảo sát, xác suất có ít hơn 20 người trong số họ thuận tay trái là bao nhiêu?

Ví dụ

p = 0.12, q = 0.88, n = 200, P($\hat{p}$ < 0.1) = ?

$$\hat{p} \sim N(0.12, \sqrt{\frac{0.12 \bullet 0.88}{200}})$$



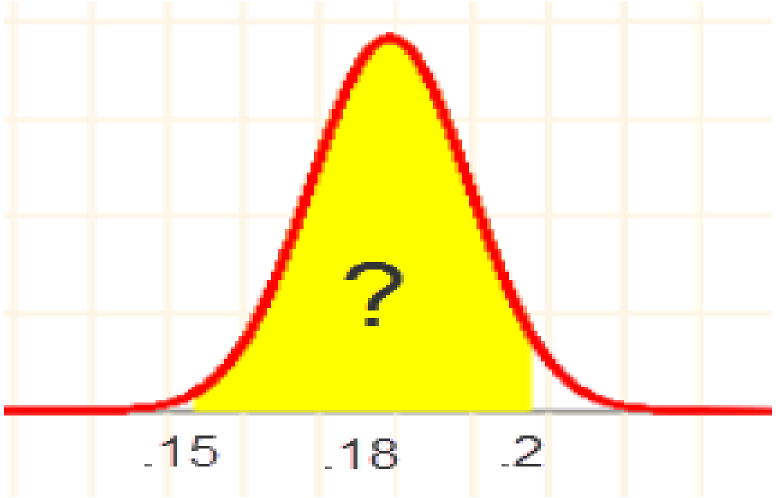
Ví dụ

Theo một cuộc thăm dò gần đây của Gallup, 18% người Mỹ đang thất làm. Nếu 150 người Mỹ được chọn ngẫu nhiên, hãy tìm xác suất từ 24 đến 30 người trong số họ bị thất nghiệp?

Ví dụ

p = 0.18, q = 0.82, n = 150, P(0.15 < $\hat{p}$ < 0.2)

$$\hat{p} \sim N(0.18, \sqrt{\frac{0.18 \bullet 0.82}{150}})$$



Chương 6
Phân phối Chuẩn

6-1 Phân phối chuẩn

6-2 Chuẩn hóa phân phối chuẩn

6-3 Phân phối mẫu và công cụ ước lượng

6-4 Định lý Giới Hạn Trung Tâm

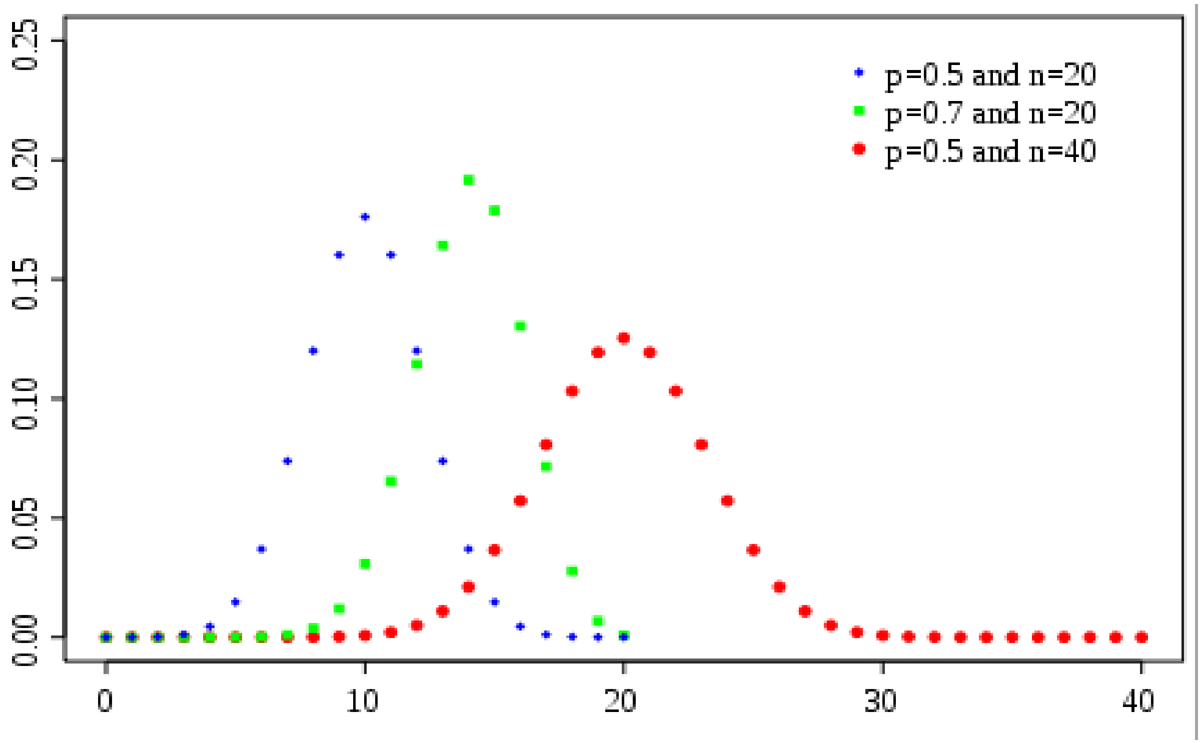
6-5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Xấp xỉ pp nhị thức bằng pp chuẩn

- Trong pp nhị thức, tính xác suất khi số phép thử lớn (ví dụ như 100) là gần như không thể
- Phân phối chuẩn có thể được dùng để xấp xỉ xác suất nhị thức khi n lớn.

Xấp xỉ pp nhị thức bằng pp chuẩn

- Khi số lần thí nghiệm lớn, xs thành công
  - p gần 0.5 → dạng chuông đối xứng
  - p gần 0 (1) → lệch trái (phải)



Xấp xỉ pp nhị thức bằng pp chuẩn

- Do phân phối nhị thức có dạng gần giống phân phối chuẩn khi n lớn → có thể sử dụng phân phối chuẩn để xấp xỉ nhị thức
- 3 câu hỏi
  - n thế nào là lớn?
  - dùng trung bình và độ lệch chuẩn nào để chuẩn hóa X về Z?
  - Phân phối nhị thức là rời rạc, còn phân phối chuẩn là liên tục. Vậy, làm thế nào để chỉnh cho liên tục?

Xấp xỉ pp nhị thức bằng pp chuẩn

- *Câu hỏi 1*: n lớn bao nhiêu?
- Quy tắc: *có thể sử dụng phối chuẩn để xấp xỉ nhị thức khi n thỏa* np $\geq$ 5 & n (1 – p) $\geq$ 5. Và *n càng lớn thì xấp xỉ càng tốt*.

Xấp xỉ pp nhị thức bằng pp chuẩn

- *Câu hỏi 2*: trung bình và phương sai cho pp chuẩn
  - $\mu = E(X) = n \times p$
  - $\sigma^2 = Var(X) = n \times p \times (1\text{-}p)$

Xấp xỉ pp nhị thức bằng pp chuẩn

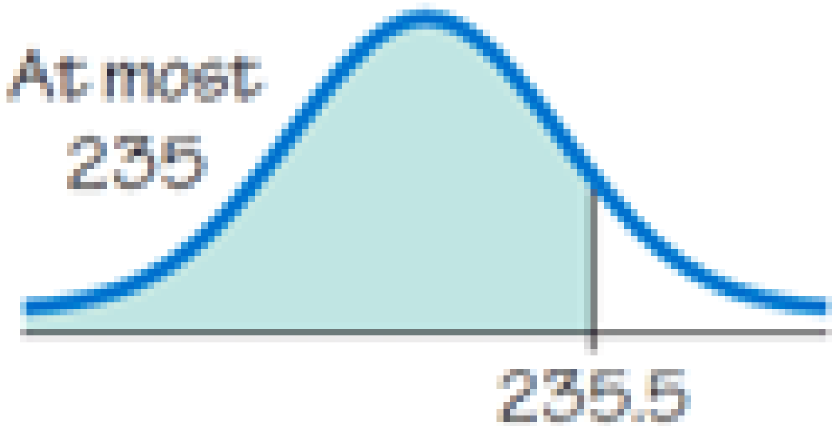
- *Câu hỏi 3*: hiệu chỉnh liên tục?
- Tại sao cần?
  - Khi dùng phân phối chuẩn để xấp xỉ phân phối nhị thức, thực chất ta đang thực hiện quá trình *làm trơn* cạnh của các thanh của phân phối nhị thức bằng một đường cong liên tục

Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh
  - Biểu diễn bài toán dưới dạng xác suất. Chẳng hạn:
    $P(X < 235)$, $P(X \leq 235)$, $P(X \geq 235)$, $P(X>235)$, $P(X=235)$.
  - Cộng thêm hay trừ đi một lượng bằng ½ tùy thuộc xác suất ta cần tìm là lớn hơn hay nhỏ hơn.

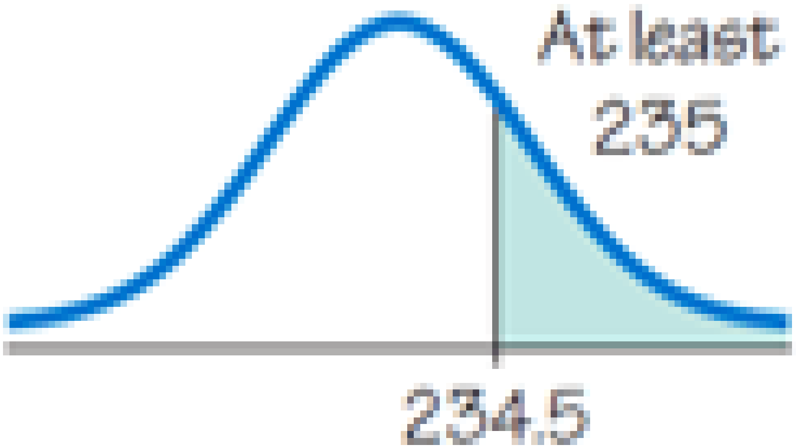
Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh
  - Nếu xác suất nhỏ hơn hoặc bằng, cộng ½ vào X trước khi lấy xác suất. Chẳng hạn, nếu X ≤ 235 sẽ được nắn thành 235.5 và tính tiếp P(X ≤235.5) theo công thức phân phối chuẩn.



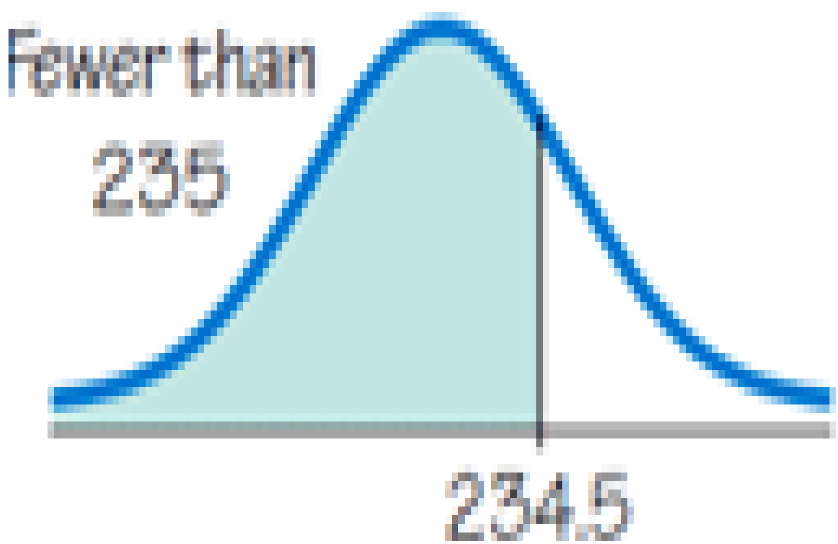
Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh (tt)
  - Nếu xác suất lớn hơn hoặc bằng, trừ ½ khỏi X trước khi lấy xác suất. Chẳng hạn, nếu X ≥ 235 sẽ được nắn thành 234.5 và tính tiếp P(X ≥234.5) theo công thức phân phối chuẩn.



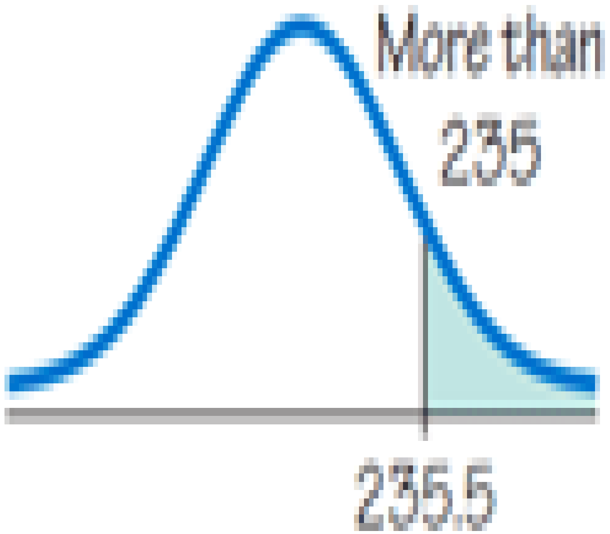
Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh (tt)
  - Nếu xác suất bé hơn hẳn thì chuyển X thành X – 0.5. Chẳng hạn, P(X < 235) thì chuyển thành P(X ≤ 234.5) và tính tiếp theo công thức phân phối chuẩn.



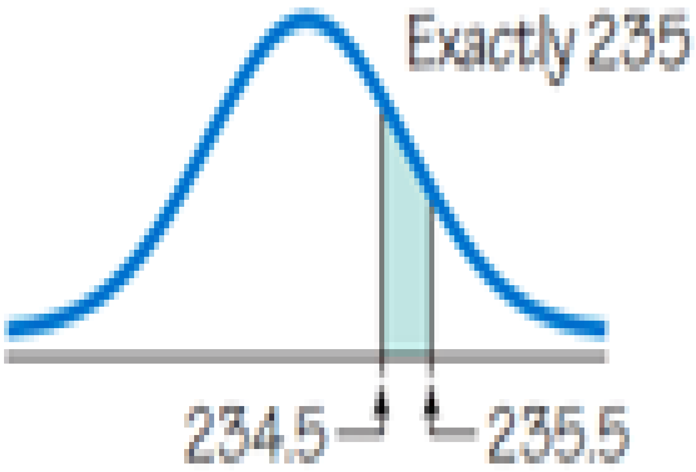
Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh (tt)
  - Nếu xác suất lớn hơn hẳn thì chuyển X thành X + 0.5. Chẳng hạn, P(X > 235) thì chuyển thành P(X ≥ 235.5) và tính tiếp theo công thức phân phối chuẩn.



Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh (tt)
  - Nếu xác suất bằng đúng 1 giá trị, nắn bằng cách vừa cộng vừa trừ. Chẳng hạn, P(X = 235) nắn thành P(234.5 ≤ X ≤ 235.5) và tính tiếp theo công thức phân phối chuẩn.



Xấp xỉ pp nhị thức bằng pp chuẩn

- Quy tắc hiệu chỉnh (tt)
  - Nếu xác suất nằm giữa 2 giá trị, chẳng hạn $2 \le X \le 5$, ta thực hiện các bước trên để nắn giá trị $X = 5$; $X = 2$ và tính tiếp $P(1.5 \le X \le 5.5)$ theo công thức phân phối chuẩn.

Ví dụ tung đồng xu NFL

Trong 431 trò chơi bóng đá NFL, các đội giành được quyền tung đồng xu tiếp tục chiến thắng 235 trong 431 trò chơi.

Nếu phương pháp tung tiền xu là công bằng, các đội thắng cược sẽ thắng khoảng 50% số trận đấu (chúng tôi hy vọng 215.5 thắng trong 431 trận đấu thêm giờ).

Giả sử xác suất thắng một ván sau khi thắng tung đồng xu là 0.5, hãy tìm xác suất nhận được ít nhất 235 trận thắng.

Ví dụ tung đồng xu NFL

Vấn đề đưa ra liên quan đến sự phân phối nhị thức với n = 431 thử nghiệm và xác suất giả định thành công của p = 0,5.

Sử dụng xấp xỉ phân phối chuẩn cho phân phối nhị thức.

Bước 1: Kiểm tra điều kiện:

$$np = 431(0.5) = 215.5 \geq 5$$

$$nq = 431(0.5) = 215.5 \geq 5$$

Ví dụ tung đồng xu NFL

Bước 2: Tìm độ lệch chuẩn và trung bình của phân phối chuẩn:

$$\mu = np = 431(0.5) = 215.5$$

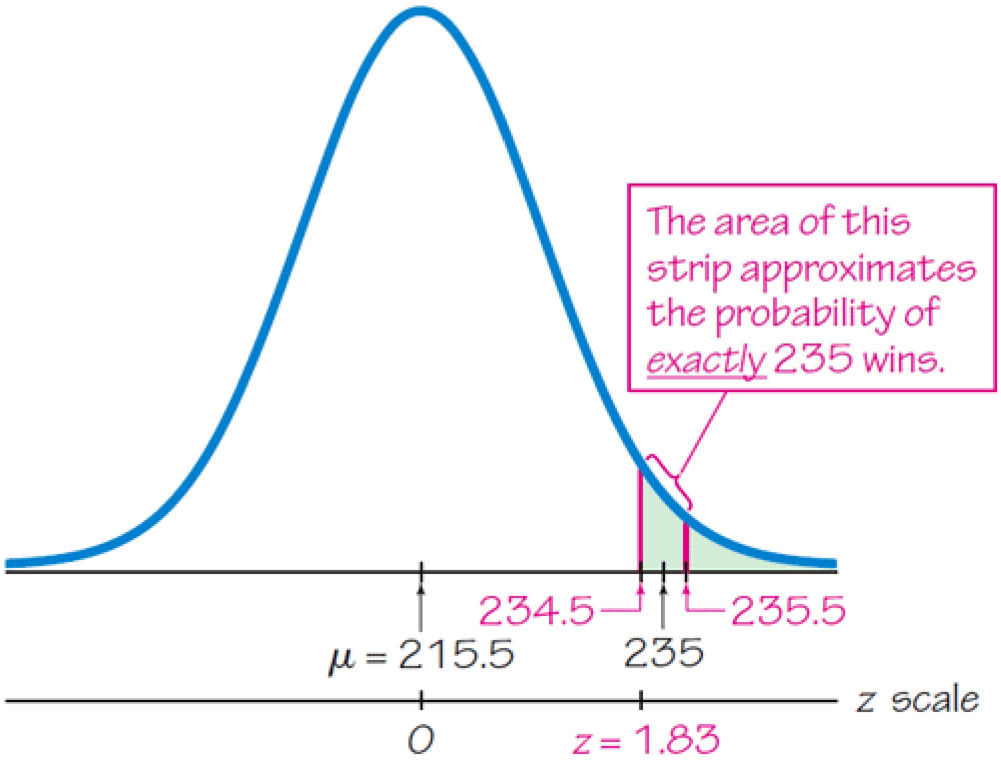
$$\sigma = \sqrt{npq} = \sqrt{431(0.5)(0.5)} = 10.38027$$

Bước 3: Chúng ta muốn xác suất ít nhất là 235 chiến thắng, vì vậy x = 235.

Bước 4: Miền giá trị đi từ 234,5 đến 235,5.

Ví dụ tung đồng xu NFL

Bước 5: Chúng ta sẽ tô đậm vùng bên phải giá trị 234.5



Ví dụ tung đồng xu NFL

Bước 6: Tìm điểm z và sử dụng bảng Z để xác định xác suất.

$$z = \frac{x - \mu}{\sigma} = \frac{234.5 - 215.5}{10.380270} = 1.83$$

Xác suất là 0,0336 cho đội giành chiến thắng lật xu để giành chiến thắng ít nhất 235 trò chơi.

Xác suất này đủ thấp để gợi ý đội giành được đồng xu lật có lợi thế không công bằng.