MONASH University

Information Technology

FIT1006
Business Information Analysis

Lecture 12
The Normal Distribution

Topics covered:

- Characteristics of the Standard Normal
- Standardising variables
- Mean and variance,
- Calculating Normal probabilities using tables and Excel
- The Normal approximation to the Binomial & Poisson Distributions

Motivating problem

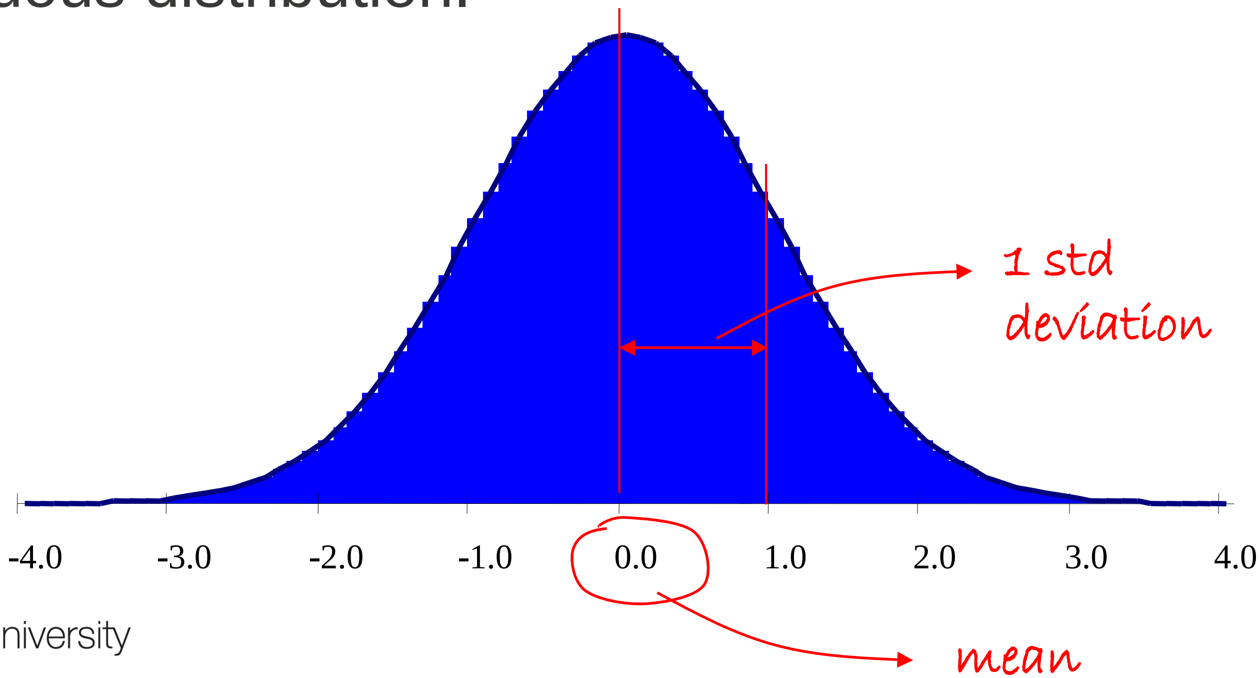
- To reduce the risk of running out of stock, retailers set a *reorder point*, which is the level of stock at which an order is raised, to arrive after a certain *lead time*.
- If a product’s demand during lead time is Normally distributed with a mean of 200 and standard deviation of 50 what should the reorder point be set to so that there is only a 2% chance of running out of stock?

The Normal Distribution

- The most important distribution in statistics.
- Arises when we measure a large number of nearly identical objects subject to random fluctuations - height, weight, response time. (Used a lot in biometrics).
- The Normal Distribution arises when we take the sum or average of a large number of observations from any probability distribution and thus provides the basis for sampling theory.

The Bell Curve

- The shape of the Normal distribution can be seen below. The shape is often referred to as the bell curve.
- The distribution below has a mean of 0 and a standard deviation of 1 and is called the *Standard Normal*. Note: continuous distribution.



https://flux.qa (Feed code: SJ6KGV)

Question 1

Using the *approximate* areas for the Standard Normal distribution below, P(Z > 1) =

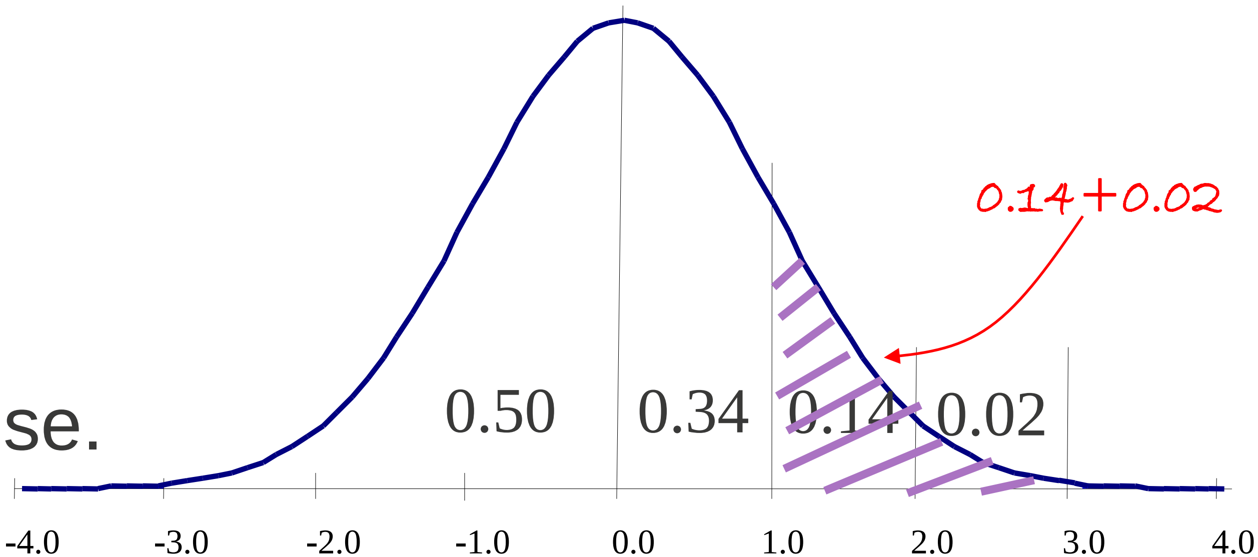
A. 0.14

✓ B. 0.16

C. 0.34

D. 0.84

E. None of these.



https://flux.qa (Feed code: SJ6KGV)

Question 2

Using the *approximate* areas for the Standard Normal distribution below, P(– 2 < Z < 2) =

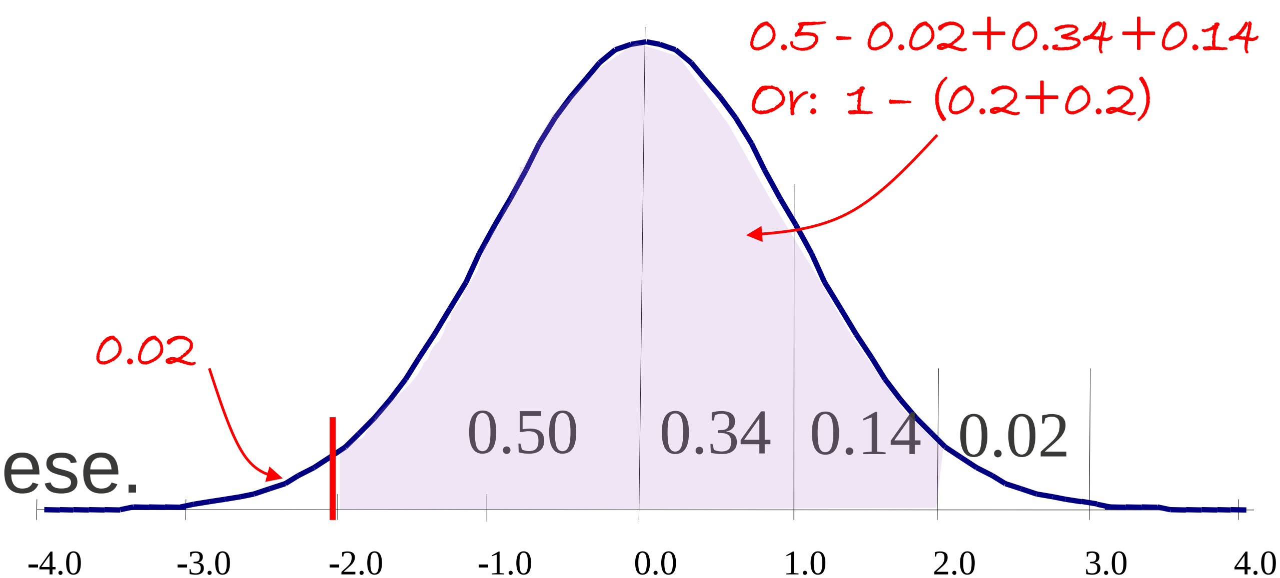
A. 0.14

B. 0.48

✓ C. 0.96

D. 0.98

E. None of these.



https://flux.qa (Feed code: SJ6KGV)

Question 3

Using the *approximate* areas for the Standard Normal distribution below, $P(Z > -2) =$

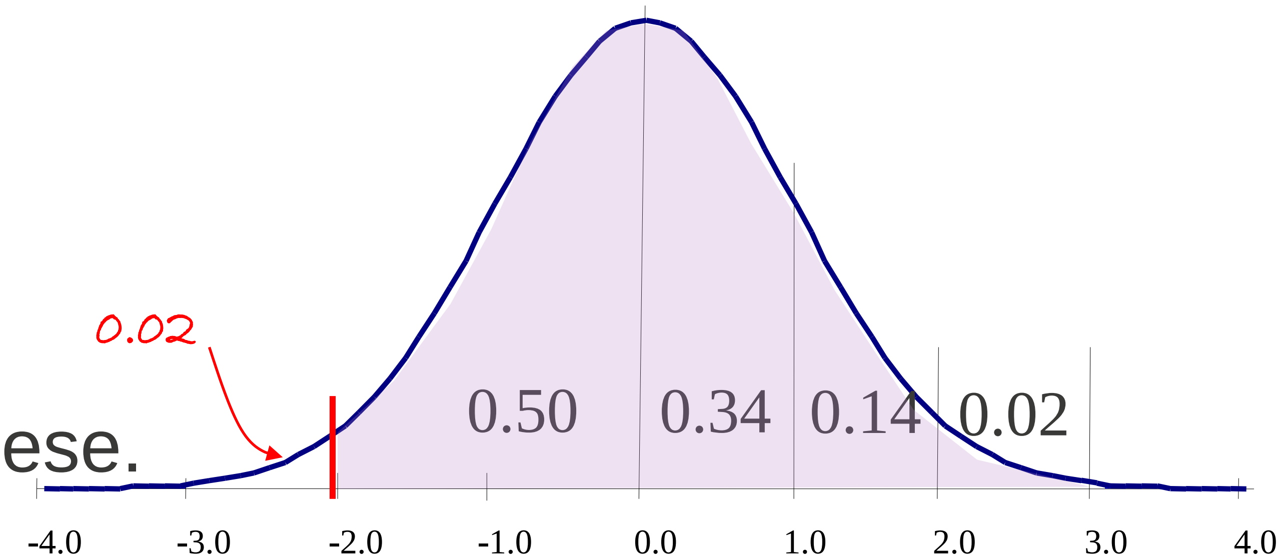
A. 0.02

B. 0.14

C. 0.48

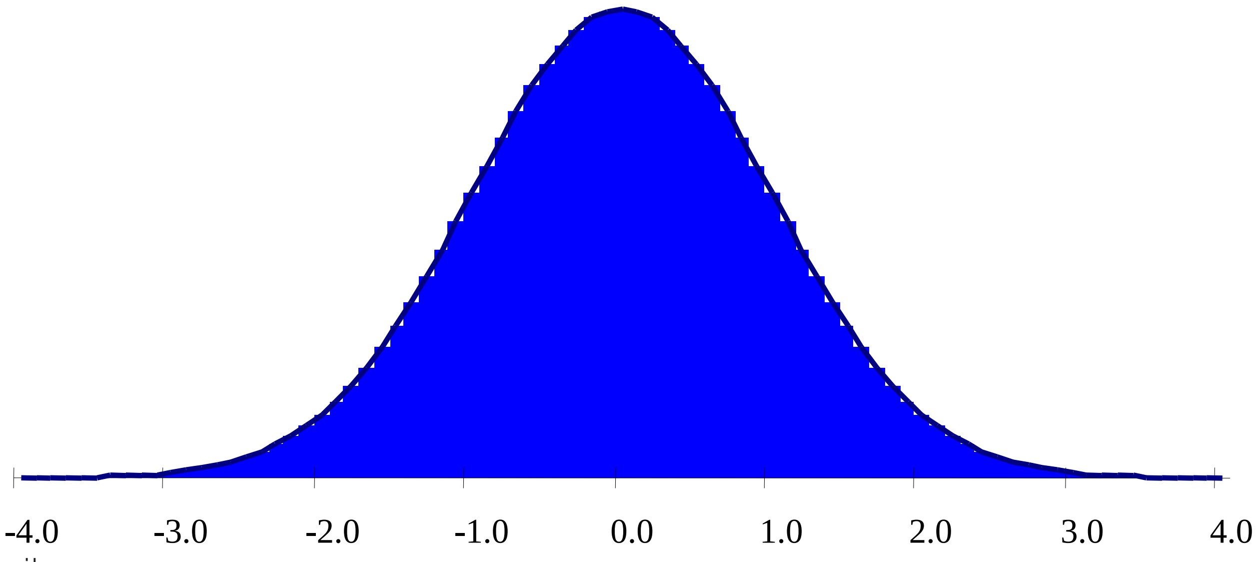
✓ D. 0.98

E. None of these.



General Properties

- The total area under the curve = 1. Approximately:
- 68% of the area is within 1 standard deviation of the mean.
- 95% of the area is within 2 standard deviations of the mean.
- 99.7% of the area is within 3 standard deviations of the mean.



Calculating Normal Probabilities

- The Normal probability function is not easily evaluated so you won't calculate probabilities directly.
- By Hand: use the table of Standard Normal cumulative probabilities, CDF.
- Excel: use the built in Excel functions (which are good approximations of the exact values).
- Excel: Normal CDF: =normdist(z,mean,stdev,true).
- Inverse Normal CDF: = norminv(prob,mean, stdev).
- Your calculator should be able to calculate Normal probabilities using similar commands.

Standard Normal CDF Table

Cumulative Probabilities for the Standard Normal Distribution

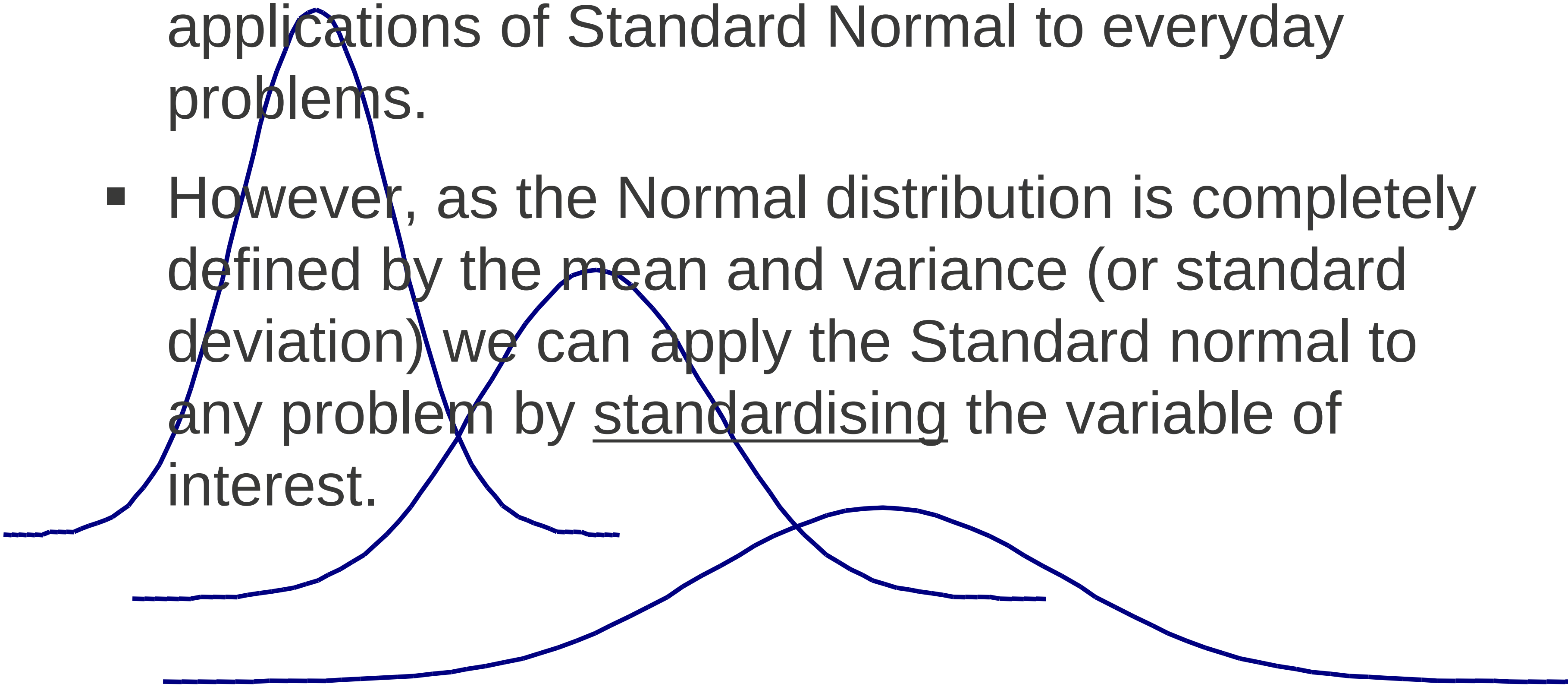
Table gives P(Z<z) for Z = N(0,1)

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	0.5040	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

Table continues…

The Normal Distribution in Practice

- The Standard Normal distribution has mean = 0 and standard deviation = 1. This tends to limit the applications of Standard Normal to everyday problems.
- However, as the Normal distribution is completely defined by the mean and variance (or standard deviation) we can apply the Standard normal to any problem by standardising the variable of interest.



Standardising Variables

Let X be a normal variate with mean μ and variance σ^2. We can write $X \approx \mathrm{N}(\mu, \sigma^2)$. We can standarise X by use of the formula: $z = \dfrac{x - \mu}{\sigma}$ and $\mathrm{Z} \approx \mathrm{N}(0,1)$. In this way we can apply the Standard Normal probabilities to any problem.

https://flux.qa (Feed code: SJ6KGV)

Question 4

If X is Normally distributed with mean 200 and standard deviation 50. Using the approximate areas for the standard normal P(X < 150) =

A. 0.14

✓ B. 0.16

C. 0.34

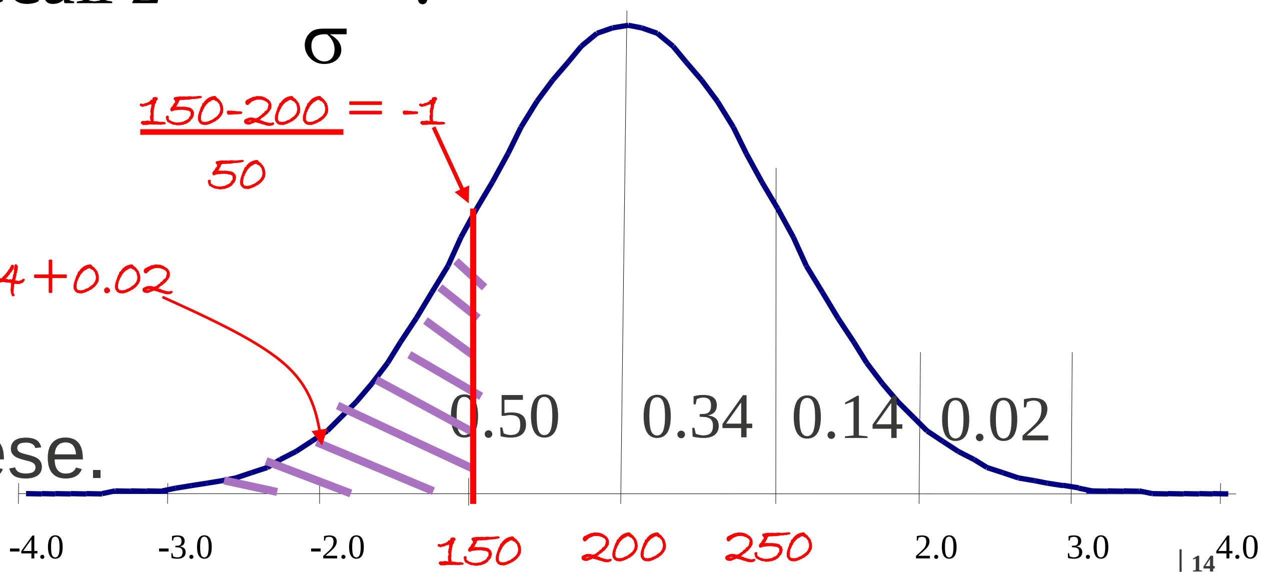
D. 0.84

E. None of these.

recall $z = \dfrac{x - \mu}{\sigma}$.

$\dfrac{150-200}{50} = -1$

$0.14 + 0.02$



https://flux.qa (Feed code: SJ6KGV)

Question 5

If X is Normally distributed with mean 200 and standard deviation 50. Using the approximate areas for the standard normal, if P(X < B) = 0.98, then

---

A. B = 200

B. B = 250

✓ C. B = 300

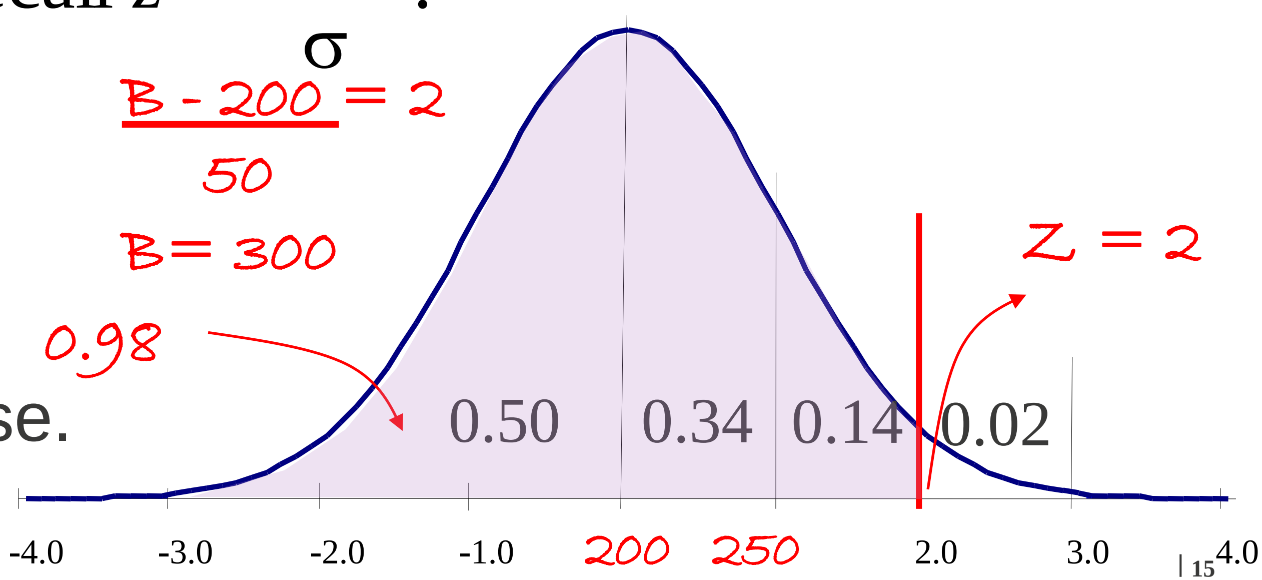
D. B = 350

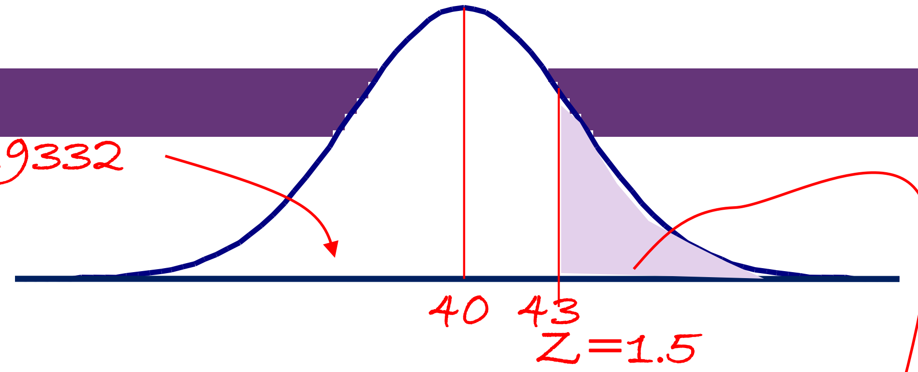
E. None of these.

recall $z = \dfrac{x-\mu}{\sigma}$.

$\dfrac{B - 200}{50} = 2$

$B = 300$





Example

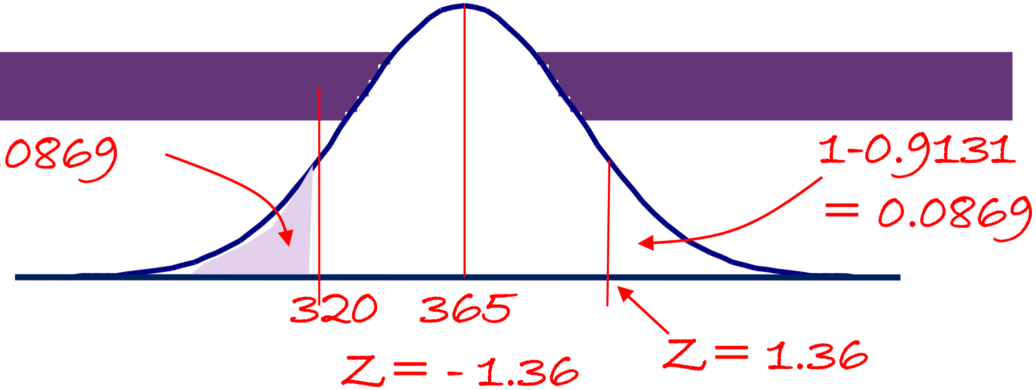
- A machine manufactures bolts which have a length of 40mm with a variance of 4mm^2 what is the probability that a bolt manufactured by the machine has a length greater than 43 mm?

From the problem, $X \approx \mathrm{N}(\mu = 40, \sigma^2 = 4)$. We standarise X using the formula: $z = \dfrac{x-\mu}{\sigma}$, thus $P(X > 43)$ becomes $\mathrm{P}\left(\mathrm{Z} > \dfrac{43-40}{2}\right)$ and $\mathrm{Z} \approx \mathrm{N}(0,1)$.

Thus we calculate $\mathrm{P}(\mathrm{Z} > 1.5) = 1 - 0.9332 = 0.0668$

From CDF table
in slide 11

Example

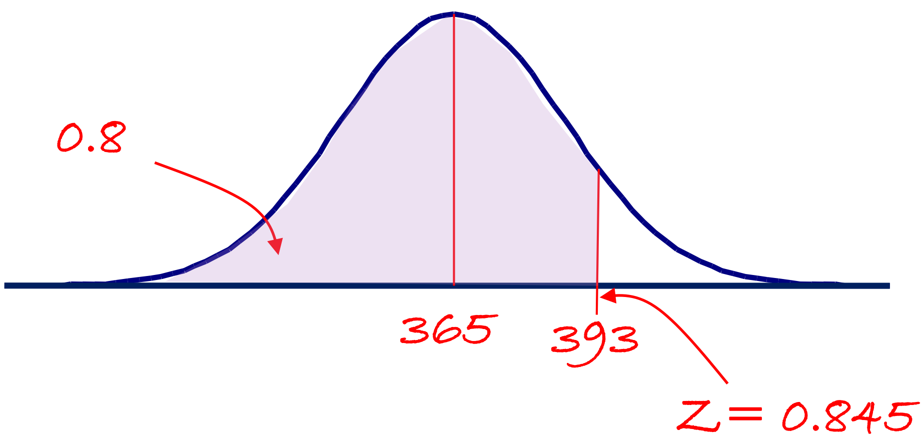


- The number of customers entering a shop is known from historical information to be normally distributed with a mean of 365 and a standard deviation of 33. What is the probability that on any given day the number of customers will be less than 320?

$$\text{From the problem, } X \approx \mathrm{N}(\mu = 365, \sigma^2 = 33^2). \text{ thus } P(X < 320)$$

$$\text{becomes } \mathrm{P}\left(\mathrm{Z} < \frac{320 - 365}{33}\right) \text{ and } \mathrm{Z} \approx \mathrm{N}(0,1).$$

$$\text{Thus we calculate } \mathrm{P}(\mathrm{Z} < -1.36) \; = \; 1 - 0.9131 = 0.0869$$



.01	0.02	0.03	0.04	0.05	0.06	0.
40	0.5080	0.5120	0.5160	0.5199	0.5239	0.52
38	0.5478	0.5517	0.5557	0.5596	0.5636	0.56
32	0.5871	0.5910	0.5948	0.5987	0.6026	0.60
17	0.6255	0.6293	0.6331	0.6368	0.6406	0.64
91	0.6628	0.6664	0.6700	0.6736	0.6772	0.68
50	0.6985	0.7019	0.7054	0.7088	0.7123	0.71
91	0.7324	0.7357	0.7389	0.7422	0.7454	0.74
11	0.7642	0.7673	0.7704	0.7734	0.7764	0.77
10	0.7939	0.7967	0.7995	0.8023	0.8051	0.80
86	0.8212	0.8238	0.8264	0.8289	0.8315	0.83

- The proprietor of the shop described in the previous question wants to set her staffing levels. She wants to be 80% sure of being able to meet customer demand. What number of customers per day should she plan for?

From the problem, $X \approx N(\mu = 365, \sigma^2 = 33^2)$. We want to find α such that $P\left(Z < \frac{\alpha - 365}{33}\right) = 0.8$. From the tables, $P(Z < 0.845) = 0.8$ and so $\alpha = 0.845 \times 33 + 365 = 392.85$, approximately 393 customers per day.

Rearrange formula: $\alpha = z\sigma + \mu$

Motivating problem

- To reduce the risk of running out of stock, retailers set a *reorder point*, which is the level of stock at which an order is raised, to arrive after a certain *lead time*.
- If demand during the lead time is Normally distributed with a mean of 200 and standard deviation of 50 what should the reorder point be set to so that there is only a 2% chance of running out of stock?

Question: Wat should the reorder point be set to so that there is only a 2% chance of running out of stock?

Given: $\mu = 200, \sigma = 50$

Solving…

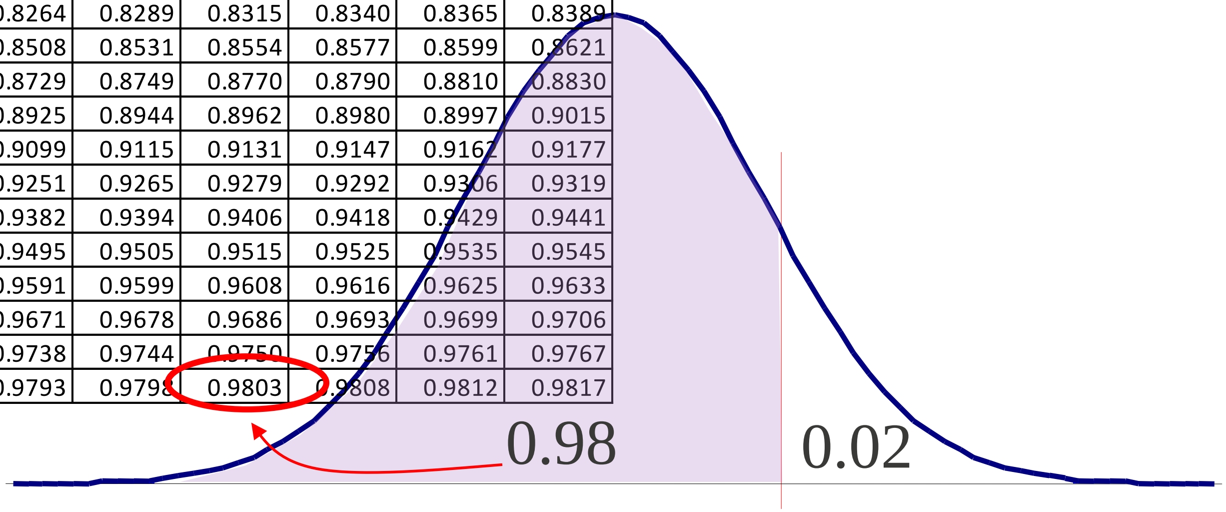
- In groups: Find $Z \approx N(0,1)$ then $X \approx N(200,50^2)$

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	0.5040	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

$X = z\sigma + \mu$

$2.06 \times 50 + 200$

$= 303$



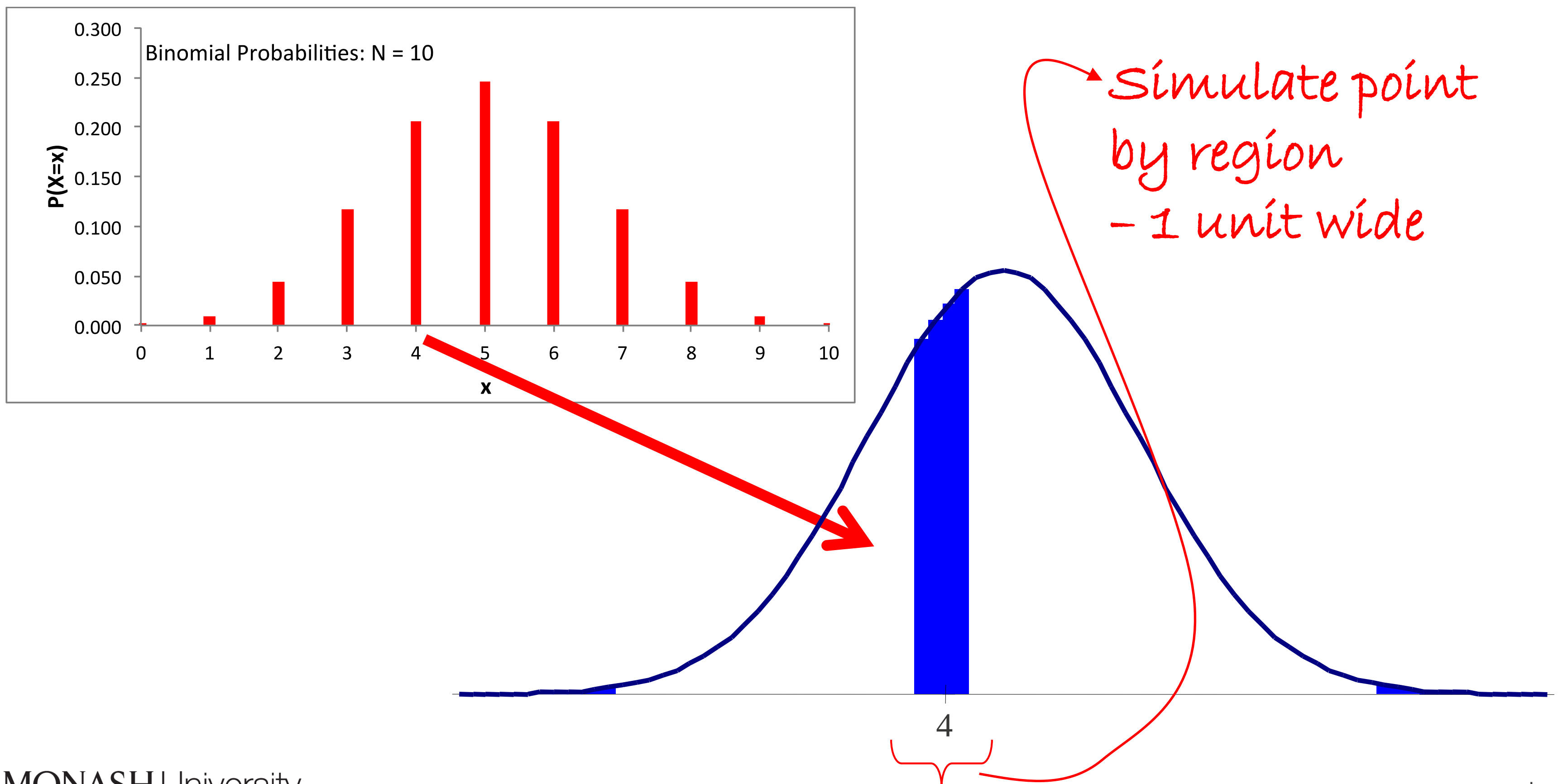
$z = 2.06$

Normal Approximation of Binomial

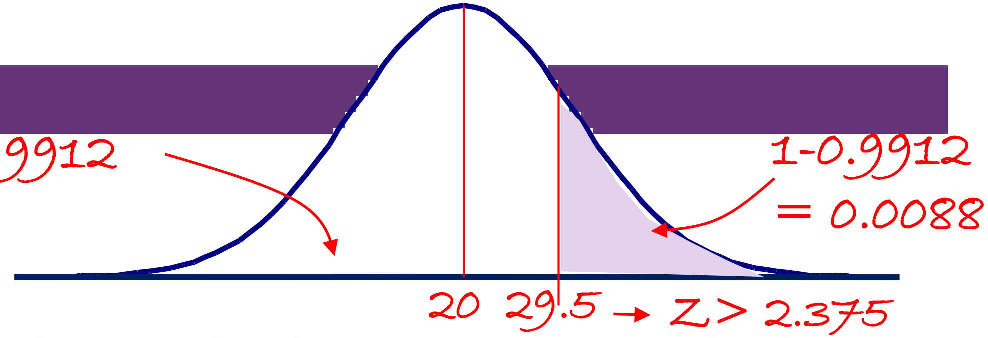
- When *n* is large and *p* is small, we can use the Normal distribution to approximate the Binomial distribution.
- We use the mean and variance of the Binomial Distribution to give us the parameters of the Normal Distribution and proceed as before.
- Because a point probability = 0 in a continuous distribution, (e.g. P(X=7) = 0) we make a <u>continuity correction</u> that assumes the probability is determined over an interval of 1 unit when we approximate a discrete distribution with a continuous one.
- If we wanted to determine P(X = 4) for a binomial problem, we would use P(3.5 < X < 4.5) as the required interval using a Normal approximation.

Continuity Correction

- *P(X = 4), X ≈ Bi <u>vs</u> P(3.5 < X < 4.5), X ≈ Normal*



Example



- Students attempt a multiple choice test consisting of 100 questions, each with 5 possible responses. What is the probability that a student will score 30 or more just by guessing?

$n = 100, p = 0.2$

Let X be the student's score. X= Bi(100,0.2)

We can approximate this with X=N(20, 16)

$\mu = np = 100 \times 0.2 = 20,$

$\sigma^2 = np(1-p) = 20 \times 0.8 = 16$

$\rightarrow \sigma = 4$

P(X>29.5), X=N(20,16)

= P(Z>2.375), Z=N(0,1)

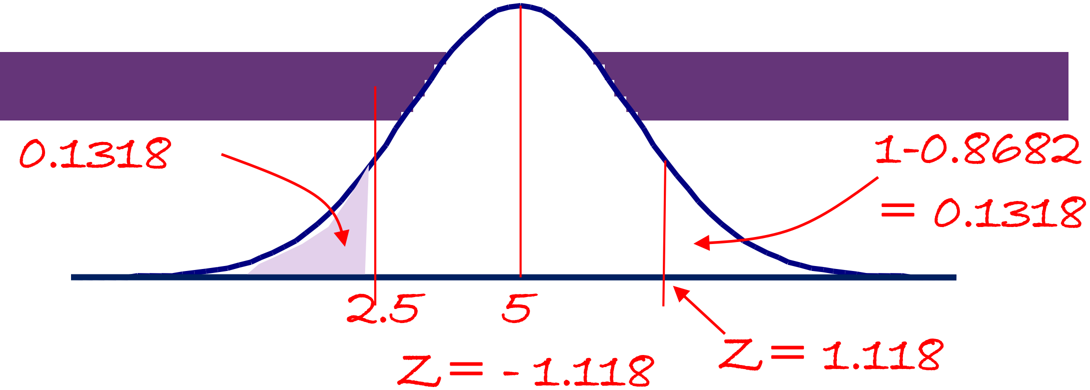
=1-0.9912 = 0.0088

$P(Z > \frac{X-\mu}{\sigma}) = P(Z > \frac{29.5-20}{4})$

Normal Approximation of Poisson

- When $\mu \geq 5$, we can use the Normal distribution to approximate the Poisson distribution.
- We use the mean and variance of the Poisson Distribution to give us the parameters of the Normal Distribution and proceed as before, using a continuity correction.

Example



- I have a 100 metre roll of carpet, with a 0.05 chance of a defect in any metre.
- What is the probability that the roll will contain two or fewer defects?

$\mu = 100 \times 0.05 = 5$

$\rightarrow \sigma^2 = 5$

Let X be the number of defects. X= Poi(5)

We can approximate this with X=N(5, 5)

P(X <= 2), X=Poi(5) ≈ P(Z < 2.5), X = N(5, 5)

= P(Z < (2.5 – 5)/sqrt(5), Z=N(0,1)

$P(Z < \frac{X-\mu}{\sigma}) = P(Z < \frac{2.5-5}{\sqrt{5}})$

= 0.1318 (exact is 0.1247)

Using Excel: =NORM.DIST(2.5,5,SQRT(5),TRUE)

$P(Z < -1.118)$

Reading/Questions

- Reading:
  - 7th Ed. Section 8.3.
- Questions:
- 7th Ed. Questions 8.8, 8.9, 8.10, 8.11, 8.12, 8.13, 8.14, 8.16, 8.19, 8.31, 8.32.

- Also file: ProbabilityDistributions.xls