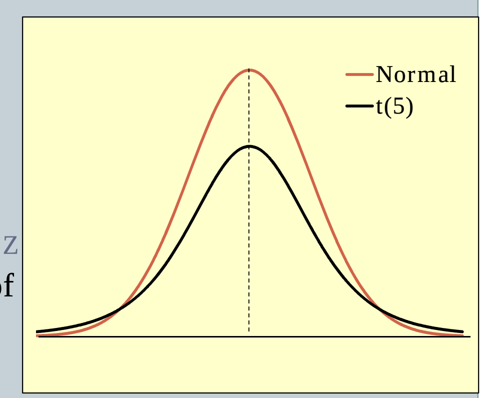
Using Excel 2010 to Find Probabilities for the Normal and t Distributions

A TUTORIAL

Using Excel with the Standard Normal, Normal, and t distributions

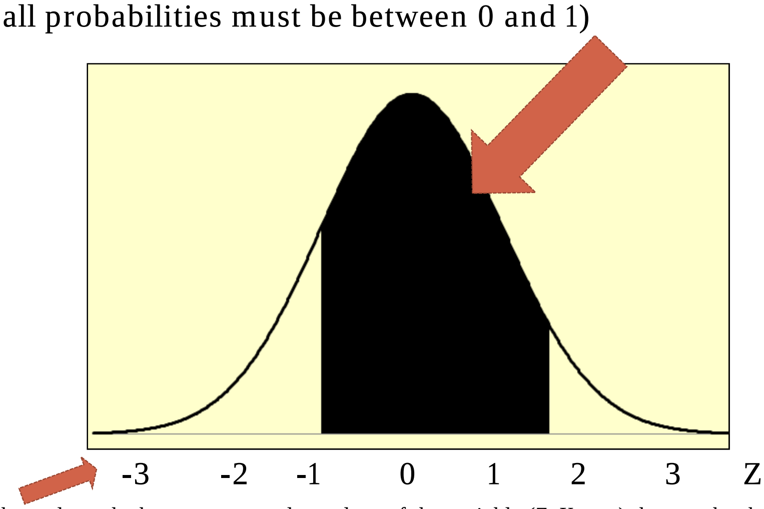
- Standard Normal ("Z")
  - Mean 0
  - Standard deviation 1
- "Generic" Normal with mean and standard deviation known
  - In "olden days" we converted these questions to ones about Z
- t (with specified degrees of freedom)
  - Looks like "normal" that is slightly "flatter"



General Information
Applies to Normal and t Distributions

- Probability corresponds to area under a "curve"
- Total area under a curve is 1
- Probability is found in one of two ways:
  - Integration
  - Using tables (paper or electronic)
- We will focus on using tables (actually electronic tables in Excel)!

Area under the curve corresponds to probability (and all probabilities must be between 0 and 1)

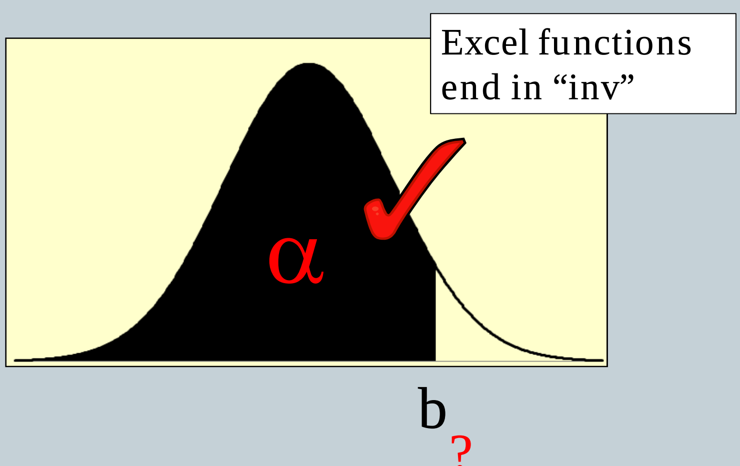


Numbers along the base correspond to values of the variable (Z, X, or t) that can be thought about as numbers on a number line. These numbers can range from -∞ to ∞. The value of the mean will appear under the peak of the distribution for these three distributions.

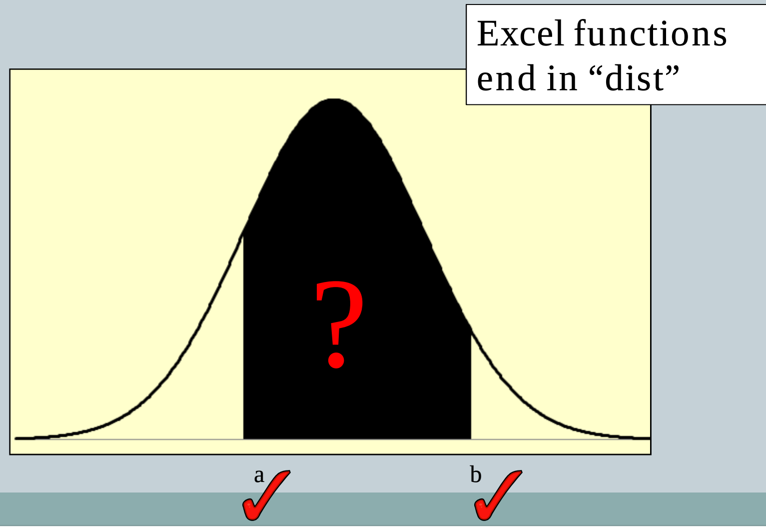
Two types of questions

- If I have (name that) distribution, what proportion of the observations will fall between ___ and ___. (or what is the likelihood that a randomly selected value will be between the same values)
  - This question asks you to find a probability from a distribution with specific parameters
- If I have (name that) distribution, find the point that will have ___% of the values above/below that point.
  - This question gives you a probability and asks you to find a value for the variable.

Situation 2: Given α, find b



Situation 1: Given a and b, find P(a<X<b)

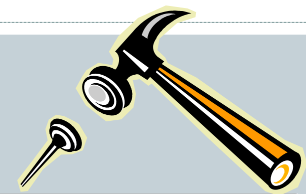


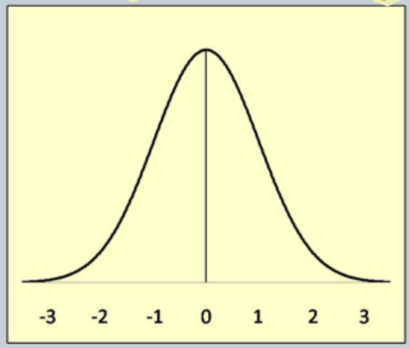
Three "Tables"

- Standard Normal (aka the Z table)
- Normal table with mean (μ) and standard deviation (σ) specified [notation X ~ N(μ =value, σ = value)]
  - Z is the special case where X ~ N(μ =0, σ = 1)
- t distribution with degrees of freedom known [notation t_{df}]
- Note: All three distributions "look" like the Bell Curve.

Standard Normal Distribution

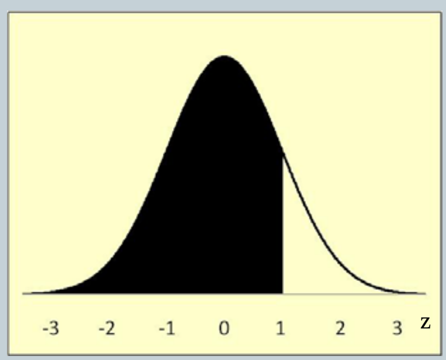
- The “Bell Curve” with a mean of 0 and standard deviation of 1.
- The curve is symmetric around 0
- Often called the Z distribution
- Z represents the number of standard deviations an observations is from the mean
- The most referenced distribution in statistics



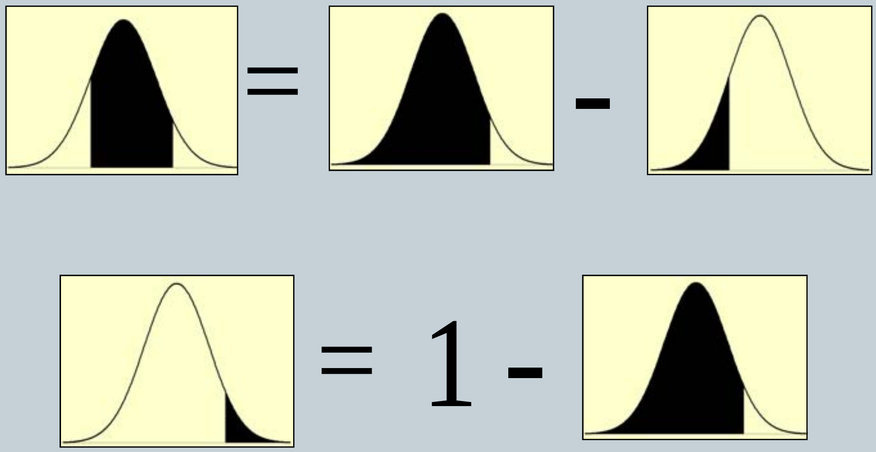


Using Excel’s Norm.s.dist Function

- Z (the number of standard deviations away from the mean) can range from –∞ to ∞ but is between -3 and 3 more than 99% of the time.
- Excel’s Standard Normal “table” provides the area under the curve to the left of a given value of Z.
- =norm.s.dist(z,true)
  - Excel will accept both positive and negative values for z
  - True can be replaced with a 1; these tell Excel to find the area under the curve

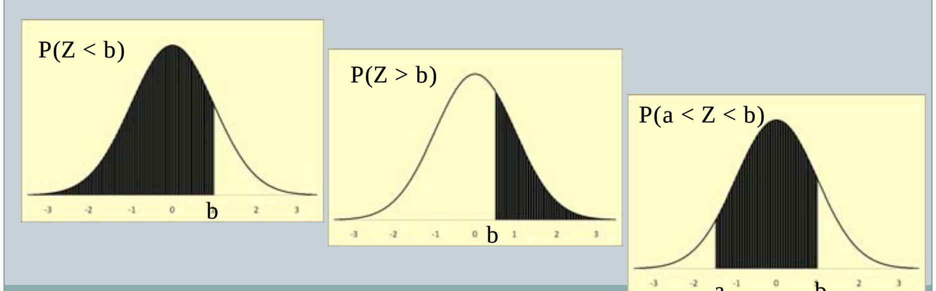


Making things work…



Finding Probabilities

- P(Z < b) = norm.s.dist(b,true) OR norm.s.dist(b,1)
- P(Z > b) = 1-norm.s.dist(b,true) OR 1-norm.s.dist(b,1)
- P(a < Z < b) = P(b > Z > a)
  =norm.s.dist(b,true) – norm.s.dist(a,true) OR
  =norm.s.dist(b,1) – norm.s.dist(a,1)



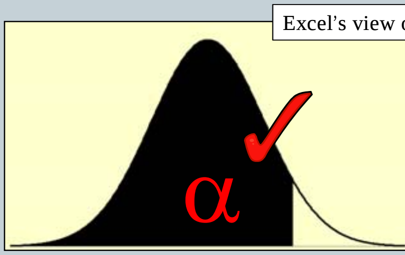
Practice

- Step 1: Draw a picture
  (to think about the logical answer)
- Step 2: Use the table (Excel function) to find the answer
- Attempt the following before moving to the next slide:
  - P(Z < 1.28)
  - P(Z > -1.64)
  - P(-1.64 < Z < 1.28)
  - P(Z < 0)
  - P(1.36 > Z > .54)

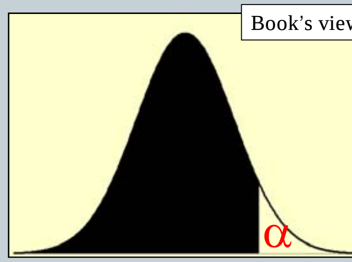
Answers

- P(Z < 1.28) = .8997
  - =norm.s.dist(1.28,true) OR =norm.s.dist(1.28,1)
- P(Z > -1.64) = .9495
  - = 1-norm.s.dist(-1.64,true) OR 1-norm.s.dist(-1.64,1)
  - Or since P(Z > -1.64) = P(Z < 1.64), you could use =norm.s.dist(1.64,true) OR norm.s.dist(1.64,1)
- P(-1.64 < Z < 1.28) = .8492
  - =norm.s.dist(1.28,true)-norm.s.dist(-1.64,true) OR =norm.s.dist(1.28,1)-norm.s.dist(-1.64,1)
- P(Z < 0) = .5
  - = norm.s.dist(0,true) OR =norm.s.dist(0,1)
  - Or recognize that by symmetry, half of the observations are below 0.
- P(1.36 > Z > .54) = .2077
  - =norm.s.dist(1.36,true)-norm.s.dist(.54,true) OR =norm.s.dist(1.36,1)-norm.s.dist(.54,1)

Situation 2: Given α, find b



α is an area under the curve—for Excel α is the area to the left of a specified value; for your book α is the area to the right of the specified value. Remember the total area under the curve is 1.



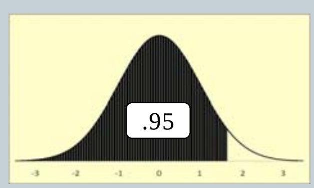
α is pronounced "alpha" and can be typed as a lower case a in Symbol font.

Reversing the Process

- Sometimes you know the distribution is normal and you know the probability, but you want to know Z.
- According to Excel, the notation z_α represents the value of z that has an area of α under the curve to the left of z. [NOTE: Your book uses the same symbol for the area to the right of z.]

Excel's command to find the value of Z with an are of α to the left:

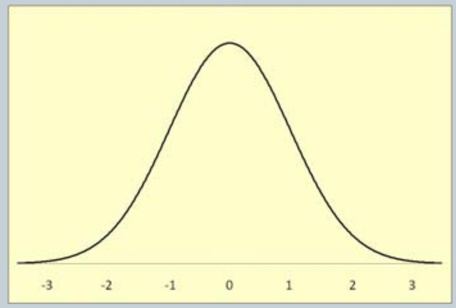
=norm.s.inv(α)



Using Excel's notation: $z_{.95} = 1.645$

Practice

- Step 1: Draw a picture
  (to think about the logical answer)
- Step 2: Use the table (Excel function) to find the answer
- Find the following where the notation matches your book's notation.
  (answers on the following slide):
  - $z_{.01}$
  - $z_{.025}$
  - $z_{.10}$
  - $z_{.10/2}$



- Answer the same questions where the notation matches that used by Excel.

Answers

- When the subscript is the area in the right tail:
  - $z_{.01}$ =norm.s.inv(.99) = 2.326
  - $z_{.025}$ =norm.s.inv(.975) = 1.96
  - $z_{.10}$ =norm.s.inv(.90) = 1.282
  - $z_{.10/2}$ =norm.s.inv(.95) = 1.645
- When the subscript is the area in the left tail:
  - $z_{.01}$ =norm.s.inv(.01) = -2.326
  - $z_{.025}$ =norm.s.inv(.025) = -1.96
  - $z_{.10}$ =norm.s.inv(.10) = -1.282
  - $z_{.10/2}$ =norm.s.inv(.05) = -1.645

Normal Distribution with Other Mean and/or Standard Deviation

- Most normal distributions don't have a mean of 0 nor a standard deviation of 1.
- The Empirical Rule tells us that approximately
  - 68% of the observations will be within one σ of μ
  - 95% of the observations will be within two σs of μ
  - 99.7% (just about all) of the observations will be within three σs of μ
- But what if we want to be more specific?... Change the question to one about Z!

$$Z = \frac{\text{Number of interest} - \text{mean for the random variable}}{\text{standard deviation for the random variable}}$$

Logical Meaning of Z

- When we are talking about individual observations:

$$Z = \frac{x - \mu}{\sigma}$$

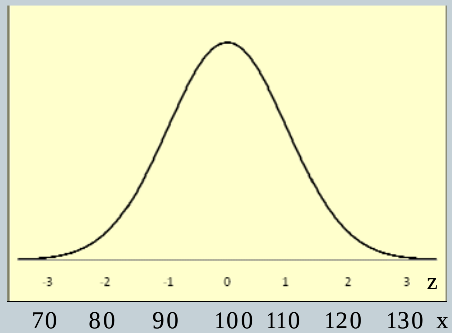
- x-μ provides the distance and direction to travel
- σ represents the size of each step
- Z represents the number of steps from the mean (positive is above the mean and negative is below the mean)

If X~N(μ = 100, σ = 10) We can change the question to one about Z

- Find
  - P(X < 110)

    $$= P\left(Z < \frac{110-100}{10}\right)$$

    =P(Z < 1)
    =norm.s.dist(1,1) [Excel]
    = .8413
- From the Empirical Rule we could have estimated the answer to be .84.



Letting Excel Do the Work

- To find the P(X < a):
  - =norm.dist(a,mean,standard deviation,true)
  - The "true" as the last value tells Excel that you are looking for the area under the curve (and can be replaced with 1).

    P(X < 110) =
    =norm.dist(110,100,10,true) OR norm.dist(110,100,10,1)
    = .8413
- To find the value of X that has an area of α to the left of it:
  - =norm.inv(α,mean,standard deviation)

Practice

- X~N(μ=85, σ=3)
  - P(X < 80)
  - P(X > 89)
  - P(80 < X < 90)
  - P(X > 100)
- X~N(μ=25, σ=.5)
  - P(X < 23.5)
  - P(X > 24.25)
  - P(24.3 < X < 25.2)
  - P(X > 25) [Do this one without using Excel.]

Answers

- X~N(μ=85, σ=3)
  - P(X < 80) =norm.dist(80,85,3,1) = .0478
  - P(X > 89) =1-norm.dist(89,85,3,1) = .0912
  - P(80 < X < 90) =norm.dist(90,85,3,1)-norm.dist(80,85,3,1) =.9044
  - P(X > 100) =1-norm.dist(100,85,3,1) = 2.87×10^{-7}
- X~N(μ=25, σ=.5)
  - P(X < 23.5) =norm.dist(23.5,25,.5,1) = .0013
  - P(X > 24.25) =1-norm.dist(24.25,25,.5,1)= .9332
  - P(24.3 < X < 25.2) =norm.dist(25.2,25,.5,1)-norm.dist(24.3,25,.5,1) = .5747
  - P(X > 25) [Do this one without using Excel]
    - μ=25 so P(X > 25) = .5

Application

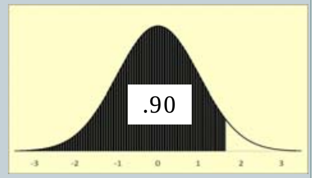
- Assume that the travel time between Dahlonega and downtown Atlanta follows a Normal distribution with a mean of 80 minutes and standard deviation of 15 minutes.
  - What proportion of the time will it take less than one and a half hours to make the trip?
  - What is the likelihood that a randomly selected trip will take more than two hours?
  - How long should be allowed for the trip if the driver wants to make sure that they arrive on time?
    - Assume the driver is willing to take some risk but not too much. The driver says, "I want to make sure that the allowed time would get me there on time or early 90% of the time."

Answers (part 1)

- X = travel time in minutes
- $X \sim N(\mu = 80, \sigma = 15)$
- What proportion of the time will it take less than one and a half hours to make the trip?
  - $P(X < 90)$ =norm.dist(90,80,15,1) = .7475
- What is the likelihood that a randomly selected trip will take more than two hours?
  - $P(X > 120)$ =1-norm.dist(120,80,15,1)= .0038

Answers (part 2)

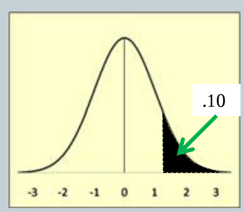
- How long should be allowed for the trip if the driver wants to make sure that they arrive on time?
  - Assume the driver is willing to take some risk but not too much. The driver says, “I want to make sure that the allowed time would get me there on time or early 90% of the time.”



- P(X < ___) = .90
  - =norm.inv(.9,80,15) = 99.223
  - Or use =norm.s.inv(.9) to find z in 80 + z(15) = 80 + 1.28155(15) = 99.223

A twist to the last problem

- How long should be allowed for the trip if the driver wants to make sure that they aren’t late?
  - Assume the driver is willing to take some risk but not too much. The driver says, “I want to make sure that the allowed time will get me there so that I won’t be late more than 10% of the time.”
  - Draw a picture (shown to the right) Notice that the picture to the right illustrates the situation. Also notice that the answer will be the same as the answer to the previous question since the point that has 10% to the right is the same as the point that has 90% to the left.



Using the t tables
With Excel 2010

- **WARNING: The functions for t in Excel 2007 and earlier are not similar to the ones from Excel 2010.**
- Think of the t tables in Excel 2010 as similar to the Z table…with the added information about degrees of freedom.
  - Like using Z with a "generic" normal distribution, you will need to standardize the value of t if you are working with a distribution that has a mean and standard deviation other than 0 and 1, respectively.
  - Generally, this is done by taking:

$$\frac{\text{(number of interest)} - \text{mean for the random variable}}{\text{standard deviation for the random variable}}$$

Finding Table Values (Excel 2010)

- To find probabilities associated with t (where b is the standardize value)
  - $P(t_{df} < b)$ = t.dist(b,df,true) OR =t.dist(b,df,1)
  - $P(t_{df} > b)$ = 1-t.dist(b,df, true) OR 1-t.dist(b,df,1)
  - $P(a < t_{df} < b) = P(b > t_{df} > a)$
    = t.dist(b,df,true) – t.dist(a,df,true) OR t.dist(b,df,1)-t.dist(a,df,1)
- To find t values associated with an area of α in the left tail of the distribution:
  - =t.inv(α,df)

Examples

- $P(t_{15} < 2.75)$
  - =t.dist(2.75,15,1) = .99256
- $P(t_{15} > -1.23)$
  - 1- t.dist(-1.23,15,1) = .88118
- $P(t_{15} <$ ____) = .96
  - =t.inv(.96,15) = 1.8777

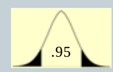
Practice

- $P(t_{18} < 2.34)$
- $P(t_{18} < -2.34)$
- $P(-1.82 < t_{18} < 2.34)$
- $P(t_{18} > 1.37)$
- $P(t_{18} > -1.37)$
- $P(t_{18} <$ ___) = .95
- $P(t_{18} >$ |___|) = .05
- $P(t_{18} >$ ___) = .95

Answers

- $P(t_{18} < 2.34)$
  - = t.dist(2.34,18,1)
    = .9845
- $P(t_{18} < -2.34)$
  - =t.dist(-2.34,18,1)
    = .0155
- $P(-1.82 < t_{18} < 2.34)$
  - =t.dist(2.34,18,1)-t.dist(-1.82,18,1)
    = .9418
- $P(t_{18} > 1.37)$
  - =1-t.dist(1.37,18,1)
    = .0938
- $P(t_{18} > -1.37)$
  - =1-t.dist(-1.37,18,1)
    = .9062

- $P(t_{18} <$ ___) = .95
  - =t.inv(.95,18)
    =1.734
- $P(t_{18} >$ |___|) = .05



  - =t.inv(.975,18)
    =2.101
- $P(t_{18} >$ ___) = .95
  - =t.inv(.05,18)
    =-1.734