Continuous Random Variables and the Normal Distribution

Dr. Tom Ilvento
FREC 408

What are Continuous Random Variables?

- Unlike Discrete Random Variables, Continuous Random Variables take on any point in the interval
- Thus the probability distribution is continuous
- It is referred to as a **Probability Density Function**
  - PDF
  - f(x)

When dealing with a PDF

- It is not particularly useful to think of a probability when a continuous random variable takes on a particular value
  - **P(x=a) = 0**
- But, we can think of areas under the curve as reflecting a probability
  - P(a ≤ x ≤ b) = some proportion of the curve
  - E.G., P(10 ≤ x ≤ 20)
  - Or the probability up to a point, or after a point

This is a key concept!!!!!

Normal Distribution

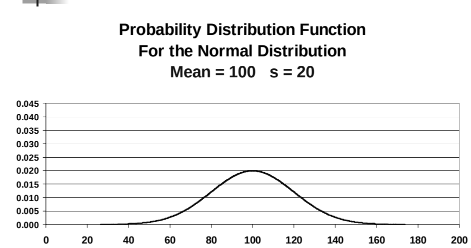
- One bell shaped, symmetrical distribution is the **normal distribution** (Def6.1 P283)
- It is defined by two parameters
  - μ the Mean
  - σ The Standard Deviation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}$$

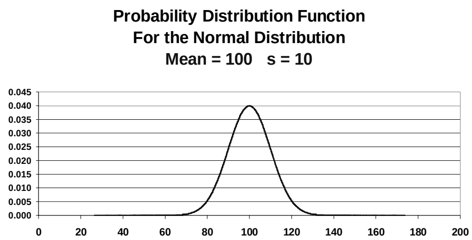
Normal Distribution

- For every distribution with a mean (μ) and a standard deviation (σ) there is a different normal curve
- Thus, there are an infinite number of normal curves
- If x is distributed as a normal variable then it is designated as:
  - **x ~ N(mean, std Dev)**

Normal Distribution



Normal Distribution



Properties of the Normal Distribution (P284)

- Symmetrical, Bell-shaped curve
- Defined by the mean and standard deviation
- Mean = Median = Mode

Standard Normal Distribution

- Since its properties are defined by a formula, we can a priori define probabilities associated with the curve
- If we convert our normally distributed variable to a z-score, we make it possible to use one table of probabilities for all normal pdf
- This is called the **Standard Normal Distribution**
  - mean = 0
  - std dev = 1

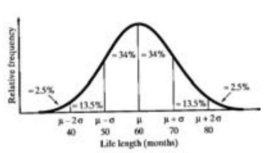
Standard Normal Distribution

- $\mu=0$
- $\sigma=1$
  - If x ~ N, then
  - $\mathbf{z = (x - \mu)/\sigma}$ is also ~ N

Finding Areas under the Curve, Table on page 737

- **Steps**

1. Draw the curve and the area we are interested in
2. Convert the values to z-scores
3. Read the proportions in the table (P737)
4. Do any calculations necessary



Look at the table B3 on page 737

- Only ½ of the curve is presented since the distribution is symmetrical
  - Each half represents p = .5
- Allows for two decimal places
  - Vertical axis is the ones and first decimal place
  - Horizontal axis is the second decimal place

Look at the table B3 on page 737

- The probabilities in the table represent the probability from z=0 up to the z value you choose (also the shaded area under the curve)
- Thus The probability associated with z = 1.00 is the area under the curve from z=0 (the mean) to z=1.00
- Or one standard deviation from the mean

Normal Curve Table using z-scores

Z	.00	.01	.02		.06	.07	.08	.09
1.0	.3413	.3438	.3461		.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686		.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888		.3962	.3980	.3997	.4015
2.0	.4772	.4778	.4783		.4803	.4808	.4812	.4817

Find the probabilities associated with

- Z = 1.00
- Z = 1.20
- Z = 1.28
- Z = 2.00
- Z = 2.09

Normal Curve Table using z-scores

Z	.00	.01	.02		.06	.07	.08	.09
1.0	.3413	.3438	.3461		.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686		.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888		.3962	.3980	.3997	.4015
2.0	.4772	.4778	.4783		.4803	.4808	.4812	.4817

Find the probabilities associated with

- Z = 1.00 **.3413**
- Z = 1.20 **.3849**
- Z = 1.28 **.3997**
- Z = 2.00 **.4772**
- Z = 2.09 **.4817**

Note: The probability from the table means the probability from Z=0 up to the calculated Z value

Notice form the Table

- Z= 1.0 is one standard deviation from the mean
  - p = .3413
  - ± 1.0 would be 2(.3413) = .6826
  - or **68.26%**

Notice form the Table

- Z = 2 is two standard deviations from the mean
  - p = .4772
  - ± 2.0 would be 2(.4772) = .9544
  - or **95.44%**

Notice from the Table

- Z = 3 is two standard deviations from the mean
  - p = .4987
  - ± 2.0 would be 2(.4987) = .9974
  - or **99.74%**

Problem

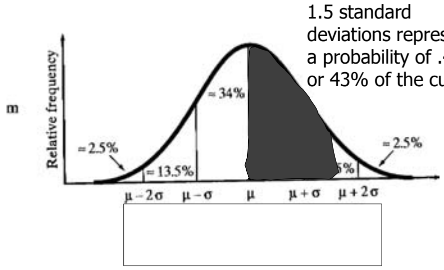
- Find the area under the standard normal curve for a z-score between 0 and 1.5.

Answer

- A z-score of zero is at the mean, with a probability of zero
- A z-score of 1.5 is 1.5 standard deviations above the mean, which corresponds to a probability of .4332
- We want the area from the mean to 1.5 standard deviations from the mean
- **Equal to .4332**

Graphic depiction

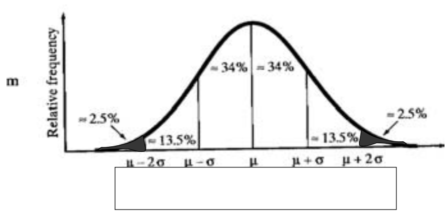
1.5 standard deviations represents a probability of .4332 or 43% of the curve



Problems in Class

- Suppose a variable is distributed normally with a mean = 300 and a standard deviation of 30
  - X ~ N μ = 300 σ = 30
- **What is the probability that a value of x is more than 2 standard deviations away from the mean?**
- **STEPS:**
  - Draw it out
  - Calculate z-score
  - Check the table
  - Do any final calculations

More than two standard deviations in the part in the tails



Problem - More than 2 std deviations

X ~ N $\mu = 300$ $\sigma = 30$

- Let's start with a z-score = 2
- In the table a z-score of 2 represents a probability up to that point of .4772
  - .5 - .4772 = **.0228** one side of curve
  - 2 x .0228 = **.0456** both sides of curve

Problem

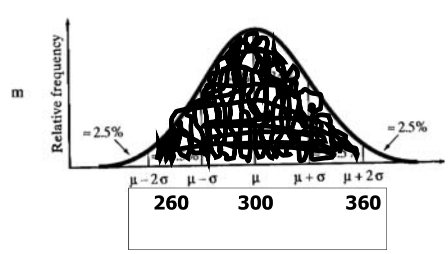
X ~ N $\mu = 300$ $\sigma = 30$

- More than 3 std deviations
  - Z = 3.00
  - In table when z = 3.00 we have a probability up to that point on one side of the curve of .4987
  - .5 - .4987 = .0013 one side of curve
  - 2 x .0013 = **.0026** both sides of curve

Problem

- X ~ N $\mu = 300$ $\sigma = 30$
- **Probability that x is between 260 and 360?**
  - Draw it
  - Calculate z-scores
  - Look up in the table
  - Do any calculations

Graphic depiction



Calculate Z-scores

- X ~ N $\mu = 300$ $\sigma = 30$
- **Probability that x is between 260 and 360?**

  X =260 z = (260 – 300)/30 = -1.33

  X = 360 z = (360 – 300)/30 = 2.00

Probability that x is between 260 and 360?

- Since the table shows only one side, use absolute value
  - Z for 1.33 = .4082
  - Z for 2.00 = .4772
  - **.4082 + .4772 = .8854**

What is the X value at the 80th percentile?

- Given X ~ N $\mu = 300$ $\sigma = 30$
- What am I looking for?
  - I am looking for the X value that corresponds to the 80th percentile
  - The 80th percentile reflects everything up to the mean (50th percentile)
  - Plus .30 more

What is the value at the 80th percentile?

- Look in the table for .30
- It is between .84 (p=.2995) and .85 (p=.3023)
- I could extrapolate, but I know it is a lot closer to .84
  - .841 is a good approximation
- So now I can solve for the value of X that corresponds to a z-value of .841

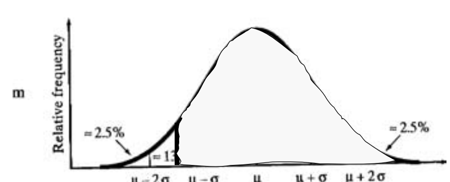
Next -

- Solve for x
  - $.841 = (x - 300)/30$
  - $30 \cdot .841 = x - 300$
  - $300 + (30 \cdot .841) = x$
  - $325.23 = x$
- **The 80th percentile is at 325.23**

Problem – you solve it!

- The tread life of a particular brand of tire is a random variable best described by a normal distribution with a mean of 60,000 miles and a standard deviation of 8,300 miles.
- **What proportion of the tires last longer than 53,775 miles?**

Graphic depiction



53,775 60,000

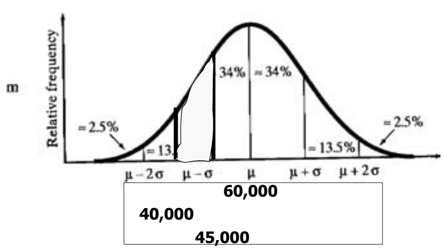
Problem – you solve it!

- **What proportion of the tires last longer than 53,775 miles?**

Problem – you solve it!

- The tread life of a particular brand of tire is a random variable best described by a normal distribution with a mean of 60,000 miles and a standard deviation of 8,300 miles.
- **What proportion of tires last between 40,000 and 45,000 miles?**

Graphic depiction



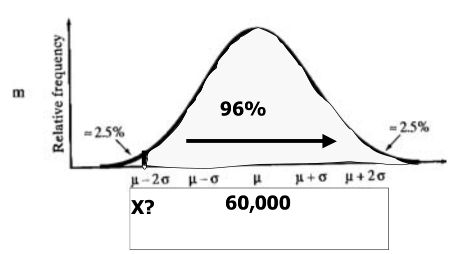
Problem – you solve it!

- **What proportion of tires last between 40,000 and 45,000 miles?**

Problem – you solve it!

- The tread life of a particular brand of tire is a random variable best described by a normal distribution with a mean of 60,000 miles and a standard deviation of 8,300 miles.
- **What warranty should the company use if they want 96% of the tires to outlast the warranty?**

Graphic depiction



Problem – value so 96% of the tires outlast the warranty?

Normal Distribution as an approximation of the Binomial Distribution – pages 307-309

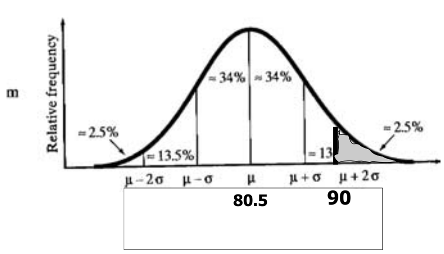
- The normal distribution can be used as an approximation for the Binomial distribution of x successes in n trials with a *continuity correction factor*
- The book provides a formula when it is ok to do this, whenever both $np \geq 5$ and $nq \geq 5$
- For our purposes we will note that when n is reasonable large (> 50), and p is not extremely small (> .10), we can generally use this approach
- **Note:** when the sample size (n) is very large, i.e., greater than 500, we can ignore the continuity correction factor.

Problem – you solve it!

- Pete Targus Enterprises is the largest privately held agricultural operations in the U.S. A recent study of baled hay at the operation reveals that baled hay is
  - **~N(80.5 lbs, 7.2 lbs)**
- Hay that is too heavy (over 90 lbs) has too much moisture and will spoil. Hay that is too light may be overly dry and not have the desired feed content.

What percent of the bales can we expect to be over 90 lbs?

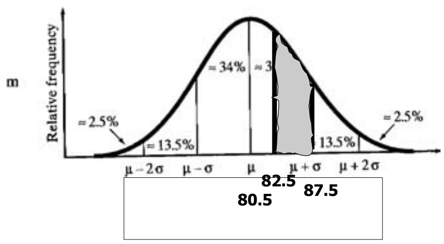
Graphic depiction



Problem – you solve it!

- The ideal weight for bales is between 82.5 lbs and 87.5 lbs.
- **What percent of the bales are expected to be within these weights?**

Graphic depiction



Bales between 82.5 lbs and 87.5 lbs.

Problem – you solve it!

- The physical fitness of a patient is measured by the maximum oxygen uptake (recorded in milliliters per kilogram, ml/kg)
- The maximum oxygen uptake for cardiac patients who regularly participate in sports or exercise programs was found to be:
  - ~N(24.1, 6.3)

Problem – you solve it!

- **What is the probability that a cardiac patient who regularly participates in sports has a maximum oxygen uptake of at least 20 ml/kg?**

Answer

Problem – you solve it!

- **What is the probability that a cardiac patient who regularly participates in sports has a maximum oxygen uptake of 10.5 ml/kg or lower?**

Answer

- **Consider a cardiac patient with a maximum oxygen uptake of 10.5 ml/kg. Is it likely that this patient participates regularly in sports or exercise programs?**