CRP 272
The Normal Distribution

Dave Swenson

Introduction to Normal Distribution

- Why Study Normal Distribution?
  - Although few data distributions are precisely normal, many resemble the normal distribution, some closely
  - The sampling distribution of *any* distribution is a normal distribution provided the sample size is large enough
  - Used in estimation and inferential statistics (e.g., hypothesis testing) as a statistical model
- Normal Distribution Characteristics
  - The curve forms a bell shape and the midpoint is where the mean, median, and mode can be found
  - The curve is symmetric
  - The curve is spread out in a particular way

Normal Distribution Examples

- Why Study Normal Distribution?
  - Many data distributions resemble a normal curve
  - The sampling distribution of any distribution is a normal distribution provided the sample size is large enough
  - Used in estimation and inferential statistics (e.g., hypothesis testing)
- Normal Distribution Characteristics
  - The curve forms a bell shape and the midpoint is where the mean, median, and mode can be found
  - The curve is symmetric
  - The curve is spread out in a particular way

Figure 5

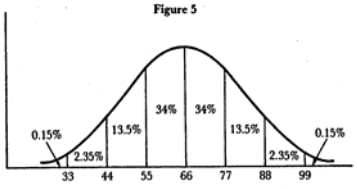
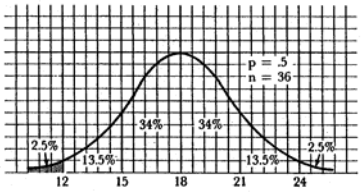


Figure 6



2

Statistical Models

- In order to make inferences about a population (based on data obtained from a sample), one needs to interpret the statistics in light of their underlying distribution and probability
- A statistical model computes the probabilities that an event will occur under the specified conditions
- The normal curve is an excellent example of a statistical model
  - In fact, it is the most commonly used statistical model

3

Statistical Models

- There are two ways to develop a statistical model:
  - Create an experimental sampling distribution
    - Draw all possible samples of size N from the population
    - Compute sample statistics of interest for each (e.g. the mean)
    - Create a frequency distribution of these statistics
    - Calculate the probability of getting each value of the statistic if chance alone is operating
    - Yikes!
  - Create or use a theoretical sampling distribution
    - Use an equation to generate it or use previously developed tables
    - Example: the normal distribution and Z-score tables

The Normal Curve As A Statistical Model

- It is generated by an equation and is theoretical
- It is a sampling distribution that approximates the sampling distribution of many different populations in the real world
  - Caution: certainly not all distributions resemble the normal distribution!
- Using the theoretical normal curve model allows one to simply determine probabilities and other needed information

Normally Distributed Phenomena

- Many types of data conform to the normal distribution:
  - Intelligence (IQ scores)
  - Heights and weights of human beings
  - Weights and lengths of fish caught in a large lake
  - Annual incomes of households
  - SAT or GRE scores
  - Grades on exams or in large classes
  - Climate (e.g. high and low temperatures over time)
  - Many psychological data items

Data Generally Thought NOT To Be Normally Distributed

- Financial indicators and economic data
  - Price changes, interest rates, stock values, commodity prices, exchange rates
- Lifetimes of humans
- Lifetimes of mechanical and electronic devices
- Waiting times (queuing)
- Safety data (e.g. car crash data)

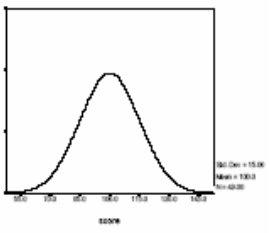
Characteristics

- Midpoint is the highest point, which corresponds to the greatest frequency (the mode)
- The curve is symetrical
- The curve has a distinctive spread and shape

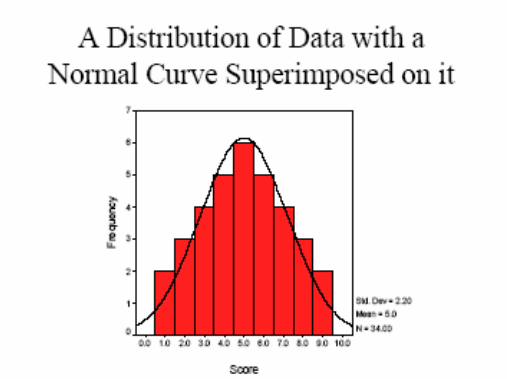
Properties Of Any Normal Curve: A Bell Shaped Curve

Properties of a Normal Curve

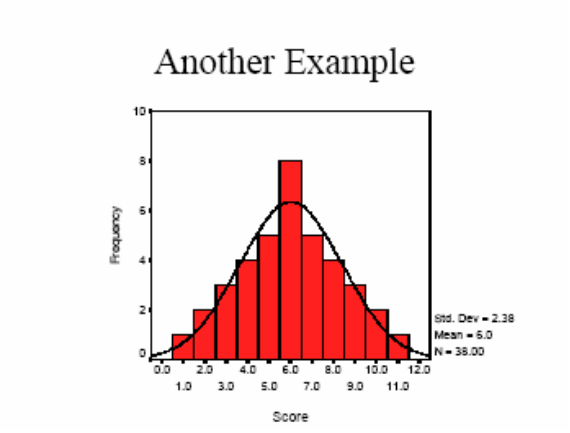
- Symmetrical, bell-shaped distribution
  - Unimodal
  - Tails don't touch X axis
- Mean = Median = Mode
- Values in the middle of the curve have a greater probability of occurring than values further away
- There are many possible normal curves with different means and standard deviations



A Dataset That Approaches Normality



And One That Departs From Normality



The Standardized Normal Curve

Because there are many possible normal curves, in order to make use of their handy properties, we need to standardize the model:

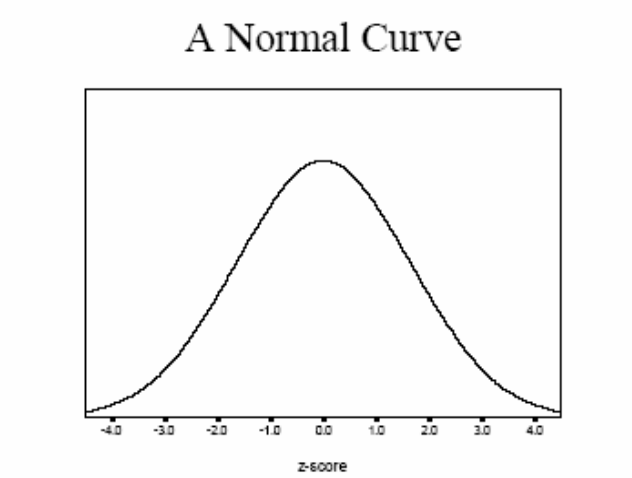
– Convert every normal distribution to a standardized (or standard) normal distribution and convert all of the scores to standard scores

– Then use a Table of Areas under the Normal Curve to interpret and make inferences about the data

Properties of a Standardized Normal Curve: Z-Scores

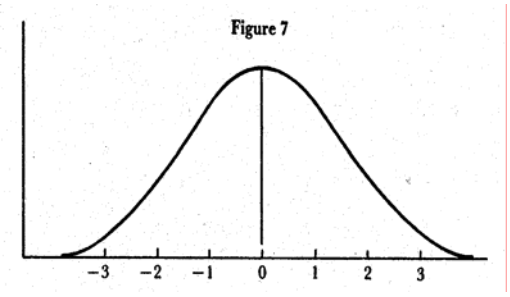
- **Convert the raw scores to z-scores**
- **Every distribution of z-scores:**
  - **Has a mean equal to 0, a median equal to 0, and a mode equal to 0**
  - **Has a standard deviation equal to 1**
  - **Has a shape that is the same as the underlying distribution of raw scores**

A Standardized Normal Curve



The Standardized Normal Distribution (Centered On Zero)

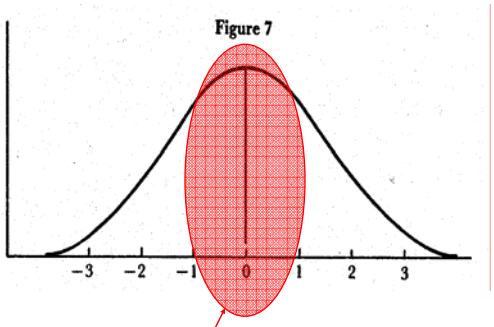
- Characteristics
  - Mean: 0
  - Median and Mode: 0
  - Standard Deviation: 1
- The z-score
  - Tells us how many standard deviations there are between the selected score and the mean



$$z = \frac{(x_i - \bar{x})}{s}$$

The Standardized Normal Distribution

- Characteristics
  - Mean: 0
  - Median and Mode: 0
  - Standard Deviation: 1
- The z-score
  - It tells us how many standard deviations there are between the selected score and the mean



About 68% of the distribution is within + or – 1 standard deviation under the curve

The Z-Score Table

z	0 z	0 z	z	0 z	0 z	z	0 z	0 z
0.00	.0000	.5000	0.55	.2088	.2912	1.10	.3643	.1357
0.01	.0040	.4960	0.56	.2123	.2877	1.11	.3665	.1335
0.02	.0080	.4920	0.57	.2157	.2843	1.12	.3686	.1314
0.03	.0120	.4880	0.58	.2190	.2810	1.13	.3708	.1292
0.04	.0160	.4840	0.59	.2224	.2776	1.14	.3729	.1271
0.05	.0199	.4801	0.60	.2257	.2743	1.15	.3749	.1251
0.06	.0239	.4761	0.61	.2291	.2709	1.16	.3770	.1230
0.07	.0279	.4721	0.62	.2324	.2676	1.17	.3790	.1210
0.08	.0319	.4681	0.63	.2357	.2643	1.18	.3810	.1190
0.09	.0359	.4641	0.64	.2389	.2611	1.19	.3830	.1170
0.10	.0398	.4602	0.65	.2422	.2578	1.20	.3849	.1151
0.11	.0438	.4562	0.66	.2454	.2546	1.21	.3869	.1131
0.12	.0478	.4522	0.67	.2486	.2514	1.22	.3888	.1112
0.13	.0517	.4483	0.68	.2517	.2483	1.23	.3907	.1093
0.14	.0557	.4443	0.69	.2549	.2451	1.24	.3925	.1075
0.15	.0596	.4404	0.70	.2580	.2420	1.25	.3944	.1056
0.16	.0636	.4364	0.71	.2611	.2389	1.26	.3962	.1038
0.17	.0675	.4325	0.72	.2642	.2358	1.27	.3980	.1020
0.18	.0714	.4286	0.73	.2673	.2327	1.28	.3997	.1003
0.19	.0753	.4247	0.74	.2704	.2296	1.29	.4015	.0985

Handy Formulas Using Z-Scores

Formulas for z-scores

- Given that you know the raw score, and the mean and standard deviation of the population, and you want to obtain the z-score:

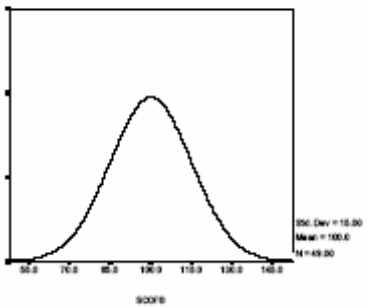
$$z = \frac{X - \mu}{\sigma}$$

- Given that you have the z-score and you want to know the corresponding raw score:

$$X = z\,\sigma + \mu$$

Example Z-Score Population

An Example



- IQ scores
- If your IQ is 110, how do you compare to the rest of the population?
- $z = \frac{110 - 100}{15}$
- = 10/15
- z = .67

Using Z-Scores: An Example

- **With an IQ of 110, your Z-Score is +.67**
- **Go to any Z-Score Table (it will list the Percent Area Under the Normal Curve) and find a z score for .67**
- **The area between the z-score and the mean is .2486**
- **We know that the area up to the mean is .50, and we know that our score was above the mean, so we add the two numbers to get a total area of .7486**
- **Therefore, your percentile rank in terms of IQ is 74.86%, which means that you scored equal to or greater than 74.86% of the population**

Remember Z Scores Are Standardized, So:

- **Percentage = Area = Probability**
- **If you have the area or probability, and want the percentage, move the decimal to the right 2 places**
- **Area = .7486, percentage = 74.86%**

Going From Percentage To Raw Score

- **What if you're given the percentile rank, or a percentage, and asked for the raw score?**
- **First, go to the Normal Distribution table and find the percent score (in the central area of the table)**
- **Find the z-score that corresponds to the percent score**
- **Then use the following formula to find the raw score:**
  - **$X = z\,\sigma + \mu$**

The Z-Score Table

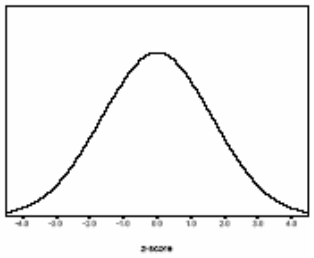
Percentiles

z	0 z	0 z	z	0 z	0 z	z	0 z	0 z
0.00	.0000	.5000	0.55	.2088	.2912	1.10	.3643	.1357
0.01	.0040	.4960	0.56	.2123	.2877	1.11	.3665	.1335
0.02	.0080	.4920	0.57	.2157	.2843	1.12	.3686	.1314
0.03	.0120	.4880	0.58	.2190	.2810	1.13	.3708	.1292
0.04	.0160	.4840	0.59	.2224	.2776	1.14	.3729	.1271
0.05	.0199	.4801	0.60	.2257	.2743	1.15	.3749	.1251
0.06	.0239	.4761	0.61	.2291	.2709	1.16	.3770	.1230
0.07	.0279	.4721	0.62	.2324	.2676	1.17	.3790	.1210
0.08	.0319	.4681	0.63	.2357	.2643	1.18	.3810	.1190
0.09	.0359	.4641	0.64	.2389	.2611	1.19	.3830	.1170
0.10	.0398	.4602	0.65	.2422	.2578	1.20	.3849	.1151
0.11	.0438	.4562	0.66	.2454	.2546	1.21	.3869	.1131
0.12	.0478	.4522	0.67	.2486	.2514	1.22	.3888	.1112
0.13	.0517	.4483	0.68	.2517	.2483	1.23	.3907	.1093
0.14	.0557	.4443	0.69	.2549	.2451	1.24	.3925	.1075
0.15	.0596	.4404	0.70	.2580	.2420	1.25	.3944	.1056
0.16	.0636	.4364	0.71	.2611	.2389	1.26	.3962	.1038
0.17	.0675	.4325	0.72	.2642	.2358	1.27	.3980	.1020
0.18	.0714	.4286	0.73	.2673	.2327	1.28	.3997	.1003
0.19	.0753	.4247	0.74	.2704	.2296	1.29	.4015	.0985

Example: Percentile To Raw Score

Assume that the mean weight of American males between 21 to 30 years old is 156 pounds with a standard deviation of 8 pounds.

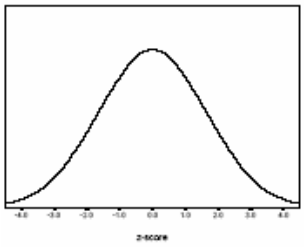
- **Tom's percentile rank is 6.68%. How much does he weigh?**
- Find the % in table to get corresponding z score:
  - 6.68% = .0668
  - z = 1.5, but below the mean, therefore -1.5
- X = -1.5(8) + 156
  = -12 + 156
  =144
- Tom weighs 144 lbs



24

Probability Calculation Using The Standardized Normal Curve

Assume that the mean weight of American males between 21 to 30 years old is 156 pounds with a standard deviation of 8 pounds.



- **If you select a male at random, what is the probability that his weight will be 160 pounds or more?**
- $z = \frac{160 - 156}{8}$
  $= 4/8 = 0.5$
- Find z in table, get percent for area above the score
- Area = .3085

25

Let's Collect Student Height Data In Inches

- Should these data resemble a normal distribution?
- Collect the data
- Prepare the data
- Analyze the data
  - Frequency distribution
    - Does the distribution appear normal
  - Mean, median, mode
  - Quartiles
  - Standard deviation
  - Selected Z-scores