**2.1 Describing Location of an Individual in a Distribution**

**Measures of Position -**

***Percentile –*** the *pth percentile* of a distribution is the value with *p* percent of the observations less than it

**Example #1**

The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2009

5 | 9

6 | 2455

7 | 00455589

8 | 0345667778

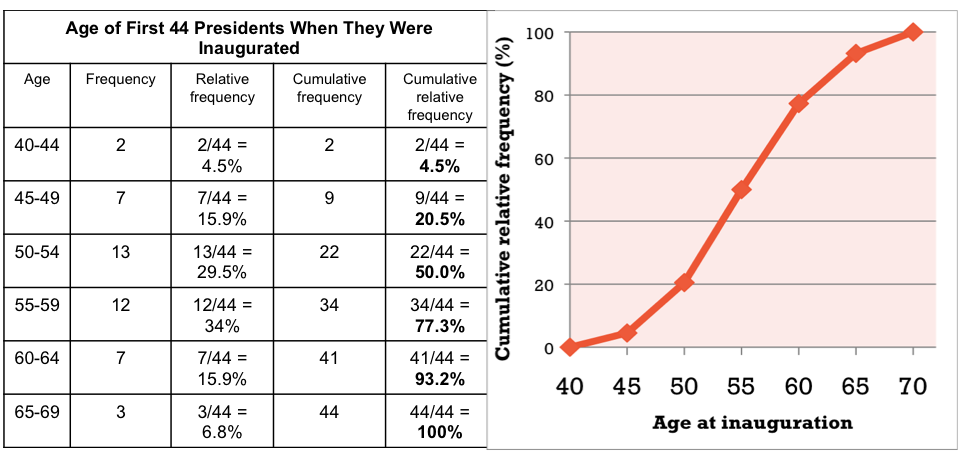
9 | 123557

10 | 3

Find the percentiles for the following teams

1. The Colorado Rockies, who won 92 games
2. The New York Yankees, who won 103 games
3. The Kansas City Royals and Cleveland Indians, who both won 65 games

***Cumulative Relative Frequency Graph (ogives)*** *–*



**Example #2**

Here is a table showing the distribution of median household incomes for the 50 states and the District of Columbia.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Median income  ($1000s | Frequency | Relative  Frequency | Cumulative  Frequency | Cumulative Relative  Frequency |
| 35 to < 40 | 1 | 1/51 = 0.020 | 1 | 1/51 = 0.020 |
| 40 to < 45 | 10 | 10/51 = 0.196 | 11 | 11/51 = 0.216 |
| 45 to < 50 | 14 | 14/51 = 0.275 | 25 | 25/51 = 0.490 |
| 50 to < 55 | 12 | 12/51 = 0.236 | 37 | 37/51 = 0.725 |
| 55 to < 60 | 5 | 5/51 = 0.098 | 42 | 42/51 = 0.824 |
| 60 to < 65 | 6 | 6/51 = 0.118 | 48 | 48/51 = 0.941 |
| 65 to < 70 | 3 | 3/51 = 0.059 | 51 | 51/51 = 1.00 |

1. Create a cumulative relative frequency graph
2. What does the steepness of the graph tell you about the distribution?
3. At what percentile is California, with a median income of $57,445?
4. Estimate and interpret the first quartile of this distribution

***Z-Scores (standardized value) –*** indicates how many standard deviations the observation is above or below the mean



1. The z-score is directional. If the z-score is positive, the observation is above average. If the z-score is negative, the observation is below average.
2. Z-score tells how many standard deviations a value is from the mean, no matter what the shape of the distribution.
3. When comparing z-scores from different distributions, it is important that the distributions be roughly the same shape.

**Example #4**

In 2009, the mean number of wins in MLB was 81 with a standard deviation of 11.4 wins.

1. Find and interpret the z-score for the Phillies with 93 wins

b) Find and interpret the z-score for the New York Mets with 70 wins

**Example #5**

The single-season home run record for MLB has been set just three times since Babe Ruth hit 60 home runs in 1927. Roger Maris hit 61 in 1961, Mark McGwire hit 70 in 1998, and Barry Bonds hit 73 in 2001. In an absolute sense, Barry Bonds had the best performance of these four players, since he hit the most home runs in a single season. However, in a relative sense, this may not be true. Baseball historians suggest that hitting a home run has been easier in some eras than others. This is due to many factors, including quality of batters, quality of pitchers, hardness of the baseball, dimensions of ballparks, and possible use of performance-enhancing drugs. The make a fair comparison, we should see how these performances rate relative to those of other hitters during the same year.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Player** | **HR** | **Mean** | **SD** |
| 1927 | Babe Ruth | 60 | 7.2 | 9.7 |
| 1961 | Roger Maris | 61 | 18.8 | 13.4 |
| 1998 | Mark McGuire | 70 | 20.7 | 12.7 |
| 2001 | Barry Bonds | 73 | 21.4 | 13.2 |

1. Compute the z-scores for each performance. Which player had the most outstanding performance relative to his peers?

**Transforming Data:**

1. Adding or subtracting the same number *a*
   1. Adds *a* to measures of mean, median, quartiles, percentiles
   2. Does not change range, IQR, or standard deviation
   3. Does not change the shape of the distribution
2. Multiplying or dividing by the same number *b*
   1. Multiplies mean, median, quartiles, and percentiles, by *b*
   2. Multiplies range, IQR, and standard deviation by |*b*|
   3. Does not change the shape of the distribution
3. Multiplying or dividing by a *variable* changes the shape of the distribution (Chapter 8)

**Example #8**

In 2010, taxi cabs in NYC charged an initial fee of $2.50 plus $2 per mile. In equation form,

*fare = 2.50 + 2(miles).* At the end of the month, a businessman collects all his taxi cab receipts and calculates some numerical summaries. The mean fare he paid was $15.45, with a standard deviation of $10.20

1. What are the mean and standard deviation of the lengths of his cab rides in miles?

**Density Curves:**

1. An idealized description of the overall pattern of a distribution that “smooths” out the irregularities in the actual data.
2. Always remains above the horizontal axis
3. Has total area 1 underneath it
4. An area under a density curve gives the proportion of observations that fall in a range of values.
5. The curve is an approximation that is easy to use and accurate enough for practical use

****

** - mean**

** - standard deviation**

**\* and  are used to distinguish them from the  and the  of the *actual* data**