Math 1000
Mathematical Literacy in Today’s World

Lecture 4
Winter 2010
Can I take this course?

- Placement by Exam:
  If you took the Mathematics Placement Exam after January 2009 and received placement into MAT 1000 or 1050.

- Placement by Course:
  If you took MAT 0993 during Winter 2009, Spring/Summer 2009 or Fall 2009 and received a grade of C- or better.
Office Hours for Lecturer

- Office in 1123 FAB
- Monday 1:00-2:00
- Tuesday 1:00-2:00
- Friday 11:45-12:40
- And by appointment
Problem of the Week

- Problems Tuesday of each week
- Turn in to Math Department
- Best three over the whole term get prizes
- Your group would be Group B
- Website at
  http://www.clas.wayne.edu/unit-inner.asp?WebPageID=1900
Green Data

Median = 37
Q1 = 32.3
Q3 = 41.5
IQR = Q3 - Q1 = 9.0
One Criterion for Outliers

- Use the IQR (Exercise 27)
- A high outlier is one which is 1.5 times IQR above the Third Quartile
- A low outlier is one which is 1.5 times IQR below the First Quartile
Green Data

\[
\{26, 27, 28, 28, 31, 32, 33, 34, 34, 35, 36, 37, 37, 38, 39, 40, 40, 41, 42, 43, 44, 45, 45, 46\}
\]

\[Q_1 = 32.5 \quad \text{IQR} = 9.0\]

So low outliers are below

\[32.5 - (9.0)(1.5)\]

\[= 32.5 - 13.5 = 19.0\]
Green Data

26 27 28 28 31 32 33 34 34 35
36 37 37 37 38 39 40 40 41 42
43 44 45 45 46

Q3 = 41.5  IQR = 9.0
So high outliers are above
41.5 + (9.0)(1.5)
= 41.5 + 13.5 = 55.0
About the Standard Deviation

- $s$ measures the spread away from the mean (not median)
- $s = 0$ only when all data values are the same (the ultimately narrow spread)
- $s$ has the same units as the data values (we squared, then took the square root)
- $s$ is strongly affected by outliers
# Review Computation

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Deviations</th>
<th>Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>5</td>
<td>-1.6</td>
<td>2.56</td>
</tr>
<tr>
<td>X2</td>
<td>6</td>
<td>-0.6</td>
<td>0.36</td>
</tr>
<tr>
<td>X3</td>
<td>6</td>
<td>-0.6</td>
<td>0.36</td>
</tr>
<tr>
<td>X4</td>
<td>7</td>
<td>+0.4</td>
<td>0.16</td>
</tr>
<tr>
<td>X5</td>
<td>9</td>
<td>+2.4</td>
<td>5.76</td>
</tr>
<tr>
<td>Sums</td>
<td>33</td>
<td>0</td>
<td>9.40</td>
</tr>
<tr>
<td>Mean</td>
<td>6.6</td>
<td>s = 1.53</td>
<td></td>
</tr>
</tbody>
</table>
Formula for Variance

If the data values are X1, X2, X3, … Xn
Then the mean is X\text{bar} = \frac{X1+X2+…+Xn}{n}
And the variance is
\[ s^2 = \frac{(X1-X\text{bar})^2+(X2-X\text{bar})^2+…+(Xn-X\text{bar})^2}{n-1} \]
Finally, the standard deviation \ s \ is the square root of the variance.
What good is it?

- We will see that, for data which is approximately symmetric without outliers, the standard deviation gives us some numerical information about the proportion of the data which is not far from the mean.
- In particular, approximately 68% of the data will be within one standard deviation from the mean.
- Other percentages apply to two or three standard deviations.
Green Data

\[
\text{Mean} = 36.72 \\
S = 5.91
\]
Book’s Website Apps

You can try the Applet Exercises at the end of the Chapter

See www.whfreeman.com/fapp8e
Try the Mean and Median applet
The Normal Distribution

The following slides attempt to make more clear the idea that the area under a histogram or line graph represents a certain percentage of the data values.

And to describe the idealized form that many distributions take in many real-life situations.
Green data Histogram

Histogram

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>22-26</td>
<td>2</td>
</tr>
<tr>
<td>26-30</td>
<td>6</td>
</tr>
<tr>
<td>30-34</td>
<td>4</td>
</tr>
<tr>
<td>34-38</td>
<td>8</td>
</tr>
<tr>
<td>38-42</td>
<td>6</td>
</tr>
<tr>
<td>42-46</td>
<td>2</td>
</tr>
</tbody>
</table>
Insert Percentages

Histogram

- <26: 4%
- 26-30: 12%
- 30-34: 20%
- 34-38: 24%
- 38-42: 20%
- 42-46: 20%
Percentage replaces Number

Percentages

- <26: 4%
- 26-30: 12%
- 30-34: 20%
- 34-38: 24%
- 38-42: 20%
- 42-46: 20%
Histogram of Percentages

Percentages

<26 26-30 30-34 34-38 38-42 42-46
Add Lines to Histogram

Line Chart Percentages

<22 22-26 26-30 30-34 34-38 38-42 42-46 >46
Change to Fraction Histogram

Line Chart Fractions
Line Chart for Green Data
Area under Line Chart

Total Area = 1 = 100%
Partial Area under Line Chart

Area = .12
Estimates for parts of classes

Area = (0.5)(0.14)

= 0.07 (= 7%)
Area above/below Median

Median = 37

Area = 0.5

Area = 0.5
Idealized version of Line Chart, with same mean and Std. Dev.
Mean (36.72) at Peak
Width to curvature change
Width to curvature change is $s$

$S = \text{standard deviation}$
One Standard Deviation includes 68% of the Area/Data
34% on each Side of Mean

36.72 - 5.91 = 31.81
36.72
42.63 = 36.72 + 5.91

0.4
0.3
0.2
0.1
0
The “Normal” Distribution

- A bell-shaped curve
- Perfectly symmetric
- Peak at the Mean = Median
- Total Area underneath = 1 (or 100%)
- Spread described by the Standard Deviation
Normal Distribution depends only on Mean and Std. Dev.
Quartiles on Normal Distribution

- Median same as Mean (symmetric)
- Median-to-Q3 should include 25% of data
- So not as far as standard deviation (mean-to-mean+s includes 34% of data)
- First quartile is at 0.67 standard deviations below the mean
- Third quartile is at 0.67 standard deviations above the mean.
Computing Quartiles on Normal Distribution

- For Green Data, Mean = 36.72 and s = 5.91.
- So 0.67 standard deviations is $(0.67)(5.91) = 3.96$
- First quartile is at $36.72 - 3.96 = Q1 = 32.76$
- Third quartile is at $36.72 + 3.96 = Q3 = 40.68$
Quartiles on the Normal Dist’n

36.72 − 3.96 = 32.76

36.72

40.68 = 36.72 + 3.96
Quartiles on the Normal Dist’n

Interquartile Range = IQR = 2(3.96) = 2(0.67)s = 7.92
One Standard Deviation includes 68% of the Area/Data

\[
36.72 - 5.91 = 31.81 \quad 36.72 \quad 42.63 = 36.72 + 5.91
\]
Two Standard Deviations include 95% of the Area/Data

36.72 – 11.82 = 25.90
36.72
48.54 = 36.72 + 11.82

95%
How much is this Area?

\[
36.72 - 11.82 = 25.90 \quad 36.72 \quad 48.54 = 36.72 + 11.82
\]
How much is this Area?

\[ 36.72 - 11.82 = 25.90 \]
\[ 36.72 \]
\[ 48.54 = 36.72 + 11.82 \]

95%
How much is this Area?

36.72 – 11.82 = 25.90  36.72  48.54 = 36.72 + 11.82

2.5%  95%  2.5%
Three Standard Deviations include 99.7% of the Area/Data

\[ 36.72 - 17.73 = 18.99 \]

\[ 36.72 + 17.73 = 54.45 \]
But still more Area “out there”
Spread of the Normal Distribution

- Area within 1 Std. Dev. = 0.68
- Area within 2 Std. Dev. = 0.95
- Area within 3 Std. Dev. = 0.997
- Area within 0.67 Std. Dev. = 0.5
- Other areas available in a table
Another Applet

www.whfreeman.com/fapp8e

Try the “Normal Curve” applet
Try an Exercise

Suppose that test scores on an exam are normally distributed with mean 80 and standard deviation 7. What percentage of test scores would be below 66?
Start of Exercise Answer

Suppose that test scores on an exam are normally distributed with mean 80 and standard deviation 7. What percentage of test scores would be below 66?

That would mean that 68% of the scores would be between $80-7=73$ and $80+7=87$.

And 95% of the scores would be between $80-14=66$ and $80+14=94$. 
Remember this picture?

\[ 36.72 - 11.82 = 25.90 \quad 36.72 \quad 48.54 = 36.72 + 11.82 \]
Exercise Answer

Suppose that test scores on an exam are normally distributed with mean 80 and standard deviation 7. What percentage of test scores would be below 66?

Since 95% of the scores would be between $80-14=66$ and $80+14=94$, 2.5% would be below 66 (and 2.5% above 94).
Exercise Discussion

Suppose that test scores on an exam are normally distributed with mean $\mu$ and standard deviation $\sigma$.

- Do you think the mean would be more than 80 or less?
- Do you think the standard deviation would be more than 7 or less?
The End for Today