A DEEPER DIVE INTO NORMALITY

Bill Navidi  
Colorado  
School of Mines

Barry Monk  
Middle Georgia  
State University

Don Brown  
Middle Georgia  
State University

BEGINNINGS
Mathematician and satirical writer John Arbuthnot used Christening records in London to analyze the number of male and female births from 1629 to 1710.

<table>
<thead>
<tr>
<th>Year</th>
<th>Male</th>
<th>Female</th>
<th>Year</th>
<th>Male</th>
<th>Female</th>
<th>Year</th>
<th>Male</th>
<th>Female</th>
<th>Year</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1629</td>
<td>5218</td>
<td>4683</td>
<td>1630</td>
<td>4858</td>
<td>4457</td>
<td>1631</td>
<td>4422</td>
<td>4102</td>
<td>1632</td>
<td>4994</td>
<td>4590</td>
</tr>
<tr>
<td>1650</td>
<td>2890</td>
<td>2722</td>
<td>1671</td>
<td>6449</td>
<td>6061</td>
<td>1692</td>
<td>7602</td>
<td>7316</td>
<td>1693</td>
<td>7676</td>
<td>7483</td>
</tr>
<tr>
<td>1633</td>
<td>5158</td>
<td>4839</td>
<td>1654</td>
<td>3441</td>
<td>3179</td>
<td>1672</td>
<td>6443</td>
<td>6120</td>
<td>1694</td>
<td>6985</td>
<td>6647</td>
</tr>
<tr>
<td>1655</td>
<td>4820</td>
<td>4655</td>
<td>1673</td>
<td>6552</td>
<td>5847</td>
<td>1695</td>
<td>7263</td>
<td>7133</td>
<td>1696</td>
<td>7612</td>
<td>7229</td>
</tr>
<tr>
<td>1634</td>
<td>5035</td>
<td>4820</td>
<td>1656</td>
<td>3668</td>
<td>3382</td>
<td>1674</td>
<td>6423</td>
<td>6203</td>
<td>1697</td>
<td>8062</td>
<td>7767</td>
</tr>
<tr>
<td>1635</td>
<td>5106</td>
<td>4928</td>
<td>1657</td>
<td>3668</td>
<td>3289</td>
<td>1675</td>
<td>6528</td>
<td>6298</td>
<td>1698</td>
<td>8426</td>
<td>7626</td>
</tr>
<tr>
<td>1636</td>
<td>4917</td>
<td>4605</td>
<td>1658</td>
<td>3668</td>
<td>3396</td>
<td>1676</td>
<td>6588</td>
<td>6439</td>
<td>1699</td>
<td>7911</td>
<td>7452</td>
</tr>
<tr>
<td>1637</td>
<td>4703</td>
<td>4457</td>
<td>1659</td>
<td>3668</td>
<td>3517</td>
<td>1677</td>
<td>6713</td>
<td>6290</td>
<td>1700</td>
<td>7378</td>
<td>7061</td>
</tr>
<tr>
<td>1638</td>
<td>4952</td>
<td>4520</td>
<td>1660</td>
<td>3668</td>
<td>3209</td>
<td>1678</td>
<td>6713</td>
<td>6290</td>
<td>1701</td>
<td>8102</td>
<td>7514</td>
</tr>
<tr>
<td>1639</td>
<td>4784</td>
<td>4520</td>
<td>1661</td>
<td>3668</td>
<td>3174</td>
<td>1679</td>
<td>6713</td>
<td>6290</td>
<td>1702</td>
<td>8031</td>
<td>7666</td>
</tr>
<tr>
<td>1640</td>
<td>5518</td>
<td>5323</td>
<td>1662</td>
<td>5216</td>
<td>4803</td>
<td>1680</td>
<td>6713</td>
<td>6290</td>
<td>1703</td>
<td>7765</td>
<td>7683</td>
</tr>
<tr>
<td>1641</td>
<td>5470</td>
<td>5200</td>
<td>1663</td>
<td>5411</td>
<td>4881</td>
<td>1681</td>
<td>6713</td>
<td>6290</td>
<td>1704</td>
<td>6113</td>
<td>5738</td>
</tr>
<tr>
<td>1642</td>
<td>5460</td>
<td>5190</td>
<td>1664</td>
<td>5411</td>
<td>4881</td>
<td>1682</td>
<td>6713</td>
<td>6290</td>
<td>1705</td>
<td>8366</td>
<td>7779</td>
</tr>
<tr>
<td>1643</td>
<td>4793</td>
<td>4617</td>
<td>1665</td>
<td>5411</td>
<td>4881</td>
<td>1683</td>
<td>6713</td>
<td>6290</td>
<td>1706</td>
<td>7952</td>
<td>7417</td>
</tr>
<tr>
<td>1644</td>
<td>4107</td>
<td>3997</td>
<td>1666</td>
<td>5114</td>
<td>4858</td>
<td>1684</td>
<td>6713</td>
<td>6290</td>
<td>1707</td>
<td>8379</td>
<td>7687</td>
</tr>
<tr>
<td>1645</td>
<td>4047</td>
<td>3919</td>
<td>1667</td>
<td>4678</td>
<td>4319</td>
<td>1685</td>
<td>6713</td>
<td>6290</td>
<td>1708</td>
<td>8239</td>
<td>7623</td>
</tr>
<tr>
<td>1646</td>
<td>3768</td>
<td>3395</td>
<td>1668</td>
<td>5166</td>
<td>5322</td>
<td>1686</td>
<td>6713</td>
<td>6290</td>
<td>1709</td>
<td>7840</td>
<td>7380</td>
</tr>
<tr>
<td>1647</td>
<td>3796</td>
<td>3536</td>
<td>1669</td>
<td>6068</td>
<td>5660</td>
<td>1687</td>
<td>7604</td>
<td>7167</td>
<td>1710</td>
<td>7649</td>
<td>7288</td>
</tr>
<tr>
<td>1648</td>
<td>3363</td>
<td>3181</td>
<td>1670</td>
<td>6506</td>
<td>5829</td>
<td>1688</td>
<td>9709</td>
<td>7302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1649</td>
<td>3079</td>
<td>2746</td>
<td>1671</td>
<td>6278</td>
<td>5719</td>
<td>1689</td>
<td>7662</td>
<td>7392</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Noting in almost all years, there were more male births than female births, Arbuthnot calculated the probability that there is an equal chance of male and female births over the 82 years to be

\[
\frac{1}{4 836 000 000 000 000 000 000 000 000 000 000 000}\]

He proposed that if it were extended for “not only 82 years, but for Ages and Ages, and not only at London, but all over the World” then the probability would be “near an infinitely small Quantity, at least less than any assignable fraction.”
In 1712, Dutch philosopher and mathematician Willem ‘sGravesande compared the notion that male and female births are equally likely against the actual births over the 82 years.

Male births:  
Low of \( \frac{7765}{15448} = 0.5027 \) in 1703

High of \( \frac{4748}{8855} = 0.5362 \) in 1661

‘sGravesande multiplied these ratios by the average number of births to obtain the bounds of 5745 and 6128 on the number of male births in each year.

He went on to estimate the probability of the surplus of male births in any year given that male and female births are equally likely.

\[
P \left( 5745 \leq x \leq 6128 \mid p = \frac{1}{2} \right) \approx 0.292
\]

The probability of this surplus over the 82 years was calculated as 0.292\(^{82} \approx 1.45 \times 10^{-44} \).
Arbuthnot and ‘sGravesande attributed the surplus of male births to \textit{divine intervention}.

‘sGravesande’s computation involves the binomial calculation
\[ \sum_{x=5745}^{6128} \binom{11,429}{x} \left(\frac{1}{2}\right)^{11,429} \]

which was an onerous task at the beginning of the 18\textsuperscript{th} century.

As early as 1721, Abraham de Moivre began searching for easier approximation methods.
By 1733, de Moivre had shown that
\[
\left( \frac{n}{n/2 + d} \right)^{1/2} \approx \frac{2}{\sqrt{2\pi n}} e^{-\frac{2d^2}{n}}
\]
allowing for the easy approximation of binomial computations by an exponential function.

De Moivre’s discovery led to approximations of binomial random variables using what is now known as the normal distribution.

\[
\sum_{k} \binom{j}{k} p^k (1 - p)^{n-k} \approx \Phi(j) - \Phi(i)
\]

where \( \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt \) with \( \mu = np \) and \( \sigma = \sqrt{np(1 - p)} \).
EXPLORING THE NORMAL CURVE

The mean ($\mu$) and standard deviation ($\sigma$) affect the shape of the normal curve.

[Graph showing three normal curves with different means and standard deviations]
DEMONSTRATING
THE CENTRAL LIMIT THEOREM

THE CENTRAL LIMIT THEOREM

Let $\bar{x}$ be the mean of a large simple random sample from a population with mean $\mu$ and standard deviation $\sigma$. Then $\bar{x}$ has an approximately normal distribution with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

How large is large?
**SYMMETRIC POPULATIONS**

For symmetric populations, a *small sample* may suffice.

- **Probability histogram** for a symmetric population
- **Probability histogram for the sampling distribution of $\bar{x}$ for samples of size 3**

**SKEWED POPULATIONS**

For skewed populations, a *larger sample* may be needed.

- **Probability histogram for a skewed population**
- **Samples of size 3**
- **Samples of size 10**
- **Samples of size 30**
The mean concentration of ammonia in water wells in the state of Iowa is 0.71 milligrams per liter, with a standard deviation of 1.09 milligrams per liter. Is it possible that these concentrations are normally distributed?

If the distribution were normal, approximately 26% of the concentrations would be less than 0. In fact, 0% of the concentrations are less than 0. Therefore the concentrations are not normally distributed.
Heights of adult women in the U.S. are approximately normally distributed with mean 64 inches and standard deviation 4 inches. If five women are chosen at random, what is the probability that exactly two of them are more than 68 inches tall?

The number of women who are more than 68 inches tall follows the binomial distribution with \( n = 5 \) and \( p = 0.1587 \). Therefore

\[
P(2) = \frac{5!}{2!3!} (0.1587)^2 (0.8413)^3 = 0.1500
\]

Speeds of automobiles on a certain stretch of freeway are normally distributed with mean 65 mph. 20% of the cars are traveling between 55 and 65 mph. What percentage of the cars are traveling faster than 75 mph?

Given \( P(55 < X < 65) = 0.2 \)

Compute \( P(X < 55) = 0.3 \)

By symmetry, \( P(X > 75) = 0.3 \)
FIND A NORMAL PROBABILITY
WITHOUT BEING GIVEN THE VARIANCE (PART 2)

Speeds of automobiles on a certain stretch of freeway are normally distributed with mean 65 mph. 20% of the cars are traveling between 55 and 65 mph. What percentage of the cars are traveling faster than 80 mph?

Compute z-score for 55:
\[ z_{55} = -0.52 \]
Compute \( \sigma \): 55 = 65 – 0.52\( \sigma \):
\[ \sigma = (65-55)/0.52 = 19.23 \]
Compute z-score for 80:
\[ z_{80} = (80-65)/19.23 = 0.78 \]
\[ P(X > 80) = 0.2177 \]

FIND A NORMAL PROBABILITY
WITHOUT BEING GIVEN THE MEAN OR VARIANCE

Scores on an exam were normally distributed. Jack got a score of 64 and Jill got a score of 81. Jack’s score was on the 10th percentile and Jill’s score was on the 80th percentile. Tom got a score of 70. What percentile was Tom’s score on?

Find z-scores for 64 and 81:
\[ z_{64} = -1.28, \quad z_{81} = 0.84 \]
Then find the mean and variance:
\[ 64 = \mu - 1.28\sigma, \quad 81 = \mu + 0.84\sigma \]
\[ \mu = 74.26, \quad \sigma = 8.02 \]
Find the z-score for 70:
\[ z = (70-74.26)/8.02 = -0.53 \]
A score of 70 is on the 30th percentile.
ALTERNATIVES TO THE NORMAL DISTRIBUTION
A DEEPER DIVE INTO NORMALITY

Thank you