Section 4.4: The Normal Distribution

1 The Normal Distribution and the 68-95-99.7 Rule

The **normal curve** (also known as the bell curve) is the most important continuous probability distribution. The normal curve is completely determined by two parameters: mean and standard deviation. The normal curve is symmetric about the mean which is also the median and the mode. Most data is clumped in close to the mean. The notation $N(\mu, \sigma)$ states that the data has a normal distribution with mean μ and a standard deviation σ. We use the notation mean μ and standard deviation σ to indicate that these are defining parameters for a statistical distribution rather than statistical values we computed from a sample. For example, $N(35, 2.3)$ indicates a normal distribution with a mean of 35 and a standard deviation of 2.3.

The pdf for the normal curve is

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

The probability that an observation falls in the interval from a to b in the normal curve is

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.$$

The function inside the integral does not have an anti-derivative. Numerical methods must be used to determine the area under the curve. In practice, this is done with technology. Without access to proper technology, tables of normal curve values are utilized. In reference books, it is typical to present a table for computing probabilities in the standard normal curve where $\mu = 0$ and $\sigma = 1$. For the standard normal curve, the cumulative distribution function is denoted $\Phi(z)$ and tables of values exist as well.

Without resorting to integration, a useful tool for computing probabilities in the normal curve is to convert to standardized units is the z-score

$$z = \frac{x - \mu}{\sigma}.$$

When using a table from the book to compute normal curve values, z-scores are needed. This is not the case when using technology.

2 The Normal Curve and the TI 83/84

The TI 83/84 series of calculators has two basic types of normal curve commands;

1. Find the percentage p of data between two values z_1 (lower bound) and z_2 (upper bound) in the normal distribution defined by μ and σ:
   normalcdf(z_1 **(lower bound)**, z_2 **(upper bound)**, μ, σ)

2. Find the value z such that p percent of data falls below (to the left of) z in the normal distribution defined by μ and σ:
   invNorm(p, μ, σ)**.** Recall that this is the p^{th} percentile.

Definition 1

3 Applications of the Normal Distribution

Problem 1 *Scores on a particular test follow a normal distribution with a mean of 75.6 and a standard deviation of 7.8.*
i. What percentage of students scored above a 90?

ii. Find the score separating the top 15% of the scores from the bottom 85% of scores.

iii. Between what two values do the central 40% of scores fall?

Problem 2 *On a particular track team mile running times follow a normal distribution with a mean of 8.5 minutes and a standard deviation of 1.2 minutes.*
i. What percentage of the track team runs a mile in between 8 and 9.3 minutes?

ii. What time separates the fastest 10% of the runners from the rest of the team?

iii. The slowest 20% of all runners will be cut from the team. How fast of a mile does one need to run in order to stay on the team?

Definition 2 *The values z_α on the standard normal curve is one such that α of the data appears to the right. Note that $z_\alpha = P_{100-\alpha}$.*

Example 1 *Determine $z_{.05}$. If 5% of the data appears to the right then we wish to compute P_{95}. $P_{95} = 1.645$.*

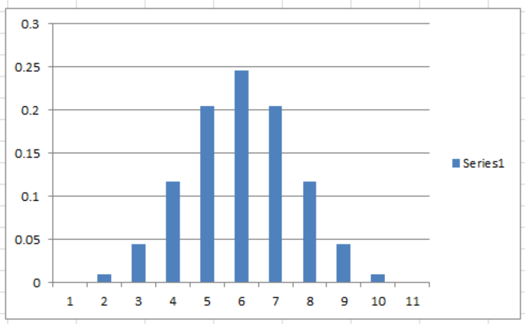
Remark 3 *This concept and notation (especially in the form $z_{\frac{\alpha}{2}}$) appears again in confidence intervals and hypothesis testing.*

4 The Normal Distribution and Discrete Populations

When plotting histograms of discrete distributions a very familiar shape can appear.

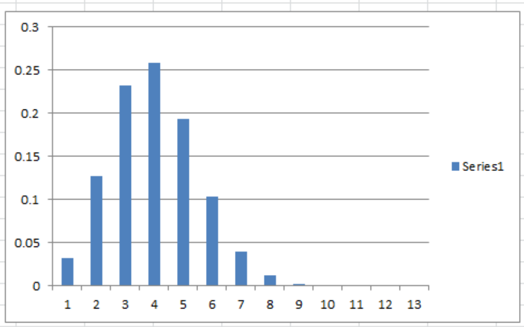
10 flips of a coin

number of heads k	0	1	2	3	4	5	6	7	8	9	10
probability of k heac	0.000977	0.009766	0.043945	0.117188	0.205078	0.246094	0.205078	0.117188	0.043945	0.009766	0.000977



12 multiple choice questions with 4 possible responses and random guessing

number correct k	0	1	2	3	4	5	6	7	8	9	10	11	12
probability of k correct	0.031676	0.126705	0.232293	0.258104	0.193578	0.103241	0.040149	0.011471	0.00239	0.000354	3.54E-05	2.15E-06	5.96E-08



Both of the above histograms emanate from binomial distributions. Under certain conditions a binomial probability distribution is approximately normal.

Theorem 4 *A binomial distribution with n fixed trials and constant probability of success p has an approximately normal distribution, $N(np, \sqrt{np(1-p)})$ if*

$np \geq 10$ and $n(1-p) \geq 10$. A continuity correction of $\frac{1}{2}$ is applied to each boundary of an interval to facilitate a better approximation.

Example 2 *Should the normal curve approximation be used when counting the number of heads that appear when flipping a fair coin 10 times? No, since $np = 10 * .5 = 5.0 \ngeqq 10$.*

Example 3 *Should the normal curve approximation be used when counting the number of correct answers with random guessing on a 50 question multiple choice test where every question has four possible responses? In this case $np = 50 * .25 = 12.5 \geq 10$ and $n(1-p) = 50 * .75 = 37.5 \geq 10$. So, yes it is appropriate.*

Example 4 *A student resorts to guessing on a 50 question multiple choice test where every question has four possible responses. What is the probability that the student answers no more than 10 questions correctly?*
We've already shown that it is appropriate to use the normal distribution. Our mean is $\mu = np = 50 * .25 = 12.5$ and standard deviation is $\sigma = \sqrt{np(1-p)} = \sqrt{50 * .25 * .75} = 3.0619$. With a continuity correction, we want to find $P(z \leq 10.5)$. Using the normal curve we get the value 0.257. The precise binomial probability is 0.2622.

Example 5 *Find the probability the same students earns at least a 50 on the test.*