CHAPTER 4

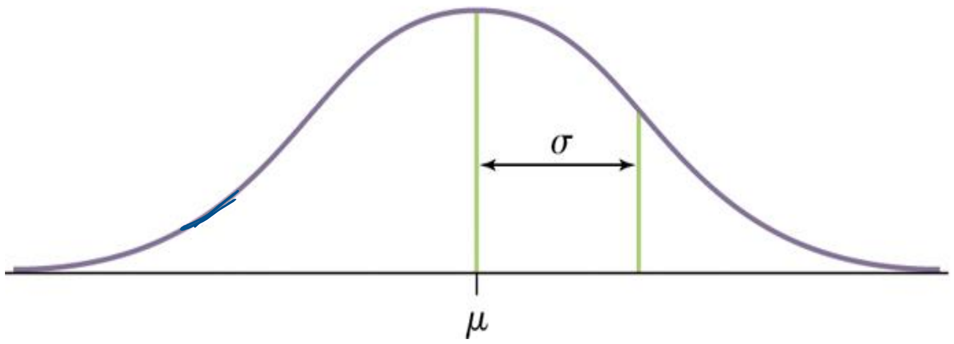
Normal Distribution

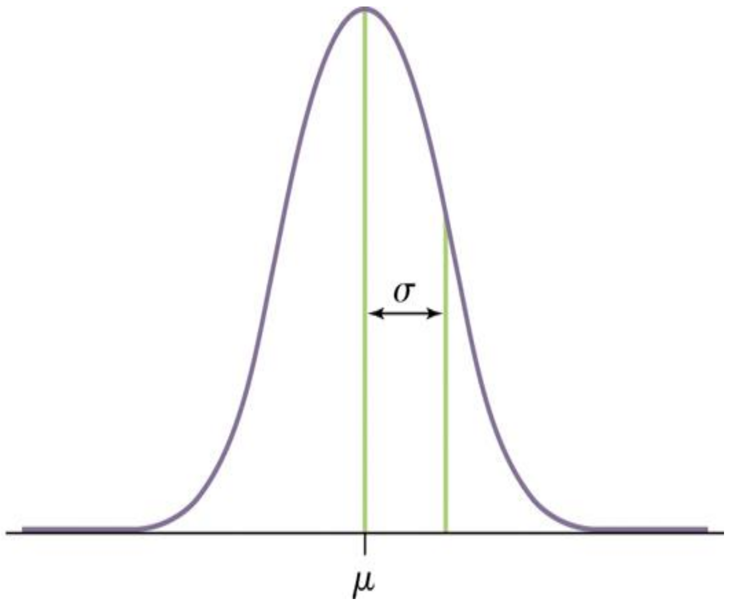
4.1 Normal Distribution

Normal Distribution Model

One particularly important class of density curves is the class of Normal curves, which describe Normal distributions.

- All Normal curves are symmetric, single-peaked, and bell-shaped.
- A specific Normal curve is described by giving its mean μ and standard deviation σ.





Normal Distributions

A Normal distribution is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers: its mean μ and standard deviation σ.

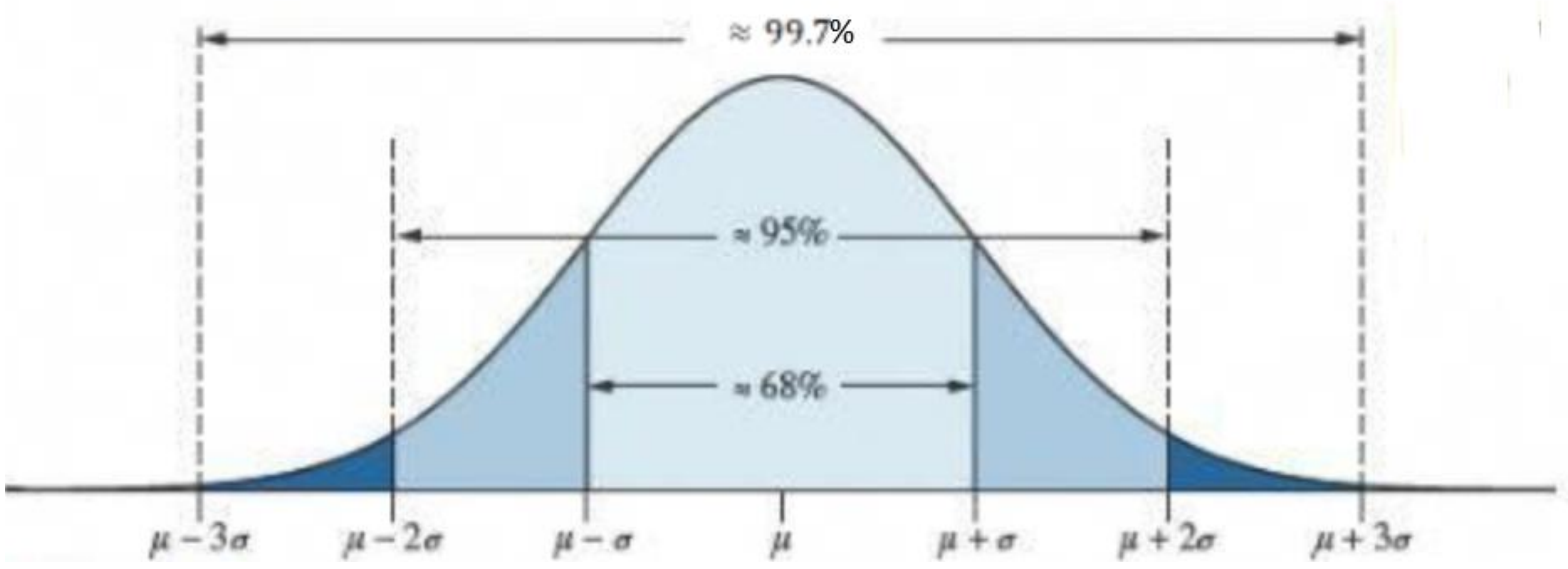
- The mean (μ) of a Normal distribution is the center of the symmetric Normal curve.
- The standard deviation (σ) is the distance from the center to the change-of-curvature points on either side. Standard deviation controls the spread of a Normal curve.
- We abbreviate the Normal distribution with mean μ and standard deviation σ as $N(\mu,\sigma)$.

The 68-95-99.7 Rule

In the Normal distribution with mean μ and standard deviation σ:

- Approximately **68%** of the observations fall within σ of μ.
- Approximately **95%** of the observations fall within 2σ of μ.
- Approximately **99.7%** of the observations fall within 3σ of μ.

This rule only describes distributions that are exactly Normal.



Example

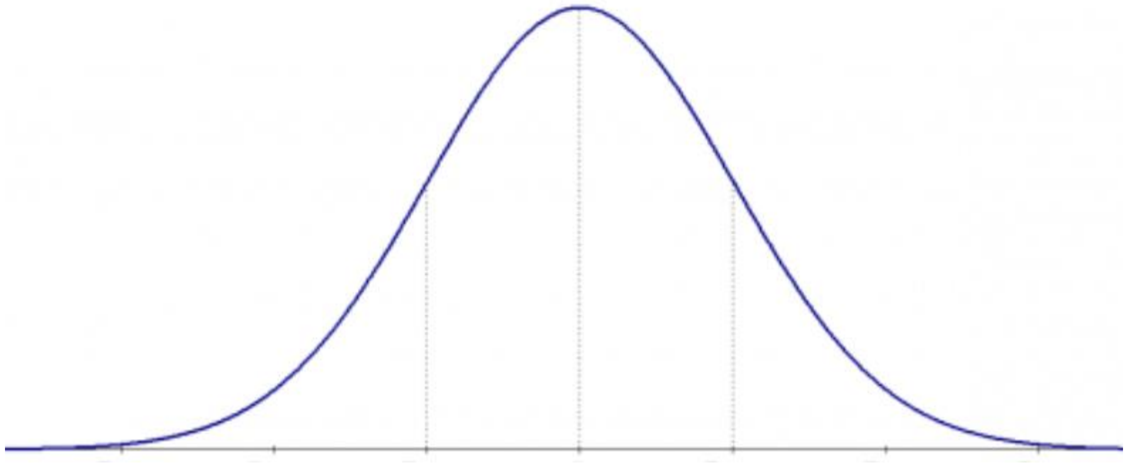
Liquid detergent carton are filled automatically on a production line. Filling weights frequently have a bell-shape distribution. If the mean filling weight is 16 ounces and the standard deviation is .25 ounces.

- What percentage of detergent carton weight between 16 and 16.25 ounces?

- What percentage of detergent carton weight between 15.5 and 16 ounces?

- What percentage of detergent carton weight less than 15.5 ounces?

- What percentage of detergent carton weight between 15.5 and 16.25 ounces?



Standardizing with Z-scores

If a variable x has Normal distribution with mean μ and standard deviation σ, then the standardized value of x, or its Z-score, is

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

A Z-score tells us how many standard deviations the original observation falls away from the mean (and in which direction).

The Standard Normal Distribution

- If a variable x has any Normal distribution $N(\mu, \sigma)$ mean μ and standard deviation σ
- then the standard normal distribution is the Normal distribution $N(0,1)$ with mean 0 and standard deviation 1
- The ability to convert scores in a normal distribution to standard scores has a very practical application. Values from two different distributions can be compared by converting both to standard scores.

Example

Suppose, for example, that ACT scores are normally distributed with a mean of 24 and standard deviation of 3.5 and that SAT scores are normally distributed with a mean of 550 and standard deviation of 80. Suppose that Jack has an ACT score of 30 and Jill has an SAT score of 690. Which student had the better score?

The Standard Normal Table

- Because all Normal distributions are the same when we standardize, we can find areas under any Normal curve from a single table.
- The Standard Normal Table is a table of areas under the standard Normal curve. The table entry for each value *z* is the area under the curve to the left of *z*.

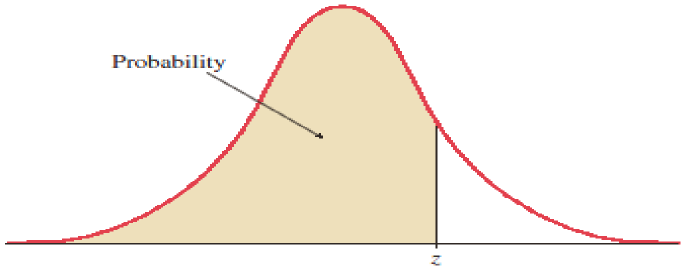


Table entry for z is the area under the standard normal curve to the left of z.

TABLE A Standard normal probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

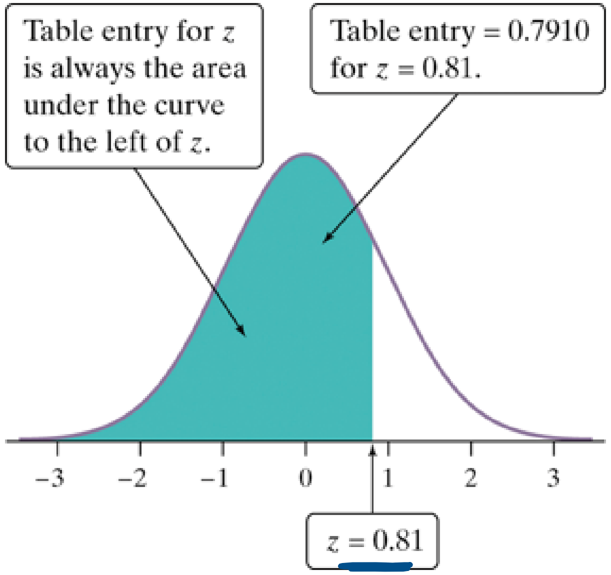
The Standard Normal Table

Suppose we want to find the proportion of observations from the standard Normal distribution that are less than 0.81.

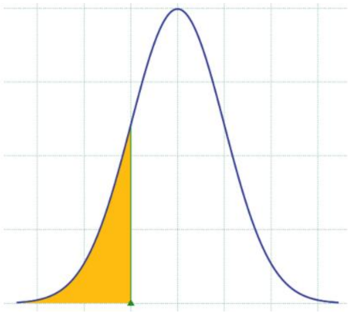
We can use

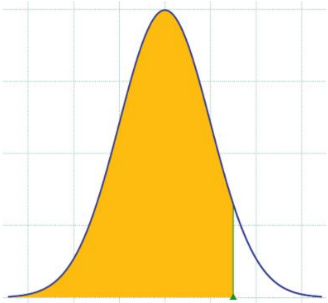
$$P(z < 0.81) = .7910$$

Z	.00	.01	.02
0.7	.7580	[illegible]	.7642
0.8	.7881	.7910	.7939
0.9	.8159	.8186	.8212



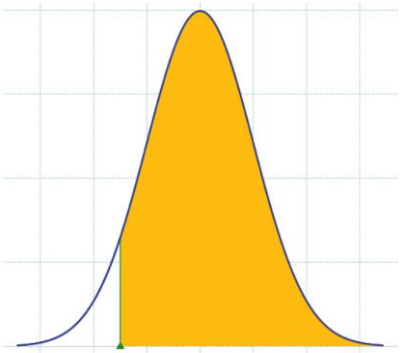
Rules to Obtain Probabilities in the Standard Normal Distribution from Table





To find $P(Z \leq a)$

- If a is positive: Use the table for positive values of z
- If a is negative: Use the table for negative values of z

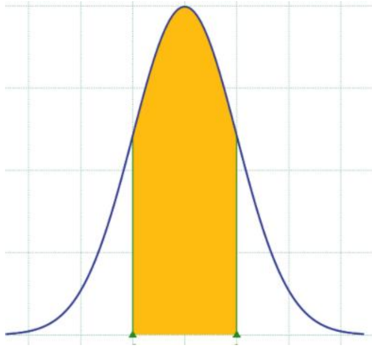


To find $P(Z \geq a)$

- Use the equation $P(Z \geq a)$=1- $P(Z \leq a)$
- Follow the same steps of 1.

To find $P(a \leq Z \leq b)$

- Use the equation $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$
- Follow the same steps of 1.



Examples

Find the following probabilities.

- P(Z $\leq$ 1.64)
- P(Z < -0.44)
- P(Z $\geq$ 0.02)
- P(-1.96 $\leq$ Z $\leq$ 1.96)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Examples:
Finding the Exact Probability Using R

$$pnorm = (x, mean = \ , sd = \ , Lower.tail = \quad)$$

- P(Z $\leq$ 1.64)

- P(Z < -0.44)

- P(Z $\geq$ 0.02)

- P(|z| < 1.96)

- P(|z| > 2)

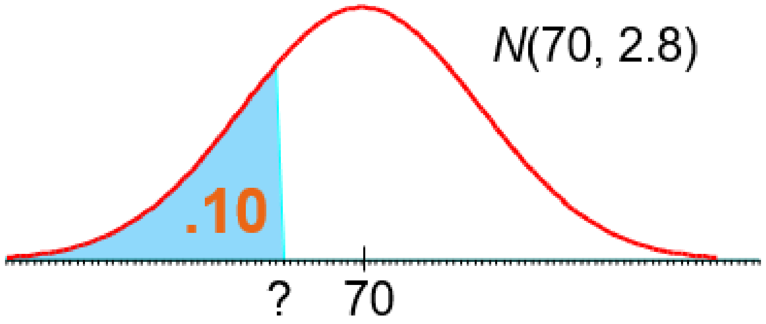
Example

Women's heights are Normally distributed with mean 64.5 and standard deviation 2.5 in. If we pick one woman at random, what is the probability that her height will be between 68 and 70 inches? Because the woman is selected at random, X is a random variable.

Example

According to the Health and Nutrition Examination Study of 1976–1980, the heights (in inches) of adult men aged 18–24 are $N(70, 2.8)$.

If exactly 10% of men aged 18-24 are shorter than a particular man, how tall is he?



z	.07	.08	.09
–1.3	.0853	.0[illegible]8	.0823
–1.2	.1020	.1003	.0985
–1.1	.1210	.1190	.1170

Z = –1.28

4.3 Binomial Distribution

Binomial Distribution

- When the same chance process is repeated several times, we are often interested in whether a particular outcome does or does not happen on each repetition.
- In other words, we are interested in the number of successes in a fixed number of independent trails.
- For example, what is the chance that no more than 4 out of 25 randomly selected parts are defective.
- The probability distribution describing this count is called the **Binomial Distribution**.

The Binomial Setting

A **binomial setting** arises when we perform several independent trials of the same chance process and record the number of times that a particular outcome occurs. The four conditions (called **BINS**) for a binomial setting are:

B	• **B**inary? The possible outcomes of each trial can be classified as "success" or "failure."
I	• **I**ndependent? Trials must be **independent**; that is, knowing the result of one trial must not have any effect on the result of any other trial.
N	• **N**umber? The number of trials n of the chance process must be fixed in advance.
S	• **S**uccess? On each trial, the probability p of success must be the same.

Examples

1. Flipping a coin three times and counting the number of head. Is this Binomial?

2. The number of correct guesses on a multiple-choice exam. Each question has 4 choices, and we have 5 questions total. Is this Binomial?

3. Stand at an intersection for 10 minutes and count the number of cars that turn left. Is this Binomial?

Binomial Distribution

Suppose the probability of a single trial being a success is p. then the probability of observing exactly k successes in n independent trials is given by:

$$\boxed{\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}}$$

Example

5 question on multiple choice exam randomly guessing. Each question has 4 possibilities.

a) What is the probability of getting four correct answers?

b) What is the probability of getting at least 3 correct answers?

Example

Each child of a particular pair of parents has probability 0.25 of having blood type O. Suppose the parents have five children.

a) Find the probability that exactly three of the children have type O blood.

b) Should the parents be surprised if more than three of their children have type O blood?

Example

Suppose 10% pf CDs have defective copy-protection schemes that can harm computers. A music distributor inspects an SRS of 10 CDs from a shipment of 10,000. Let X = number of defective CDs in the SRS of size 10. What is $p(X = 0)$?

Binomial Mean and Standard Deviation

If a count X has the binomial distribution with number of trials n and probability of success p, the *mean* and *standard deviation* of X are:

$$\mu_x = np$$

$$\sigma_x = \sqrt{np(1-p)}$$

Example

If we toss a fair coin 3 times, the number of heads is a random variable that is binomial. Find the mean and the standard deviation of X.

Normal Approximation for Binomial Distribution

Suppose that X has the binomial distribution with n trials and success probability p. When <u>n is large</u>, the distribution of X is approximately Normal with mean and standard deviation

$$\boxed{x \sim N\left(\mu_x = np, \sigma_x = \sqrt{np(1-p)}\right)}$$

As a rule of thumb, we will use the Normal approximation when n is so large that $np \geq 10$ and $n(1 - p) > 10$.

Example

Sample surveys show that fewer people enjoy shopping than in the past. A survey asked a nationwide random sample of 2500 adults if they agreed or disagreed that “I like buying new clothes, but shopping is often frustrating and time-consuming.” Suppose that exactly 60% of all adult U.S. residents would say “Agree” if asked the same question. Let X = the number in the sample who agree. Estimate the probability that 1520 or more of the sample agree.

Example:

The National Vaccine Information Center estimates that 90% of Americans have had chickenpox by the time they reach adulthood.

1. Suppose we take a random sample of 100 American adults. Is the use of the binomial distribution appropriate for calculating the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood? Explain.

2. Calculate the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood.

Example Cont.

3. What is the probability that exactly 3 out of a new sample of 100 American adults have not had chickenpox in their childhood?

4. What is the probability that at least 1 out of 10 randomly sampled American adults have had chickenpox?

5. What is the probability that at most 3 out of 10 randomly sampled American adults have not had chickenpox?

Example Cont.

We now consider a random sample of 120 American adults.

6. How many people in this sample would you expect to have had chickenpox in their childhood? And with what standard deviation?

7. Would you be surprised if there were 105 people who have had chickenpox in their childhood?

8. What is the probability that 105 or fewer people in this sample have had chickenpox in their childhood?