Lecture Two

Section 2.1 – Scatter Diagrams and Correlation

Draw and Interpret Scatter Diagrams

Definition

The ***response variable*** is the variable whose value can be explained by the value of the ***explanatory*** or ***predictor variable***.

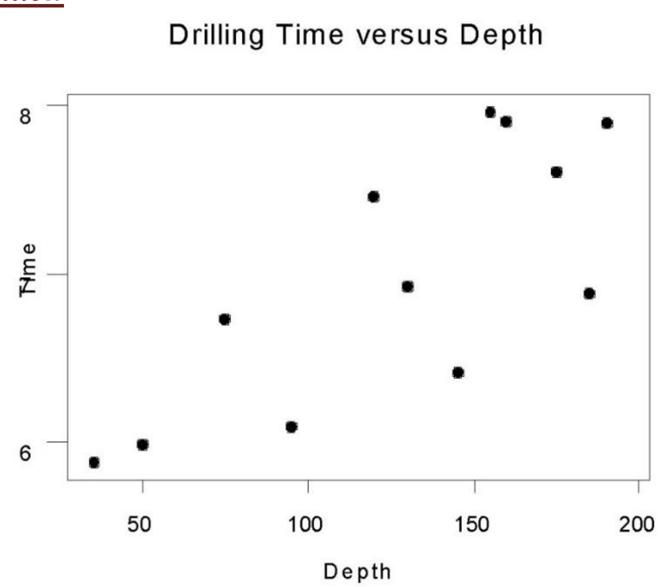
A ***scatter diagram*** is a graph that shows the relationship between two quantitative variables measured on the same individual. Each individual in the data set is represented by a point in the scatter diagram. The explanatory variable is plotted on the horizontal axis, and the response variable is plotted on the vertical axis.

Example

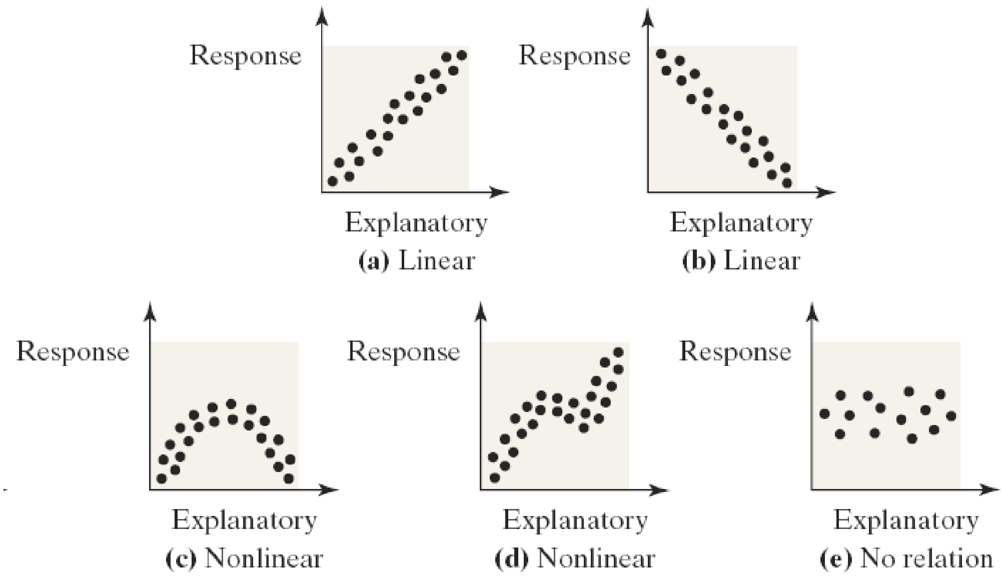
The data shown to the right are based on a study for drilling rock. The researchers wanted to determine whether the time it takes to dry drill a distance of 5 feet in rock increases with the depth at which the drilling begins. So, depth at which drilling begins is the explanatory variable, x, and time (in minutes) to drill five feet is the response variable, y. Draw a scatter diagram of the data.

Depth at Which Drilling Begins x (in feet)	**Time to Drill 5 feet** y (in feet)
35	5.88
50	5.99
75	6.74
95	6.1
120	7.47
130	6.93
145	6.42
155	7.97
160	7.92
175	7.62
185	6.89
190	7.9

Solution



Various Types of Relations in a Scatter Diagram



- Two variables that are linearly related are ***positively associated*** when above-average values of one variable are associated with above-average values of the other variable and below-average values of one variable are associated with below-average values of the other variable. That is, two variables are positively associated if, whenever the value of one variable increases, the value of the other variable also increases.
- Two variables that are linearly related are ***negatively associated*** when above-average values of one variable are associated with below-average values of the other variable. That is, two variables are negatively associated if, whenever the value of one variable increases, the value of the other variable decreases.

Definitions

A ***correlation*** exists between two variables when the values of one are somehow associated with the values of the other in some way.

The ***linear correlation coefficient* (*r*)** or ***Pearson product moment correlation coefficient*** is a measure of the strength and direction of the linear relation between two quantitative variables. The Greek letter ρ (rho) represents the population correlation coefficient, and r represents the sample correlation coefficient. We present only the formula for the sample correlation coefficient

Requirements

1. The sample of paired (x, y) data is a simple random sample of quantitative data.
2. Visual examination of the scatterplot must confirm that the points approximate a straight-line pattern.
3. The outliers must be removed if they are known to be errors. The effects of any other outliers should be considered by calculating r with and without the outliers included.

Notation for the Linear Correlation Coefficient

n number of pairs of sample data

$\sum$ denotes the addition of the items indicated.

$\sum x$ denotes the sum of all x-values.

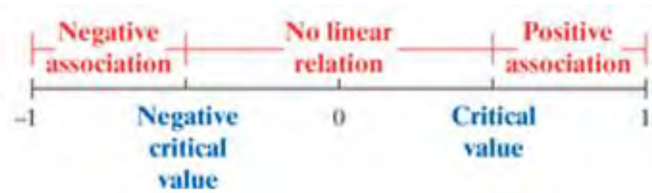
$\sum x^2$ indicates that each x-value should be squared and then those squares added.

$\left(\sum x\right)^2$ indicates that the x-values should be added and then the total squared

$\sum xy$ indicates that each x-value should be first multiplied by its corresponding y-value. After obtaining all such products, find their sum.

r = linear correlation coefficient for ***sample*** data.

ρ = linear correlation coefficient for **population** data.



Formula

The linear correlation coefficient r measures the strength of a linear relationship between the paired values in a sample.

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

- Computer software or calculators can compute r

Sample Linear Correlation Coefficient

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}{n-1}$$

where

$\bar{x}$ is the sample mean of the explanatory variable

s_x is the sample standard deviation of the explanatory variable

$\bar{y}$ is the sample mean of the response variable

s_y is the sample standard deviation of the response variable

n is the number of individuals in the sample

✓ *Know that the methods of this section apply to a linear correlation. If you conclude that there does not appear to be linear correlation, know that it is possible that there might be some other association that is not linear.*

Properties of the Linear Correlation Coefficient *r*

1. The value of r is always between -1 and 1, inclusive. That is $-1 \le r \le 1$
2. If $r = +1$, then a perfect positive linear relation exists between the two variables.
3. If $r = -1$, then a perfect negative linear relation exists between the two variables.
4. The closer r is to $+1$, the stronger is the evidence of positive association between the two variables.
5. The closer r is to -1, the stronger is the evidence of negative association between the two variables.
6. If r is close to 0, then little or no evidence exists of a linear relation between the two variables. So r close to 0 does not imply no relation, just no linear relation.
7. The linear correlation coefficient is a unitless measure of association. So the unit of measure for x and y plays no role in the interpretation of r.
8. The correlation coefficient is not resistant. Therefore, an observation that does not follow the overall pattern of the data could affect the value of the linear correlation coefficient.

(a) Perfect positive linear relation, $r = 1$

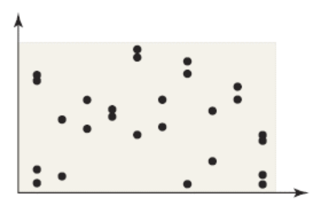
(b) Strong positive linear relation, $r \approx 0.9$

(c) Moderate positive linear relation, $r \approx 0.4$

(d) Perfect negative linear relation, $r = -1$

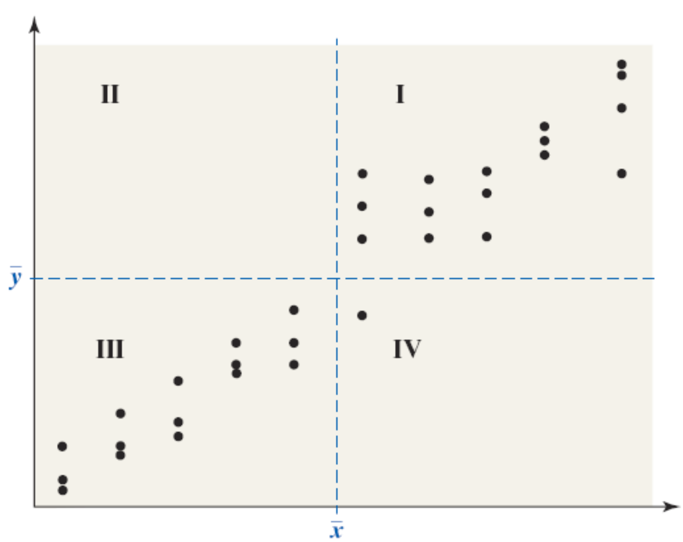
(e) Strong negative linear relation, $r \approx -0.9$

(f) Moderate negative linear relation, $r \approx -0.4$



(g) No linear relation, r close to 0.

(h) No linear relation, r close to 0.



Example

Determine the linear correlation coefficient of the drilling data

Depth at Which Drilling Begins x (in feet)	**Time to Drill 5 feet** y (in feet)
35	5.88
50	5.99
75	6.74
95	6.1
120	7.47
130	6.93
145	6.42
155	7.97
160	7.92
175	7.62
185	6.89
190	7.9

Solution

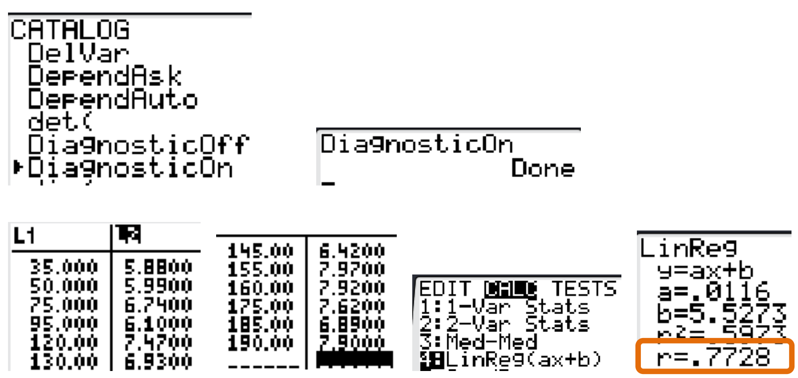
x	y	$\frac{x_i - \bar{x}}{s_x}$	$\frac{y_i - \bar{y}}{s_y}$	
35	5.88	−1.74712	−1.41633	2.474501
50	5.99	−1.45992	−1.27544	1.862051
75	6.74	−0.98126	−0.31486	0.308958
95	6.1	−0.59833	−1.13456	0.678839
120	7.47	−0.11967	0.620111	−0.07421
130	6.93	0.0718	−0.07151	−0.00513
145	6.42	0.358998	−0.72471	−0.26017
155	7.97	0.550463	1.260501	0.693859
160	7.92	0.646196	1.196462	0.77319
175	7.62	0.93394	0.812228	0.758129
185	6.89	1.12486	−0.12274	−0.13807
190	7.9	1.220592	1.170846	1.429126
$\bar{x} = 126.25$	$\bar{y} = 6.9858$			8.50104
$s_x = 52.23$	$s_y = 0.781$			

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}{n-1}$$

$$= \frac{8.501037}{12-1}$$

$$= 0.7728$$

OR



Example

The paired pizza/subway fare costs are shown in the table below. Use computer software with these paired sample values to find the value of the linear correlation coefficient r for the paired sample data.

Table – *Cost of a Slice of Pizza, subway Fare, and the CPI*						
Year	**1960**	**1973**	**1986**	**1995**	**2002**	**2003**
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00
Subway Fare	0.15	0.35	1.00	1.35	1.50	2.00
CPI	30.2	48.3	112.3	162.2	191.9	197.8

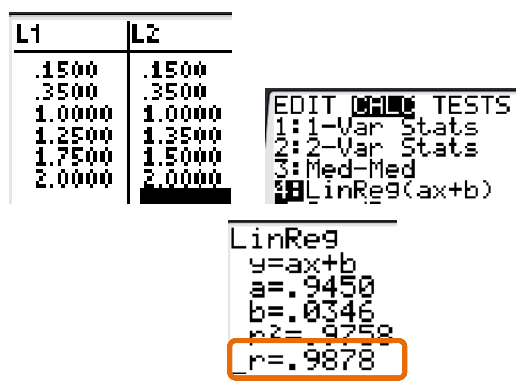
Solution

x (Pizza)	y (Subway)	x^2	y^2	xy
0.15	0.15	0.0225	0.0225	0.0225
0.35	0.35	0.1225	0.1225	0.1225
1.00	1.0	1.0	1.0	1.0
1.25	1.35	1.5625	1.8225	1.6875
1.75	1.50	3.0625	2.250	2.6250
2.0	2.0	4.0000	4.0000	4.0000
$\sum x = 6.50$	$\sum y = 6.35$	$\sum x^2 = 9.77$	$\sum y^2 = 9.2175$	$\sum xy = 9.4575$

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

$$= \frac{6(9.4575) - (6.5)(6.35)}{\sqrt{6(9.77) - (6.5)^2} \cdot \sqrt{6(9.2175) - (6.35)^2}}$$

$$= 0.9878$$



Difference between Correlation and Causation

According to data obtained from the Statistical Abstract of the United States, the correlation between the percentage of the female population with a bachelor's degree and the percentage of births to unmarried mothers since 1990 is 0.940.

Does this mean that a higher percentage of females with bachelor's degrees causes a higher percentage of births to unmarried mothers?

Certainly not! The correlation exists only because both percentages have been increasing since 1990. It is this relation that causes the high correlation. In general, time series data (data collected over time) may have high correlations because each variable is moving in a specific direction over time (both going up or down over time; one increasing, while the other is decreasing over time).

When data are observational, we cannot claim a causal relation exists between two variables. We can only claim causality when the data are collected through a designed experiment.

Another way that two variables can be related even though there is not a causal relation is through a *lurking variable*.

A ***lurking variable*** is related to both the explanatory and response variable.

For ***example***, ice cream sales and crime rates have a very high correlation. Does this mean that local governments should shut down all ice cream shops? No! The lurking variable is temperature. As air temperatures rise, both ice cream sales and crime rates rise.

Example

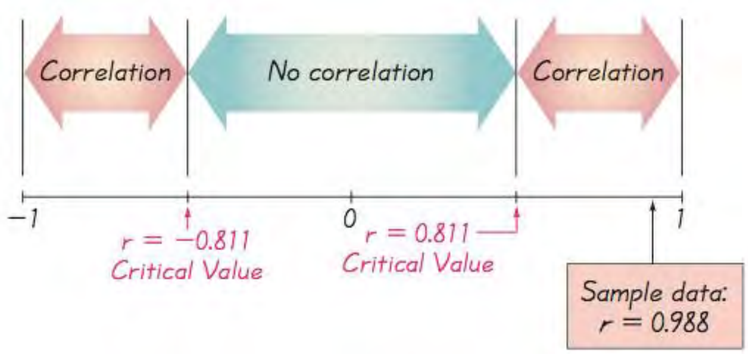
In prospective cohort studies, data are collected on a group of subjects through questionnaires and surveys over time. Therefore, the data are observational. So the researchers cannot claim that increased cola consumption causes a decrease in bone mineral density.
Some lurking variables in the study that could confound the results are:

- body mass index
- height
- smoking
- alcohol consumption
- calcium intake
- physical activity

The authors were careful to say that increased cola consumption is associated with lower bone mineral density because of potential lurking variables. They never stated that increased cola consumption causes lower bone mineral density.

Using Table Critical Values of Spearman's Rank Correlation Coefficient r_S to Interpret r:

If $|r|$ exceeds the value in Critical Values of Spearman's Rank Correlation Coefficient **Table**, conclude that there is a linear correlation. Otherwise, there is not sufficient evidence to support the conclusion of a linear correlation.



Interpreting *r*: Explained Variation

The value of r^2 is the proportion of the variation in *y* that is explained by the linear relationship between *x* and *y*.

Example

Using the pizza subway fare costs in Table below, we have found that the linear correlation coefficient is $r = 0.988$. What proportion of the variation in the subway fare can be explained by the variation in the costs of a slice of pizza?

***Table** – Cost of a Slice of Pizza, subway Fare, and the CPI*						
Year	**1960**	**1973**	**1986**	**1995**	**2002**	**2003**
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00
Subway Fare	0.15	0.35	1.00	1.35	1.50	2.00
CPI	30.2	48.3	112.3	162.2	191.9	197.8

Solution

With $r = 0.988$, we get $r^2 = 0.976$.

We conclude that 0.976 (or about 98%) of the variation in the cost of a subway fares can be explained by the linear relationship between the costs of pizza and subway fares. This implies that about 2% of the variation in costs of subway fares cannot be explained by the costs of pizza.

Common Errors Involving Correlation

1. ***Causation***: It is wrong to conclude that correlation implies causality.
2. ***Averages***: Averages suppress individual variation and may inflate the correlation coefficient.
3. ***Linearity***: There may be some relationship between *x* and *y* even when there is no linear correlation.

TI-83/84 PLUS Enter the paired data in lists L1 and L2, then press **STAT** and select **TESTS.** Using the option of **LinRegTTest** will result in several displayed values, including the value of the linear correlation coefficient *r*. To obtain a scatterplot, press **2nd,** then **Y =** (for STAT PLOT). Press **Enter** twice to turn Plot 1 on, then select the first graph type, which resembles a scatterplot. Set the *X* list and *Y* list labels to L1 and L2 and press the **ZOOM** key, then select **ZoomStat** and press the **Enter** key.

Exercises *Section* 2.1 – Correlation

1. For each of several randomly selected years, the total number of points scored in the Super Bowl football game and the total number of new cars sold in The U.S. are recorded. For this sample of paired data

a) What does r represent?

b) What does ρ represent?

c) With our doing any research or calculations, estimate the value of r.

2. The heights (in inches) of a sample of eight mother/daughter pairs of subjects measured. Using Excel with the paired mother/daughter heights, the linear correlation coefficient is found to be 0.693. Is there sufficient evidence to support the claim that there is a linear correlation between the heights of mothers and the heights of their daughters? Explain.

3. The heights and weights of a sample of 9 supermodels were measured. Using a TI calculator, the linear correlation coefficient is found to be 0.360. Is there sufficient evidence to support the claim that there is a linear correlation between the heights and weights of supermodels? Explain.

4. Given the table below

x	10	8	13	9	11	14	6	4	12	7	5
y	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74

a) Construct a scatterplot

b) Find the value of linear correlation coefficient r and then determine whether there is sufficient evidence to support the claim of a linear correlation between the 2 variables.

c) Identify the feature of the data that would be missed if part (*b*) was completed without constructing the scatterplot.

5. Given the table below

x	10	8	13	9	11	14	6	4	12	7	5
y	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73

a) Construct a scatterplot

b) Find the value of linear correlation coefficient r and then determine whether there is sufficient evidence to support the claim of a linear correlation between the 2 variables.

c) Identify the feature of the data that would be missed if part (*b*) was completed without constructing the scatterplot.

6. The paired values of the Consumer Price Index (CPI) and the cost of a slice of pizza are shown below

CPI	30.2	48.3	112.3	162.2	191.9	197.8
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00

a) Construct a scatterplot

b) Find the value of linear correlation coefficient r.

7. Listed below are systolic blood pressure measurements (in mm HG) obtained from the same woman.

Right Arm	102	101	94	79	79
Left Arm	175	169	182	146	144

a) Construct a scatterplot
b) Find the value of linear correlation coefficient r.

8. Listed below are costs (in dollars) of air fares for different airlines from NY to San Francisco. The costs are based on tickets purchased 30 days in advance and one day in advance.

30 *Days*	244	260	264	264	278	318	280
One Day	456	614	567	943	628	1088	536

a) Construct a scatterplot
b) Find the value of linear correlation coefficient r.

9. Listed below are repair costs (in dollars) for cars crashed at 6 mi/h in full-front crash tests and the same cars crashed at 6 mi/f in full-rear crash tests.

Front	936	978	2252	1032	3911	4312	3469
Rear	1480	1202	802	3191	1122	739	2767

a) Construct a scatterplot
b) Find the value of linear correlation coefficient r .

10. For the following data:

x	2	4	7	7	9
y	1.6	2.1	2.4	2.6	3.2

a) Draw a scatter diagram
b) Compute the correlation coefficient
c) Comment on the type of the relation that appears to exist between x and y.

11. A pediatrician wants to determine the relation that may exist between a child's height and head circumference. She randomly selects 8 children, measures their height and head cieecumference, and obtains the data shown in the table.

Height (in.)	27.25	25.75	26.25	25.75	27.75	26.75	25.75	26.75
Head Circumference (in.)	17.6	17	17.2	16.9	17.6	17.3	17.2	17.4

a) Draw a scatter diagram
b) Compute the correlation coefficient
c) If the pediatrician wants to use height to predict head circumference, determine which variable is the explanatory variable and which is the response variable.
d) Does a linear relation exist between height and head circumference?

12. An engineer wanted to determine how the weight of a car affects gas mileage. The accompanying data represent the weight of various domestic cars and their mileages in the city for the 2008 model year. Suppose that we add Car 12 to the original data. Car 12 weights 3,305 *lbs*. and gets 19 miles per gallon.

Car	***Weight (lbs)***	***Miles / Gal***
1	3,775	21
2	3,964	17
3	3,470	21
4	3,175	22
5	2,580	27
6	3,730	18
7	2,605	26
8	3,772	17
9	3,310	20
10	2,991	25
11	2,752	26

a) Draw a scatter diagram
b) Compute the correlation coefficient.
c) Compute the correlation coefficient with Car 12 included
d) Compare the correlation coefficient in part (*b*) & (*c*), and why are the results reasonable.
e) Suppose that we add Car 13 (a hybrid car) to the original data (remove the Car 12). Car 13 weighs 2,890 *lbs*. and gets 60 miles per gallon. Compute the linear coefficient with Car 13 included,

Section 2.2 – Least–Squares Regression

Basic Concept of Regression

Two variables are sometimes related in a ***deterministic*** way, meaning that given a value for one variable, the value of the other variable is exactly determined from a given equation.
The regression equation expresses a relationship between x (called the ***explanatory*** variable, ***predictor*** variable or ***independent*** variable), and y (called the ***response*** variable or ***dependent*** variable).

The typical equation of a straight line
$y = mx + b$ is expressed in the form
$\hat{y} = b_0 + b_1 x$, where b_0 is the y-intercept and b_1 is the slope.

Definitions

- **Regression Equation**
  Given a collection of paired data, the regression equation algebraically describes the relationship between the two variables.

$$\hat{y} = b_1 x + b_0$$

- **Regression Line**
  The graph of the regression equation is called the regression line (or line of best fit, or least squares line).

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} \qquad b_0 = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

Notation for Regression Equation

	Population Parameter	***Sample Statistic***
Slope of regression equation	β_1	$b_1 = r \cdot \frac{s_y}{s_x}$
y-intercept of regression equation	β_0	$b_0 = \overline{y} - b_1 x$
Equation of the regression line	$y = \beta_0 + \beta_1 x$	$\hat{y} = b_0 + b_1 x$

Requirements

1. The sample of paired (x, y) data is a random sample of quantitative data.
2. Visual examination of the scatterplot shows that the points approximate a straight-line pattern.
3. Any outliers must be removed if they are known to be errors. Consider the effects of any outliers that are not known errors.

Formulas for b_0 and b_1

Slope: $b_1 = r\frac{s_y}{s_x}$

y-intercept: $b_0 = \bar{y} - b_1\bar{x}$

Where r is the linear correlation coefficient,

s_y is the standard deviation of the y values, and

s_x is the standard deviation of the x values

Special Property: The regression line fits the sample points best.

Rounding the *y*-intercept b_0 and the Slope b_1

Example

Using the following sample data

x	0	2	3	5	6	6
y	5.8	5.7	5.2	2.8	1.9	2.2

a) Find a linear equation that relates x and y by selecting 2 points and finding the equation of the line containing the points.

b) Graph the equation on the scatter diagram

c) Use the equation to predict y if $x = 3$

Solution

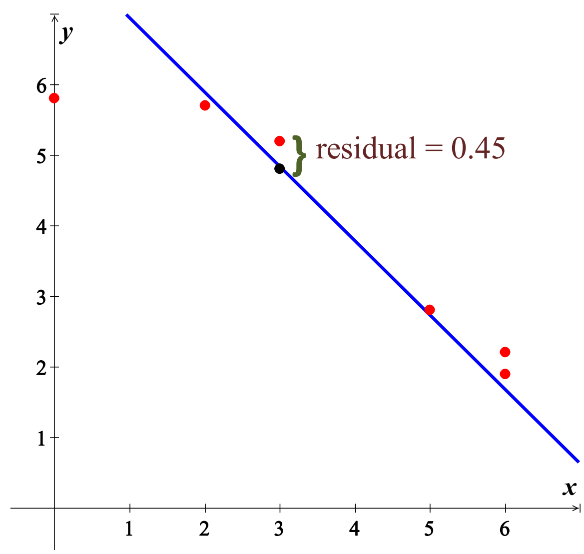
a) Using the 2 points (2, 5.7) and (6, 1.9)

$$m = \frac{5.7 - 1.9}{2 - 6} = -0.95$$

$$y = m(x - x_1) + y_1$$

$$y = -0.95(x - 2) + 5.7 \underline{= -0.95x + 7.6}$$

b) Graph →



c) $y = -0.95(3) + 7.6 \underline{= 4.75}$

The difference between the observed value of y and the predicted value of y is the error, or ***residual***. Using the line from the last example, and the predicted value at $x = 3$:

residual = observed y – predicted y

= 5.2 – 4.75

= 0.45

Example

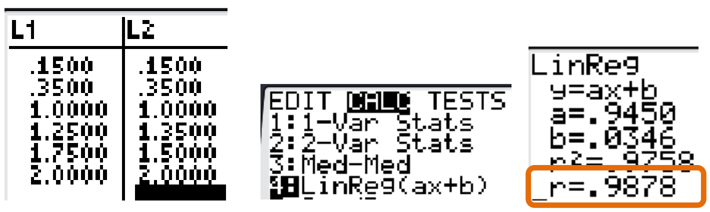
Using the pizza subway fare costs in Table below, Use technology to find the equation of the regression line in which the explanatory variable (or *x* variable) is the cost of a slice of pizza and the response variable (or *y* variable) is the corresponding cost of a subway fare. What proportion of the variation in the subway fare can be explained by the variation in the costs of a slice of pizza?

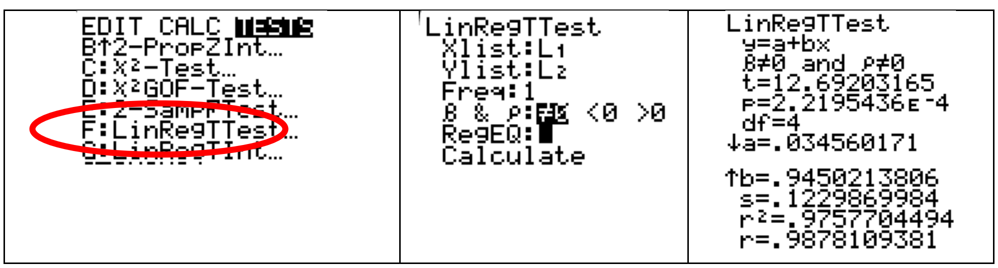
Table *– Cost of a Slice of Pizza, subway Fare, and the CPI*						
Year	**1960**	**1973**	**1986**	**1995**	**2002**	**2003**
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00
Subway Fare	0.15	0.35	1.00	1.35	1.50	2.00
CPI	30.2	48.3	112.3	162.2	191.9	197.8

Solution

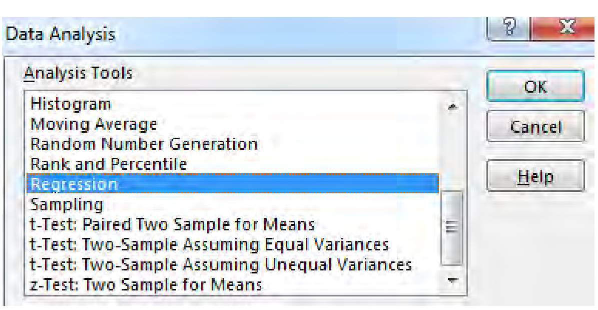
Requirements are satisfied: simple random sample; scatterplot approximates a straight line; no outliers

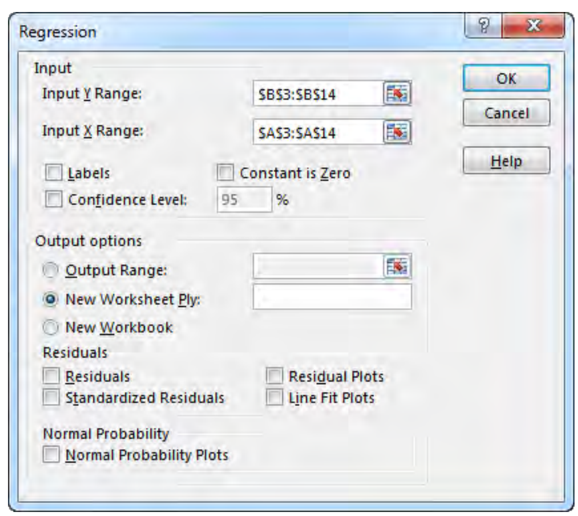
Here are results from four different technologies





Excel





SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9878109
R Square	0.9757704
Adjusted R	0.9697131
Standard E	0.122987
Observatic	6

ANOVA

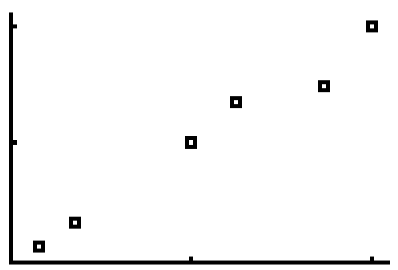
	df	SS	MS	F	ignificance F
Regressior	1	2.436580126	2.4365801	161.08767	0.000222
Residual	4	0.060503207	0.0151258		
Total	5	2.497083333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	.ower 95.0%	Jpper 95.0%
Intercept	0.0345602	0.095012806	0.3637422	0.7344608	-0.229238	0.298358	-0.229238	0.298358
X Variable	0.9450214	0.074457849	12.692032	0.000222	0.7382932	1.1517495	0.7382932	1.1517495

Example

Graph the regression equation $\hat{y} = 0.0346 + 0.945x$ (from the preceding Example) on the scatterplot of the pizza/subway fare data and examine the graph to subjectively determine how well the regression line fits the data.

Solution



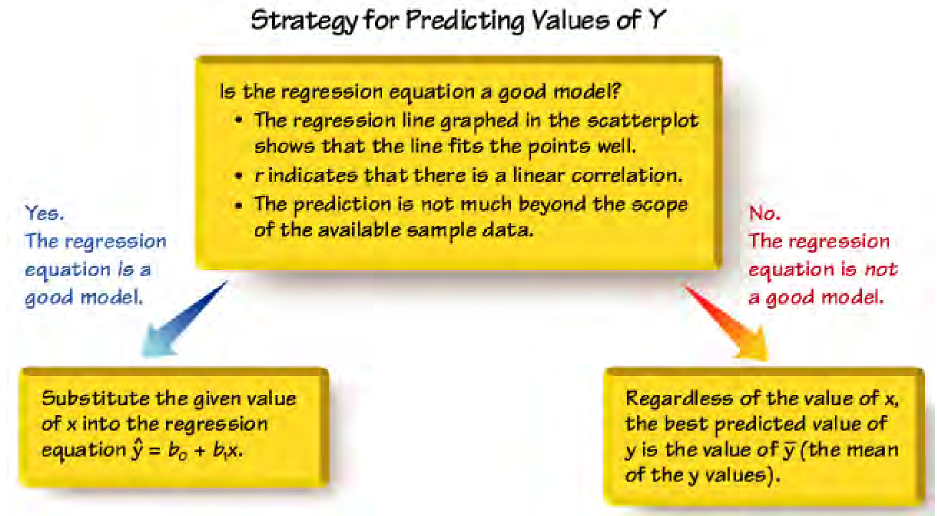
Using the Regression Equation for Predictions

1. Use the regression equation for predictions only if the graph of the regression line on the scatterplot confirms that the regression line fits the points reasonably well.
2. Use the regression equation for predictions only if the linear correlation coefficient r indicates that there is a linear correlation between the two variables.
3. Use the regression line for predictions only if the data do not go much beyond the scope of the available sample data. (Predicting too far beyond the scope of the available sample data is called *extrapolation*, and it could result in bad predictions.)
4. If the regression equation does not appear to be useful for making predictions, the best predicted value of a variable is its point estimate, which is its sample mean.

If the regression equation is not a good model, the best predicted value of y is simply y, the mean of the y values.

Remember, this strategy applies to linear patterns of points in a scatterplot.

If the scatterplot shows a pattern that is not a straight-line pattern, other methods apply.



Definitions

In working with two variables related by a regression equation, the marginal change in a variable is the amount that it changes when the other variable changes by exactly one unit. The slope b_1 in the regression equation represents the marginal change in y that occurs when x changes by one unit.

In a scatterplot, an outlier is a point lying far away from the other data points.

Paired sample data may include one or more influential points, which are points that strongly affect the graph of the regression line.

Beyond the Basics of Regression

Definitions

In working with two variables related by a regression equation, the marginal change in a variable is the amount that it changes when the other variable changes by exactly one unit. The slope b_1 in the regression equation represents the marginal change in y that occurs when x changes by one unit.

In a scatterplot, an ***outlier*** is a point lying far away from the other data points. Paired sample data include one or more ***influential points***, which are points that strongly affect the graph of the regression line.

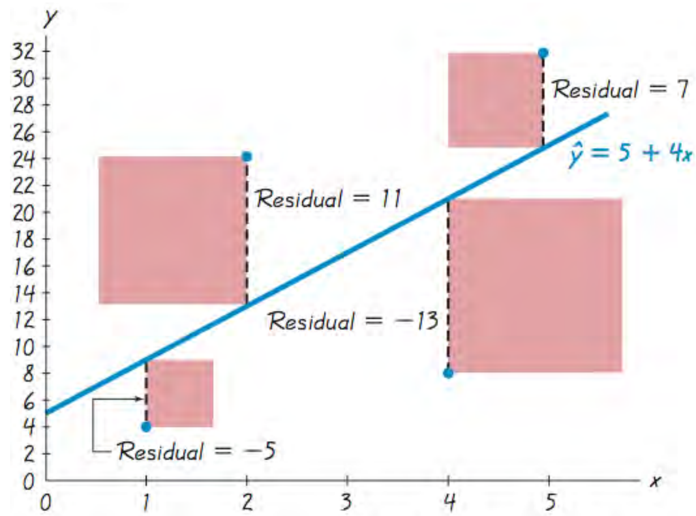
Residuals and the Least-Squares Property

Definition

For a pair of sample *x* and *y* values, the residual is the difference between the *observed* sample value of *y* and the *y*-value that is *predicted* by using the regression equation. That is,

$$residual = observed\ y - predicted\ y = y - \hat{y}$$

Residuals



Definitions

A straight line satisfies the least-squares property if the sum of the squares of the residuals is the smallest sum possible.

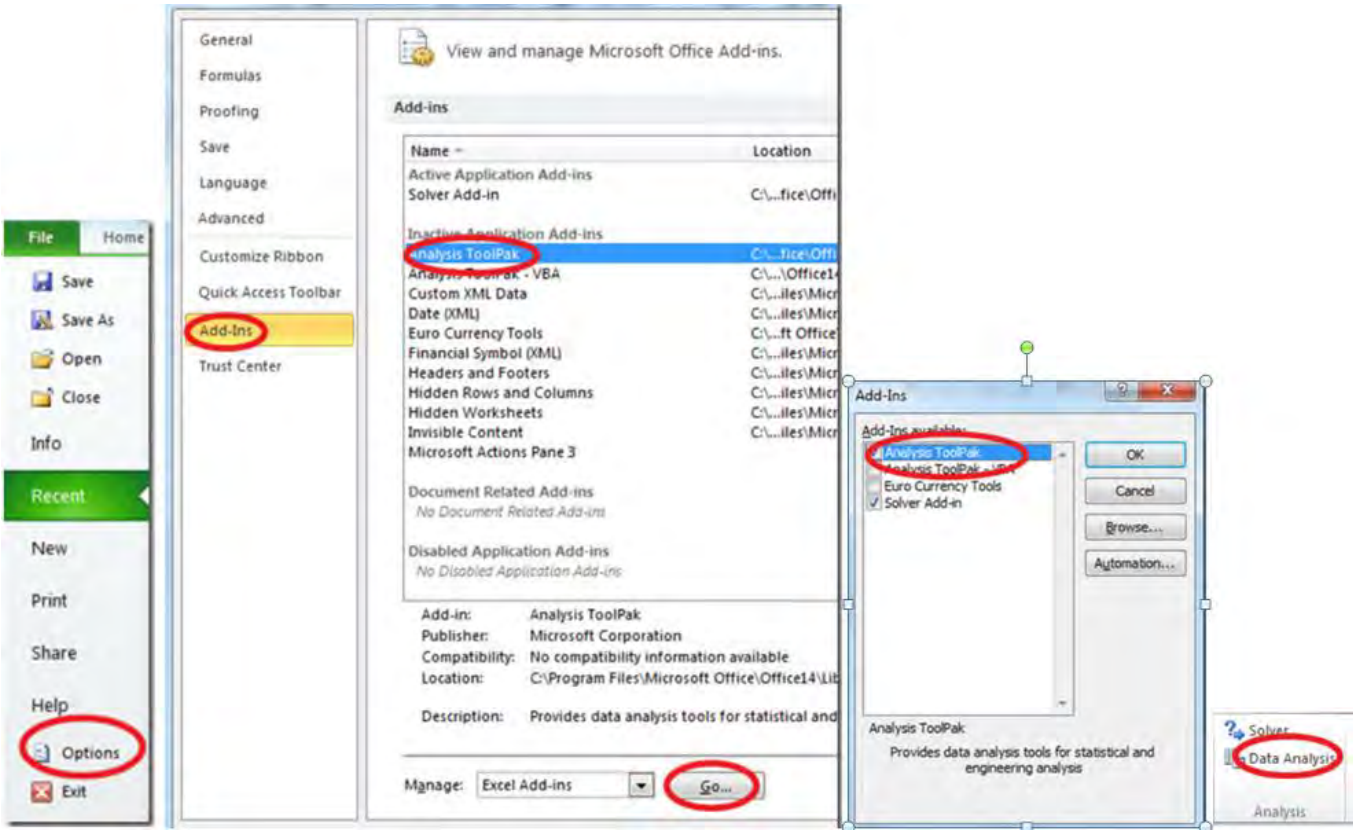
A ***residual plot*** is a scatterplot of the (*x*, *y*) values after each of the *y*-coordinate values has been replaced by the residual value $y - \hat{y}$ (where *y* denotes the predicted value of *y*). That is, a residual plot is a graph of the points $(x,\ y - \hat{y})$.

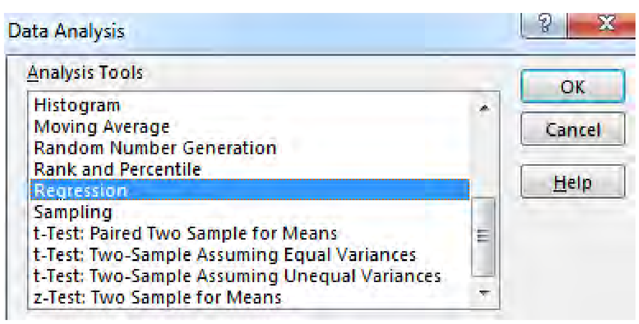
Residual Plot Analysis

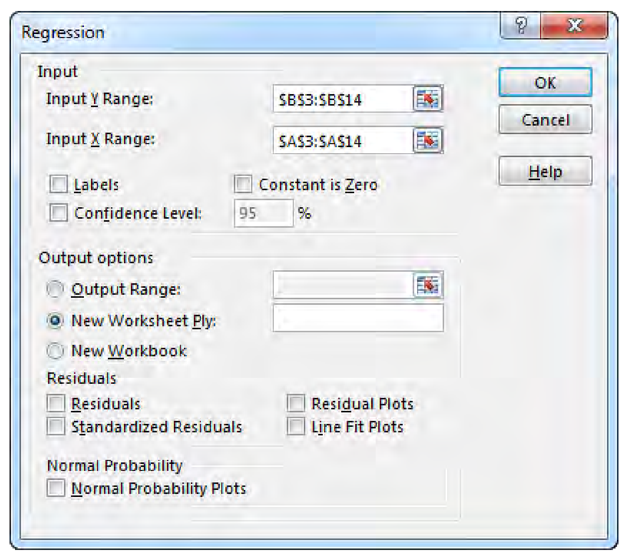
When analyzing a residual plot, look for a pattern in the way the points are configured, and use these criteria:

- The residual plot should not have an obvious pattern that is not a straight-line pattern.
- The residual plot should not become thicker (or thinner) when viewed from left to right.

Installation analysis package to use regression







Exercises *Section* 2.2 – Least–Squares Regression

1. A physician measured the weights and cholesterol levels of a random sample of men. The regression equation is $\hat{y} = -116 + 2.44x$, where x represents weight (in pounds). What does the symbol $\hat{y}$ represent? What does the predictor variable represent? What does the response variable represent?

2. In what sense is the regression line the straight line that "best" fits the points in a scatterplot?

3. In a study, the total weight (in pounds) of garbage discarded in one week and the household size were recorded for 62 households. The linear correlation coefficient is $r = 0.759$ and the regression equation $\hat{y} = 0.445 + 0.119x$, where x represents the total weight of discarded garbage. The mean of the 62 garbage weights is 27.4 lb. and the 62 households have a mean size of 3.71 people. What is the best predicted number of people in a household that discards 50 lb. of garbage?

4. A sample of 8 mother/daughter pairs of subjects was obtained, and their heights (in inches) were measured. The linear correlation coefficient is 0.693 and the regression equation $\hat{y} = 69 - 0.0849x$, where x represents the height of the mother. The mean height of the mothers is 63.1 in. and the mean height of the daughters is 63.3 in. Find the best predicted height of a daughter given that the mother has a height of 60 in.

5. A sample of 40 women is obtained, and their heights (in inches) and pulse rates (in beats per minute) are measured. The linear correlation coefficient is 0.202 and the equation of the regression line is $\hat{y} = 18.2 + 0.920x$, where x represents height. The mean of the 40 heights is 63.2 in. and the mean of the 40 pulse rates is 76.3 beats per minute. Find the best predicted pulse rate of a woman who is 70 in. tall.

6. Heights (in inches) and weights (in pounds) are obtained from a random sample of 9 supermodels. The linear correlation coefficient is 0.360 and the equation of the regression line is $\hat{y} = 31.8 + 1.23x$, where x represents height. The mean of the 9 heights is 69.3 in. and the mean of the 9 weights is 117 lb. Find the best predicted weight of a supermodel with a height of 72 in.?

7. Find the equation of the regression line for the given data below

x	10	8	13	9	11	14	6	4	12	7	5
y	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74

Examine the scatterplot and identify a characteristic of the data that is ignored by the regression line

8. Find the equation of the regression line for the given data below

x	10	8	13	9	11	14	6	4	12	7	5
y	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73

Examine the scatterplot and identify a characteristic of the data that is ignored by the regression line

9. Find the equation of the regression line for the given data below

CPI	30.2	48.3	112.3	162.2	191.9	197.8
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00

Let the first variable be the predictor (x) variable. Find the best indicated predicted cost of a slice of pizza when the Consumer Price Index (CPI) is 182.5 (in the year 2000).

10. Find the equation of the regression line for the given data below

CPI	30.2	48.3	112.3	162.2	191.9	197.8
Subway fare	0.15	0.35	1.00	1.35	1.5	2.00

Let the first variable be the predictor (x) variable. Find the best indicated predicted cost of a slice of pizza when the Consumer Price Index (CPI) is 182.5 (in the year 2000).

11. Listed below are systolic blood pressure measurements (in mm *HG*) obtained from the same woman.

Right Arm	102	101	94	79	79
Left Arm	175	169	182	146	144

Find the best predicted systolic blood pressure in the left arm given that the systolic blood pressure in the right arm is 100 mm Hg.

12. Find the best predicted height of runner-up Goldwater, given that the height of the winning presidential candidate is 75 in. Is the predicted height of Goldwater close to his actual height of 72 in.?

Winner	69.5	73	73	74	74.5	74.5	71	71
Runner-Up	72	69.5	70	68	74	74	73	76

13. Find the best predicted amount of revenue (in millions of dollars), given that the amount has a size 87 thousand ft^2. How does the result compare to the actual revenue of $65.1 million?

Size	160	227	140	144	161	147	141
Revenue	189	157	140	127	123	106	101

14. Find the best predicted new mileage rating of a jeep given that old rating is 19 mi/gal. Is the predicted value close to the actual value of 17 mi/gal?

Old	16	27	17	33	28	24	18	22	20	29	21
New	15	24	15	29	25	22	16	20	18	26	19

15. Find the best predicted temperature for a recent year in which the concentration (in parts per million) of CO_2 is 370.9. Is the predicted temperature close to the actual temperature of 14.5° C??

CO_2	314	317	320	326	331	339	346	354	361	369
Temperature	13.9	14.0	13.9	14.1	14.0	14.3	14.1	14.5	14.5	14.4

16. Find the best predicted IQ score of someone with a brain size of 1275 cm^3

Brain Size	965	1029	1030	1285	1049	1077	1037	1068	1176	1105
IQ	90	85	86	102	103	97	124	125	102	114

17. Listed below are the word counts for men and women.

Male

27531	15684	5638	27997	25433	8077	21319	17572	26429	21966	11680	10818
12650	21683	19153	1411	20242	10117	20206	16874	16135	20734	7771	6792
26194	10671	13462	12474	13560	18876	13825	9274	20547	17190	10578	14821
15477	10483	19377	11767	13793	5908	18821	14069	16072	16414	19017	37649
17427	46978	25835	10302	15686	10072	6885	20848				

Female

20737	24625	5198	18712	12002	15702	11661	19624	13397	18776	15863	12549
17014	23511	6017	18338	23020	18602	16518	13770	29940	8419	17791	5596
11467	18372	13657	21420	21261	12964	33789	8709	10508	11909	29730	20981
16937	19049	20224	15872	18717	12685	17646	16255	28838	38154	25510	34869
24480	31553	18667	7059	25168	16143	14730	28117				

Find the best predicted word count of a woman given that her male partner speaks 6,000 words in a day.

18. According the least-squares property, the regression line minimizes the sum of the squares of the residuals. Listed below are the paired data consisting of the first 6 pulse and the first systolic blood pressures of males.

Pulse (x)	68	64	88	72	64	110
Systolic (y)	125	107	126	110	72	107

a) Find the equation of the regression line.
b) Identify the residuals, and find the sum of squares of the residuals.
c) Show that the equation $\hat{y} = 70 + 0.5x$ results in a larger sum of squares of residuals.

19. The scatter diagram for the data set below

x	0	2	3	5	5	5
y	7.3	5.1	6	4	5.3	3.6

Given that $\bar{x} = 3.333$, $s_x = 2.0655911$, $\bar{y} = 5.217$, $s_y = 1.3467244$, and $r = -0.8363944$, determine the least squares regression line.

20. The scatter diagram for the data set below

x	0	0	0	2	4	6
y	7.1	5.9	6.5	6.2	4.9	2.2

a) Determine the least squares regression line.
b) Graph the least-squares regression line on the scatter diagram

21. The scatter diagram for the data set below

x	4	5	6	7	9
y	5	8	9	11	14

a) Determine the least squares regression line.
b) Graph the least-squares regression line on the scatter diagram.
c) Compute the sum of the squared residuals for the least-squares regression line found in part (*a*).

22. A student at a junior college conducted a survey of 20 randomly selected full-time students to determine the relation between the number of hours of video game playing each week, x, and grade-point average, y. She found that a linear relation exists between the two variables. The least-squares regression line that describes this relation is $\hat{y} = -0.0531x + 2.9213$.

a) Predict the grade-point average of a student who plays video games 8 hours per week.
b) Interpret the slope
c) Interpret the appropriate y-intercept.
d) A student who plays video games 7 hours per week has a grade-point average of 2.67. Is the student grade-point average above or below average among all students who play video games 7 hours per week.

Section 2.3 – Probability Rules, Addition Rule and Complements

Definitions

Probability is a measure of the likelihood of a random phenomenon or chance behavior. Probability describes the long-term proportion with which a certain ***outcome*** will occur in situations with short-term uncertainty.

Use the probability applet to simulate flipping a coin 100 times. Plot the proportion of heads against the number of flips. Repeat the simulation.

Probability deals with experiments that yield random short-term results or outcomes, yet reveal long-term predictability.

The long-term proportion in which a certain outcome is observed is the probability of that outcome.

The Law of Large Numbers

As the number of repetitions of a probability experiment increases, the proportion with which a certain outcome is observed gets closer to the probability of the outcome.

Definitions

In probability, an ***experiment*** is any process that can be repeated in which the results are uncertain.

An ***event*** is any collection of results or outcomes of a procedure

A ***simple event*** is an outcome or an event that cannot be further broken down into simpler components

The ***sample space*** for a procedure consists of all possible simple events; that is, the sample space consists of ***all outcomes*** that cannot be broken down any further

Example

We use "f" to denote a female baby and "m" to denote a male baby.

Procedure	***Example of Event***	***Complete Sample Space***
Single birth	1 female (simple event)	$\{f, m\}$
3 births	2 females and 1 male (ffm, fmf, mff) are simple events	$\{fff, ffm, fmf, mff, mfm, mmf, mmm\}$

Notation for Probabilities

P – denotes a probability

A, B, and C – denote specific events

$P(A)$ - denotes the probability of event A occurring.

Basic Rules for Computing Probability

***Rule* 1: Relative Frequency Approximation of Probability**

Conduct (or observe) a procedure, and count the number of times event A actually occurs. Based on these actual results, $P(A)$ is approximated as follows:

$$P(A) = \frac{\#\ of\ times\ A\ occured}{\#\ of\ times\ procedure\ was\ repeated}$$

***Rule* 2: Classical Approach to Probability (**Requires Equally Likely Outcomes**)**

Assume that a given procedure has n different simple events and that each of those simple events has an equal chance of occurring. If event A can occur in s of these n ways, then

$$P(A) = \frac{m}{n} = \frac{\#\ of\ ways\ A\ can\ occur}{\#\ of\ different\ simple\ events} = \frac{N(E)}{N(S)}$$

***Rule* 3: Subjective Probabilities**

$P(A)$, the probability of event A, is estimated by using knowledge of the relevant circumstances.

Rules of probabilities

1. The probability of any event E, $P(E)$, must be greater than or equal to 0 and less than or equal to 1. That is, $0 \le P(E) \le 1$.

2. The sum of the probabilities of all outcomes must equal 1. That is, if the sample space

$$S = \{e_1, e_2, \ldots, e_n\} \quad then \quad P(e_1) + P(e_2) + \ \ldots + P(e_n) = 1$$

Example

Find the probability that a randomly selected car in U.S. will be in a crash this year. There were 6,511,100 cars that crashed among the 135,670,000 cars registered.

Solution

$$P(crash) = \frac{\#\ of\ cars\ that\ crashed}{Total\ number\ of\ cars} = \frac{6,511,100}{135,670,000} \ \underline{= 0.048}$$

Example

When studying the effect of heredity on height, we can express each individual genotype, AA, Aa, aA, and aa, on an index card and shuffle the four cards and randomly select one of them. What is the probability that we select a genotype in which the two components are different?

Solution

$$P(outcome\ with\ different\ components) = \frac{2}{4} \ \underline{= 0.5}$$

Example

What is the probability that you will get stuck in the next elevator that you ride?

Solution

There are 2 possible outcomes (stuck or not becoming stuck). But that are not equally likely, so we cannot use the classical approach.
The leaves us with a subjective estimate, in this case, say 0.0001 (equivalent to 1 chance in ten thousand). This is likely to be in general ballpark of the true probability.

Example

Find the probability that when a couple has 3 children, they will have exactly 2 boys. Assume that boys and girls are equally likely and that the gender of any child is not influenced by the gender of any other child.

Solution

$S=\{bbb,\ bbg,\ bgb,\ gbb,\ bgg,\ gbg,\ ggb,\ ggg\}$

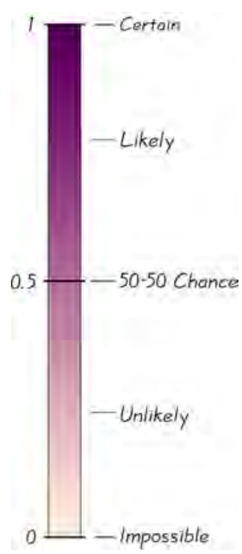
$P(2\ boys)=\frac{3}{8}\ \underline{=0.375}$

Probability Limits

Always express a probability as a fraction or decimal number between 0 and 1.

- ✓ The probability of an impossible event is 0.
- ✓ The probability of an event that is certain to occur is 1.
- ✓ For any event A, the probability of A is between 0 and 1 inclusive.

That is, $0 \le P(A) \le 1$.



Example

If a year is selected at random, find the probability that Thanksgiving Day will be

a) On a Wednesday
b) On a Thursday

Solution

a) $P(On\ a\ Wednesday)\ \underline{=0}$

b) It is certain that Thanksgiving Day will be on a Thursday. When an event certain to occur:

$P(On\ a\ Thursday)\ \underline{=1}$

Addition Rule and Complements

Definition

A ***compound event*** is any event combining 2 or more simple events

Notation for Addition Rule

$P(A \text{ or } B) = P(\text{in a single trial, event } A \text{ occurs or event } B \text{ occurs or they both occur})$

Example

If 1 subject is randomly selected from the 98 subjects given the polygraph, find the probability of selecting a subject who had a positive test result or lied.

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

Solution

There are 66 subjects who had a positive test result or lied.

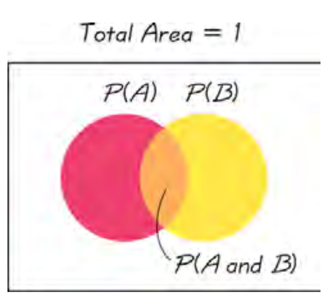
$P(\text{positive test result or lied}) = \frac{66}{98} = 0.673$

➢ When finding the probability that event *A* occurs or event *B* occurs, find the total number of ways *A* can occur and the number of ways *B* can occur, but *find that total in such a way that no outcome is counted more than once*.

Formal Addition Rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

where $P(A \text{ and } B)$ denotes the probability that *A* and *B* both occur at the same time as an outcome in a trial of a procedure.

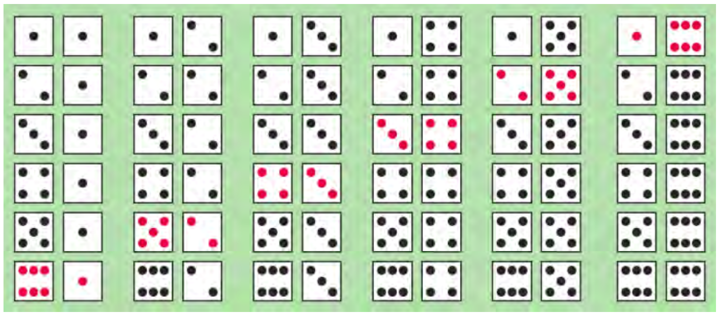


Intuitive Addition Rule

To find $P(A$ or $B)$, find the sum of the number of ways event A can occur and the number of ways event B can occur, *adding in such a way that every outcome is counted only once*. $P(A$ or $B)$ is equal to that sum, divided by the total number of outcomes in the sample space.

Example

Suppose that a pair of dice are thrown. Let E = "the first die is a two" and let F = "the sum of the dice is less than or equal to 5". Find $P(E$ ***or*** $F)$



Solution

$$P(E)=\frac{N(E)}{N(S)}=\frac{6}{36}$$

$$P(F)=\frac{N(F)}{N(S)}=\frac{10}{36}$$

$$P(E \text{ and } F)=\frac{N(E \text{ and } F)}{N(S)}=\frac{3}{36}$$

$$\begin{aligned} P(E \text{ or } F) &= P(E)+P(F)-P(E \text{ and } F) \\ &= \frac{6}{36}+\frac{10}{36}-\frac{3}{36} \\ &= \frac{13}{36} \end{aligned}$$

Disjoint or Mutually Exclusive

Definition

Events A and B are ***disjoint*** (or ***mutually exclusive***) if they cannot occur at the same time. (That is, disjoint events do not overlap.)

Example

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) Consider the procedure of randomly selecting 1 of the 98 subjects. Determine whether the following event are disjoint:
  A: Getting a subject with a negative test result.
  B: Getting a subject who did not lie.

b) Assuming that 1 subject is randomly selected from the 98 that were tested, find the probability of selecting a subject who had a negative test result or did not lie.

c) Assuming that one of the 98 test results summarized in the table above is randomly selected, find the probability that it is a positive test result.

Solution

a) There are 41 subjects with negative test results and are 47 subjects who did lie. The event of getting a subject with a negative test result and getting a subject who did not lie can occur at the same time, there are 32 subjects. Therefore, the events are not disjoint.

b) $P(\text{negative test result or did not lie}) = \frac{56}{98} \underline{= 0.571}$

c) $P(\text{positive result}) = \frac{\#\ \text{of positive results}}{\text{Total number of results}} = \frac{42+15}{15+42+32+9} = \frac{57}{98} \underline{= 0.582}$

Complementary Events

The complement of event A, denoted by $\overline{A}$, consists of all outcomes in which the event A ***does not*** occur.

$P(A)$ and $P\left(\overline{A}\right)$ are disjoint

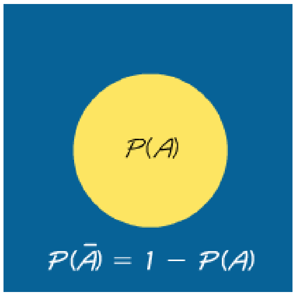
It is impossible for an event and its complement to occur at the same time.

Rule of Complementary Events

$$P(A)+P\left(\overline{A}\right)=1$$

$$P\left(\overline{A}\right)=1-P(A)$$

$$P(A)=1-P\left(\overline{A}\right)$$



Example

FBI data show that 62.4% of murders are cleared by arrests. We can express the probability of a murder being cleared by an arrest as $P(cleared)=0.624$. For a randomly selected murder, find $P\left(\overline{cleared}\right)$

Solution

$$\begin{aligned} P\left(\overline{cleared}\right) &= 1-P(cleared) \\ &= 1-.624 \\ &= 0.376 \end{aligned}$$

Example

A typical question on a SAT test requires the test taker to select one of five possible choices: A, B, C, D, or E. Because only one answer is correct, if you make a random guess, your probability of being correct is $\frac{1}{5}$ *or* 0.2. Find the probability of making a random guess and not being correct (or being incorrect).

Solution

$$P(\text{not guessing the correct answer})=P\left(\overline{correct}\right)=\frac{4}{5}=0.8$$

or $\quad P\left(\overline{correct}\right)=1-P(correct)=1-\frac{1}{5}=\frac{4}{5}=0.8$

Odds

Definition

The ***actual odds against*** event A occurring are the ratio $\frac{P(\bar{A})}{P(A)}$, usually expressed in the form of a:b (or "a to b"), where a and b are integers having no common factors.

The ***actual odds in favor*** of event A occurring are the ratio $\frac{P(A)}{P(\bar{A})}$, which is the reciprocal of the actual odds against the event. If the odds against A are a:b, then the odds in favor of A are b:a.

The ***payoff odds*** against event A occurring are the ratio of the net profit (if you win) to the amount bet.

$$payoff\ \ odds\ against\ event\ A = \frac{net\ profit}{amount\ bet}$$

Example

If you bet \$5 on the number 13 in roulette, your probability of winning is $\frac{1}{38}$ and the payoff odds are given by the casino as 35:1.

a) Find the actual odds against the outcome of 13.
b) How much net profit would you make if you win by betting on 13?
c) If the casino was not operating for profit, and the payoff odds were changed to match the actual odds against 13, how much would you win if the outcome were 13?

Solution

a) With odds: $P(13) = \frac{1}{38}$ and $P(not\ 13) = \frac{37}{38}$

Actual odds against 13 $\frac{P(not\ 13)}{P(13)} = \frac{37/38}{1/38} = \frac{37}{1}$ or 37:1

b) Because the payoffs odds against 13 are 35:1, we have:

35:1 = (*net profit*) : (*amount bet*)

So there is a \$35 profit for each \$1 bet. For \$5 bet, the net profit is $5 \times 35 = \$175$.
The winning bettor would collect \$175 plus the original \$5 bet. That is, the total amount collected would be \$180, for the net profit of \$175.

c) If the casino were not operating for profit, the payoff odds would be equal to the actual odds against the outcome of 13, or 37:1. So there is a net profit of \$37 for each \$1 bet.
For a \$5 bet the net profit would be \$185. (The casino makes its profit by paying only \$175 instead of the \$185 that would be paid with a roulette game that is fair instead of favoring the casino.)

Rounding Off Probabilities

When expressing the value of a probability, either give the ***exact*** *fraction* or decimal or round off final decimal results to three significant digits. (*Suggestion*: When a probability is not a simple fraction such as $\frac{2}{3}$ or $\frac{5}{9}$, expresses it as a decimal so that the number can be better understood.)

- ✓ The probability of $\frac{1}{3}$ can be left as a fraction, or rounded to 0.333. (Do ***not*** round to 0.3)
- ✓ The probability of $\frac{2}{4}$ can be expressed as $\frac{1}{2}$ or 0.5; because is exact, there is no need to express as 0.500.
- ✓ $\frac{1941}{3405} = \underline{0.570|}$

Exercises *Section* 2.3 – Probability Rules, Addition Rule and Complements

1. Based on recent results, the probability of someone in the U.S. being injured while using sports or recreation equipment is $\frac{1}{500}$ (based on data from Statistical Abstract of the U.S.). What does it mean when we say that the probability is $\frac{1}{500}$? Is such an injury unusual?

2. When a baby is born, there is approximately a 50−50 chance that the baby is a girl. Indicate the degree of likelihood as a probability value between 0 and 1.

3. When a rolling a single die, there are 6 chances in 36 that the outcome is a 7. Indicate the degree of likelihood as a probability value between 0 and 1.

4. Identify probability values

a) What is the probability of an event that is certain to occur?
b) What is the probability of an impossible event?
c) A sample space consists of 10 separate events that are equally likely. What is the probability of each?
d) On a true/false test, what is the probability of answering a question correctly if you make a random guess?
e) On a multiple-choice test with five possible answers for each question, what is the probability of answering a question correctly if you make a random guess?

5. When a couple has 3 children, find the probability of each event.

a) There is exactly one girl. *b)* There are exactly 2 girls. *c)* All are girls

6. The 110^{th} Congress of the U.S. included 84 male Senators and 16 female Senators. If one of these Senators is randomly selected, what is the probability that a woman is selected? Does this probability agree with a claim that men and women have the same chance of being elected as Senators?

7. When Mendel conducted his famous genetics experiments with peas, one sample of offspring consisted of 428 green peas and 152 yellow peas. Based on those results, estimate the probability of getting an offspring pea that is green, Is the result reasonably close to the expected value of $\frac{3}{4}$, as claimed by Mendel?

8. A single fair die is rolled. Find the probability of each event

a) Getting a 2
b) Getting an odd number
c) Getting a number less than 5
d) Getting a number greater than 2
e) Getting any number except 3

9. A jar contains 3 white, 4 orange, 5 yellow, and 8 black marbles. If a marble is drawn at random, find the probability that it is the following.

a) White *b)* Orange *c)* Yellow *d)* Black *e)* Not black

10. The student sitting next to you in class concludes that the probability of the ceiling falling down on both of you before class ends is 1/2, because there are two possible outcomes - the ceiling will fall or not fall. What is wrong with this reasoning?

11. Let consider rolling 2 dice. Find the probabilities of the following events

a) ***E*** = Sum of 5 turns up

b) ***F*** = a sum that is a prime number greater than 7 turns up

12. A poll was conducted preceding an election to determine the relationship between voter persuasion concerning a controversial issue and the area of the city in which the voter lives. Five hundred registered voters were interviewed from three areas of the city. The data are shown below. Compute the probability of having no opinion on the issue or living in the inner city.

Area of city	***Favor***	***Oppose***	***No Opinion***
East	30	40	55
North	25	45	50
Inner	95	65	85

13. Suppose a single fair die is rolled. Use the sample space $S = \{1,2,3,4,5,6\}$ and give the probability of each event.

a) E: the die shows an even number

b) F: the die show a number less than 10

c) G: the die shows an 8

14. A solitaire game was played 500 times. Among the 500 trials, the game was won 77 times. (The results are from the solitaire game, and the Vegas rules of "draw 3" with $52 bet and a return of $5 per card are used). Based on these results, find the odds against winning.

15. A roulette wheel has 38 slots. One slot is 0, another is 00, and the others are numbered 1 through 36, respectively. You place a bet that the outcome is an odd number.

a) What is your probability of winning?

b) What are the actual odds against winning?

c) When you bet that the outcome is an odd number, the payoff odds are 1:1. How much profit do you make if you bet $18 and win?

d) How much profit would you make on the $18 bet if you could somehow convince the casino to change its payoff odds so that they are the same as the actual odds against winning?

16. Women have a 0.25% rate of red/green color blindness. If a woman is randomly selected, what is the probability that she does not have red/green color blindness?

17. A pew Research center poll showed that 79% of Americans believe that it is morally wrong to not report all income on tax returns. What is the probability that an American does not have that belief?

18. When the author observed a sobriety checkpoint conducted by the Dutchess County Sheriff Department, he saw that 676 drivers were screened and 6 were arrested for driving while intoxicated. Based on those results, we can estimate that $P(I) = 0.00888$, where *I* denotes the event of screening a driver and getting someone who is intoxicated. What does $P(\bar{I})$ denote and what is its value?

19. Use the polygraph test data

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) If one of the test subjects is randomly selected, find the probability that the subject had a positive test result or did not lie

b) If one of the test subjects is randomly selected, find the probability that the subject did not lie

c) If one of the test subjects is randomly selected, find the probability that the subject had a true negative test result

d) If one of the test subjects is randomly selected, find the probability that the subject had a negative test result or lied.

20. Use the data

Was the challenge to the call successful?		
	Yes	***No***
Men	201	288
Women	126	224

a) If *S* denotes the event of selecting a successful challenge, find $P(\bar{S})$

b) If *M* denotes the event of selecting a challenge made by a man, find $P(\bar{M})$

c) Find the probability that the selected challenge was made by a man or was successful.

d) Find the probability that the selected challenge was made by a woman or was successful.

e) Find $P(\textit{challenge was made by a man or was not successful})$

f) Find $P(\textit{challenge was made by a woman or was not successful})$

21. Refer to the table below

Age

	18 – 21	22 – 29	30 – 39	40 – 49	50 – 59	60 and over
Responded	73	255	245	136	138	202
Refused	11	20	33	16	27	49

a) What is the probability that the selected person refused to answer? Does that probability value suggest that refusals are a problem for pollsters? Why or why not?

b) A pharmaceutical company is interested in opinions of the elderly, because they are either receiving Medicare or will receive it soon. What is the probability that the selected subject is someone 60 and over who responded?
c) What is the probability that the selected person responded or is in the 18–21 age bracket?
d) What is the probability that the selected person refused or is over 59 years of age?
e) A market researcher is interested in responses, especially from those between the ages of 22 and 39, because they are the people more likely to make purchases. Find the probability that a selected responds or is aged between the ages of 22 and 39.
f) A market researcher is not interested in refusals or subjects below 22 years of age or over 59. Find the probability that the selected person refused to answer or is below 22 or is older than 59.

22. Two dice are rolled. Find the probabilities of the following events.
a) The first die is 3 or the sum is 8
b) The second die is 5 or the sum is 10.

23. One card is drawn from an ordinary of 52 cards. Find the probabilities of drawing the following cards
a) A 9 or 10
b) A red card or a 3
c) A 9 or a black 10
d) A heart or a black card
e) A face card or a diamond

24. One card is drawn from an ordinary of 52 cards. Find the probabilities of drawing the following cards
a) Less than a 4 (count aces as ones)
b) A diamond or a 7
c) A black card or an ace
d) A heart or a jack
e) A red card or a face card

25. Pam invites relatives to a party: her mother, 2 aunts, 3 uncles, 2 brothers, 1 male cousin, and 4 female cousins. If the chances of any one guest first equally likely, find the probabilities that the first guest to arrive is as follows.
a) A brother or an uncle
b) A brother or a cousin
c) A brother or her mother
d) An uncle or a cousin
e) A male or a cousin
f) A female or a cousin

26. Suppose $P(E)=0.26$, $P(F)=0.41$, and $P(E\cap F)=0.16$. Find the following
a) $P(E\cup F)$
b) $P(E'\cap F)$
c) $P(E\cap F')$
d) $P(E'\cup F')$

27. Suppose $P(E)=0.42$, $P(F)=0.35$, and $P(E\cup F)=0.59$. Find the following
a) $P(E'\cap F')$
b) $P(E'\cup F')$
c) $P(E'\cup F)$
d) $P(E\cap F')$

28. From survey involving 1,000 people in the certain city, it was found that 500 people had tried a certain brand of diet cola, 600 had tried a certain brand of regular cola, and 200 had tried both types of cola. If a resident of the city is selected at random, what is the empirical probability that
a) The resident has not tried either cola? What are the empirical odds for this event?
b) The resident has tried the diet or has not tried the regular cola? What are the empirical odds against this event?

29. In a poll, respondents were asked whether they had ever been in a car accident. 329 respondents indicated that they had been in a car accident and 322 respondents said that they had not been in a car accident. If one of these respondents is randomly selected, what is the probability of getting someone who has been in a car accident?

30. Refer to the table which summarizes the results of testing for a certain disease

	Positive Test Result	***Negative Test Result***
Subject has the disease	114	5
Subject does not have the disease	12	177

If one of the results is randomly selected, what is the probability that it is a false negative (test indicates the person does not have the disease when in fact they do)?. What is the probability suggest about the accuracy of the test?

31. In a certain town, 2% of people commute to work by bicycle. If a person is selected randomly from the town, what are the odds against selecting someone who commutes by bicycle?

32. Suppose you are playing a game of chance, if you bet $4 on a certain event, you will collect $176 (including your $4 bet) if you win. Find the odds used for determining the payoff.

33. The odds in favor of a particular horse winning a race are 4:5.

a) Find the probability of the horse winning.
b) Find the odds against the horse winning.

34. Consider the sample space of equally likely events for the rolling of a single fair die.

a) What is the probability of rolling an odd number **and** a prime number?
b) What is the probability of rolling an odd number **or** a prime number?

35. Suppose that 2 fair Dice are rolled

a) What is the probability of that a sum of 2 or 3 turns up?
b) What is the probability of that both dice turn up the same or that a sum greater than 8 turns up?

36. A single card is drawn from an ordinary of 52 cards. Calculate the probabilities of and odds for each event

a) A face card or a club is drawn
b) A king or a heart is drawn
c) A black card or an ace is drawn
d) A heart or a number less than 7 (count an ace as 1) is drawn.

37. What is the probability of getting at least 1 black card in a 7-card hand dealt from a standard 52-card deck?

38. What is the probability that a number selected at random from the first 600 positive integers is (exactly) divisible by 6 or 9?

39. What is the probability that a number selected at random from the first 1,000 positive integers is (exactly) divisible by 6 or 8?

40. From a survey involving 1,000 students at a large university, a market research company found that 750 students owned stereos, 450 owned cars, and 350 owned cars and stereos. If a student at the university is selected at random, what is the (empirical) probability that

a) The student owns either a car or a stereo?

b) The student owns neither a car nor a stereo?

41. In order to test a new car, an automobile manufacturer wants to select 4 employees to test drive the car for 1 year. If 12 management and 8 union employees volunteer to be test drivers and the selection is made at random, what is the probability that at least 1 union employee is selected.

42. A shipment of 60 inexpensive digital watches, including 9 that are defective, is sent to a department store. The receiving department selects 10 at random for testing and rejects the whole shipment if 1 or more in the sample are found defective. What is the probability that the shipment will be rejected?

43. If you bet \$5 on the number 13 in roulette, your probability of winning is $\frac{1}{38}$ and the payoff odds are given by the casino as 35:1.

a) Find the actual odds against the outcome of 13.

b) How much net profit would you make if you win by betting on 13?

c) If the casino was not operating for profit, and the payoff odds were changed to match the actual odds against 13, how much would you win if the outcome were 13?

Section 2.4 – Multiplication Rule and Conditional

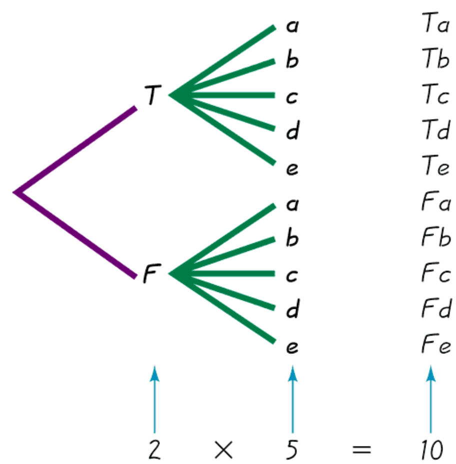
Definition

Two events A and B are ***independent*** if the occurrence of one does not affect the *probability* of the occurrence of the other. (Several events are similarly independent if the occurrence of any does not affect the probabilities of the occurrence of the others.)
If A and B are not independent, they are said to be ***dependent***.

Two events are ***dependent*** if the occurrence of one of them affects the *probability* of the occurrence of the other, but this does not necessarily mean that one of the events is a ***cause*** of the other.

Tree Diagrams

A tree diagram is a picture of the possible outcomes of a procedure, shown as line segments emanating from one starting point. These diagrams are sometimes helpful in determining the number of possible outcomes in a sample space, if the number of possibilities is not too large.



The figure summarizes the possible outcomes for a true/false question followed by a multiple choice question.
Note *that there are 10 possible combinations.*

Multiplication Rule for Independent Events

If E and F are independent events, then $\quad P(E\ and\ F)=P(E)\cdot P(F)$

Example

Assume that we have a batch of 100,000 heart pacemakers, including 99,950 that are good (G) and 50 that are defective (D).

a) If two of those 100,000 pacemakers are randomly selected without replacement, find the probability that they are both good.

b) If 20 of those 100,000 pacemakers are randomly selected without replacement, find the probability that they are both good.

Solution

a) $P(1st\ good)=\dfrac{99{,}950}{100{,}000} \qquad P(2nd\ good)=\dfrac{99{,}949}{100{,}000}$

$$P(1st\ good\ and\ 2nd\ good)=\frac{99{,}950}{100{,}000}\cdot\frac{99{,}949}{100{,}000}\ \underline{=0.999}$$

b) $P(all\ 20\ pacemakers\ are\ good)=\dfrac{99{,}950}{100{,}000}\cdot\dfrac{99{,}949}{100{,}000}\cdot\dfrac{99{,}948}{100{,}000}\cdot\ldots\cdot\dfrac{99{,}931}{100{,}000}$

Example

Assume that two people are randomly selected and also assume that birthdays occur on the same days of the week with equal frequencies.

a) Find the probability that the two people are born on the same day of the week.

b) Find the probability that the two people are born on Monday.

Solution

a) Because no particular day of the week is specified, the first person can be born on any one of the seven week days.

The probability that the second person is born on the same day as the first person is $\frac{1}{7}$.

Probability that 2 people are born on the same day of the week is $\frac{1}{7}$

b) $P(1st\ born\ on\ Monday) = \frac{1}{7}$

$P(2nd\ born\ on\ Monday) = \frac{1}{7}$

$$P(both\ born\ on\ Monday) = \frac{1}{7} \cdot \frac{1}{7}$$

$$= \boxed{\frac{1}{49}}$$

Example

A geneticist developed a procedure for increasing the likelihood of female babies. In an initial test, 20 couples use the method and the results consist of 20 females among 20 babies. Assuming that the gender-selection procedure has no effect, find the probability of getting 20 females among 20 babies by chance. Does the resulting provide strong evidence to support the geneticist's claim that the procedure is effective in increasing the likelihood that babies will be females?

Solution

$$\begin{aligned} P(all\ 20\ are\ female) &= P(1st\ is\ female\ and\ 2nd\ female\ \cdots\ and\ 20th\ is\ female) \\ &= P(female) \bullet P(female) \bullet \cdots \bullet P(female) \\ &= (0.5) \bullet (0.5) \bullet \cdots \bullet (0.5) \\ &= (0.5)^{20} \\ &= \boxed{0.000000954} \end{aligned}$$

The low probability of 0.000000954 indicates that instead of getting 20 females by chance, a more reasonable explanation is that females appear to be more likely with the gender-selection procedure. Because there is such a small probability of getting 20 females in 20 births, we do have to support the geneticist's claim that the gender-selection procedure is effective in increasing the likelihood that babies will be female.

Example

Modern aircraft engines are now highly reliable. One design feature contributing to that reliability is the use of redundancy, whereby critical components are duplicated so that if one fails, the other will work. For example, single-engine aircraft now have two independent electrical systems so that is one electrical system fails, the other can continue to work so that the engine does not fail. For the purposes of this example, we will assume that the probability of an electrical system failure is 0.001.

a) If the engine in an aircraft has one electrical system, what is the probability that it will work?

b) If the engine in an aircraft has 2 independent electrical systems, what is the probability that the engine can function with a working electrical system?

Solution

a) $P(\text{electrical system failure}) = 0.001 \qquad P(\text{doe not fail}) = 1 - 0.001 = 0.999$

$$P(\text{working electrical system}) = P(\text{electrical system does not fail})$$
$$= 0.999$$

b) $P(\text{both electrical system fail}) = P(1st\ \text{electrical system fails and}\ 2nd\ \text{electrical system fails})$

$$= (.001)(.001)$$
$$= 0.000001$$

There is a 0.000001 probability of both electrical systems failing, so the probability that the engine can function with a working electrical system is 1 – 0.000001 = 0.999999

Complements

Finding the Probability of "*At Least One*"

To find the probability of at least one of something, calculate the probability of none, then subtract that result from 1. That is,

$$P(\text{at least one}) = 1 - P(\text{none})$$

- "At least one" is equivalent to "one or more."
- The complement of getting at least one item of a particular type is that you get no items of that type.

Example

Find the probability of a couple having at least 1 girl among 3 children. Assume that boys and girls are equally likely and that the gender of a child is independent of any other child.

Solution

Let A = at least 1 of the 3 children is a girl.

$\overline{A}$ = not getting at least 1 girl among 3 children
= all 3 children are boys
= boy and boy and boy

$$P(\bar{A}) = P(\textit{boy and boy and boy})$$
$$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$
$$= \frac{1}{8}$$

$$P(A) = 1 - P(\bar{A})$$
$$= 1 - \frac{1}{8}$$
$$= \frac{7}{8}$$

There is $\frac{7}{8}$ probability that if a couple has 3 children; at least 1 of them is a girl.

Example

Assume that the probability of a defective Firestone tire is 0.0003 (based on data from Westgard (QC). If the retail outlet Car Stuff buys 100 Firestone tires, find the probability that they get at least 1 that is defective. If that probability is high enough, plans must be made to handle defective tires returned by consumers. Should they make those plans?

Solution

Let A = at least 1 of the 100 tires is defective.

$\bar{A}$ = not getting at least 1 defective among 100 tires
= all 100 tires are good

$$P(\bar{A}) = (1-.0003)(1-.0003)\ldots(1-.0003)$$
$$= (0.9997)(0.9997)\ldots(0.9997)$$
$$= (0.9997)^{100}$$
$$= .9704$$

$$P(A) = 1 - P(\bar{A})$$
$$= 1 - .9704$$
$$= .0296$$

There is 0.0296 probability of at least 1 defective tire among the 100 tires. Because the probability is so low, it is not necessary to make plans for dealing with defective tires returned by consumers.

Conditional Probability *Key Point*

We must adjust the probability of the second event to reflect the outcome of the first event.

The probability for the second event B should take into account the fact that the first event A has already occurred.

Notation for Conditional Probability

$P(B|A)$ represents the probability of event *B* occurring after it is assumed that event *A* has already occurred (read $B|A$ as "*B* given *A*.")

Conditional Probability

A ***conditional probability*** of an event is a probability obtained with the additional information that some other event has already occurred. $P(B|A)$ denotes the conditional probability of event *B* occurring, given that event *A* has already occurred, and it can be found by dividing the probability of events *A* and *B* both occurring by the probability of event *A*:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

The conditional probability of *B* given *A* can be found by assuming that event *A* has occurred, and then calculating the probability that event *B* will occur.

Notation

$$P(A \text{ and } B) = P(\text{event } A \text{ occurs in the first trial and event } B \text{ occurs in a second trial})$$

$$P(A \text{ and } B) = P(A) \bullet P(B)$$

Formal Multiplication Rule

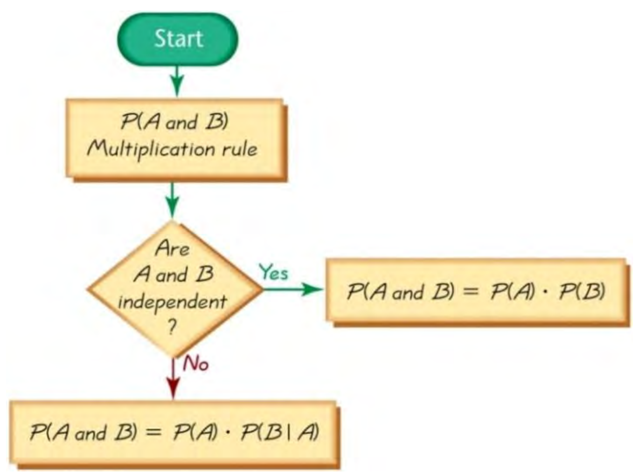
$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

if *A* and *B* are independent events, $P(B|A)$ is really the same as $P(B)$.

Intuitive Multiplication Rule

When finding the probability that event *A* occurs in one trial and event *B* occurs in the next trial, multiply the probability of event *A* by the probability of event *B*, but be sure that the probability of event *B* takes into account the previous occurrence of event *A*.

✓ *When applying the multiplication rule, always consider whether the events are independent or dependent, and adjust the calculations accordingly.*



Example

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) If 1 of the 98 test subjects is randomly selected, find the probability that the subject had a positive test result, given that the subject actually lied. That is, find *P*(positive test result | subject lied)

b) If 1 of the 98 test subjects is randomly selected, find the probability that the subject actually lied, given that he or she had a positive test result. That is, find *P* (subject lied | positive test result)

c) If two of the subjects are randomly selected without replacement, find the probability that the first selected person had a positive test result and the second selected person had a negative test result.

Solution

a) There are (42 + 9 =) 51 subjects that they lied, 42 has a positive test results.

$$P(\text{positive test result}|\text{lied}) = \frac{42}{51} \ \underline{= 0.824|}$$

Or

$$= \frac{\frac{42}{98}}{\frac{51}{98}} \ \underline{= 0.824|}$$

This indicates that a subject who lies has a 0.824 probability of gettin

$$P(\text{positive test result}|\text{lied}) = \frac{P(\text{lied and had a positive test result})}{P(\text{subject lied})}$$ g a positive test result.

b) There are (42 + 15 =) 57 with a positive test results subjects among that 42 lied.

$$P(\text{subject lied}|\text{positive test result}) = \frac{42}{57} \ \underline{= 0.737|}$$

This indicates that for a subject who gets a positive test result, there is a 0.737 probability that this subject actually lied.

c) $P(\text{positive test result}) = \frac{57}{98}$

After the first selection of a subject, there are 97 subjects remaining

$$P(\text{negative test result}) = \frac{41}{97}$$

$$P(1st\ \text{positive and 2nd negative test result}) = \frac{57}{98} \cdot \frac{41}{97} \ \underline{= 0.246|}$$

Confusion of the Inverse

To incorrectly believe that $P(A|B)$ and $P(B|A)$ are the same, or to incorrectly use one value for the other, is often called confusion of the inverse.

Bayes' Theorem

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

Bayes' Formula

$$P\left(U_1 \mid E\right) = \frac{P\left(U_1 \cap E\right)}{P(E)}$$

$$= \frac{P\left(U_1 \cap E\right)}{P\left(U_1 \cap E\right) + P\left(U_2 \cap E\right) + \cdots}$$

$$= \frac{P\left(E \mid U_1\right)P\left(U_1\right)}{P\left(E \mid U_1\right)P\left(U_1\right) + P\left(E \mid U_2\right)P\left(U_2\right) + \cdots}$$

Exercises Section 2.4 – Multiplication Rule and Conditional

1. Use the data below:

	No (Did Not Lie)	***Yes*** (Lied)
Positive test result	15 (*false positive*)	42 (*true positive*)
Negative test result	32 (*true negative*)	9 (*false negative*)

a) If 2 of the 98 test subjects are randomly selected without replacement find the probability that they both had false positive results. Is it unusual to randomly select 2 subjects without replacement and get 2 results that are both false positive results? Explain.

b) If 3 of the 98 test subjects are randomly selected without replacement, find the probability that all had false positive results. Is it unusual to randomly select 3 subjects without replacement and get 3 results that are all false positive results? Explain.

c) If 4 of the test subjects are randomly selected without replacement find the probability that, in each case, the polygraph indicated that the subject lied. Is such an event unusual?

d) If 4 of the test subjects are randomly selected without replacement find the probability that they all had incorrect test result (either false positive or false negative). Is such an event Likely?

e) Assume that 1 of the 98 test subjects is randomly selected. Find the probability of selecting a subject with a negative test result, given that the subject lied. What does this result suggest about the polygraph test?

f) Find $P(\text{negative test result} \mid \text{subject did not lie})$

g) Find $P(\text{subject did not lie} \mid \text{negative test result})$

2. Use the data in the table below

	Group			
Type	***O***	***A***	***B***	***AB***
Rh^+	39	35	8	4
Rh^-	6	5	2	1

a) If 2 of the 100 subjects are randomly selected, find the probability that they are both group O and type Rh^+

i. Assume that the selections are made with replacement.

ii. Assume that the selections are made without replacement.

b) If 3 of the 100 subjects are randomly selected, find the probability that they are both group B and type Rh^-

i. Assume that the selections are made with replacement.

ii. Assume that the selections are made without replacement.

c) People with blood that is group O and type Rh^{-} are considered to be universal donors, because they can give blood to anyone. If 4 of the 100 subjects are randomly selected, find the probability that they are all universal donors.
   i. Assume that the selections are made with replacement.
   ii. Assume that the selections are made without replacement.

d) People with blood that is group AB and type Rh^{+} are considered to be universal donors, because they can give blood to anyone. If 3 of the 100 subjects are randomly selected, find the probability that they are all universal recipients.
   i. Assume that the selections are made with replacement.
   ii. Assume that the selections are made without replacement.

3. With one method of a procedure called acceptance sampling, a sample of items is randomly selected without replacement and the entire batch is accepted if every item in the sample is okay. The Telektronics Company manufactured a batch of 400 backup power supply units for computers, and 8 of them are defective. If 3 of the units are randomly selected for testing, what is the probability that the entire batch will be accepted?

4. It is common for public opinion polls to have a "confidence level" of 95% meaning that there is a 095 probability that the poll results are accurate within the claimed margins of error. If each of the following organizations conducts an independent poll, find the probability that all of them are accurate within the claim margins of error: Gallup, Roper, Yankelovich, Harris, CNN, ABC, CBS, and NBC, New York Times. Does the result suggest that with a confidence level of 95%, we can expect that almost all polls will be within the claimed margin of error?

5. The principle of redundancy is used when system reliability is improved through redundant or back up components. Assume that your alarm clock has a 0.9 probability of working on any given morning.

a) What is the probability that your alarm clock will not work on the morning of an important final exam?

b) If you have 2 such alarm clocks, what is the probability that they both fail on the morning of an important final exam?

c) With one alarm clock, you have a 0.9 probability of being awakened. What is the probability of being awakened if you use two alarm clocks?

d) Does a second alarm clock result in greatly improved reliability?

6. The wheeling Tire Company produced a batch of 5,000 tires that includes exactly 200 that are defective.

a) If 4 tires are randomly selected for installation on a car, what is the probability that they are all good?

b) If 100 tires are randomly selected for shipment to an outlet, what is the probability that they are all good? Should this outlet plan to deal with defective tires returned by consumers?

7. When the 15 players on the LA Lakers basketball team are tested for steroids, at least one of them tests positive. Provide a written description of the complement of this event.

8. If a couple plans to have 6 children, what is the probability that they will have at least one girl? Is that probability high enough for the couple to be very confident that they will get at least one girl in six children?

9. If a couple plans to have 8 children (it could happen), what is the probability that they will have at least one girl? Is the couple eventually has 8 children and they are all boys, what can the couple conclude?

10. If you make guesses for 4 multiple-choice test questions (each with 5 possible answers), what is the probability of getting at least one correct? If a very lenient instructor says that passing test occurs if there is at least one correct answer, can you reasonably expect to pass by guessing?

11. Find the probability of a couple having a baby girl when their fourth child is born, given that the first 3 children were all girls. Is the result the same as the probability of getting 4 girls among 4 children?

12. In China, the probability of a baby being a boy is 0.5845. Couples are allowed to have only one child. If relatives give birth to 5 babies, what is the probability that there is at least one girl? Can that system continue to work indefinitely?

13. An experiment with fruit flies involves one parent with normal wings and one parent with vestigial wings. When these parents have an offspring, there is a $\frac{3}{4}$ probability that the offspring has normal wings and a $\frac{1}{4}$ probability of vestigial wings. If the parents give birth to 10 offspring, what is the probability that at least 1 of the offspring has vestigial wings? If researchers need at least one offspring with vestigial wings, can they be reasonably confident of getting one?

14. According to FBI data, 24.9% of robberies are cleared with arrests. A new detective is assigned to 10 different robberies.

a) What is the probability that at least 1 of them is cleared with an arrest?
b) What is the probability that the detective clears all 10 robberies with arrests?
c) What should we conclude if the detective clears all 10 robberies with arrests?

15. A statistics student wants to ensure that she is not late for an early statistics class because of a malfunctioning alarm clock. Instead of using one alarm clock, she decides to use three. What is the probability that at least one of her alarm clocks works correctly if each individual alarm clock has a 90% chance of working correctly? Does the student really gain much by using three alarm clocks instead on only one? How are the results affected if all of the alarm clocks run on electricity instead of batteries?

16. In a batch of 8,000 clock radios 8% are defective. A sample of 5 clock radios is randomly selected without replacement from the 8,000 and tested. The entire batch will be rejected if at least one of those tested is defective. Find the probability that the entire batch will be rejected.

17. In a blood testing procedure, blood samples from 3 people are combined into one mixture. The mixture will only test negative if all the individual samples are negative. If the probability that an individual sample tests positive is 0.1, find the probability that the mixture will test positive.

18. A sample of 4 different calculators is randomly selected from a group containing 16 that are defective and 36 that have no effects. Find the probability that at least one of the calculator is defective.

19. Among the contestants in a competition are 46 women and 29 men. If 5 winners are randomly selected, find the probability that they are all men?

20. A bin contains 60 lights bubs of which 7 are defective. If 4 light bulbs are randomly selected from the bin with replacement, find the probability that all the bulbs selected are good ones.

21. You are dealt two cards successively (without replacement) from a shuffled deck of 52 playing cards. Find the probability that both cards are black. Express your answer as a simplified fraction.

Section 2.5 – Counting Techniques

Fundamental Counting Rule

For a sequence of two events in which the first event can occur *m* ways and the second event can occur *n* ways, the events together can occur a total of $m \bullet n$ ways.

Example

It's wise not to disclose social security numbers, because they are often used by criminals attempting identity theft. Assume that a criminal is found using your social security number and claims that all of the digits were randomly generated. What is the probability of getting your social security number when randomly generated 9 digits? Is the criminal's claim that your number was randomly generated likely to be true?

Solution

Each of the 9 digits has 10 possible outcomes: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

$$10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 1{,}000{,}000{,}000$$

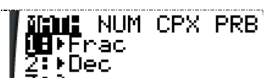
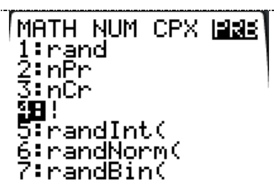
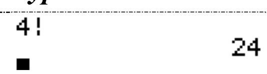
Notation

The ***factorial symbol*** (!) denotes the product of decreasing positive whole numbers.

For example,

$$0! = 1$$

$$4! = 4 \bullet 3 \bullet 2 \bullet 1 = 24$$

4	**Math**	***Select* PRB**	***Type* 4**	
				

Factorial Rule

A collection of *n* different items can be arranged in order *n*! different ways. (This factorial rule reflects the fact that the first item may be selected in *n* different ways, the second item may be selected in *n – 1* ways, and so on.)

Example

During the summer, you are planning to visit these 6 national parks: Glacier, Yellowstone, Yosemite, Arches, Zion, and Grand Canyon. You would like to plan the most efficient route and you decide to list all of the possible routes. How many different routes are possible?

Solution

There 6 different parks can be arranged in order 6! different ways.

$$6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720 \ \ \textit{ways}$$

Permutations ***Rule*** (*when items are all different*)

Requirements:

1. There are ***n* different** items available. (This rule does not apply if some of the items are identical to others.)
2. We select ***r*** of the ***n*** items (without replacement).
3. We consider rearrangements of the same items to be different sequences. (The permutation of ABC is different from CBA and is counted separately.)

If the preceding requirements are satisfied, the number of permutations (or sequences) of r items selected from n available items (without replacement) is

$$_nP_r = \frac{n!}{(n-r)!}$$

Example

In horse racing, a bet on an exacta in a race is won by correctly selecting the horses that finish first and second, and you must select those 2 horses in the correct order. The 132nd running of the Kentucky Derby has a field of 20 horses. If a bettor randomly selects 2 of those horses for an exacta bet, what is the probability of winning?

Solution

We have $n = 20$ horses and we must select $r = 2$ of them without replacement.

..

Permutations Rule (*when some items are identical to others*)

Requirements:

1. There are n items available, and some items are identical to others.
2. We select all of the n items (without replacement).
3. We consider rearrangements of distinct items to be different sequences.

If the preceding requirements are satisfied, and if there are n_1 alike, n_2 alike, . . . , n_k alike, the number of permutations (or sequences) of all items selected without replacement is

$$\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$$

Example

In a preliminary test of the MicroSort gender selection method developed by the Genetics and IVF Institute, 14 couples tried to have baby girls. Analysis of the effectiveness of the MicroSort method is based on a probability value, which in turn is based on numbers of permutations. Let's consider this simple problem: How many ways can 11 girls and 3 boys be arranged in sequence? That is, find the number of permutations of 11 girls and 3 boys.

Solution

$n = 14; \quad n_1 = 11; \quad n_2 = 3$

$$\frac{n!}{n_1! \cdot n_2!} = \frac{14!}{11! \cdot 3!}$$

$$= \frac{12 \cdot 13 \cdot 14}{1 \cdot 2 \cdot 3}$$

$$\underline{= 364}$$

There are 364 different ways to arrange 11 girls and 3 boys.

Example

In how many ways can the letters in the word *Mississippi* be arranged?

Solution

$$\frac{11!}{1!4!4!2!} \underline{= 34{,}650 \;\; ways}$$

m	i	s	p
1	4	4	2

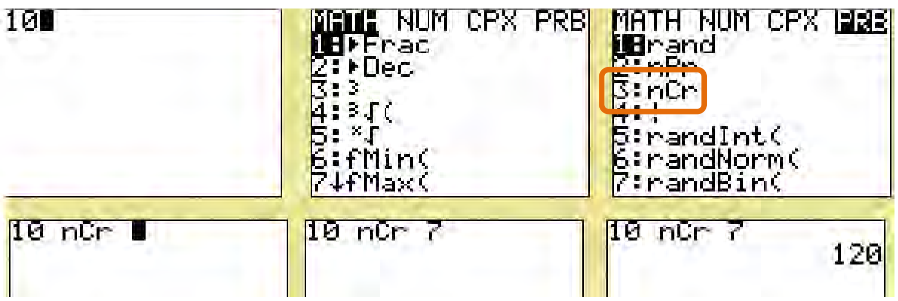
Combinations Rule

Requirements:

1. There are *n* different items available.
2. We select *r* of the *n* items (without replacement).
3. We consider rearrangements of the same items to be the same. (The combination of ABC is the same as CBA.)

If the preceding requirements are satisfied, the number of combinations of *r* items selected from *n* different items is

$$ {}_nC_r = \binom{r}{n} = \frac{n!}{(n-r)!r!} $$



Permutations versus Combinations

When different orderings of the same items are to be counted separately, we have a permutation problem, but when different orderings are not to be counted separately, we have a combination problem.

Permutation: *order matter*.

Combination: *Order doesn't matter*.

Examples

For each problem, tell whether permutations or combinations should be used to solve the problem.

a) How many 4-digit code numbers are possible if no digits are repeated?

Permutation

b) A sample of 3 light bulbs is randomly selected from batch of 15. How many different samples are possible?

Combination

c) In a baseball conference with 8 teams, how many games must be played so that each team plays every other team exactly once?

Combination

d) In how many ways can 4 patients be assigned to 6 different hospital rooms so that each patient has a private room?

Permutation

Example

A clinical rest on humans of a new drug is normally done in 3 phases. Phase I is conducted with a relatively small number of healthy volunteers. Let's assume that we want to treat 8 healthy humans with a new drug, and we have 10 suitable volunteers available.

a) If the subjects are selected and treated in sequence, so that the trial is discontinued if anyone displays adverse effects, how many different sequential arrangements are possible if 8 people are selected from the 10 that are available?

b) If 8 subjects are selected from the 10 that are available, and the 8 selected subjects are all treated at the same time, how many different treatment groups are possible?

Solution

a) Because order does count, we want the number of permutations of $r = 8$ people selected from the $n = 10$.

$$_{10}P_8 = 1,814,400$$

b) Because order does ***not*** count, we want the number of combinations of $r = 8$ people selected from the $n = 10$.

$$_{10}C_8 = 45$$

Example

The Florida Lotto game is typical of state lotteries. You must select 6 different numbers between 1 and 53. You win the jackpot if the same 6 numbers are drawn in any order. Find the probability of winning the jackpot.

Solution

Because order does ***not*** count, we want the number of combinations of $r = 6$ people selected from the $n = 53$

$$_{53}C_6 = 22,957,480$$

With 1 winning combination and 22,957,480 different possible combinations, the probability of winning the jackpot is

$$P = \frac{1}{22,957,480}$$

Exercises *Section* 2.5 – Counting Techniques

1. Decide whether the situation involves ***permutations*** or ***combinations***

a) A batting order for 9 players for a baseball game
b) An arrangement of 8 people for a picture
c) A committee of 7 delegates chosen from a class of 30 students to bring a petition to the administration
d) A selection of a chairman and a secretary from a committee of 14 people
e) A sample of 5 items taken from 71 items on an assembly line
f) A blend of 3 spices taken from 7 spices on a spice rack
g) From the 7 male and 10 female sales representatives for an insurance company, team of 8 will be selected to attend a national conference on insurance fraud.
h) Marbles are being drawn without replacement from a bag containing 15 marbles.
i) The new university president named 3 new officers a vice-president of finance, a vice-president of academic affairs, and a vice-president of student affairs.
j) A student checked out 4 novels from the library to read over the holiday.
k) A father ordered an ice cream cone (chocolate, vanilla, or strawberry) for each of his 4 children.

2. Find the number of different ways that five test questions can be arranged in order by evaluating 5!

3. In the game of blackjack played with one deck, a player is initially dealt 2 cards. Find the number of different two-card initial hands by evaluating ${}_{52}C_{2}$

4. A political strategist must visit state capitols, but she has time to visit only 3 of them. Find the number of different possible routes by evaluating ${}_{50}P_{3}$

5. Select the six winning numbers from 1, 2, …, 54. Find the probability of winning lottery by buying one ticket. $\left(of\ \ winning\ this\ lottery\ \ \dfrac{1}{575,757} \right)$

6. Select the five winning numbers from 1, 2, …, 36. Find the probability of winning lottery by buying one ticket. $\left(of\ \ winning\ this\ lottery\ \ \dfrac{1}{575,757} \right)$

7. In a club with 9 male and 11 female members, how many 5-member committees can be chosen that have

a) All men? *b)* All women? *c)* 3 men and 2 women?

8. In a club with 9 male and 11 female members, how many 5-member committees can be selected that have

a) At least 4 women?
b) No more than 2 men?

9. In how many ways can 5 out of 9 plants be arranged in a row on a windowsill?

10. From a pool of 8 secretaries, 3 are selected to be assigned to 3 managers, one per manager. In how many ways can they be selected and assigned?

11. A group of 9 workers decides to send a delegation of 3 to their supervisor to discuss their grievances.

a) How many delegations are possible?
b) If it is decided that a particular worker must be in the delegation, how many different delegations are possible?
c) If there are 4 women and 5 men in the group, how many delegations would include at least 1 woman?

12. Hamburger Hut sells regular hamburgers as well as a larger burger. Either type can include cheese, relish, lettuce, tomato, mustard, or catsup.

a) How many different hamburgers can be ordered with exactly three extras?
b) How many different regular hamburgers can be ordered with exactly three extras?
c) How many different regular hamburgers can be ordered with at least five extras?

13. In an experiment on plant hardiness, a researcher gathers 6 wheat plants, 3 barley plants, and 2 rye plants. She wishes to select 4 plants at random.

a) In how many ways can this be done?
b) In how many ways can this be done if exactly 2 wheat plants must be included?

14. A legislative committee consists of 5 Democrats and 4 Republicans. A delegation of 3 is to be selected to visit a small Pacific island republic.

a) How many different delegations are possible?
b) How many delegations would have all Democrats?
c) How many delegations would have 2 Democrats and 1 Republican?
d) How many delegations would have at least 1 Republican?

15. Five cards are chosen from an ordinary deck to form a hand in poker. In how many ways is it possible to get the following results?

a) 4 queens
b) No face card
c) Exactly 2 face cards
d) At least 2 face cards
e) 1 heart, 2 diamonds, and 2 clubs

16. The student sitting next to you in class concludes that the probability of the ceiling falling down on both of you before class ends is 1/2, because there are two possible outcomes - the ceiling will fall or not fall. What is wrong with this reasoning?

17. Identity theft often begins by someone discovering your 9-digit social security number. Answer each of the following. Express probabilities as fractions.

a) What is the probability of randomly generating 9 digits and getting your social security number?
b) In the past, many teachers posted grades along with the last 4 digits of your social security number, what is the probability that if they randomly generated the order digits, they would match yours? Is that something to worry about?

18. Credit card numbers typically have 16 digits, but not all of them are random. Answer the following and express probabilities as fractions.

a) What is the probability of randomly generating 16 digits and getting your MasterCard number?

b) Receipts often show the last 4 digits of a credit card number. If those last 4 digits are known, what is the probability of randomly generating the order digits of your MasterCard number?

c) Discover cards begin with the digits 6011. If you also know the last 4 digits, what is the probability of randomly generating the other digits and getting all of them correct? Is this something to worry about?

19. When testing for current in a cable with five color-coded wires, the author used a meter to test two wires at a time. How many different tests are required for every possible pairing of two wires?

20. The starting 4 players for the Boston Celtics basketball team have agreed to make charity appearances tomorrow night. If you must send three players to the United Way event and the other 2 to a Heart Fund event, how many different ways can you make the assignments?

21. In phase I of a clinical trial with gene therapy used for treating HIV, 5 subjects were treated (based on data from Medical News Today). If 20 people were eligible for the Phase I treatment and a simple random of 5 is selected, how many different simple random samples are possible? What is the probability of each simple random sample?

22. Many newspapers carry "Jumble" a puzzle in which the reader must unscramble letters to form words. The letters BUJOM were included in newspapers. How many ways can the letters if BUJOM be arranged? Identify the correct unscrambling, then determine the probability of getting that result by randomly selecting one arrangement of the given letters.

23. There are 11 members on the board of directors for the Coca Cola Company.

a) If they must select a chairperson, first vice chairperson, second vice chairperson, and secretary, how many different slates of candidates are possible?

b) If they must form an ethics subcommittee of 4 members, how many different subcommittees are possible?

24. The author owns a safe in which he stores his book. The safe combination consists of 4 numbers between 0 and 99. If another author breaks in and tries to steal this book, what is the probability that he or she will get the correct combination on the first attempt? Assume that the numbers are randomly selected. Given the number of possibilities, does it seem feasible to try opening the safe by making random guesses for the combination?

25. In a preliminary test of the MicroSort gender selection method, 14 babies were born and 13 of them were girls

a) Find the number of different possible sequences of genders that are possible when 14 babies are born.

b) How many ways can 13 girls and 1 boy be arranged in a sequence?

c) If 14 babies are randomly selected, what is the probability that they consist of 13 girls and 1 boy?

d) Does the gender-selection method appear to yield a result that is significantly different from a result that might be expected by random chance?

26. You become suspicious when a genetics researcher randomly selects groups of 20 newborn babies and seems to consistently get 10 girls and 10 boys. The researchers claims that it is common to get 10 girls and 10 boys in such cases,

a) If 20 newborn babies are randomly selected, how many different gender sequences are possible.

b) How many different ways can 10 girls and 10 boys be arranged in sequence?

c) What is the probability of getting 10 girls and 10 boys when 10 babies are born?

d) Based on the preceding results, do you agree with the researcher's explanation that ir is common to get 10 girls and 10 boys when 20 babies are randomly selected?

27. The Powerball lottery is run in 29 states. Winning the jackpot requires that you select the correct five numbers between 1 and 55 and, in a separate drawing, you must also select the correct single number between 1 and 42. Find the probability of winning the jackpot.

28. The Mega Millions lottery is run in 12 states. Winning the jackpot requires that you select the correct 5 numbers between 1 and 56 and, in a separate drawing, you must also select the correct single number between 1 and 46. Find the probability of winning the jackpot.

29. A state lottery involves the random selection of six different numbers between 1 and 31. If you select one six number combination, what is the probability that it will be the winning combination?

30. How many ways can 6 people be chosen and arranged in a straight line if there are 8 people to choose from?

31. 12 wrestlers compete in a competition. If each wrestler wrestles one match with each other wrestler, what are the total numbers of matches?

32. Wing has different books to arrange on a shelf: 4 blue, 3 green, and 2 red.

a) In how many ways can the books be arranged on a shelf?

b) If books of the same color are to be grouped together, how many arrangements are possible?

c) In how many distinguishable ways can the books be arranged if books of the same color are identical but need not be grouped together?

d) In how many ways can you select 3 books, one of each color, if the order in which the books are selected does not matter?

e) In how many ways can you select 3 books, one of each color, if the order in which the books are selected matters?

33. A child has a set of differently shaped plastic objects. There are 3 pyramids, 4 cubes, and 7 spheres.

a) In how many ways can she arrange the objects in a row if each is a different color?

b) How many arrangements are possible if objects of the same shape must be grouped together and each object is a different color?

c) In how many distinguishable ways can the objects be arranged in a row if objects of the same shape are also the same color, but need not be grouped together?

d) In how many ways can you select 3 objects, one of each shape, if the order in which the objects are selected does not matter and each object is a different color?

e) In how many ways can you select 3 objects, one of each shape, if the order in which the objects are selected matters and each object is a different color?

34. Twelve drugs have been found to be effective in the treatment of a disease. It is believed that the sequence in which the drugs are administered is important in the effectiveness of the treatment. In how many different sequences can 5 of the 12 drugs be administered?

35. In a club with 16 members, how many ways can a slate of 3 officers consisting of president, vice-president, and secretary/treasurer be chosen?

36. In how many ways can 7 of 11 monkeys be arranged in a row for a genetics experiment?

37. In an experiment on social interaction, 6 people will sit in 6 seats in a row. In how many ways can this be done?

38. In an election with 3 candidates for one office and 6 candidates for another office, how many different ballots may be printed?

39. A business school gives courses in typing, shorthand, transcription, business English, technical writing, and accounting. In how many ways can a student arrange a schedule if 3 courses are taken? Assume that the order in which courses are schedules matters.

40. If your college offers 400 courses, 25 of which are in mathematics, and your counselor arranges your schedule of 4 courses by random selection, how many schedules are possible that do not include a math course? Assume that the order in which courses are scheduled matters.

41. A baseball team has 19 players. How many 9-player batting orders are possible?

42. A chapter of union Local 715 has 35 members. In how many different ways can the chapter select a president, a vice-president, a treasurer, and a secretary?

43. A concert to raise money for an economics prize is to consist of 5 works; 2 overtures, 2 sonatas, and a piano concerto.

a) In how many ways can the program be arranged?

b) In how many ways can the program be arranged if an overture must come first?

44. A zydeco band from Louisiana will play 5 traditional and 3 original Cajun compositions at a concert. In how many ways can they arrange the program if

a) it begins with a traditional piece?

b) An original piece will be played last?

45. Given the set $\{A, B, C, D\}$, how many permutations are there of this set of 4 object taken 2 at a time?

a) Using the multiplication principle

b) Using the Permutation

46. Find the number of permutations of 30 objects taken 4 at a time.

47. Five cards are marked with the numbers 1, 2, 3, 4, and 5, then shuffled, and 2 cards are drawn.

a) How many different 2-card combinations are possible?

b) How many 2-card hands contain a number less than 3?

48. An economics club has 31 members.

a) If a committee of 4 is to be selected, in how many ways can the selection be made?

b) In how many ways can a committee of at least 1 and at most 3 be selected?

49. Use a tree diagram for the following

a) Find the number of ways 2 letters can be chosen from the set $\{L, M, N\}$ if order is important and repetition is allowed.

b) Reconsider part a if no repeats are allowed

c) Find the number of combinations of 3 elements taken 2 at a time. Does this answer differ from part *a* or *b*?

For each problem, decide whether permutations or combinations should be used to solve the problem.

50. In a club with 9 male and 11 female members, how many 5-member committees can be chosen that have

a) All men? *b)* All women? *c)* 3 men and 2 women?

51. In a club with 9 male and 11 female members, how many 5-member committees can be selected that have

a) At least 4 women? *b)* No more than 2 men?

52. In a game of musical chairs, 12 children will sit in 11 chairs arranged in a row (one will be left out). In how many ways can this happen, if we count rearrangements of the children in the chairs as different outcomes?

53. A group of 3 students is to be selected from a group of 14 students to take part in a class in cell biology.

a) In how many ways can this be done?

b) In how many ways can the group who will not take part be chosen?

54. Marbles are being drawn without replacement from a bag containing 16 marbles.

a) How many samples of 2 marbles can be drawn?

b) How many samples of 2 marbles can be drawn?

c) If the bag contains 3 yellow, 4 white, and 9 blue marbles, how many samples of 2 marbles can be drawn in which both marbles are blue?

55. There are 7 rotten apples in a crate of 26 apples

a) How many samples of 3 apples can be drawn from the crate?

b) How many samples of 3 could be drawn in which all 3 are rotten?

c) How many samples of 3 could be drawn in which there are two good apples and one rotten one?

56. A bag contains 5 black, 1 red, and 3 yellow jelly beans; you take 3 at random. How many samples are possible in which the jelly beans are

a) All black?
b) All red?
c) All yellow?
d) 2 black and 1 red?
e) 2 black and 1 yellow?
f) 2 yellow and 1 black?
g) 2 red and 1 yellow?

57. In how many ways can 5 out of 9 plants be arranged in a row on a windowsill?

58. From a pool of 8 secretaries, 3 are selected to be assigned to 3 managers, one per manager. In how many ways can they be selected and assigned?

59. A salesperson has the names of 6 prospects.

a) In how many ways can she arrange her schedule if she calls on all 6?
b) In how many ways can she arrange her schedule if she can call on only 4 of the 6?

60. A group of 9 workers decides to send a delegation of 3 to their supervisor to discuss their grievances.

a) How many delegations are possible?
b) If it is decided that a particular worker must be in the delegation, how many different delegations are possible?
c) If there are 4 women and 5 men in the group, how many delegations would include at least 1 woman?

61. Hamburger Hut sells regular hamburgers as well as a larger burger. Either type can include cheese, relish, lettuce, tomato, mustard, or catsup.

a) How many different hamburgers can be ordered with exactly three extras?
b) How many different regular hamburgers can be ordered with exactly three extras?
c) How many different regular hamburgers can be ordered with at least five extras?

62. Five items are to be randomly selected from the first 50 items on an assembly line to determine the defect rate. How many different samples of 5 items can be chosen?

63. From a group of 16 smokers and 22 nonsmokers, a researcher wants to randomly select 8 smokers and 8 nonsmokers for a study. In how many ways can the study group be selected?

64. In an experiment on plant hardiness, a researcher gathers 6 wheat plants, 3 barley plants, and 2 rye plants. She wishes to select 4 plants at random.

a) In how many ways can this be done?
b) In how many ways can this be done if exactly 2 wheat plants must be included?

65. A legislative committee consists of 5 Democrats and 4 Republicans. A delegation of 3 is to be selected to visit a small Pacific island republic.

a) How many different delegations are possible?
b) How many delegations would have all Democrats?
c) How many delegations would have 2 Democrats and 1 Republican?
d) How many delegations would have at least 1 Republican?

66. From 10 names on a ballot, 4 will be elected to a political party committee. In how many ways can the committee of 4 be formed if each person will have a different responsibility, and different assignments of responsibility are considered different committees?

67. How many different 13-card bridge hands can be selected from an ordinary deck?

68. Five cards are chosen from an ordinary deck to form a hand in poker. In how many ways is it possible to get the following results?

f) 4 queens

g) No face card

h) Exactly 2 face cards

i) At least 2 face cards

j) 1 heart, 2 diamonds, and 2 clubs

69. In poker, a flush consists of 5 cards with the same suit, such as 5 diamonds.

a) Find the number of ways of getting a flush consisting of cards with values from 5 to 10 by listing all the possibilities.

b) Find the number of ways of getting a flush consisting of cards with values from 5 to 10 by using combinations

70. If a baseball coach has 5 good hitters and 4 poor hitters on the bench and chooses 3 players at random, in how many ways can be choose at least 2 good hitters?

71. The coach of a softball team has 6 good hitters and 8 poor hitters. He chooses 3 hitters at random.

a) In how many ways can he choose 2 good hitters and 1 poor hitter?

b) In how many ways can he choose 3 good hitters?

c) In how many ways can he choose at least 3 good hitters?

72. How many 5 card hands will have 3 aces and 2 kings?

73. How many 5 card hands will have 3 hearts and 2 spades?

74. 2 letters follow by 3 numbers; 2 letters out of 8 & 3 numbers out of 10

75. Serial numbers for a product are to be made using 3 letters follow by 2 digits (0 – 9 no repeats). If the letters are to be taken from the first 8 letters of the alphabet with no repeats, how many serial numbers are possible?

76. A company has 7 senior and 5 junior officers. An ad hoc legislative committee is to be formed.

a) How many 4-officer committees with 1 senior officer and 3 junior officers can be formed?

b) How many 4-officer committees with 4 junior officers can be formed?

c) How many 4-officer committees with at least 2 junior officers can be formed?

77. From a committee of 12 people,

a) In how many ways can we choose a chairperson, a vice-chairperson, a secretary, and a treasurer, assuming that one person can’t hold more than on position

b) In how many ways can we choose a subcommittee of 4 people?

78. Find the number of combinations of 30 objects taken 4 at a time.

79. How many different permutations are the of the set $\{a, b, c, d, e, f, g\}$?

80. How many permutations of $\{a, b, c, d, e, f, g\}$ end with a?

81. Find the number of 5-permutations of a set with nine elements

82. In how many different orders can five runners finish a race if no ties are allowed?

83. A coin flipped eight times where each flip comes up either heads or tails. How many possible outcomes

a) Are there in total?
b) Contain exactly three heads?
c) Contain at least three heads?
d) Contain the same number of heads and tails?

84. In how many ways can a set of two positive integers less than 100 be chosen?

85. In how many ways can a set of five letters be selected from the English alphabet?

Section 2.6 – Discrete Random Variables

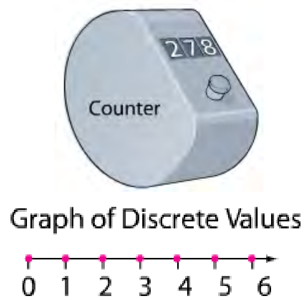
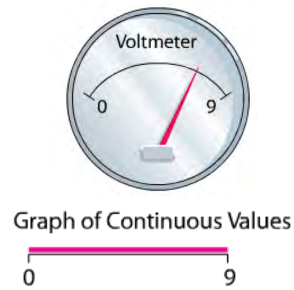
Defintions

A ***random variable*** is a variable (typically represented by x) that has a single numerical value, determined by chance, for each outcome of a procedure

A ***discrete random*** variable hass either a finite number of values or countable number of values, where "countable" refers to the fact that there might be infinitely many values, but they result from a counting process

A ***continuous random*** variable has infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions

A ***Probability distribution*** is a description that gives the probability for each value of the random variable; often expressed in the format of a graph, table, or formula

	
Discrete Random Variable: Count of the number of movies patrons.	***Continuous Random Variable***: The measured voltage of a smoke detector battery

Example

The following are example of discrete and continuous radom variables.

1. ***Discrete***: Let x = the number of eggs that a hen lays in a day. This is a *discrete* random variable because its only possible values are 0, or 1, or 2, and so on. No han can lay 2.343115 eggs, which would have been possible if the data had come from a continuous scale.
2. ***Discrete***: The count of the number of statistics students present in class on a given day is a whole number and is therefore a discrete random variable. The counting device is capable of indicating only a finite number of values, so it is used to obtain values for a *discrete* random variable.
3. ***Continuous***: Let x = the amount of milk a cow produces in one day. This is a *continuous* random variable because it can have any value over a continuous span. During a single day, a cow might yield an amount of milk that can be any value between 0 gallon and 5 gallons. It owould be 4.123 gallons, because the coe is not restricted to the discrete amounts of 0, 1, 2, 3, 4, or 5 gallons.
4. ***Continuous***: The measure of voltage for a particular smoke detector battery can be any value between 0 volt and 9 volts. It is therefore a *continuous* random variable.

Requirements for Probability Distribution

- ✓ $\sum P(x)=1$ where x assumes all possible values. The sum of all probailities must be 1. (but sunch 0.999 ot 1.001 are acceptable)
- ✓ $0\le P(x)\le 1$ for every individual value of x

Example

Is the following a probability distribution?

x	$P(x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.30
5	-0.01

No. $P(x=5)=-0.01$

x	$P(x)$
0	0.001
1	0.015
2	0.088
3	0.264
4	0.396
5	0.237

Yes. $\sum P(x)=1.001\approx 1$

x	$P(x)$
0	0.16
1	0.18
2	0.22
3	0.10
4	0.30
5	0.01

No. $\sum P(x)=0.97<1$

Example

Does $P(x)=\frac{x}{10}$ (where x can be 0, 1, 2, 3, or 4) determine a probability distribution?

Solution

$P(0)=\frac{0}{10}=0;\quad P(1)=\frac{1}{10};\quad P(2)=\frac{2}{10};\quad P(3)=\frac{3}{10};\quad and \quad P(4)=\frac{4}{10}$

$$\begin{aligned}\sum P(x)&=P(0)+P(1)+P(2)+P(3)+P(4)\\&=\frac{0}{10}+\frac{1}{10}+\frac{2}{10}+\frac{3}{10}+\frac{4}{10}\\&=\frac{10}{10}\\&=1\end{aligned}$$

Each value of the $P(x)$ is beteween 0 and 1.

Because both requirements are satisfied, the formula given is a probability distribution.

Mean, Variance and Standard Deviation of a Probability Distribution

$\mu = \sum\left[x \cdot P(x)\right]$	Mean of Discrete random variable
$\sigma^2 = \sum\left[(x-\mu)^2 \cdot P(x)\right]$	Variance for a probability distribution (easier to understand)
$\sigma^2 = \left[\sum x^2 \cdot P(x)\right] - \mu^2$	Variance for a probability distribution (easier computations)
$\sigma = \sqrt{\left[\sum x^2 \cdot P(x)\right] - \mu^2}$	Standard deviation for a probability distribution

Example

Find the mean, variance, and standard deviation for the probaiblity distribution described in the table.

x (# of peas)	$P(x)$
0	0.001
1	0.015
2	0.088
3	0.264
4	0.396
5	0.237

Solution

Mean: $\mu = \sum\left[x \cdot P(x)\right]$

$$= 0(0.001) + 1(0.015) + 2(0.088) + 3(0.264) + 4(0.396) + 5(0.237)$$

$$\underline{= 3.752}$$

Variance: $\sigma^2 = \sum\left[(x-\mu)^2 \cdot P(x)\right]$

$$= (0-3.8)^2(0.001) + (1-3.8)^2(0.015) + (2-3.8)^2(0.088) + (3-3.8)^2(0.264)$$

$$+ (4-3.8)^2(0.396) + (5-3.8)^2(0.237)$$

$$\underline{= 0.940574}$$

Standard deviation: $\sigma = \sqrt{0.940574}\ \underline{= 0.9698}$

Expected Value

The expected value of a discrete random variable is denoted by E, and it represents the mean value of the outcomes. It is obtained by finding the value of $\sum\left[x \cdot P(x)\right]$

$$E = \sum\left[x \cdot P(x)\right]$$

Example

You areconsidering placing a bet either on the number 7 in roulette or on the "pass line" in the dice game of craps at the casino.

a) If you bet \$5 on the number 7 in roulette, the probability of losing \$5 is $\frac{37}{38}$ and the probability of making a net gain of \$175 is $\frac{1}{38}$. (The prize id \$180, including your \$5 bet, so the net gain is \$175.) Find your expected value if you bet \$5 on the number 7 in roulette.

b) If you bet \$5 on the pass line in the dice game of craps, the probability of losing \$5 is $\frac{251}{495}$ and the probability of making a net gain of \$5 is $\frac{244}{495}$. (If you bet \$5 on the Pass Line and win, you are given \$10 that includes your bet, so the net gain is \$5.) Find your expected value if you bet \$5 on the pass line in the dice game? Why?.

Solution

a)

Event	x	$P(x)$	$x \cdot P(x)$
Lose	$-\$5$	$\frac{37}{38}$	$-\$4.87$
Gain (net)	\$175	$\frac{1}{38}$	\$4.61
Total			$-\$0.26$ ***Or*** $-26¢$

You can expect to lose an average of **26¢**.

b)

Event	x	$P(x)$	$x \cdot P(x)$
Lose	$-\$5$	$\frac{251}{495}$	$-\$2.54$
Gain (net)	\$5	$\frac{244}{495}$	\$2.46
Total			$-\$0.08$ ***Or*** $-8¢$

You can expect to lose an average of 8¢.

Exercises Section 2.6 – Discrete Random Variables

1. Determine whether or not a probability distribution is given. If a probability is given, find its mean and standard deviation. If the probability is not given, identify the requirements that are not satisfied.

a)

x	$P(x)$
0	0.125
1	0.375
2	0.375
3	0.125

b)

x	$P(x)$
0	0.22
1	0.16
2	0.21
3	0.16

c)

x	$P(x)$
0	0.528
1	0.360
2	0.098
3	0.013
4	0.001
5	0^+

0^+ *denotes a positive probability value that is very small.*

d)

x	$P(x)$
0	0.02
1	0.15
2	0.29
3	0.26
4	0.16
5	0.12

2. Based on past results found in the *Information Please Almanac*, there is a 0.1919 probability that a baseball World Series context will last 4 games, is a 0.2121 probability that it will last 5 games, a 0.2222 probability that it will last 6 games, a 0.3737 probability that us will last 7 games.

a) Does the given information describe a probability distribution?

b) Assuming that the given information describes a probability distribution, find the mean and standard deviation for the numbers of games in World Series contests.

c) Is it unusual for a team to "sweep" by winning in four games? Why or why not?

3. Based on information from MRI Network, some job applicants are required to have several interviews before a decision is made. The number of required interviews and the corresponding probabilities are: 1 (0.09); 2 (0.31); 3 (0.37); 4 (0.12); 5 (0.05); 6 (0.05).

a) Does the given information describe a probability distribution?

b) Assuming that a probability distribution is described, find its mean and standard deviation.

c) Use the range rule of thumb to identify the range of values for usual numbers of interviews.

d) Is it unusual to have a decision after just one interview? Explain?

4. Based on information from Car dealer, when a car is randomly selected the number of bumper stickers and the corresponding probabilities are: 0 (0.824); 1 (0.083); 2 (0.039); 3 (0.014); 4 (0.012); 5 (0.008); 6 (0.008); 7 (0.004); 8 (0.004); 9 (0.004).

a) Does the given information describe a probability distribution?

b) Assuming that a probability distribution is described, find its mean and standard deviation.
c) Use the range rule of thumb to identify the range of values for usual numbers of bumper stickers.
d) Is it unusual for a car to have more than one bumper sticker? Explain?

5. A Company hired 8 employees from a large pool of applicants with an equal numbers of males and females. If the hiring is done without regard to sex, the numbers of females hired and the corresponding probabilities are: 0 (0.004); 1 (0.0313); 2 (0.109); 3 (0.219); 4 (0.273); 5 (0.219); 6 (0.109); 7 (0.031); 8 (0.004).
a) Does the given information describe a probability distribution?
b) Assuming that a probability distribution is described, find its mean and standard deviation.
c) Use the range rule of thumb to identify the range of values for usual numbers of females hired in such groups of 8.
d) If the most recent group of 8 newly hired employees does not include any females, does there appear to be discrimination based on sex? Explain?

6. Let the random variable x represent the number of girls in a family of 4 children. Construct a table describing the probability distribution; then find the mean and the standard deviation. (Hint: List the different possible outcomes.) Is it unusual for a family of 3 children to consist of 3 girls?

7. In 4 lottery game, you pay 50¢ to select a sequence of 4 digits, such 1332. If you select the same sequence of 4 digits that are drawn, you win and collect \$2788.
a) How many different selections are possible?
b) What is the probability of winning?
c) If you win, what is your net profit?
d) Find the expected value.

8. When playing roulette at casino, a gambler is trying to decide whether to bet \$5 on the number 13 or bet \$5 that the outcomes any one of these 5 possibilities: 0 or 00 or 1 or 2 or 3.the expected value of the \$5 bet for a single number is –26¢. For the \$5 bet that the outcome 0 or 00 or 1 or 2 or 3, there is a probability of $\frac{5}{38}$ of making a net profit of \$30 and a $\frac{33}{38}$ probability of losing \$5.
a) Find the expected value for the \$5 bet that the outcome is 0 or 00 or 1 or 2 or 3.
b) Which bet is better: A \$5 bet on the number 13 or a \$5 bet the outcome is 0 or 00 or 1 or 2 or 3? Why?

9. There is a 0.9986 probability that a randomly selected 30-year-old male lives through the year. As insurance company charges \$161 for insuring that the male will live through the year. If the male does not survive the year, the policy pays out \$100,000 as a death benefit.
a) From the perspective of the 30-year-old male, what are the values corresponding to the 2 events of surviving the year and not surviving?
b) If a 30-year-old male purchases the policy, what is his expected value?
c) Can the insurance company expect to make a profit from many such policies? Why?

10. An insurance company charges a 21-year-old male a premium of $500 for a one-year $100,000 life insurance policy. A 21-year-old male has a 0.9985 probability of living for a year.

a) From the perspective of a 21-year-old male (or estate), what are the values of the two different outcomes?

b) What is the expected value for a 21-year-old male who buys the insurance?

c) What would be the cost of the insurance if the company just breaks even (in the long run with many such policies), instead of making a profit?

d) Given that the expected value is negative (so the insurance company can make a profit), why

Section 2.7 – Binomial Probability Distribution

Definition

A binomial probability distribution results from a procedure that meets all the following ***requirements***:

1. The procedure has a fixed number of trials.
2. The trials must be independent. (The outcome of any individual trial doesn't affect the probabilities in the other trials.)
3. Each trial must have all outcomes classified into two categories (commonly referred to as success and failure).
4. The probability of a success remains the same in all trials.

Notation for Binomial Probability Distributions

S and *F* (*success* and *failure*) denote the two possible categories of all outcomes; *p* and *q* will denote the probabilities of *S* and *F*, respectively, so

$P(\text{S}) = p$ (p = probability of success)

$P(\text{F}) = 1 - p = q$ (q = probability of failure)

n denotes the fixed number of trials.

x denotes a specific number of successes in *n* trials, so *x* can be any whole number between 0 and *n*, inclusive.

p denotes the probability of success in *one* of the *n* trials.

q denotes the probability of failure in one of the *n* trials.

***P*(*x*)** denotes the probability of getting exactly *x* successes among the *n* trials.

Example

Consider an experiment in which 5 offspring peas are generated each having the green/yellow combination of genes for pod color. The probability of an offspring pea with a green pod is 0.75. That is, $P(\textit{green pod}) = 0.75$. Suppose we want to find the probability that exactly 3 of the 5 offspring peas have a green pod.

a) Does this procedure result in a binomial distribution?

b) If this procedure result in a binomial distribution, identify the values of *n, x, p*, and *q*.

Solution

a) 1. The number of trials (5) is fixed.

2. The 5 trials are independent, because the probability of any offspring pea having a green pod is not affected by the outcome of any other offspring pea.

3. Each of the 5 trials has 2 categories of outcomes: The pea has a green pod or it does not.

4. For each offspring pea, the probability that it has a green pod is 0.75, and that probability remains the same for each of the 5 peas.

b) Having concluded that the given procedure does result in a binomial distribution, we now proceed to identify the values of n, x, p, and q.

1. With 5 offspring peas, we have $n = 5$.
2. We want the probability of exactly 3 peas with green pods, so $x = 3$.
3. The probability of success (getting a pea with a green pod) for one selection is 0.75, so $p = 0.75$.
4. The probability of failure (not getting a green pod) is 0.25, so $q = 0.25$..

Important Hints

- Be sure that x and p both refer to the same category being called a success.
- When sampling without replacement, consider events to be independent if $n < 0.05N$.

Example

Which of the following are binomial experiments?

a) A player rolls a pair of fair die 10 times. The number X of 7's rolled is recorded.

b) The 11 largest airlines had an on-time percentage of 84.7% in November, 2001 according to the Air Travel Consumer Report. In order to assess reasons for delays, an official with the FAA randomly selects flights until she finds 10 that were not on time. The number of flights X that need to be selected is recorded.

c) In a class of 30 students, 55% are female. The instructor randomly selects 4 students. The number X of females selected is recorded.

Solution

a) Binomial experiment

b) Not a binomial experiment – not a fixed number of trials.

c) Not a binomial experiment – the trials are not independent.

Method 1: Using the Binomial Probability Formula

$$P(x) = \frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x} \quad \text{for} \quad x = 0,\ 1,\ 2,\ \ldots,\ n$$

$$P(x\ success) = C_{n,x}\, p^x q^{n-x} \quad or \quad P(x\ success) = C_{n,x}\, p^x (1-p)^{n-x}$$

where

n = number of trials

x = number of successes among n trials

p = probability of success in any one trial

q = probability of failure in any one trial ($q = 1 - p$)

Example

The probability of an offspring pea with a green pod is 0.75, use the binomial probability formula to find the probability that exactly 3 peas with green pods when of the 5 offspring peas have a green pod when 5 offspring peas are generated. That is, find $P(3)$ given that $n = 5$, $x = 3$, $p = 0.75$, and $q = 0.25$.

Solution

$$P(x=3) = {}_5C_3 \cdot (0.75)^3 \cdot (0.25)^{5-3}$$

$$\underline{= 0.263671875}$$

The probability of getting exactly 3 peas with green pods among 5 offspring peas is 0.264,

TI-83/84 PLUS Press **2nd VARS** (to get **DISTR,** which denotes "distributions"), then select the option identified as **binompdf(.** Complete the entry of **binompdf(n, p, x)** with specific values for *n*, *p*, and *x*, then press **ENTER.** The result will be the probability of getting *x* successes among *n* trials.

You could also enter **binompdf(n, p)** to get a list of *all* of the probabilities corresponding to $x = 0, 1, 2, \ldots, n$. You could store this list in L2 by pressing **STO** → **L2.** You could then manually enter the values of 0, 1, 2, . . . , *n* in list L1, which would allow you to calculate statistics (by entering **STAT, CALC,** then **L1, L2**) or view the distribution in a table format (by pressing **STAT,** then **EDIT**).

The command **binomcdf** yields *cumulative* probabilities from a binomial distribution. The command **binomcdf(n, p, x)** provides the sum of all probabilities from $x = 0$ through the specific value entered for *x*.

Example

The fast food chain McDonald's has a brand name recognition rate of 95% around the world. Assuming that we randomly select 5 people, use the table to find the following

a) The probability that exactly 3 of the 5 people recognize McDonald's
b) The probability that the number of people who recognize McDonald's is 3 or fewer

n	*x*	*P(x)* =.95
5	0	0+
4	1	0+
3	2	0.001
2	3	0.021
1	4	0.204
0	5	0.774

Solution

a) $P(x=3) = .021$

b) $P(3 \text{ or fewer}) = P(3) + P(2) + P(1) + P(0)$

$$= 0.021 + 0.001 + 0 + 0$$

$$\underline{= 0.02259}$$

Phrase	*Math Symbol*
"more than" ***or*** "greater than"	$>$
"fewer than" ***or*** "less than"	$<$
"no more than" ***or*** "at most" ***or*** "less than or equal to	$\leq$
"at least" ***or*** "no less than" ***or*** "greater than or equal to"	$\geq$
"exactly" ***or*** "equals" ***or*** "is"	$=$

Example

According to the Experian Automotive, 35% of all car-owning households have three or more cars.

a) In a random sample of 20 car-owning households, what is the probability that exactly 5 have three or more cars?

b) In a random sample of 20 car-owning households, what is the probability that less than 4 have three or more cars?

c) In a random sample of 20 car-owning households, what is the probability that at least 4 have three or more cars?

Solution

a) $P(5) = {}_{20}C_5 (.35)^5 (1-.35)^{20-5} \ \underline{= 0.1272}$

b) $P(X < 4) = P(0) + P(1) + P(2) + P(3)$

$$= {}_{20}C_0 (.35)^0 (.65)^{20} + {}_{20}C_1 (.35)^1 (.65)^{19} + {}_{20}C_2 (.35)^2 (.65)^{18} + {}_{20}C_3 (.35)^3 (.65)^{17}$$

$$\underline{= 0.0444}$$

c) $P(X \geq 4) = 1 - P(X \leq 3)$

$$= 1 - .0444$$

$$\underline{= 0.9556}$$

Example

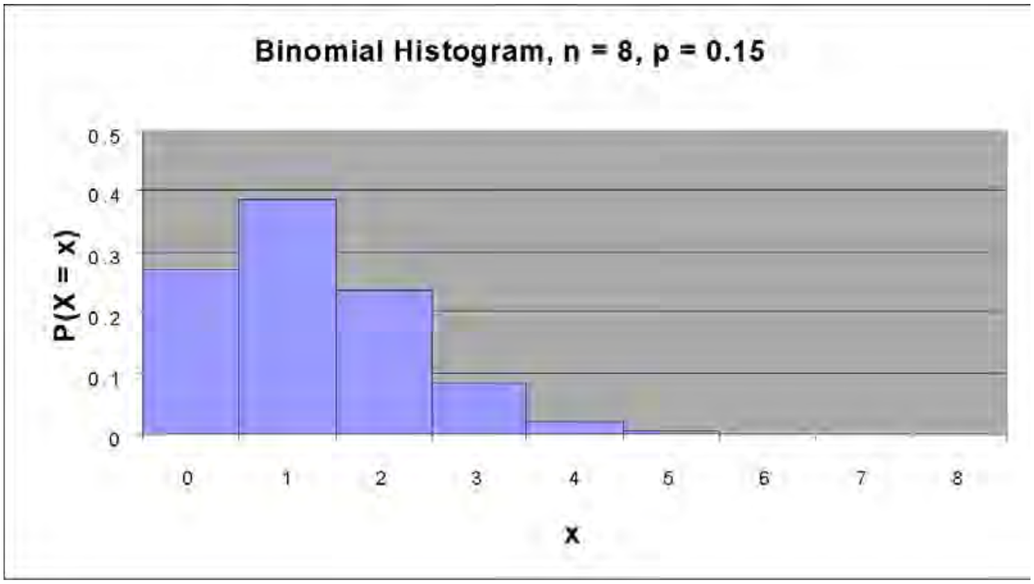
a) Construct a binomial probability histogram with $n = 8$ and $p = 0.15$.

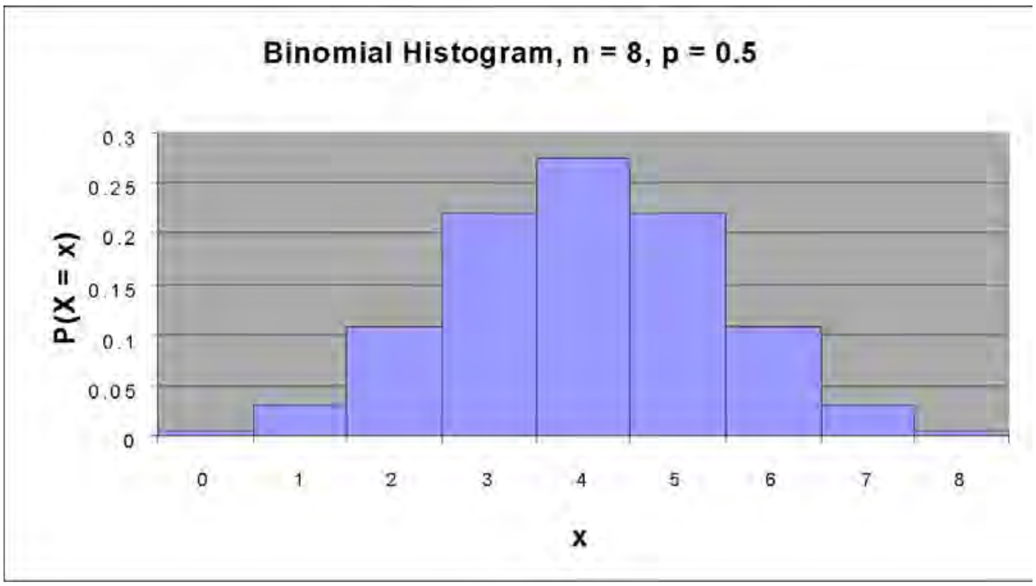
b) Construct a binomial probability histogram with $n = 8$ and $p = 0.\ 5$.

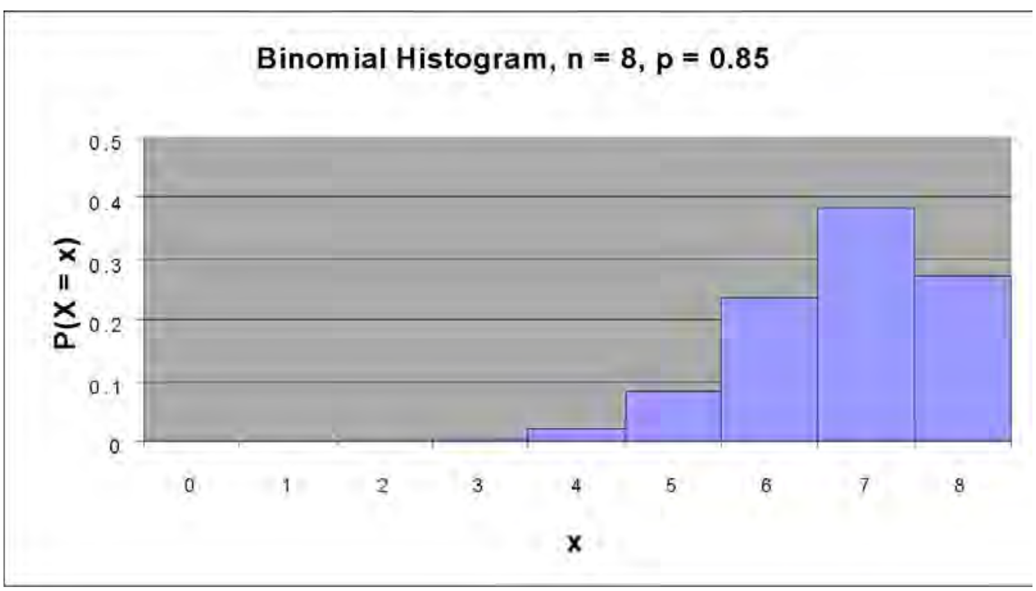
c) Construct a binomial probability histogram with $n = 8$ and $p = 0.85$.

For each histogram, comment on the shape of the distribution.

Solution


Binomial Histogram, n = 8, p = 0.15
P(X = x)
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8
x


Binomial Histogram, n = 8, p = 0.5
P(X = x)
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6 7 8
x


Binomial Histogram, n = 8, p = 0.85
P(X = x)
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8
x

Mean, Variance, and Standard Deviation

We consider important characteristics of a binomial distribution including center, variation and distribution. That is, given a particular binomial probability distribution we can find its mean, variance and standard deviation.

A strong emphasis is placed on *interpreting* and *understanding* those values.

	Discrete Probability Distribution	***Binomial Distribution***
Mean	$\mu = \sum \left[x \cdot P(x) \right]$	$\mu = np$
Variance	$\sigma^2 = \left[\sum x^2 \cdot P(x) \right] - \mu^2$	$\sigma^2 = npq$
Std. Dev.	$\sigma = \sqrt{\left[\sum x^2 \cdot P(x) \right] - \mu^2}$	$\sigma = \sqrt{npq}$

Where

$\boldsymbol{n}$ = number of fixed trials

$\boldsymbol{p}$ = probability of ***success*** in one of the $\boldsymbol{n}$ trials

$\boldsymbol{q}$ = probability of ***failure*** in one of the $\boldsymbol{n}$ trials

Interpretation of Results

It is especially important to interpret results. The range rule of thumb suggests that values are unusual if they lie outside of these limits:

$$\textit{Maximum usual values} = \boldsymbol{\mu + 2\sigma}$$
$$\textit{Minimum usual values} = \boldsymbol{\mu - 2\sigma}$$

Example

Find the mean and standard deviation for the numbers of peas with green pods when groups of 5 offspring peas are generated. Assume that there is 0.75 probability that an offspring pea has a green pod.

Solution

Given: $n = 5; \quad p = 0.75; \quad q = 1 - p = 1 - 0.75 = 0.25$

Mean: $\mu = np = 5(0.75) = 3.75 \ \underline{\approx 3.8}$

Standard deviation: $\sigma = \sqrt{npq} = \sqrt{(5)(0.75)(0.25)} = 0.9375 \ \underline{\approx 1}$

Example

Mendel generated 580 offspring peas. He claimed that 75% or 4.35, of them would have green pods. The actual experiment resulted in 428 peas with green pods.

a) Assuming that groups of 580 offspring peas are generated find the mean and standard deviation for the numbers of peas with green pods.

b) Use the range rule of thumb to find the minimum usual number and the maximum usual number of peas with green pods. Based on those numbers, can we conclude that Mendel's actual result of 428 peas with green pods is unusual? Does this suggest that Mendel's value of 75% wrong?

Solution

a) Given: $n = 580; \quad p = 0.75; \quad q = 0.25$

Mean: $\mu = np = (580)(0.75) \underline{= 435}$

Standard deviation: $\sigma = \sqrt{npq} = \sqrt{(580)(0.75)(0.25)} \underline{\approx 10.4}$

b) *Maximum usual values* $= \mu + 2\sigma = 435 + 2(10.4) \underline{= 455.8}$

Minimum usual values $= \mu - 2\sigma = 435 - 2(10.4) \underline{= 414.2}$

If Mendel generated groups of 580 offspring peas and if his 75% rate is correct, the numbers of peas with green pods should usually fall between 414.2 and 455.8.
Mendel actually got 428 peas with green pods, and that value does fall within the range of usual values, so the experimental results are consistent with the 75% rate.
The results do not suggest that Mendel's claimed rate of 75% is wrong.

Exercises *Section* 2.7 – Binomial Probability Distribution

1. 20 different Senators are randomly selected from the 100 Senators in the current Congress, and each was asked whether he or she is in favor of abolishing estate taxes. Does this procedure result in a binomial distribution, if it is not binomial, identify at least one requirement that is not satisfied?

2. 15 different Governors are randomly selected from the 50 Governors in the currently office and the sex of each Governor is recorded. Does this procedure result in a binomial distribution, if it is not binomial, identify at least one requirement that is not satisfied?

3. 200 statistics students are randomly selected and each asked if he or she owns a TI-84 Plus calculator. Does this procedure result in a binomial distribution, if it is not binomial, identify at least one requirement that is not satisfied?

4. Multiple choice questions on the SAT test have 5 possible answers (*a, b, c, d, e*), 1 of which is correct. Assume that you guess the answers to 3 such questions.

a) Use the multiplication rule to find the probability that the first 2 guesses are wrong and the third is correct. That is, find $P(WWC)$, where C denotes a correct answer and W denotes a wrong answer.

b) Beginning with *WWC*, make a complete list of the different possible arrangements of 2 wrong answers and 1 correct answer, then find the probability for each entry in the list.

c) Based on the proceeding results, what is the probability of getting exactly 1 correct answer when 3 guesses are made?

5. A psychology test consists of multiple choice questions, each having 4 possible answers (*a, b, c, d*), 1 of which is correct. Assume that you guess the answers to 6 such questions.

a) Use the multiplication rule to find the probability that the first 2 guesses are wrong and the last 4 guesses are correct. That is, find $P(WWCCCC)$, where C denotes a correct answer and W denotes a wrong answer.

b) Beginning with *WWCCCC*, make a complete list of the different possible arrangements of 2 wrong answers and 4 correct answers, then find the probability for each entry in the list.

c) Based on the proceeding results, what is the probability of getting exactly 4 correct answers when 6 guesses are made?

6. Use the Binomial Probability Table to find the probability of x success given the probability p of success on a single trial

$a)\quad n=2,\quad x=1,\quad p=.30$ $\qquad b)\quad n=5,\quad x=1,\quad p=0.95$

$c)\quad n=15,\quad x=11,\quad p=0.99$ $\qquad d)\quad n=14,\quad x=4,\quad p=0.60$

$e)\quad n=10,\quad x=2,\quad p=0.05$ $\qquad f)\quad n=12,\quad x=12,\quad p=0.70$

7. Use the Binomial Probability Formula to find the probability of x success given the probability p of success on a single trial

$a)\ \ n=12,\quad x=10,\quad p=\frac{3}{4}$ $\qquad b)\ \ n=9,\quad x=2,\quad p=0.35$

$c)\ \ n=20,\quad x=4,\quad p=0.15$ $\qquad d)\ \ n=15,\quad x=13,\quad p=\frac{1}{3}$

8. In the US, 40% of the population have brown eyes. If 14 people are randomly selected, find the probability that at least 12 of them have brown eyes. Is it unusual to randomly select 14 people and find that at least 12 of them have brown eyes? Why or why not?

9. When blood donors were randomly selected, 45% of them had blood that is Group *O*. The display shows that the probabilities obtained by entering the values of $n = 5$ and $p = 0.45$.

x	$P(x)$
0	0.050328
1	0.205889
2	0.336909
3	0.275653
4	0.112767
5	0.018453

a) Find the probability that at least 1 of the 5 donors has Group *O* blood. If at least 1 Group *O* donor is needed, is it reasonable to expect that at least 1 will be obtained?

b) Find the probability that at least 3 of the 5 donors have Group *O* blood. If at least 3 Group *O* donors are needed, is it very likely to expect that at least 3 will be obtained?

c) Find the probability that all donors have Group *O* blood. Is it unusual to get 5 Group *O* donors from 5 randomly selected donors? Why or Why not?

d) Find the probability that at most 2 of the 5 donors have Group *O* blood.

10. There is 1% delinquency rate for consumers with FICO credit rating scores above 800. If a bank provides large loans 12 people with FICL scores above 800, what is the probability that at least one of them becomes delinquent? Based on that probability, should the bank plan on dealing with a delinquency?

11. Ten peas are generated from parents having the green/yellow pair of genes, so there is a 0.75 probability that an individual pea will have a green pod. Find the probability that among the 10 offspring peas, at least 9 have green pods. Is it unusual to get at least 9 peas with green pods when 10 offspring peas are generated? Why or why not?

12. You purchased a slot machine configured so that there is a $\frac{1}{2{,}000}$ probability of winning the jackpot on any individual trial. Although no one would seriously consider tricking the author, suppose that a guest claims that she played the slot machine 5 times and hit the jackpot twice

a) Find the probability of exactly 2 jackpots in 5 trials.

b) Find the probability of at least 2 jackpots in 5 trials.

c) Does the guest's claim of hitting 2 jackpots in 5 trials seem valid? Explain.

13. In a survey of 320 college graduates, 36% reported that they stayed on their first full-time job less than one year.

a) If 15 of those survey subjects are randomly selected without replacement for a follow-up survey, find the probability that 5 of them stayed on their first full-time job less than one year.

b) If part (a) is changed so that 20 different survey subjects are selected, explain why the binomial probability formula *cannot* be used.

14. In a survey of 150 senior executives, 47% said that the most common job interview mistake is to have little or no knowledge of the company.

a) If 6 of those surveyed executives are randomly selected without replacement for a follow-up survey, find the probability that 3 of them said that the most common job interview mistake is to have little or no knowledge of the company.

b) If part (a) is changed so that 9 different surveyed executives are selected, explain why the binomial probability formula *cannot* be used.

15. In a Gallup poll of 1236 adults, it was found that 5% of those polled said that bad luck occurs after breaking a mirror. Based on these results, such randomly selected groups of 1236 adults will have a mean of 61.8 people with that belief, and a standard deviation of 7.7 people. What is the variance?

16. Random guesses are made for 50 SAT multiple choice questions, so $n = 50$ and $p = 0.2$.

a) Find the mean μ and standard deviation σ.

b) Use the range rule of thumb to find the minimum usual number and the maximum usual number.

17. In an analysis of test result from the YSORT gender selection method, 152 babies are born and it is assumed that boys and girls are equally likely, so $n = 152$ and $p = 0.5$.

a) Find the mean μ and standard deviation σ.

b) Use the range rule of thumb to find the minimum usual number and the maximum usual number.

18. In a Gallup poll of 1236 adults, it showed that 145% believe that bad luck follows if your path is crossed by a black car, so $n = 1236$ and $p = 0.14$.

a) Find the mean μ and standard deviation σ.

b) Use the range rule of thumb to find the minimum usual number and the maximum usual number.

19. The midterm exam in a nursing course consists of 75 true/false questions. Assume that an unprepared student makes random guesses for each of the answers.

a) Find the mean and standard deviation for the number of correct answers for such students.

b) Would it be unusual for a student to pass this exam by guessing and getting at least 45 correct answers? Why or why not?

20. The final exam in a nursing course consists of 100 multiple-choice questions. Each question has 5 possible answers, and only 1 of them is correct. An unprepared student makes random guesses for all of the answers.

a) Find the mean and standard deviation for the number of correct answers for such students.

b) Would it be unusual for a student to pass this exam by guessing and getting at least 60 correct answers? Why or why not?

21. In a test of the XSORT method of gender selection, 574 babies are born to couples trying to have baby girls, and 525 of those babies are girls.

a) If the gender-selection method has no effect and boys and girls are equally likely, find the mean and standard deviation for the numbers of girls born in groups of 574.

b) Is the result of 525 girls unusual? Does it suggest that the gender-selection method appears to be effective?

22. In a test of the YSORT method of gender selection, 152 babies are born to couples trying to have baby boys, and 127 of those babies are boys.

a) If the gender-selection method has no effect and boys and girls are equally likely, find the mean and standard deviation for the numbers of boys born in groups of 152.

b) Is the result of 127 boys unusual? Does it suggest that the gender-selection method appears to be effective?

23. A headline in USA Today states that "most stay at first job less than 2 years." That headline is based on a poll of 320 college graduates. Among those polled, 78% stayed at their full-time job less than 2 years.

a) Assuming that 50% is the true percentage of graduates who stay at their first job less than 2 years, find the mean and the standard deviation of the numbers of such graduates in randomly selected groups of 320 graduates.

b) Assuming that the 50% rate in part (*a*) is correct; find the range of usual values for the numbers of graduates among 320 who stay at their first job less than 2 years.

c) Find the actual number of surveyed who stayed at their first job less 2 years. Use the range of values from part (*b*) to determine whether that number is unusual. Does the result suggest that the headline is not justified?

d) This statement was given as part of the description of the survey methods used: "Alumni who opted-in to receive communications from Experience were invited to participate in the online poll, and 320 of them completed the survey." What does that statement suggest about the result?

24. In a study of 420,095 cell phone users in Denmark, it was found that 135 developed cancer of the brain or nervous system. If we assume that the use of cell phones has no effect on developing such cancer, then the probability of a person having such cancer is 0.000340.

a) Assuming that the cell phones have no effect on developing cancer, find the mean and the standard deviation of the numbers of people in groups of 420,095 that can be expected to have cancer of the brain or nervous system.

b) Based on the result from part (*a*), is it unusual to find that among 420,095 people, there are 135 cases of cancer of the brain or nervous system? Why or why not?

c) What do these results suggest about the publicized concern that cell phones are a health danger because they increase the risk of cancer of the brain or nervous system?

25. Mario's Pizza Parlor has just opened. Due to a lack of employee training, there is only a 0.8 probability that a pizza will be edible. An order for 5 pizzas has just been placed. What is the minimum number of pizzas that must be made in order to be at least 99% sure that there will be 5 that are edible?

Section 2.8 – Properties of the Normal Distribution

The *standard normal distribution* has three properties:

1. It is graph is bell-shaped.
2. It is mean is equal to 0 ($\mu = 0$).
3. It is standard deviation is equal to 1 ($\sigma = 1$).

Definitions

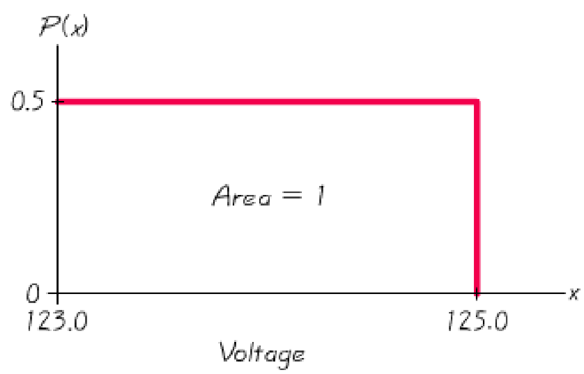
A continuous random variable has a ***uniform distribution*** if its values are spread evenly over the range of probabilities. The graph of a uniform distribution results in a rectangular shape.

A ***probability density function*** (***pdf***) is an equation used to compute probabilities of continuous random variables. It must satisfy the following two properties:

1. The total area under the graph of the equation over all possible values of the random variable must equal 1.
2. The height of the graph of the equation must be greater than or equal to 0 for all possible values of the random variable.

Example

The Power and Light Company provides electricity with voltage levels that are uniformly distributed between 123.0 volts and 125.0 volts. That is, any voltage amount between 123.0 volts and 125.0 volts is possible, and all of the possible values are equally likely. If we randomly select one of the voltage levels and represent its value by the random variable *x*, than *x* has a distribution that can me graphed as below



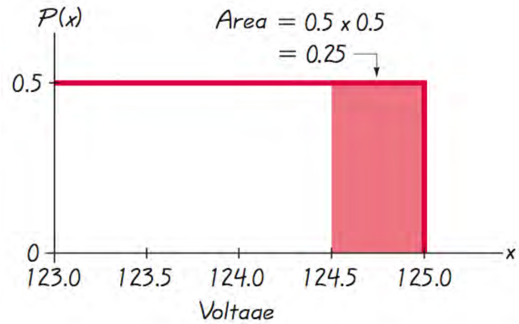
Uniform Distribution of Voltage Levels

Area and Probability

Because the total area under the density curve is equal to 1, there is a correspondence between ***area*** and ***probability***.

Example

Given the uniform distribution illustrated in the figure below, find the probability that a randomly voltage level is greater than 124.5 volts.



Shaded area represents voltage levels greater than 124.5 volts.

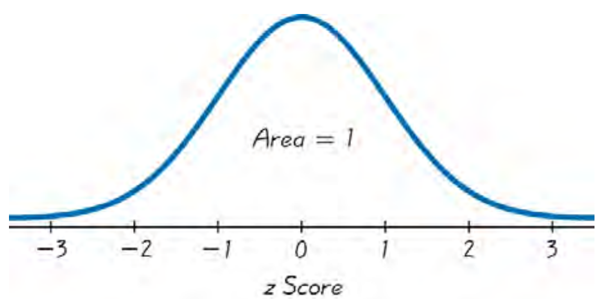
$$P(\textit{voltage greater than } 124.5 \textit{ volts}) = \textit{area of the shaded region}$$
$$= 0.5 \times 0.5$$
$$= 0.25$$

The probability of randomly selecting a voltage level greater than 124.5 volts is 0.25.

Standard Normal Distribution

Definition

The standard normal distribution is a normal probability distribution with $\mu = 0$ and $\sigma = 1$. The total area under its density curve is equal to 1.



Finding Probabilities: When Given z-scores

➢ Standard Normal distribution Table.

Using Table (Normal Distribution Table):

1. It is designed only for the *standard* normal distribution, which has a mean of 0 and a standard deviation of 1.
2. It is on two pages, with one page for *negative* z-scores and the other page for *positive* z-scores.
3. Each value in the body of the table is a *cumulative area from the left* up to a vertical boundary above a specific z-score.
4. When working with a graph, avoid confusion between z-scores and areas.

   z Score: Distance along horizontal scale of the standard normal distribution; refer to the leftmost column and top row of Normal Distribution Table.

   Area Region under the curve; refer to the values in the body of Normal Distribution Table.

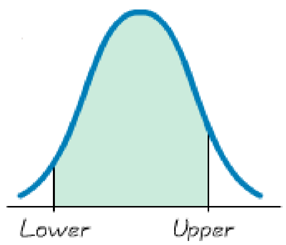
5. The part of the z-score denoting hundredths is found across the top.

- Formulas and Tables insert card
- Find areas for many different regions

Methods for Finding Normal Distribution Areas

TI-83/84 Plus Calculator

Gives area bounded on the left and bounded on the right by vertical lines above any specific values.



TI-83/84 Press 2ND VARS **[2: normal cdf ( ],** then enter the two *z* scores separated by a comma, as in (left *z* score, right *z* score).

Example

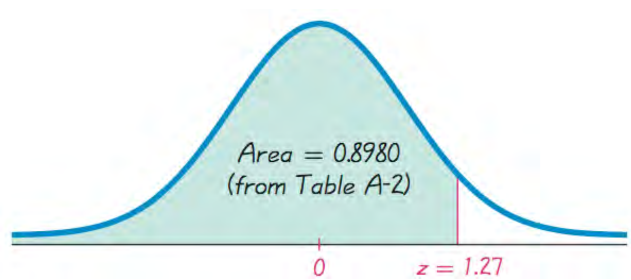
The Precision Scientific Instrument Company manufactures thermometers that are supposed to give readings of 0ºC at the freezing point of water. Tests on a large sample of these instruments reveal that at the freezing point of water, some thermometers give readings below 0º (denoted by negative numbers) and some give readings above 0º (denoted by positive numbers). Assume that the mean reading is 0ºC and the standard deviation of the readings is 1.00ºC. Also assume that the readings are normally distributed. If one thermometer is randomly selected, find the probability that, at the freezing point of water, the reading is less than 1.27º.

Solution

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292

$P(z<1.27) \underline{=0.8980}$

The *probability* of randomly selecting a thermometer with a reading less than 1.27º is 0.8980. Or 89.80% will have readings below 1.27º.



Example

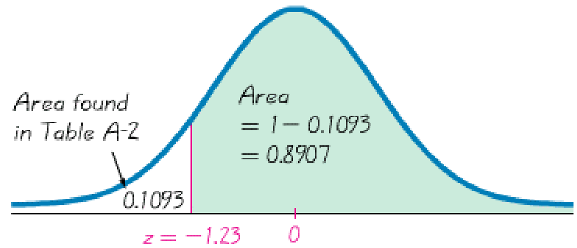
If thermometers have an average (mean) reading of 0 degrees and a standard deviation of 1 degree for freezing water, and if one thermometer is randomly selected, find the probability that it reads (at the freezing point of water) above –1.23 degrees.

Solution

$P(z > -1.23) = 1 - 0.1093 = \underline{= 0.8907}$

Probability of randomly selecting a thermometer with a reading above –1.23° is 0.8907.

89.07% of the thermometers have readings above –1.23 degrees.

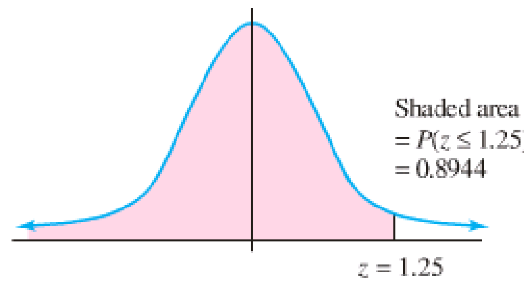


Example

Find the areas under the standard normal curve

a) The area to the ***left*** of $z = 1.25$

$A = P(z \le 1.25)\ \underline{= 0.8944}$ **(*Using left curve table*)**

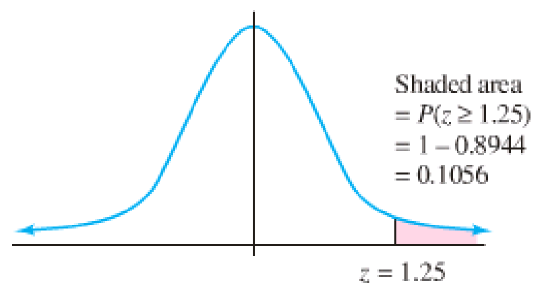


b) The area to the ***right*** of $z = 1.25$

$A = P(z \ge 1.25)$ **(*Using left curve table*)**

$= 1 - 0.8944$

$\underline{= 0.1056}$

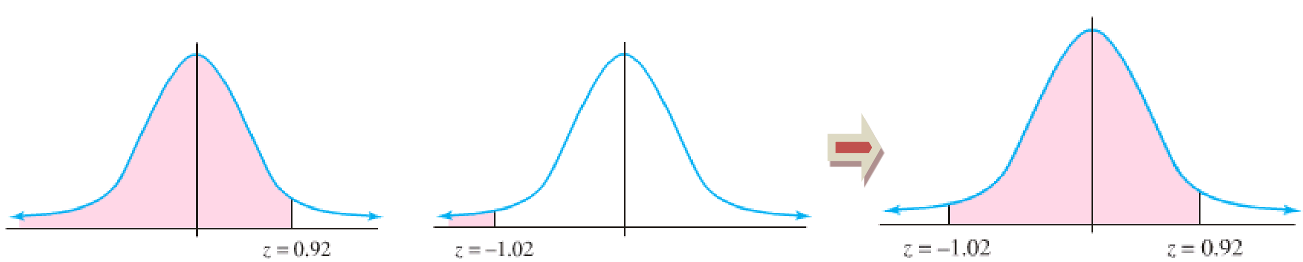


c) The area ***between*** $z = -1.02$ and

$A = P(-1.02 \le z \le 0.92)$ **(*Using left curve table*)**

$= 0.8212 - 0.1539$

$\underline{= 0.6673}$



Notation

$P(a < z < b)$ denotes the probability that the z score is between a and b.

$P(z > a)$ denotes the probability that the z score is greater than a.

$P(z < a)$ denotes the probability that the z score is less than a.

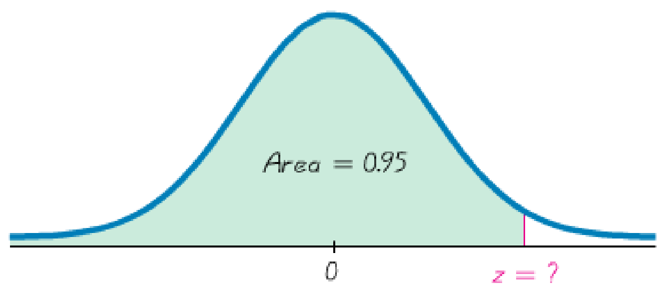
Finding a *z*–Score When Given a Probability Using Table

1. Draw a bell-shaped curve and identify the region under the curve that corresponds to the given probability. If that region is not a cumulative region from the left, work instead with a known region that is a cumulative region from the left.
2. Using the cumulative area from the left, locate the closest probability in the body of Table and identify the corresponding *z* score.

Example

With temperature readings at the freezing point of water that are normally distributed with a mean 0°C and a standard deviation of 1.00°C. Find the temperature corresponding to P_{95} , the 95th percentile. That is, find the temperature separating the bottom 95 from the top 5%.

Solution

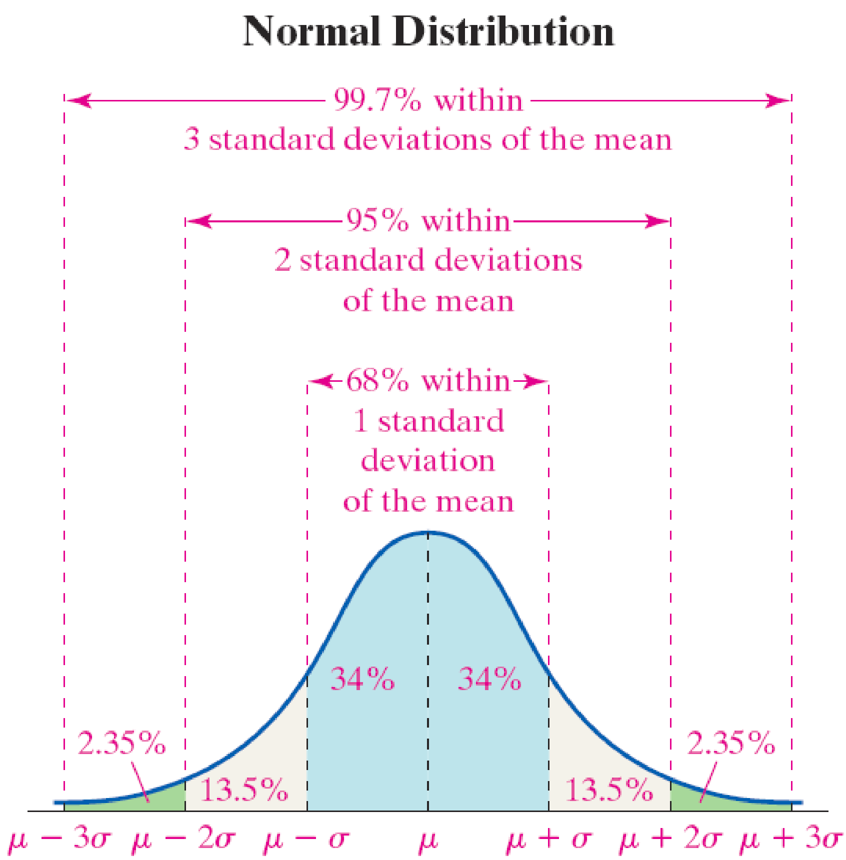


Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449

From the table, we find the areas of 0.9495 and 0.9505. The area 0.95 corresponds to a *z*-score of 1.645. When tested at freezing, 95% of the readings will be less than or equal to 1.645°C, and 5% of them will be greater than or equal to 1.645°C.

Properties of the Normal Density Curve

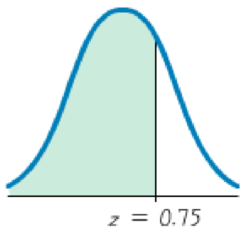
1. It is symmetric about its mean, μ.
2. Because ***mean*** = ***median*** = ***mode***, the curve has a single peak and the highest point occurs at $x = \mu$.
3. It has inflection points at $\mu - \sigma$ and $\mu - \sigma$
4. The area under the curve is 1.
5. The area under the curve to the right of μ equals the area under the curve to the left of μ, which equals 1/2.
6. As x increases without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As x decreases without bound (gets more and more negative), the graph approaches, but never reaches, the horizontal axis.
7. The Empirical Rule: Approximately 68% of the area under the normal curve is between $x = \mu - \sigma$ and $x = \mu + \sigma$; approximately 95% of the area is between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$; approximately 99.7% of the area is between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$.



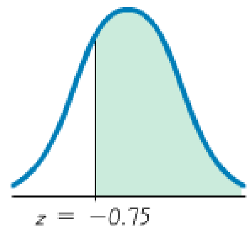
Exercises *Section* 2.8 – Properties of the Normal Distribution

1. Find the area shaded region. The graph depicts the standard distribution with mean 0 and standard deviation 1.

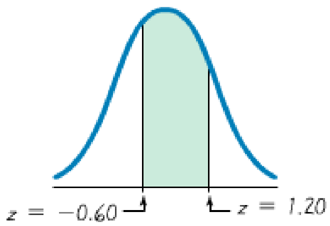
a)



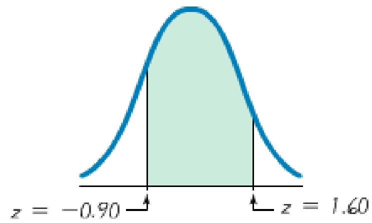
b)



c)

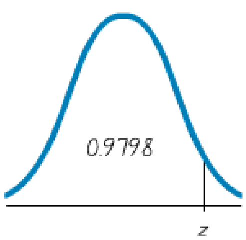


d)

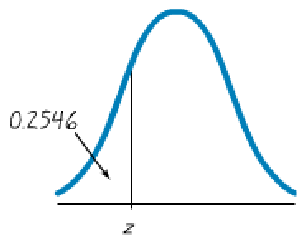


2. Find the indicated *z*-score. The graph depicts the standard distribution with mean 0 and standard deviation 1.

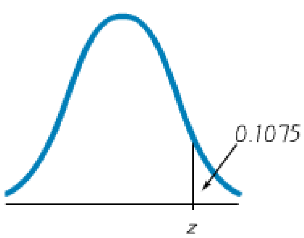
a)



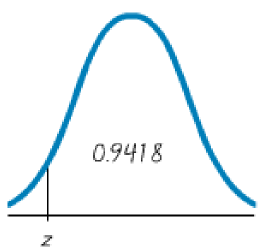
b)



c)



d)



3. Assume that thermometer readings are normally distributed with a mean of 0ºC and the standard deviation of the readings is 1.00ºC. A thermometer is randomly selected and tested. In each case, draw a sketch, and find the probability of each reading.

a) Less than –1.50
b) Less than –2.75
c) Less than 1.23
d) Greater than 2.22
e) Greater than 2.33
f) Greater than –1.75
g) Between 0.50 and 1.00
h) Between –3.00 and –1.00
i) Between –1.20 and 1.95
j) Between –2.50 and 5.00
k) Greater than 0
l) Less than 0

4. Assume that thermometer readings are normally distributed with a mean of 0ºC and the standard deviation of the readings is 1.00ºC. A thermometer is randomly selected and tested. In each case, draw a sketch, and find the temperature reading corresponding to the given information.

a) Find P_{95}, the 95th percentile. This is the temperature separating the bottom 95% from the top 5%.

b) Find P_1, the 1st percentile. This is the temperature separating the bottom 1% from the top 99%.

c) If 2.5% of the thermometers are rejected because they have readings that are too high and another 2.5% are rejected because they have readings that are too low, find the 2 readings that are cutoff values separating the rejected thermometers from the others.

d) If 0.5% of the thermometers are rejected because they have readings that are too high and another 0.5% are rejected because they have readings that are too low, find the 2 readings that are cutoff values separating the rejected thermometers from the others.

5. For a standard normal distribution, find the percentage of data that are

a) Within 2 standard deviations of the mean.

b) More than 1 standard deviation away from the mean.

c) More than 1.96 standard deviations away from the mean.

d) Between $\mu - 3\sigma$ and $\mu + 3\sigma$.

e) More than 3 standard deviations away from the mean.

Section 2.9 – Applications of the Normal Distribution

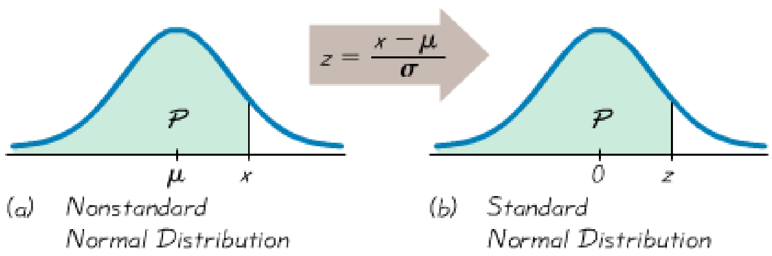
Working with normal distributions that are not standard, that is, the mean is not 0 or the standard deviation is not 1, or both.

The key concept is that we can use a simple conversion that allows us to standardize any normal distribution so that the same methods of the previous section can be used.

Conversion Formula

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

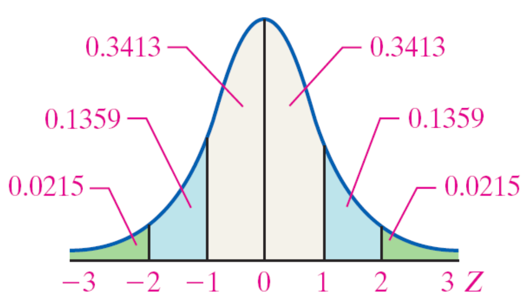
Round z scores to 2 decimal places



Finding Areas with a nonstandard normal distribution

1. Sketch a normal curve, label the mean and the specific x values, then *shade* the region representing the desired probability.
2. For each relevant value x that is a boundary for the shaded region, use the formula to convert that value to the equivalent z-score.
3. Refer the Normal Distribution Table or use the calculator to find the area of the shaded region. This area is the desired probability.

Areas Under the Standard Normal Curve

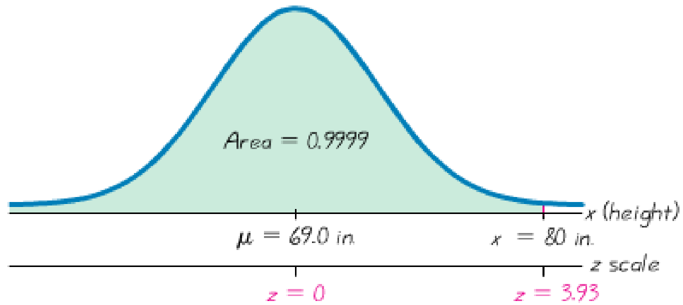


Example

The typical home doorway has a height of 6 ft. 8 in., or 80 in. Because men tend to be taller than women, we will consider only men as we investigate the limitations of that standard doorway height. Given that heights of men are normally distributed with a mean of 69.0 in. and a standard deviation of 2.8 in., find the percentage of men who can fit through the standard doorway without bending or bumping their head. Is that percentage high enough to continue using 80 in. as the standard height? Will a doorway height of 80. Be sufficient in future years?

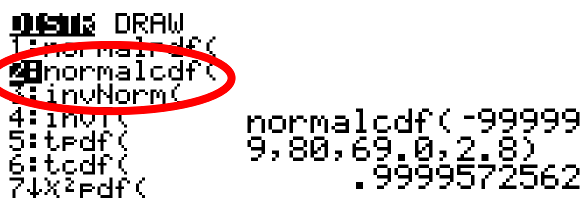
Solution

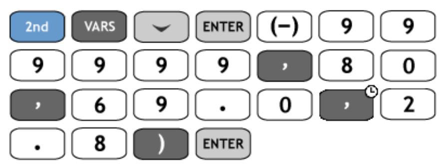
Men have heights that are normally distributed with a mean of 69.0 in. and a standard deviation of 2.8 in. The shaded region represents the men who can fir through a doorway that has a height 80 in.

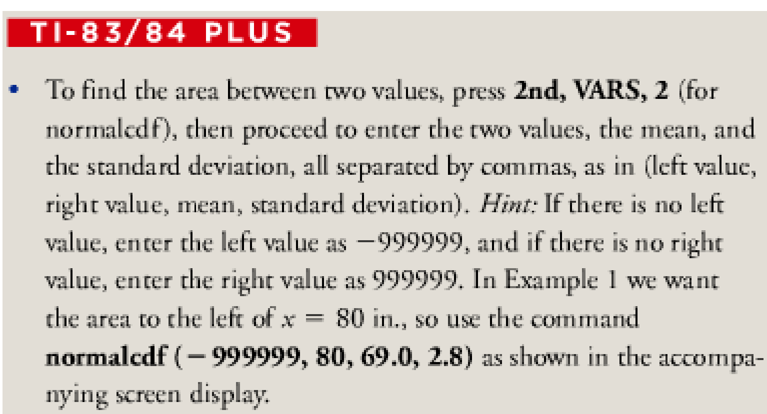


The z-score: $z = \dfrac{x-\mu}{\sigma} = \dfrac{80-69}{2.8} = 3.93$ $\qquad (80-69)/2.8$

Referring to the table, the z-score values are less than 3.5, therefore; if we use calculator:







TI-83/84 PLUS

- To find the area between two values, press **2nd, VARS, 2** (for normalcdf), then proceed to enter the two values, the mean, and the standard deviation, all separated by commas, as in (left value, right value, mean, standard deviation). *Hint:* If there is no left value, enter the left value as −999999, and if there is no right value, enter the right value as 999999. In Example 1 we want the area to the left of x = 80 in., so use the command **normalcdf (− 999999, 80, 69.0, 2.8)** as shown in the accompanying screen display.

$\boxed{normalcdf\left(-999999,\ x,\ \mu,\ \sigma\right)}$

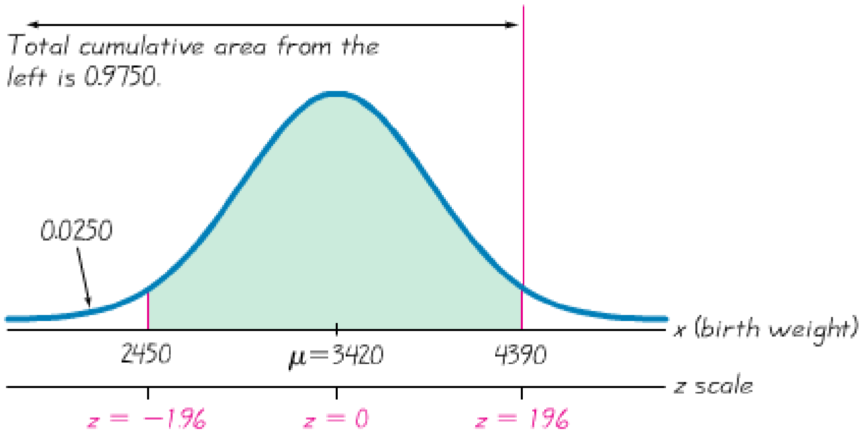
✓ The proportion of men who can fit through the standard doorway height of 80 in. is 0.9999, or 99.99%. Very few men will not be able to fit through the doorway without bending their head. This percentage id high enough to justify the use of 80 in. as the standard doorway height. However, heights of men and women have been increasing gradually but steady over the past decades.

Example

Birth weights in the U.S. are normally distributed with a mean of 3420 g and a standard deviation of 495 g. A hospital requires special treatment for babies that are less than 2450 g (unusual light) or more than 4390 g (unusually heavy). What is the percentage of babies who do not require special treatment because they have birth weights between 2450 g and 4390 g? Under these conditions, do many babies require special treatment?

Solution

Given: $\mu = 3420, \quad \sigma = 495$



The area to the ***left*** of $x = 2450 \quad \Rightarrow \quad z = \frac{2450 - 3420}{495} = -1.96$

$Area = A_1 \ (< 2450) = 0.0250$

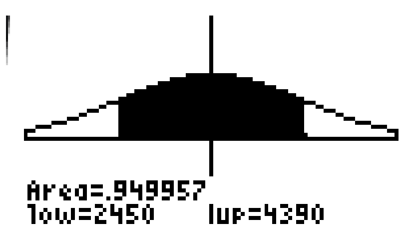
The area to the ***left*** of $x = 4390 \quad \Rightarrow \quad z = \frac{4390 - 3420}{495} = 1.96$

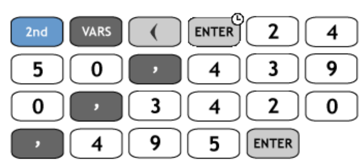
$Area = A_2 \ (< 4390) = 0.9750$

The total shaded area:

$Area = A_2 - A_1 = 0.9750 - 0.025 = 0.950$

[2ND] [VARS]





$\boxed{ShadeNorm(low\ x,\ up\ x,\ \sigma)}$

✓ We conclude that 95.00% of the babies do not require special treatment because they have birth weights between 2450 g and 4390 g. It follows that 5.00 % of the babies do require special treatment.

Helpful *Hints*

1. Don't confuse z scores and areas. z-scores are distances along the horizontal scale, but areas are regions under the normal curve. Table lists z-scores in the left column and across the top row, but areas are found in the body of the table.
2. Choose the correct (right/left) side of the graph.
3. A z-score must be negative whenever it is located in the left half of the normal distribution.
4. Areas (or probabilities) are positive or zero values, but they are never negative.

Procedure for Finding the Value of a Normal Random Variable

***Step* 1**: Draw a normal curve and shade the area corresponding to the proportion, probability, or percentile.
***Step* 2**: Use Table V to find the z-score that corresponds to the shaded area.

***Step* 3**: Obtain the normal value from the formula $x = \mu + z\sigma$.

Example

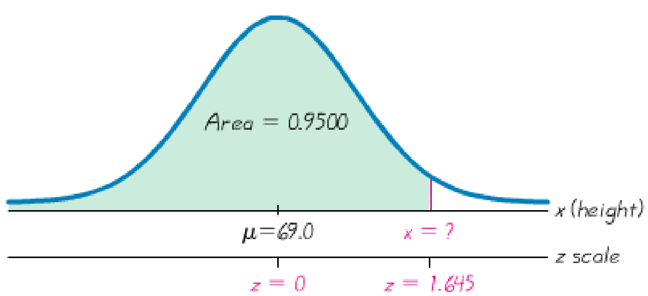
When designing an environment, one common criterion is to use a design that accommodates 95% of the population. How high should doorways be if 95% of men will fit through without bending or bumping their head? That is, find the 95th percentile of heights of men. Heights of men normally distributed with a mean of 69.0 in. and a standard deviation of 2.8 in.

Solution

Given: $\mu = 69.0, \quad \sigma = 2.8$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449

From the table, we find the areas of 0.9495 and 0.9505. The area 0.95 corresponds to a z-score of 1.645.



$$x = \mu + z \cdot \sigma = 69.0 + (1.645)(2.8) = 73.606$$

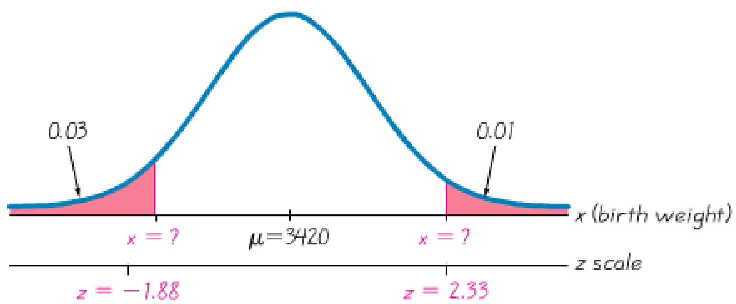
✓ A doorway height of 73.6 in. would allow 95% of men to fit without bending their head.

Example

Hospital wants to redefine the minimum and maximum birth weights that require special treatment because they are unusual low or unusual high. After considering relevant factors, a committee recommends special treatment for birth weights in the lowest 3% and the highest 1%. The committee members soon realize that specific birth weights need to be identified. Help this committee by finding the birth weights that separate the lowest 3% and the highest 1%. Birth weights in the U.S. are normally distributed with a mean of 3420 *g* and a standard deviation of 495 *g*.

Solution

Given: $\mu = 3420, \quad \sigma = 495$



For the leftmost value of x:

The cumulative area from the left is 0.03, from the table: $z = -1.88$

$x = \mu + z \cdot \sigma = 3420.0 + (-1.88)(495) = 2489.4$

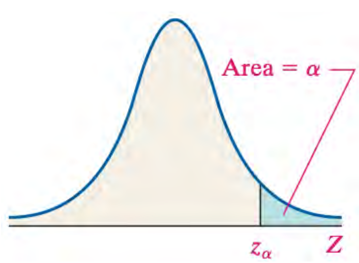
For the rightmost value of x:

The cumulative area from the left is $1 - .01 = .99$, from the table: $z = 2.33$

$x = \mu + z \cdot \sigma = 3420.0 + (2.33)(495) = 4573.35$

✓ The birth weight of 2489 *g* separates the lowest 3% of birth weights, and 4573 *g* separates the lowest 1% of birth weights. The hospital now has well-defined criteria for determining whether a newborn baby should be given special treatment for a birth weight that is unusual lo row high.

The notation z_α (pronounced "*z* sub alpha") is the *z*-score such that the area under the standard normal curve to the right of z_α is α.

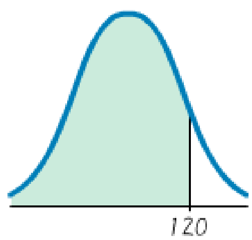


Exercises *Section* 2.9 – Applications of the Normal Distribution

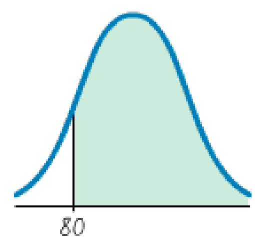
1. The distribution of IQ scores is a nonstandard normal distribution with mean of 100 and standard deviation of 15. What are the values of the mean and standard deviation after all IQ scores have been standardized by converting them to *z*-scores using $z = \dfrac{x - \mu}{\sigma}$?

2. Find the area of the shaded region. The graphs depict IQ scores adults, and those scores are normally distributed with mean of 100 and standard deviation of 15.

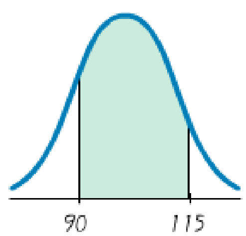
a)



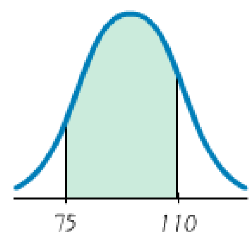
b)



c)

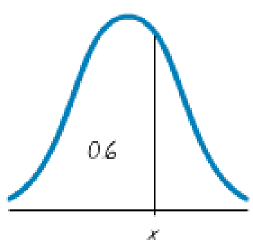


d)

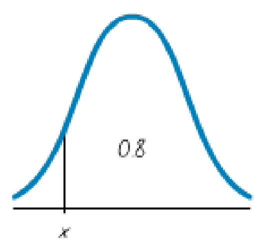


3. Find the Indicated IQ scores. The graphs depict IQ scores adults, and those scores are normally distributed with mean of 100 and standard deviation of 15.

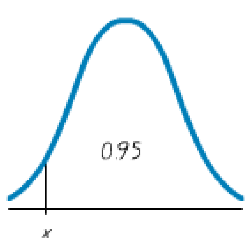
a)



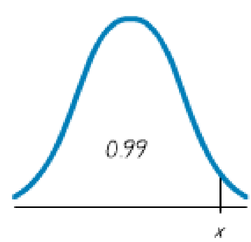
b)



c)



d)



4. Assume that adults have IQ scores that are normally distributed with mean of 100 and standard deviation of 15

a) Find the probability that a randomly selected adult has an IQ that is less than 115.

b) Find the probability that a randomly selected adult has an IQ that is greater than 131.5.

c) Find the probability that a randomly selected adult has an IQ that is between 90 and 110.

d) Find the probability that a randomly selected adult has an IQ that is between 110 and 120.

e) Find P_{30} which is the IQ score separating the bottom 30% from the top 70%.

f) Find the first quartile Q_1 which is the IQ score separating the bottom 25% from the top 75%.

5. The Gulfstream 100 is an executive jet that seats six, and it has a doorway height of 51.6

- ***Men's*** heights are normally distributed with mean 69.0 in. and standard deviation 2.8 in.
- ***Women's*** heights are normally distributed with mean 63.6 in. and standard deviation 2.5 in.

a) What percentage of adult men can fit through the door without bending?

b) What percentage of adult women can fit through the door without bending?

c) Does the door design with a height of 51.6 in. appear to be adequate? Why didn't the engineers design larder door?

d) What doorway height would allow 60% of men to fit without bending?

6. Assume that human body temperatures are normally distributed with a mean of 98.20°F and a standard deviation of 0.62°F.

a) A Hospital uses 100.6°F as the lowest temperature considered to be a fever. What percentage of normal and healthy persons would be considered to have fever? Does this percentage suggest that a cutoff of 100.6°F is appropriate?

b) Physicians want to select a minimum temperature for requiring further medical tests. What should that temperature be, if we want only 5.0% of healthy people to exceed it? (Such a result is a false positive, meaning that the test result is positive, but the subject is not really sick.)

7. The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days.

a) One classical use of the normal distribution is inspired by a letter to "Dear Abby" in which a wife claimed to have given birth 308 days after a brief visit from her husband. Given this information, find the probability of a pregnancy lasting 308 days or longer. What does the result suggest?

b) If we stipulate that a baby is premature if the length of pregnancy is the lowest 4%, find the length that separates premature babies from those who are not premature. Premature babies often require special care, and this result could be helpful to hospital administrators in planning for that care.

8. A statistics professor gives a test and finds that the scores are normally distributed with a mean of 25 and a standard deviation of 5. She plans to curve the scores.

a) If the curves by adding 50 to each grade, what is the new mean? What is the new standard deviation?

b) Is it fair to curve by adding 50 to each grade? Why or why not?

c) If the grades are curved according to the following scheme (instead of adding 50), find the numerical limits for each letter grade.

A: Top 10%

B: Scores above the bottom 70% and below the top 10%.

C: Scores above the bottom 30% and below the top 30%.
D: Scores above the bottom 10% and below the top 70%.
F: Bottom 10%.

d) Which method of curving the grades is fairer: Adding 50 to each grade or using the scheme given in part (*c*)? Explain.