Excel

Supplement

By Matt Regan Department of Statistics The University of Auckland

Contents

1.	Introductory Exercises Entering Information 1 Copying and Pasting Information Entering Functions 2 Entering Formula 3 Sorting Data 3 Creating Cell Names 5	1
2.	Constructing a Dotplot	7
3.	Calculating Binomial ProbabilitiesIndividual Probabilities9Lower Tail Probabilities10Upper Tail Probabilities10 $pr(a \le X \le b)$ 11	9

4.	Calculating Poisson Probabilities Individual Probabilities 13	13
	Lower Tail Probabilities 14	
	Upper Tail Probabilities 14	
	$\operatorname{pr}(a \le X \le b)$ 15	
5.	Calculating Normal Probabilities	17
	Lower Tail Probabilities 17	
	Upper Tail Probabilities 18	
	$pr(a \le X \le b) = 18$	
	Calculating the Inverse of the Normal Distribution 20	
6.	Generating Random Samples	23
7.	Distribution of the Sample Mean	25
8.	Calculating Student <i>t</i> -Probabilities	31
	Lower Tail Probabilities 31	
	Upper Tail Probabilities 32	
	$pr(a \le T \le b)$ 32 Calculating the Inverse of the <i>t</i> -Distribution 34	
9.	t-Test of Means for Two Independent Samples	37
10.	t-Test of Means for Paired Data	39
11.	One-Way ANOVA — F-Test	41
12.	Chi-square Test	43
	One-Dimensional Tables 43	
	Two-Dimensional Tables 44	
13.	Simple Linear Regression	47
14.	Correlation	51
15.	Copying and Pasting Excel into Word	53

© 2000 Matt Regan, Department of Statistics, The University of Auckland

Introductory Exercises

Gross Domestic Product (GDP) is the total market value of goods and services produced in a country within a given period. The percentage growth of GDP is usually used as an important measurement of the economic growth of a country. Shown below is an estimate of percentage growth in GDP for 1998 and the forecast for 2000 for some South-East Asian and Pacific countries.

	A	В	С	D	Е	F	G	Н		J	K
1		Australia	China	Hong Kong	Japan	Malaysia	New Zealand	Philippines	Singapore	South Korea	Taiwan
2	1998	2.8	6.1	3.5	-2.3	-0.6	0.5	1.4	-0.5	-1.2	4.6
3	2000	2.5	7.4	5.3	1	2.1	1.8	3.6	3.1	7.1	4.9

Figure 1

Task 1 ENTERING INFORMATION

1. Type the data as in Figure 1.

Task 2 COPYING AND PASTING INFORMATION

- 2. Select cells from A1 to K3 and click Copy on the Standard toolbar.
- **3.** Click **Sheet 2** on the Sheet tabs. (This is a new worksheet in the same workbook as the original worksheet. To go back the original worksheet, click **Sheet 1** on the Sheet tab.)
- 4. Click cell A1.
- 5. On the Menu bar, click Edit **E** Paste Special.
- **6.** Select the **Transpose** check box. (Figure 2)
- 7. Click **OK**. It will give the table as shown in Figure 3.

Paste Special	? ×		A	В	C
		1		1998	2000
Paste		2	Australia	2.8	2.5
 All 	C <u>C</u> omments	3	China	6.1	7.4
C Eormulas	🔿 Validatio <u>n</u>	4	Hong Kong	3.5	5.3
C Values	C All except borders	5	Japan	-2.3	1
O Formats		6	Malaysia	-0.6	2.1
		7	New Zealand	0.5	1.8
Operation		8	Philippines	1.4	3.6
💿 N <u>o</u> ne	C Multiply	9	Singapore	-0.5	3.1
C Add	🔿 Divide	10	South Korea	-1.2	7.1
C Subtract	-	11	Taiwan	4.6	4.9
		12			
Skip blanks	Transnosa	13			
	i manspose	14			
	1 1	15			
OK	Cancel Paste Link	16			
		17			

Figure 2

Figure 3

Task 3 ENTERING FUNCTION

- **8.** In cell A12, type **Total**.
- 9. In cell B12, type =SUM(B2:B11) and press Enter key on the keyboard.

- Note: In *Excel*, a formula starts with an equal sign (=) and it will not be calculate until the Enter key has been pressed.
 - **10.** Select cell B12.
 - **11.** Click its **Fill Handle** (Figure 4) and hold down do NOT release the mouse button.





- 12. Drag the **Fill Handle** to the cell C12 and release the mouse button. Notice that the cell C12 contains the formula =**SUM(C2:C11)** now.
- Note: This skill can also be used to fill in a series of numbers, dates, same content, or other items. Use **Office Assistance** or **Help** file to learn them.
 - **13.** In cell A13, type **Mean**.
 - 14. In cell B13, type AVERAGE(B2:B11).
 - **15.** In cell C13, use the skill described in step 11 and 12 to find the average GDP growth of all countries for year 2000.

Task 4 ENTERING FORMULA

- **16.** In cell D1, type **Difference**.
- 17. In cell D2, type =C2 C3.
- **18.** Use the skill described in step 11 and 12 to find the difference in GDP growth between 1998 and 2000 for each country. The table should look like in Figure 5.

	A	В	C	D
1		1998	2000	Difference
2	Australia	2.8	2.5	-0.3
3	China	6.1	7.4	1.3
4	Hong Kong	3.5	5.3	1.8
5	Japan	-2.3	1	3.3
6	Malaysia	-0.6	2.1	2.7
7	New Zealand	0.5	1.8	1.3
8	Philippines	1.4	3.6	2.2
9	Singapore	-0.5	3.1	3.6
10	South Korea	-1.2	7.1	8.3
11	Taiwan	4.6	4.9	0.3
12	Total	14.3	38.8	
13	Mean	1.43	3.88	

Figure 5

Task 5 SORTING DATA

- **19.** Select cells from A1 to D11.
- 20. On the Menu bar, click Data **L** Sort.
- 21. In **Sort** dialog box, click the arrow button and choose **Difference** from the list.
- 22. In Sort By box, choose Descending. (Figure 6)
- **23.** Click **OK**. Now the table is arranged according to the difference of GDP growth in descending order and should look like in Figure 7.

Note: If the column specified in the **Sort By** box has duplicate items, the values can be sorted by specifying another column in the **Then By** boxes. In this example, China and New Zealand have the same difference of GDP growth. If after step 22, we set the first **Then By** box as **2000** and **Ascending**, then click **OK**. We will have table as in Figure 7 except that New Zealand is arranged before China.

Sort	? ×
Sort by	
Difference	C Ascending O Descending
Then by	
	Ascending
,	C Descending
Then by	
	 Ascending
,	C Descending
My list has	
• Header <u>r</u> ow	C No header ro <u>w</u>
Options	OK Cancel
	Figure 6

	A	В	C	D
1		1998	2000	Difference
2	South Korea	-1.2	7.1	8.3
3	Singapore	-0.5	3.1	3.6
4	Japan	-2.3	1	3.3
5	Malaysia	-0.6	2.1	2.7
6	Philippines	1.4	3.6	2.2
7	Hong Kong	3.5	5.3	1.8
8	China	6.1	7.4	1.3
9	New Zealand	0.5	1.8	1.3
10	Taiwan	4.6	4.9	0.3
11	Australia	2.8	2.5	-0.3
12	Total	14.3	38.8	
13	Mean	1.43	3.88	



Task 6 CREATING CELL NAMES:

Sometimes it is useful to refer to cells by a name like 'sample_mean' rather than a row column reference, eg like B24.

Naming One Cell:

Eg, name the cell B24 'sample_mean'.

- 1. Select the cell B24.
- 2. Move the cursor to the **Name Box** in the top left of your screen and click the mouse button when the cursor is over the box.





- 3. The Name Box should now be highlighted. Type sample_mean.
- **4.** Press the Enter key. The **Name Box** should now contain **sample_mean** and will look like Figure 9 below.



Figure 9

Naming More than One Cell:

Eg, naming cells B2:B4, 'Bobs_age', 'Marys_age', and 'Jacks_age', as these cells contain the age of the respective person.

1. Type the entries into cells A1:B4 and select cells A2:B4, as seen in Figure 10.

	•	В	С
1	names	cells	-
2	Bobs_aqe	23	
3	Marys_age	32	
4	Jacks_age	26	
5			



- 2. On the menu bar click Insert **L** Name **L** Create...
- **3.** A dialog box is brought up, as in Figure 11. The names to be created are in the left column, select the **Left Column** check box.





4. Click OK. The cells have now been named.

Example: Construct a dotplot for the Cancer Data.

- Input the data into *Excel*. Notice that 1, 2, and 3 in Column A represent "Stomach", "Bronchus", "Colon" respectively. (See Figure 1 on the right)
- 2. Highlight the cells A1:B30 and click Chart Wizard.
- 3. Choose XY(Scatter) in Chart type box and first option in Chart sub-type box.
- 4. Click Next. (See Figure 2 below)



Figure 2

	A	B
1	1	1.18
2	1	-0.41
3	1	-0.64
4	1	1.32
5	1	1.39
6	1	0.14
7	1	2.29
8	1	-0.1
9	1	2.2
10	1	3.83
11	2	0.81
12	2	3.16
13	2	-0.22
14	2	2.05
15	2	1.9
16	2	0.85
17	2	0.27
18	2	0.73
19	2	1.84
20	2	3.86
21		2.01
22	- 3	0.51
23	3	1.07
24	3	4.31
25	3	1
26	3	3.51
27	3	3
28	3	1.79
29	3	1.67
30	3	2.16

Figure 1

5. Click Next.



Figure 3

- **6**. Enter the titles in the **Chart Title**, **Value** (**X**) **Axis**, and **Value** (**Y**) **axis**. (Figure 3 on the right hand side).
- 7. Click Finish. Your plot should look like Figure 4 below.



The Dotplot of Cancer Data

8. By using textboxes and by double clicking each axis, it is possible to format the axes to produce a plot as in Figure 5 below.



The Dotplot of Cancer Data

Figure 5

Calculating Binomial Probabilities

A. Individual probabilities: pr(X = x)

Example: Find pr(X = 5) where $X \sim Binomial(8, 0.6)$

- 1. Click cell A1.
- 2. Click the Paste Function button
- **3.** Choose **Statistical** from the dialog box.

from the tool bar.

Function category list box in the Paste Function

4. Choose **BINOMDIST** from the **Function name** list box (Figure 1).

Paste Function	? ×			
Function category:	Function name:			
Most Recently Used All Financial Date & Time Math & Trig Statistical Lookup & Reference Database Text Logical Information	AVEDEV AVERAGE AVERAGEA BETADIST BETAINV BINOMDIST CHIDIST CHIINV CHITEST CONFIDENCE CORFIDENCE CORREL			
BINOMDIST(number_s,trials,probability_s,cumulative)				
Returns the individual term binomia	al distribution probability.			
2	OK Cancel			

Figure 1

- 5. Click OK.
- 6. Fill in the **BINOMDIST** dialog box (Figure 2).

BINOMDIST			
Number_s		🗾 = number	
Trials		🗾 = number	
Probability_s		🗾 = number	
Cumulative		📑 = logical	
Returns the individu	al term binomial distribution probability.	=]
Number_s	is the number of successes in trials.		
Formul	a result =	ОК	Cancel

Figure 2

where:

- **Number_s** is the number of successes in trials. It is equivalent to *x* in our manual. In this example, we put 5 in this box.
 - **Trials** is the number of trials. It is equivalent to *n* in our manual. In this example, we put 8 in this box.

Probability is the probability of success in each trial. It is equivalent to p in our manual. In this s example, we put 0.6 in this box.

Cumulative indicates whether the number of successes is cumulative (TRUE or 1) or not(FALSE or 0). Since the individual probability is not cumulative, we put FALSE in this box.

7. Click **OK**. (The value of 0.279 should appear in cell A1).

B. Lower tail probabilities: $pr(X \le x)$

Example: Find $pr(X \le 3)$ where $X \sim Binomial(8, 0.6)$

- 1. Click cell B1.
- 2. Follow Steps 1 to 4 in A.
- **3.** Set: Number_s = 3
 - **Trials** = 8

```
Probability_s = 0.6
```

4. The lower tail probability is cumulative, hence we set

```
Cumulative = TRUE.
```

5. Click OK. (The value of 0.174 should appear in cell B1).

C. Upper tail probabilities: $pr(X \ge x)$

Example: Find $pr(X \ge 3)$ where $X \sim Binomial(8, 0.6)$

Note: $pr(X \ge 3) = 1 - pr(X \le 2)$

- **1.** Evaluate $pr(X \le 2)$ in cell C1 (use steps in B).
- **2.** In cell C2, type: **=1 C1**.
- **3.** Press **Enter**. (The value of 0.950 should appear in cell C2).

D. $pr(a \le X \le b)$

Example: Find $pr(3 \le X < 7)$ where $X \sim \text{Binomial}(8, 0.6)$

Note: $pr(3 \le X < 7) = pr(3 \le X \le 6) = pr(X \le 6) - pr(X \le 2)$

- **1.** Evaluate $pr(X \le 6)$ in cell D1 (use steps in B).
- **2.** Evaluate $pr(X \le 2)$ in cell D2 (use steps in B).
- **3.** In cell D3, type: **=D1 D2**.
- **4.** Press **Enter**. (The value of 0.844 should appear in cell D3).

Note:

Another way to calculate Binomial probabilities is to type the function "=BINOMDIST(x, n, p, c)" directly into the cell, where:

- x is the number of success in trials.
- n is the number of trials.
- p is the probability of success in each trial.
- c indicates whether the function is cumulative (TRUE or 1) or not cumulative(FALSE or 0).

For example:

To evaluate $pr(X = 5)$	In cell A1, type: =BINOMDIST(5, 8, 0.6, 0)
To evaluate $pr(X \le 3)$	In cell B1, type: =BINOMDIST(3, 8, 0.6, 1)
To evaluate $pr(X \ge 3)$	In cell C1, type: =BINOMDIST(2, 8, 0.6, 1) In cell C2, type: =1 - C1
To evaluate $pr(3 \le X \le 6)$	In cell D1, type: =BINOMDIST(6, 8, 0.6, 1) In cell D2, type: =BINOMDIST(2, 8, 0.6, 1) In cell D3, type: =D1 - D2

	A	В	С	D
1	= BINOMDIST(5, 8, 0.6, 0)	= BINOMDIST(3, 8, 0.6, 1)	= BINOMDIST(2, 8, 0.6, 1)	= BINOMDIST(6, 8, 0.6, 1)
2			=1-C1	= BINOMDIST(2, 8, 0.6, 1)
3				=D1-D2
Figure 3				

Calculating Poisson Probabilities

A. Individual probabilities: pr(X = x)

Example: Find pr(X = 5) where $X \sim Poisson(\lambda = 3)$

- 1. Click cell A1.
- 2. Click the Paste Function button f_{π}

from the tool bar.

- 3. Choose Statistical from the **Function category** list box in the **Paste Function** dialog box.
- 4. Choose **POISSON** from the **Function name** list box (Figure 1).

Paste Function	? ×
Paste Function Function category: Most Recently Used All Financial Date & Time Math & Trig Statistical Lookup & Reference Database Text	Punction name: NORMSINV PEARSON PERCENTILE PERCENTRANK PERMUT POISSON PROB QUARTILE PANK
Logical Information	RAINK RSQ SKEW
POISSON(x,mean,cumulative)	
Returns the Poisson distribution.	
	OK Cancel

Figure 1

- 5. Click OK.
- 6. Fill in the **POISSON** dialog box (Figure 2).

POISSON		
	X	🗾 = number
	Mean	🛐 = number
Cum	ulative	🗾 = logical
<u> </u>		
Returns th	e Poisson distribution.	=
	\boldsymbol{X} is the number of events.	
2	Formula result =	OK Cancel

Figure 2

where:

X is the number of events. It is equivalent to x in our manual. In this example, we put 5 in this box.

Mean is the mean of the distribution. It is equivalent to λ in our manual. In this example, we put 3 in this box.

Cumulative indicates whether the Poisson probability is cumulative (TRUE or 1) or not(FALSE or 0). Since the individual probability is not cumulative, we put FALSE in this box.

7. Click OK. (The value of 0.101 should appear in cell A1).

B. Lower tail probabilities: $pr(X \le x)$

Example: Find $pr(X \le 5)$ where $X \sim Poisson(\lambda = 3)$

- 1. Click cell B1.
- 2. Follow Steps 1 to 4 in A.
- **3.** Set: X = 5

Mean = 3

- **4.** The lower tail probability is cumulative, hence we set **Cumulative** = TRUE.
- 5. Click OK. (The value of 0.916 should appear in cell B1).

C. Upper tail probabilities: $pr(X \ge x)$

Example: Find $pr(X \ge 5)$ where $X \sim Poisson(\lambda = 3)$

Note: $pr(X \ge 5) = 1 - pr(X \le 4)$

- **1.** Evaluate $pr(X \le 2)$ in cell C1 (use steps in B).
- **2.** In cell C2, type: **=1 C1**.
- **3.** Press **Enter**. (The value of 0.185 should appear in cell C2).

D. $pr(a \le X \le b)$

Example: Find $pr(2 \le X < 9)$ where $X \sim Poisson(\lambda = 3)$

Note: $pr(2 \le X < 9) = pr(2 \le X \le 8) = pr(X \le 8) - pr(X \le 1)$

- 1. Evaluate $pr(X \le 8)$ in cell D1 (use steps in B).
- **2.** Evaluate $pr(X \le 1)$ in cell D2 (use steps in B).
- **3.** In cell D3, type: **=D1 D2**.
- 4. Press Enter. (The value of 0.797 should appear in cell D3).

Note:

Another way to calculate Poisson probabilities is to type the function "=POISSON(x, λ , c)" directly into the cell, where:

- x is the number of events.
- λ is the mean of the distribution.
- c indicates whether the function is cumulative (TRUE or 1) or not cumulative(FALSE or 0).

For example:

	In cell D3, type: =D1 - D2
To evaluate $pr(2 \le X \le 8)$	In cell D1, type: =POISSON(8, 3, 1) In cell D2, type: =POISSON(1, 3, 1)
To evaluate $pr(X \ge 5)$	In cell C1, type: =POISSON(4, 3, 1) In cell C2, type: =1 - C1
To evaluate $pr(X \le 5)$	In cell B1, type: =POISSON(5, 3, 1)
To evaluate $pr(X = 5)$	In cell A1, type: =POISSON(5, 3, 0)

	A	В	С	D
1	=POISSON(5,3,0)	=POISSON(5,3,1)	=POISSON(4,3,1)	=POISSON(8,3,1)
2			=1-C1	=POISSON(1,3,1)
3				=D1-D2

Figure 3

A. Lower tail probabilities: $pr(X \le x)$

Example: Find $pr(X \le 5)$ where $X \sim Normal(\mu = 7, \sigma = 6)$

- **1.** Click in cell A1.
- 2. Click the **Paste Function** button

from the tool bar.

- 3. Choose Statistical from the Function category list box in the Paste Function dialog box.
- 4. Choose NORMDIST from the Function name list box (Figure 1).

 f_{x}

Paste Function		
Function category: Most Recently Used All Financial Date & Time Math & Trig Statistical Lookup & Reference Database Text Logical Information	Function name: MIN MINA MODE NEGBINOMDIST NORMINY NORMSDIST NORMSINY PEARSON PERCENTILE PERCENTRANK	
NURMDIST(x,mean,standard.	_dev,cumulative)	
Returns the normal cumulative dis standard deviation.	stribution for the specified mean and	
2	Cancel OK	

Figure 1

- 5. Click OK.
- 6. Fill in the NORMDIST dialog box (Figure 2).

NODMDIST		
NUKTIDISI		
X	5	*: = 5
Mean	7	1 = 7
Standard_dev	6	1 = 6
Cumulative	TRUE	🔩 = TRUE
		= 0.369441404
Returns the normal	cumulative distribution for the specified	mean and standard deviation.
Cumulative	e is a logical value: for the cumulative dis for the probability mass function, use F	tribution function, use TRUE; ALSE.
🛛 Formula	a result = 0.369441404	Cancel OK

Figure 2

where:

is the value for which we want the distribution. It is equivalent to x in our manual. In this example, we put 5 in this box.

- **Mean** is the mean of the distribution. It is equivalent to μ in our manual. In this example, we put 7 in this box.
- **Standard_dev** is the standard deviation of the distribution. It is equivalent to σ in our manual. In this example, we put 6 in this box.
 - **Cumulative** indicates whether we want it is cumulative distribution function(TRUE or 1) or probability mass function(FALSE or 0). We will always put TRUE or 1 in this box.
- 7. Click OK. (The value of 0.369 should appear in cell A1.)

B. Upper tail probabilities: $pr(X \ge x)$

Example: Find $pr(X \ge 9)$ where $X \sim Normal(\mu = 7, \sigma = 6)$

Note: $pr(X \ge 9) = 1 - pr(X \le 9)$

- **1.** Evaluate $pr(X \le 9)$ in cell B1 (use steps in section A. above.).
- **2.** In cell B2, type: **=1 B1**.
- 3. Press Enter. (The value of 0.369 should appear in cell B2.)

2. C. pr(a $\leq X \leq b$)

Х

Example: Find $pr(5 \le X \le 11)$ where $X \sim Normal(\mu = 7, \sigma = 6)$

Note: $pr(5 \le X \le 11) = pr(X \le 11) - pr(X \le 5)$

- **1.** Evaluate $pr(X \le 11)$ in cell C1 (use steps in section A. above.).
- **2.** Evaluate $pr(X \le 5)$ in cell C2 (use steps in section A. above).
- **3.** In cell C3, type: =C1 C2.
- 4. Press Enter. (The value of 0.378 should appear in cell C3.)

Another way to calculate Normal probabilities is to type the function "=NORMDIST(x, μ, σ, c)" directly into a cell, where:

- x is the value for which we want the distribution.
- μ is the mean of the distribution.
- σ is the standard deviation of the distribution
- c is always set the value at 1.

For example:

To evaluate $pr(X \le 5)$:	In cell A1, type: =NORMDIST(5, 7, 6, 1) .
To evaluate $pr(X \ge 9)$:	In cell B1, type: =NORMDIST(9, 7, 6, 1) In cell B2, type: 1 – B1 .
To evaluate $pr(5 \le X \le 11)$:	In cell C1, type: =NORMDIST(11, 7, 6, 1). In cell C2, type: =NORMDIST(5, 7, 6, 1). In cell C3, type: C1 – C2.

	A	В	С
1	=NORMDIST(5,7,6,1)	=NORMDIST(9,7,6,1)	=NORMDIST(11,7,6,1)
2		=1-B1	=NORMDIST(5,7,6,1)
- 3			=C1-C2

Figure 3

Calculating the Inverse of the Normal Distribution

Sometimes the *x*-value for a specified probability is required.

Example: What mark would a student have to get more than, in order to be in the top 25% of the class, if the mean mark was 65 out of 100 and the standard deviation is 9.

1. Click in cell A1.

- 2. Click the **Paste Function** button from the tool bar.
- 3. Choose Statistical from the Function category list box in the Paste Function dialog box.
- 4. Choose NORMINV from the Function name list box (Figure 4).

Paste Function			
Function category: Most Recently Used All Financial Date & Time Math & Trig Statistical Lookup & Reference Database Text Logical Logical	Function name: MEDIAN MIN MINA MODE NEGBINOMDIST NORMDIST NORMSDIST NORMSDIST NORMSINY PEARSON DEDCENTILE		
NODMINV(probability mean	etenderd dev)		
Noki i Ni (pi ubabi i (g, mean	NORTHNY (probability, mean, standard_dev)		
Returns the inverse of the normal mean and standard deviation.	cumulative distribution for the specified		
	Cancel OK		

Figure 4

- 5. Click OK.
- **6.** Fill in the **NORMINV** dialog box (Figure 5).

Proba	ibility 0.75 💽 🔩 = 0.75	
	Mean 65 💽 = 65	
Standar	rd_dev 9 💽 = 9	
Returns the standard dev Standa	= 71.07041329 inverse of the normal cumulative distribution for the specified mean ar viation. ard_dev is the standard deviation of the distribution, a positive numbe	9 nd r.
2	Formula result = 71.07041329	ОК



- **Note** the Excel function **NORMINV** determines the *x*-value for the probability that is to the left of the
- : required *x*-value. In this example we want the *top* 25% therefore we use 1 0.25, or 0.75 for the probability. If instead we wanted the *bottom* 25% the probability is 0.25.
- 7. Click OK. (The value of 71.0704 should appear in cell A1.)

The formula can be directly entered into the cell by typing "=NORMINV(p, μ, σ)", where:

- *p* is the probability to the left of the *x*-value being calculated
- μ is the mean of the distribution
- σ is the standard deviation of the distribution

Note:

Excel has two other functions that work in the same manner as the functions explained above. These two functions are **NORMSDIST** and **NORMSINV**. These two functions calculate the value for a standard normal distribution, ie. $X \sim Normal(0, 1)$.

Generating Random Samples

Example:

Generate a random sample of 20 numbers that are Normally distributed with mean, μ =3, and standard deviation, σ =1.5.

- 1. Make A1 the active cell. On the menu bar Click Edit L Data Analysis
- 2. The Data Analysis dialog box is brought up. Select Random Number Generation as shown in Figure 1 below. Click OK.

Data Analysis	
Analysis Tools Exponential Smoothing F-Test Two-Sample for Variances Fourier Analysis Histogram Moving Average Random Number Generation Rank and Percentile Regression	OK Cancel Help

Figure 1

3. The Random Number Generation dialog box appears. The Number of variables is 1. The Number of random numbers is 20, and use your ID number as the Random seed. The Distribution is the Normal distribution with a Mean of 3 and Standard deviation of 1.5. The Output range is A1.

Random	Number Generation 📃	
Input Number of variables: Number of random numbers: Random seed: Distribution: Mean = Standard deviation =	1 20 9512345 Normal ♦ 3 1.5	OK Cancel Help
Output Options Output range: A1 New worksheet ply: New workbook		

Figure 2

- **4.** There should be 20 numbers in cells A1:A20. The same tool can be used to generate random samples from the Poisson and the Binomial distributions.
- **5.** To generate 20 numbers from a Binomial distribution with 10 trials and a probability of success equal to 0.25, fill in the Random Number Generation dialog box as seen below in Figure 3.

Random	Number Generation 📃	
Input Number of variables: Number of random numbers: Random seed: Distribution: Number of trials = p Value =	1 20 9512345 Binomial 10 0.25	OK Cancel Help
Output Options Output range: A1 New worksheet ply: New workbook		

Figure 3

6. To generate 20 numbers that come from a Poisson distribution with lambda equal to 2, fill in the Random Number Generation dialog box as shown below.

Random	Number Generation 📃	
Number of variables: Number of random numbers: Random seed: Distribution: Lambda =	1 20 9512345 Poisson € 2	OK Cancel Help
Output Options Output range: A1 New worksheet ply: New workbook		

Figure 3

Example:

To investigate the distribution of the sample mean we will use *Excel* to simulate the rolling of a fair sixsided die. We will find the average (mean) outcome in a sample consisting of 4 throws. We will repeat this 300 times giving us 300 samples from which we will have 300 sample means. We will also use *Excel* to calculate some summary statistics and to construct a histogram of these 300 sample means.

1. Enter the headings shown below in Figure 1 into an *Excel* worksheet.

	A	В	С	D	E	F	G	Н
1	Sample		omes	S	ample Mear	IS		
2	Number	1st Roll	2nd Roll	3rd Roll	4th Roll			
Figure 1								

2. Enter 1 in cell A3 and 2 in cell A4. Highlight these two cells. Move the mouse cursor over the black box in the bottom right hand corner of the highlighted cells so that a black plus sign is seen, as shown in Figure 2. Hold the mouse buttom down and drag down to cell A302.

	A	В
1	Sample	
2	Number	1st Roll
3	1	
4	2	l
5	I	•

Figure 2

- **3.** In cells A3:A302 should be the numbers 1-300.
- 4. In cell B3 enter the formula =**TRUNC**(**RAND**()*6+1)
- **5.** In the same manner as Step 2 above, drag the formula in cell B3 across cells C3, D3 to E3. This can be seen in Figure 3.

	A	В	С	D	E
1	Sample		Outo	omes	
2	Number	1 st Roll	2nd Roll	3rd Roll	4th Roll
3	1	=TRUNC(RAND()*6+1)	=TRUNC(RAND()*6+1)	=TRUNC(RAND()*6+1)	=TRUNC(RAND()*6+1)

Figure 3

6. Enter the formula =AVERAGE(B3:E3) in cell G3. (See Figure 4.)

F	G			
	Sample Means			
	= AVER AGE(B3 :E3)			
Figure 4				

Select cells B3:G3. Using the method in Step 2, drag the formulas in the selected cells down to row 302. The first few rows will look similar to the figure below. The values should be different due to the fact that a random number generator is being used.

	A	В	С	D	Е	F	G	Н
1	Sample		Outco	omes		S	ample Mean	IS
2	Number	1st Roll	2nd Roll	3rd Roll	4th Roll			
- 3	1	6	1	4	5		4	
4	2	5	3	3	3		3.5	
5	3	2	6	2	6		4	
6	4	1	2	4	5		3	
7	5	4	5	3	5		4.25	
8	6	6	5	5	6		5.5	
9	7	3	4	1	2		2.5	
10	8	6	4	5	1		4	

Figure 5

8. Set up the heading in the cells shown below in figure 6.

	1	J	К	L	М
1	Mean	Frequency		Summary Stat	tistics
2				Mean	
- 3 -				Std Dev	
4					

Figure	6
riguit	U

9. In cell M2 enter the formula =AVERAGE(G3:G302) and in cell M3 enter the formula =STDEV(G3:G302) (Figure 7.)

	L	М
1	Summary Statistics	
2	Mean	= AVER AGE(G3:G302)
3	Std Dev	=STDEV(63:6302)

Figure 7

10. Enter 1 and 1.25 in cells I2 and I3 respectively. Highlight cells I2and I3. In the same manner as that described in Step 2, drag the values down to cell I22, so that the worksheet looks like the figure below.

	I	J
1	Mean	Frequency
2	1	
3	1.25	
4	1.5	
5	1.75	
6	2	
7	2.25	
8	2.5	
9	2.75	
10	3	
11	3.25	
12	3.5	
13	3.75	
14	4	
15	4.25	
16	4.5	
17	4.75	
18	5	
19	5.25	
20	5.5	
21	5.75	
22	6	

Figure 8

11. Select cells J2:J22. Type =FREQUENCY(G3:G302,I2:I22) while the cells are still selected, but *do not press the Enter key*. Then while holding down the Shift and the Control key, press the Enter key. The selected cells will now contain the counts (yours will be different) of the corresponding mean in column I, as seen in Figure 9.

	1	J
1	Mean	Frequency
2	1	0
3	1.25	0
4	1.5	2
5	1.75	10
6	2	4
7	2.25	13
8	2.5	17
9	2.75	24
10	3	36
11	3.25	33
12	3.5	34
13	3.75	38
14	4	26
15	4.25	20
16	4.5	24
17	4.75	7
18	5	6
19	5.25	5
20	5.5	0
21	5.75	1
22	6	0

Figure 9



. The Chart Wizard dialog box 12. Highlight cells J2:J22. Click the Chart Wizard button pops up. Choose the Column Chart type and the Chart sub-type indicated in the Figure 10 below. Click Next.

Chart Wizard - Step 1 of 4 - Chart Type				
Standard Types Custom Types				
Chart type: Column Column Column Column Chart sub-type: Chart sub-ty				
Press and Hold to Yiew Sample Cancel < Back Next > Finish				

Figure 10

13. The next Chart Wizard dialog box appears. Click on the Series tab. In the Category (X) axis labels box type =Sheet1!\$I\$2:\$I\$22 Click Next.

Series Series1 ①	Name:	<u></u>
Add Remove	Values: =Sheet1!\$J\$2:	\$J\$22 💽
Category (X) axis labels:	=Sheet1!\$I\$2:\$I\$22	<u></u>
Cancel Cancel Finish		

Figure 11

14. Another Chart Wizard dialog box is brought up (Figure 12). The Chart title is Distribution of Sample Means, the Category (X) axis is Sample Mean, and the Value (Y) axis is frequency. Remove the Gridlines and the Legend. Click Next.

Chart Wizard -
Titles Axes Gridlines
Chart title:
Distribution of Sample Means
Category (X) axis:
Sample Mean
Value (Y) axis:
frequency
Second category (X) axis:
Second value (Y) axis:
[]]
<u> </u>

Figure 12

15. The last dialog box is shown. Ensure that the chart will be put on the current sheet. Then click Finish.

A. Upper tail probabilities: $pr(T \ge t)$

Example: Find $pr(T \ge 2)$ where $T \sim \text{Student}(df = 9)$

- **1.** Click in cell A1.
- 2. Click the **Paste Function** button

from the tool bar.

3. Choose Statistical from the Function category list box in the Paste Function dialog box.

 f_{x}

4. Choose **TDIST** from the **Function name** list box (Figure 1).

Paste Function		
Function category: Most Recently Used All Financial Date & Time Math & Trig Sitistical Lookup & Reference Database Text Logical Information TDIST(x,deg_freedom,tails Returns the Student's t-distribu	Function name: STANDARDIZE STDEV STDEVA STDEVPA STDEVPA STEYX TOIST TINV TREND TRIMMEAN TTEST	
2	Cancel OK	

Figure 1

- 5. Click OK.
- **6.** Fill in the **TDIST** dialog box (Figure 2).

TDIST		
:	X 2 3 = 2	
Deg_freedor	m 9 💽 = 9	
Tail	s 1 📑 = 1	
Returns the Stude	nt's t-distribution. = 0.0	38276412
Tai	ils specifies the number of distribution tails to return distribution = 1; two-tailed distribution = 2.	:one-tailed
2 Form	ula result = 0.038276412	cel OK

Figure 2

where:

- **X** is the value for which we want the distribution. It is equivalent to t in our manual. In this example, we put 2 in this box.
- Deg_freedo is the number of degrees of freedom. In this example, we put 29 in this box.
 m
 - Tails is the number of tails required either 1 or 2. In this example, we put 1 in this box.
- 7. Click OK. (The value of 0.038 should appear in cell A1.)

B. Lower tail probabilities: $pr(T \le t)$

Example: Find $pr(T \le 1)$ where $T \sim \text{Student}(df = 9)$

Note: $pr(T \le 1) = 1 - pr(T \ge 1)$

- **1.** Evaluate $pr(T \ge 1)$ in cell B1 (use steps in section A. above.).
- **2.** In cell B2, type: **=1 B1**.
- 3. Press Enter. (The value of 0.828 should appear in cell B2.)

3. C. $pr(a \le T \le b)$

Example: Find $pr(0 \le T \le 1)$ where $T \sim \text{Student}(df = 9)$

Note: $pr(0 \le T \le 1) = pr(T \ge 0) - pr(T \ge 1)$

- **1.** Evaluate $pr(T \ge 0)$ in cell C1 (use steps in section A. above.).
- **2.** Evaluate $pr(T \ge 1)$ in cell C2 (use steps in section A. above.).
- **3.** In cell C3, type: =**C1 C2**.
- 4. Press Enter. (The value of 0.328 should appear in cell C3.)

Another way to calculate Student *t*-probabilities is to type the function "=TDIST(x, df, tails)" directly into a cell, where:

- x is the value for which we want the distribution.
- *df* is the number of degrees of freedom of the distribution.
- tails specifies whether one tail or two tails are to be calculated

For example:

To evaluate $pr(X \le 2)$:	In cell A1, type: =TDIST(2, 9, 1) .
To evaluate $pr(X \ge 1)$:	In cell B1, type: =TDIST(1, 9, 1) In cell B2, type: 1 – B1 .
To evaluate $pr(0 \le X \le 1)$:	In cell C1, type: =TDIST(0, 9, 1) . In cell C2, type: =TDIST(1, 9, 1) . In cell C3, type: C1 – C2 .

	A	В	С
1	=TDIST(2,9,1)	=TDIST(1,9,1)	=TDIST(0,9,1)
2		=1-B1	=TDIST(1,9,1)
3			=C1-C2

Figure 3

Calculating the Inverse of the Student *t*-distribution

Example: Find the *t*-value $t_{30}(0.025)$ ie. Probability 0.025 and 30 degrees of freedom for use in a 95% confidence interval as the *t*-multiplier.

- 1. Click on cell A1.
- 2. Click the **Paste Function** button from the tool bar.
- 3. Choose Statistical from the Function category list box in the Paste Function dialog box.
- 4. Choose **TINV** from the **Function name** list box (Figure 4).

Paste Function			
Function category: Most Recently Used All Financial Date & Time Math & Trig Statistical Lookup & Reference Database Text Logical Information TINY (probability,deg_free Returns the inverse of the Studen	Function name: STDEVPA STEVYA STEYX TDIST TINV TREND TRIMMEAN TREST VAR VAR VAR VAR VAR VAR VAR VAR		
	Cancel OK		

Figure 4

5. Click OK

6. Fill in the **TINV** dialog box (figure 5).

TINY		
Probability	0.05	1 = 0.05
Deg_freedon	30	1 = 30
Returns the invers	e of the Student's t-distribution. m is a positive integer indicating f characterize the distribution.	= 2.042270353
Formu	la result = 2.042270353	Cancel OK

Figure 5

Note:

The *Excel* function **TINV** calculates the *t*-value for two–tailed *t*-distribution. So if we want to find the *t*-value whose probability to the right is 0.1, then in the **TINV** function the value for the probability is entered as 0.2, because of the two-tailed nature of the function.

7. Click OK. (The value 2.042 should appear in cell A1.)

Note:

The examples can be solved by directly typing the formula "=TINV(p, df)" into the cell, where:

- p is the probability for the two-tailed distribution
- df is the number of degrees of freedom for the distribution

t-Test of Means for Two Independent Samples

Example:

Perform a *t*-test on the Urinary Androsterone Levels Data. (Refer to example 10.2.1 in your textbook.)

1. Enter the data into *Excel* as shown in Figure 1 below.

	A	В
1	heterosexual	homosexual
2	3.9	2.5
- 3	4	1.6
4	3.8	3.9
5	3.9	3.4
6	2.9	2.3
7	3.2	1.6
8	4.6	2.5
9	4.3	3.4
10	3.1	1.6
11	2.7	4.3
12	2.3	2
13		1.8
14		2.2
15		3.1
16		1.3

Figure 1

- 2. From the menu bar select Tools Ł Data Analysis...
- **3.** Select the **t-Test: Two-Sample Assuming Unequal Variances** from the **Analysis Tools** list box seen in Figure 2.

Data Analysis	
Analysis Tools Random Number Generation Rank and Percentile Regression Sampling t-Test: Paired Two Sample for Means t-Test: Two-Sample Assuming Equal Variances t-Test: Two-Sample Assuming Unequal Variances z-Test: Two Sample for Means	OK Cancel Help

Figure 2

4. Click OK.

5. The dialog box below is brought up. In the Variable 1 range box enter A1:A12, and in the Variable 2 range box enter B1:B16. The Hypothesized mean difference is 0. Mark the Labels checkbox. And the Output range is A20.

t-Test: Two-Sa	mple Assuming Unequal Vari	ances 📃 🗌
Input Variable 1 range: Variable 2 range: Hypothesized mean differe ☑ Labels Alpha:	A1:A12	OK Cancel Help
Output Options Output range: New worksheet ply: New workbook	A20	

Figure 3

6. Click OK. The results will appear on the worksheet as in Figure 4.

	A	В	С
20	t-Test: Two-Sample Assuming U	nequal Variance	95
21			
22		heterosexual	homosexual
23	Mean	3.51818182	2.5
24	Variance	0.51963636	0.85142857
25	Observations	11	15
26	Hypothesized Mean Difference	0	
27	df	24	
28	t Stat	3.1572239	
-29	P(T<=t) one-tail	0.00212947	
-30	t Critical one-tail	1.71088232	
31	P(T<=t) two-tail	0.00425893	
32	t Critical two-tail	2.06389814	

Figure 4

t-Test of Means for Paired Data

Example:

Perform a *t*-test on the Airforce Headsizes Data. (Refer to Example 10.1.3 in your textbook.)

	A	В	С	
1	recruit	cardboard	metal	
2	1	146	145	
3	2	151	153	
4	3	163	161	
5	4	152	151	
6	5	151	145	
- 7 -	6	151	150	
8	7	149	150	
9	8	166	163	
10	9	149	147	
11	10	155	154	
12	11	155	150	
13	12	156	156	
14	13	162	161	
15	14	150	152	
16	15	156	154	
17	16	158	154	
18	17	149	147	
19	18	163	160	

1. Enter the data seen in Figure 1 into *Excel*.



- 2. From the menu bar select Tools & Data Analysis...
- 3. The Data Analysis dialog box is brought up. From the Analysis Tools list box select t-Test: Paired Two Sample for Means.(See Figure 2). Click OK.

Data Analysis	
Analysis Tools Random Number Generation Rank and Percentile Regression Sampling t-Test: Paired Two Sample for Means t-Test: Two-Sample Assuming Equal Variances t-Test: Two-Sample Assuming Unequal Variances z-Test: Two Sample for Means	OK Cancel Help

Figure 2

4. The following dialog box is shown (Figure 3). In the Variable 1 range box enter B1:B19, and in the Variable 2 range box enter C1:C19. Mark the Labels checkbox. The Hypothesized mean difference is 0. The Output range is A23.

t-Test: Pa	ired Two Sample for M	leans 📃 📃	
Input Variable 1 range: Variable 2 range: Hypothesized mean differen ∑ Labels Alpha:	B1:B19 C1:C19 nce: 0	Cance	
Output Options Output range: New worksheet ply: New workbook	A23		

Figure 3

5. Click **OK**. The results will appear as in Figure 4 below.

	A	В	С
23	t-Test: Paired Two Sample for Means		
24			
25		cardboard	metal
26	Mean	154.555556	152.944444
27	Variance	33.9084967	30.6437908
-28	Observations	18	18
- 29	Pearson Correlation	0.92985902	
-30	Hypothesized Mean Difference	0	
31	df	17	
32	t Stat	3.1854219	
-33	P(T<=t) one-tail	0.00270736	
-34	t Critical one-tail	1.73960643	
35	P(T<=t) two-tail	0.00541472	
- 36	t Critical two-tail	2.10981852	

Figure 4

Example: Construct one-way ANOVA table for the Cancer Data.

- В С A. 1 Stomach Bronchus Colon 2 1.18 0.81 2.01 3 -0.41 3.16 0.51 4 -0.64 -0.22 1.07 5 1.32 2.05 4.31 6 1.39 1.9 1 7 3.51 0.14 0.85 8 2.29 0.27 3 9 -0.1 0.73 1.79 10 2.2 1.84 1.67 11 3.83 3.86 2.16
- **1.** Enter the information as Figure 1 below.

4. Figure 1

- 2. From the Menu bar, select Tool Data Analysis.
- 3. Choose Anova: Single Factor from the Data Analysis dialog box. (Figure 2)

Data Analysis	? ×
<u>A</u> nalysis Tools	OK
Anova: Single Factor	
Anova: Two-Factor With Replication	Cancel
Anova: Two-Factor Without Replication	Cancor
Correlation	11-1-
Covariance	
Descriptive Statistics	
Exponential Smoothing	
F-Test Two-Sample for Variances	
Fourier Analysis	
Histogram	

5. Figure 2

4. Click OK.

Input Range:	A1:C11
Grouped By:	Columns
Label in First Row	\checkmark
Alpha:	0.05

6. In the **Output options**, choose **New Worksheet Ply**. The ANOVA table will be produced on a new worksheet. The **Anova: Single Factor** dialog box should look like Figure 3.

Anova: Single Factor		? ×
Input Input Range: Grouped By: Labels in First Row Alpha: 0.05	A1:C11 [●] <u>C</u> olumns [●] <u>R</u> ows	OK Cancel <u>H</u> elp
Output options C Output Range: Kew Worksheet Ply: New Workbook		

Figure 3

7. Click **OK**. The ANOVA table in Figure 4 should appear on a new worksheet.

	A	В	С	D	E	F	G
1	Anova: Sin	gle Factor					
2							
3	SUMMAR	(
4	Groups	Count	Sum	Average	Variance		
5	Stomach	10	11.2	1.12	1.987022		
6	Bronchus	10	15.25	1.525	1.651094		
7	Colon	10	21.03	2.103	1.419312		
8							
9							
10	ANOVA						
11	rce of Varia	SS	df	MS	F	P-value	F crit
12	Between G	4.881327	2	2.440663	1.447769	0.25277	3.354131
13	Within Gro	45.51686	27	1.68581			
14							
15	Total	50.39819	29				

Figure 4

Chi-square Test

One-Dimensional Tables

Example: Perform a Chi-square test for goodness of fit on the Rolls of a Die Data. (Refer to example 11.1.1 in your textbook.)

1. Enter the data into *Excel* as done in Figure 1. Also enter the formula =SUM(B2:B7) into cell B8.

	A	В
1	outcome	count
2	1	26
- 3	2	40
4	3	37
5	4	26
6	5	43
- 7 -	6	38
8		=SUM(B2:B7)



2. Enter expected into cell C1. Then enter the formula =\$B\$8/6 into cells C2:C7.

С
expected
=\$B\$8/6



3. In cell A10 type =CHITEST(B2:B7,C2:C7) as seen in Figure 3.

9	
10	=CHITEST(B2:B7,C2:C7)
11	

Figure 3

4. Once all the formulas have been entered, the *Excel* worksheet will look like the figure below. Cell A10 contains the *P-value* for the goodness of fit test.

	A	В	С
1	outcome	count	expected
2	1	26	35
- 3	2	40	35
4	3	37	35
5	4	26	35
6	5	43	35
7	6	38	35
8		210	
9			
10	0.18329453		



Two-Dimensional Tables

Example:

Perform a Chi-square test of homogeneity on the Phenotype Regional Data. The method described below can also be used to perform a test of independence. (Refer to example 11.2.2 in your textbook.)

		Phenotype			
Region	А	В	0	AB	Totals
Nithsdale	98	35	115	5	253
Cree	38	9	79	6	132
Rhinns	36	9	47	7	99
Totals	172	53	241	18	484

1. Enter the above data into *Excel* in the manner shown in the figure below.

	A	В	С	D	E	F
1	observed	A	В	0	AB	Totals
2	Nithsdale	98	35	115	5	253
- 3 -	Cree	38	9	79	6	132
4	Rhinns	36	9	47	7	99
5	Totals	172	53	241	18	484



2. In cells B8:E10, enter the formulas seen below.

7	expected	A	В	0	AB
8	Nithsdale	=(\$F\$2*B5)/\$F\$5	=(\$F\$2*C5)/\$F\$5	=(\$F\$2*D5)/\$F\$5	=(\$F\$2*E5)/\$F\$5
9	Cree	=(\$F\$3*B\$5)/\$F\$5	=(\$F\$3*C\$5)/\$F\$5	=(\$F\$3*D\$5)/\$F\$5	=(\$F\$3*E\$5)/\$F\$5
10	Rhinns	=(\$F\$4*B\$5)/\$F\$5	=(\$F\$4*C\$5)/\$F\$5	=(\$F\$4*D\$5)/\$F\$5	=(\$F\$4*E\$5)/\$F\$5

Figure 6

3. The results will then be displayed in the worksheet as shown in Figure 7.

7	expected	A	В	0	AB
8	Nithsdale	89.9091	27.7045	125.977	9.40909
9	Cree	46.9091	14.4545	65.7273	4.90909
10	Rhinns	35.1818	10.8409	49.2955	3.68182



4. Select cell A13. Then on the tool bar click the Paste Function button *I*. The Paste Function dialog box is brought up. Choose Statistical from the Function category box, and then choose CHITEST from the Function name box. Click OK.

Pa	aste Function	
Function category: Most Recently Used All Financial Date & Time Math & Trig Statistical Lookup & Reference Database Text Logical Information	Function name: AVEDEV AVERAGE AVERAGEA BETADIST BETAINV BINOMDIST CHIDIST CHIINV CHITEST CONFIDENCE CORREL	
CHITEST(actual_range,ex	xpected_range)	-
Returns the test for independen distribution for the statistic an	nce: the value from the chi-squared nd the appropriate degrees of freedom.	
2	Cancel OK	

Figure 8

The CHITEST dialog box appears. The Actual_range is B2:E4, and the Expected_range is B8:E10. Click OK.

Actual_ra	nge B2:E4 💽 = {0.828281801897
Expected_ra	nge B8:E10 🔩 = {0,0,0,0;0,0,0;0,0,0;(
Returns the tes statistic and the Expected_r	= CHITEST(B2:E4,B8:E10 t for independence: the value from the chi-squared distribution for the appropriate degrees of freedom. range is the range of data that contains the ratio of the product of row totals and column totals to the grand total.
🕐 For	rmula result = CHITEST(B2:E4,B8:E10)

Figure 9

6. Finally the worksheet should look as in Figure 10. The *P-value* for the Chi-square test is found in cell A13 and the value is 0.015.

	A	В	С	D	E	F
1	observed	A	В	0	AB	Totals
2	Nithsdale	98	35	115	5	253
- 3 -	Cree	38	9	79	6	132
4	Rhinns	36	9	47	7	99
5	Totals	172	53	241	18	484
6						
7	expected	A	В	0	AB	
8	Nithsdale	89.9091	27.7045	125.977	9.40909	
9	Cree	46.9091	14.4545	65.7273	4.90909	
10	Rhinns	35.1818	10.8409	49.2955	3.68182	
11						
12						
13	0.0150207					

Figure 10

Simple Linear Regression

Example:

Perform simple linear regression on the Chernobyl Data. (Refer to example 3.1.1 in your textbook.)

1. Enter the Chernobyl data into *Excel* as seen in Figure 1.

	A	В
1	percent	radiation
2	2.2	23
- 3	2.4	20
4	1.9	22
5	3.9	29
6	3.6	32
- 7	2.6	21
8	0	16
9	4.2	37
10	5	44

Figure 1

- 2. From the menu bar choose Tools then the Data Analysis option.
- 3. The Data Analysis dialog box appears (See Figure 2). Select Regression and then click OK.

Data Analysis	
Analysis Tools Histogram Moving Average Random Number Generation Rank and Percentile Regression Sampling t-Test: Paired Two Sample for Means t-Test: Two-Sample Assuming Equal Variances	OK Cancel Help

Figure 2

4. Fill in the Regression dialog box as shown in Figure 3 below. The Input Y range is A1:A10, and the Input X range is B1:B10. Mark the Labels checkbox. Mark the circle next to New workbook for the Output Options. Also mark the Residual plots checkbox.

	Regression	
Input Input Y range: Input X range: Input X range: Confidence level:	A1:A10 B1:B10 Constant is zero 95 %	OK Cancel Help
Output Options Output range: New worksheet ply: New workbook		
Residuals Residuals Standardized residuals Normal probability Normal probability plo	☑ Residual plots □ Line fit plots ts	

Figure 3

5. Click OK. The regression analysis tool produces the output seen below in Figure 4.

	Α	В	С	D	E	F	G	Н	I
1	SUMMARY OUTPUT								
2									
3	Regression	Statistics							
4	Multiple R	0.91165222							
5	R Square	0.83110976							
6	Adjusted R Sq	0.80698259							
- 7 -	Standard Erro	0.65570289							
8	Observations	9							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	14.810376	14.810376	34.4470379	0.00061847			
13	Residual	7	3.009624	0.42994629					
14	Total	8	17.82						
15									
16		Coefficients	itandard Error	t Stat	P-yalue	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	-1.1796123	0.72323094	-1.6310313	0.14690466	-2.8897805	0.5305559	-2.8897805	0.5305559
18	radiation	0.14924799	0.02542919	5.8691599	0.00061847	0.08911755	0.20937844	0.08911755	0.20937844

Figure 4

6. *Excel* also produces a Residual plot. This plot needs to be enlarged to look like the plot shown below in Figure 5.



Figure 5

Note:

The Normal probability plots are not the plots we are looking for as they are not normal plots for the residuals. So we ignore this option.

Correlation

Example:

Determine the sample correlation coefficient of the Chernobyl Data. (Refer to example 3.1.1 in your textbook.)

1. Enter the data into *Excel* as seen below in Figure 1.

	A	В
1	percent	radiation
2	2.2	23
- 3	2.4	20
4	1.9	22
5	3.9	29
6	3.6	32
7	2.6	21
8	0	16
9	4.2	37
10	5	44



- 2. From the menu bar select Tools **L** Data Analysis...
- **3.** The **Data Analysis** dialog box is brought up. Select **Correlation** from the **Analysis Tools** list box seen in the figure below.

Data Analysis	
Analysis Tools Anova: Single Factor Anova: Two-Factor With Replication Anova: Two-Factor Without Replication Correlation Covariance Descriptive Statistics Exponential Smoothing F-Test Two-Sample for Variances	OK Cancel Help



- 4. Click OK.
- 5. Then the Correlation dialog box is brought up. The Input range is A1:B10. Mark the check box for Labels in the first row, and ensure that the circle next to Columns is marked as well. The Output range is A15.

	Correlatio	n	
Input Input range: Grouped by: Infirst row	\$A\$1:\$B\$10 Columns	Rows	OK Cancel
Output Options Output range: New worksheet ply: New workbook	A15		Help



6. Click OK. The results that *Excel* outputs can be seen below in Figure 4.

15		percent	radiation
16	percent	1	
17	radiation	0.91165222	1

Figure 4

7. The sample correlation coefficient is contained in cell B17.

Copying and Pasting Excel into Word

It is useful to be able to include parts of a *Excel* spreadsheet or an *Excel* graph in a *Word* document as part of a report or assignment.

Example:

The *Excel* regression output and the graph below needs to be placed into a report.

	A	В	С	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	Regression	Statistics					
4	Multiple R	0.91165222					
5	R Square	0.83110976					
6	Adjusted R Sq	0.80698259					
- 7	Standard Erro	0.65570289					
8	Observations	9					
9							
10	ANOVA						
11		đf	SS	MS	F	Significance F	
12	Regression	1	14.810376	14.810376	34.4470379	0.00061847	
13	Residual	7	3.009624	0.42994629			
14	Total	8	17.82				
15							
16		Coefficients	itandard Error	t Stat	P-yalue	Lower 95%	Upper 95%
17	Intercept	-1.1796123	0.72323094	-1.6310313	0.14690466	-2.8897805	0.5305559
18	radiation	0.14924799	0.02542919	5.8691599	0.00061847	0.08911755	0.20937844
19							

Figure 1



Figure 2

- 1. Select cells A1:G18. Click on the menu bar Edit & Copy.
- 2. Then move over to *Word*. Place the blinking cursor at the spot that you wish the copied data to go. Click on the menu bar Edit & Paste.
- 3. Once copied and pasted the regression output would look like the table below.

Regressior	Statistics					
Multiple R	0.91165222					
R Square	0.83110976					
Adjusted R Square	0.80698259					
Standard	0.65570289					
Error						
Observations	9					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	14.810376	14.810376	34.4470379	0.00061847	
Residual	7	3.009624	0.42994629			
Total	8	17.82				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-1.1796123	0.72323094	-1.6310313	0.14690466	-2.8897805	0.5305559
radiation	0.14924799	0.02542919	5.8691599	0.00061847	0.08911755	0.20937844

SUMMARY OUTPUT

- 4. Activate *Excel*. Click on the white part of the graph near the edge. Small black squares appear around the edge of the graph. Copy the graph using the method described in Step 1.
- 5. Activate Word and place the blinking cursor in the spot that you want put the graph. Paste the copied graph as in Step 2.
- 6. The graph will then appear in the *Word* document.