Department of Statistics and Operations Research

College of Science

King Saud University



Exercises

STAT 328 (Statistical Packages)







nashmiah r.alshammari ^_^

Excel and Minitab

MATHEMATICAL FUNCTIONS

Write the commands of the following:

		By Excel (using (fx))	By Minitab calc \rightarrow calculator
Absolute value	-4 =4	ABS(-4)	
Combinations	$\binom{10}{6}$ =10C6=210	COMBIN(10;6)	
The exponential function	$e^{-1.6} = 0.201897$	EXP(-1.6)	
Factorial	110! =1.5882E+178	FACT(110)	
Floor function	[-3.15]= -4	INT(-3.15)	
Natural logarithm	ln(23)= 3.135494216	LN(23)	
Logarithm with respect to any base	$\log_{9}(4) = 0.630929754$	LOG(4;9)	
Logarithm with respect to base 10	log(12) = 1.079181246	LOG10(12)	
Multinomial Coefficient	$\begin{pmatrix} 9\\2&2&5 \end{pmatrix} = 756$	MULTINOMIAL(2;2;5)	
Square root	√85= 9.219544457	SQRT(85)	
Summation	Summation of: 450,11,20,5 = 486	SUM(450;11;20;5)	
Permutations	10P6=151200	PERMUT(10;6)	
Product	Product of: 450,11,20,5 = 495000	PRODUCT(450;11;20;5)	
Powers	$10^{-4} = 0.0001$	POWER(10;-4)	

MATRICES

Write the commands of the following:

			By Minitab
		By Excel (using (fx))	 data → copy → columns in matrix display data calc → matrices → arithmetic invers The name of matrices in <u>columns</u> in matrix keeps their names + × Names of matrix containing The name of new matrices in arithmetic and invers is (M#).
Addition of Matrices	$A = \begin{bmatrix} -5 & 0 \\ 4 & 1 \end{bmatrix}, B = \begin{bmatrix} 6 & -3 \\ 2 & 3 \end{bmatrix}$ $\Rightarrow A + B = \begin{bmatrix} -5 + 6 & 0 + -3 \\ 4 + 2 & 1 + 3 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 6 & 4 \end{bmatrix}$		
Subtract of Matrices	$C = \begin{bmatrix} 1 & 2 \\ -2 & 0 \\ -3 & -1 \end{bmatrix}, D = \begin{bmatrix} 1 & -1 \\ 1 & 3 \\ 2 & 3 \end{bmatrix}$ $\Rightarrow C - D = \begin{bmatrix} 1 - 1 & 2 - (-1) \\ -2 - 1 & 0 - 3 \\ -3 - 2 & -1 - 3 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ -3 & -3 \\ -5 & -4 \end{bmatrix}$		
Additive Inverse of Matrix	$A = \begin{bmatrix} 1 & 0 & 2 \\ 3 & -1 & 5 \end{bmatrix}$ $\Rightarrow -A = \begin{bmatrix} -1 & 0 & -2 \\ -3 & 1 & -5 \end{bmatrix}$		
Scalar Multiplication of Matrices	$D = \begin{bmatrix} -3 & 0 \\ 4 & 5 \end{bmatrix}$ $\Rightarrow 3D = \begin{bmatrix} -9 & 0 \\ 12 & 15 \end{bmatrix}$		
Matrix Multiplication	$E = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}, F = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$ $\implies E \times F = \begin{bmatrix} 30 & 66 \\ 36 & 81 \\ 42 & 96 \end{bmatrix}$		
Determinant and Inverse Matrices	$G = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$ $\Rightarrow \det(G) = 1 \text{ and } G^{-1} = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$		

CONDITIONAL FUNCTION (IF) AND COUNT CONDITIONAL FUNCTION

By Excel

(using (fx))

We have grades of 10 students

73	45	32	85	98	78	82	87	60	25	64	72	12	90

1. Print student case being successful (Mark >=60) and being a failure (Mark <60).

2. How many successful students?

3. How many students whose grades are less than or equal to 80?

DESCRIPTIVE STATISTICS

We have students' weights as follows: 44, 40, 42, 48, 46, 44. Find:

	By Excel (using (fx) and (Data Analysis))	By Minitab stat → basic statistics → display descriptive statistics + See Appendix -1-
Mean=44	AVERAGE(C2:C7)	
Median=44	MEDIAN(C2:C7)	
Mode=44	MODE.SNGL(C2:C7)	
Sample standard deviation=2.828	STDEV.S(C2:C7)	
Sample variance=8	VAR.S(C2:C7)	
Kurtosis=-0.3	KURT(C2:C7)	
Skewness=4.996E-17	SKEW(C2:C7)	
Minimum=40	MIN(C2:C7)	
Maximum=48	MAX(C2:C7)	
Range=8	MAX(C2:C7)-MIN(C2:C7)	
Count=6	COUNT(C2:C7)	
Coefficient of variation=6.428%	STDEV.S(C2:C7)/AVERAGE(C2:C7)*100	

- * Range= Maximum-Minimum
- ** Coefficient of variation= $\frac{\text{Sample standard deviation}}{\text{Mean}} \times 100\%$

PEARSON CORRELATION COEFFICIENT

We have the table illustrates the age X and blood pressure Y for eight female.

08	49	60	42	55	63	36	42	Х
152	145	155	140	150	140	118	125	Y

Find:

	By Excel (using (fx) and (Data Analysis))	By Minitab stat \rightarrow basic statistics \rightarrow correlation + \checkmark Display p-value
Correlation=0.791832	CORREL(M3:M10;N3:N10)	

PROBABILITY DISTRIBUTION FUNCTIONS

Discrete Distributions

Notes If X is discrete random variable, then 1) $P(a < X \le b) = P(X \le b) - P(X \le a)$ and so, if $P(a \le X < b) = P((a-1) < X \le (b-1)) = P(X \le (b-1)) - P(X \le (a-1))$ or $P(a \le X \le b) = P((a-1) < X \le b) = P(X \le b) - P(X \le (a-1))$ or $P(a \le X \le b) = P((a-1) < X \le b) = P(X \le b) - P(X \le (a-1))$ or $P(a < X < b) = P(a < X \le (b-1)) = P(X \le (b-1)) - P(X \le a).$ 2) $P(X > a) = 1 - P(X \le a)$, $P(X \ge a) = 1 - P(X \le a)$, $P(X \ge a) = 1 - P(X \le a) = 1 - P(X \le (a-1))$, $P(X < a) = P(X \le (a-1))$

1. Binomial Distribution

A biased coin is tossed 6 times. The probability of heads on any toss is 0.3. Let X denote the number of heads that come up. Calculate:

(i) If we call heads a success then this X has a binomial distribution with parameters n = 6 and p = 0.3.

$$P(X = 2) = {6 \choose 2} (0.3)^2 (0.7)^4 = 0.324135$$

(ii)

$$P(X = 3) = {6 \choose 3} (0.3)^3 (0.7)^3 = 0.18522.$$

(iii) We need $P(1 < X \le 5)$

P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)= 0.324 + 0.185 + 0.059 + 0.01 = 0.578



- 8 -

2. Poisson Distribution

Births in a hospital occur randomly at an average rate of 1.8 births per hour.

What is the probability of observing 4 births in a given hour at the hospital?

Let X = No. of births in a given hour

(i) Events occur randomly (ii) Mean rate $\lambda = 1.8$ $\Rightarrow X \sim Po(1.8)$

We can now use the formula to calculate the probability of observing exactly 4 births in a given hour

 $P(X = 4) = e^{-1.8} \frac{1.8^4}{4!} = 0.0723$

What about the probability of observing more than or equal to 2 births in a given hour at the hospital?

We want $P(X \ge 2) = P(X = 2) + P(X = 3) + \dots$

i.e. an infinite number of probabilities to calculate

but

$$P(X \ge 2) = P(X = 2) + P(X = 3) + \dots$$

= 1 - P(X < 2)
= 1 - (P(X = 0) + P(X = 1))
= 1 - (e^{-1.8} \frac{1.8^0}{0!} + e^{-1.8} \frac{1.8^1}{1!})
= 1 - (0.16529 + 0.29753)
= 0.537



- 9 -

Continuous Distributions

Notes

If X is continuous symmetric random variable (as Normal distribution and Student's tdistribution), then

1) $P(X \ge x) = 1 - P(X \le x)$ and $P(X \le x) = 1 - P(X \ge x)$ 2) $P(X \le x) = 1 - P(X \le -x)$ and $P(X \ge x) = 1 - P(X \ge -x)$

1. Exponential Distribution

On the average, a certain computer part lasts 10 years. The length of time the computer part lasts is exponentially distributed.

What is the probability that a computer part lasts more than 7 years?

Solution

Let X = the amount of time (in years) a computer part lasts.

$$\mu = 10$$
 so $m = \frac{1}{\mu} = \frac{1}{10} = 0.1$

$$P(X > 7) = 1 - P(X < 7).$$

 $P(X > 7) = e^{-0.1 \cdot 7} = 0.4966$. The probability that a computer part lasts more than 7 years is 0.4966.

By Excel	By Minitab calc \rightarrow probability distribution
$\times \sqrt{f_x} = 1$ -EXPON.DIST(7;1/10;1)	
وسيطات الدالة 🔹 🗴	
EXPON.DIST	
V = 🎼 7 X	
•,) = 🎫 1/10 Lambda	
TRUE = 🔣 1	
، ٥٠٢٤/٢٩٦ = إرجاع التوزيع الأسى. (رجاع التوزيع الأسى، قيمة منطقية للدالة التى سيتم إرجاعها: دالة التوزيع التراكمى = TRUE؛ دالة كتافة الاحتمال = FALSE.	
ناتج الصيغة = ٤٠٢٥/٥٢٩٤,٠ تعليمات حول هذه الدالة تعليمات حول هذه الدالة	

2. Normal Distribution



- 11 -

3. Student's t Distribution

\bigcap	Notes in 1	Excel	$\overline{}$
	 1) =T.DIST(t, ν, 0) 2) =T.DIST(t, ν, 1) 3) =T.DIST.RT(t, ν) 	$ \begin{array}{l} \leftrightarrow \ f_{\mathcal{T}_v}(t) \\ \leftrightarrow \ P(\mathcal{T}_v \leq t) \\ \leftrightarrow \ P(\mathcal{T}_v \geq t) \end{array} $	
	4) =T.DIST.2T(t , ν)	$\leftrightarrow 2P(T_v \ge t)$	
	5) =T.INV(p, v)	$\leftrightarrow P(T_v \leq t_0) = p$	
	6) =T.INV.2T(p, ν)	$\leftrightarrow 2P(T_v \geq t_0) = p$	

Find:

 $(a)t_{0.025}$ when v = 14

 $(b)t_{0.01}$ when v = 10

 $(c)t_{0.995}$ when v = 7



Given a random sample of size $\mathbf{24}$ from a normal distribution, find \mathbf{k} such that:

$$(a) P(-1.7139 < T < k) = 0.90$$
$$(b) P(k < T < 2.807) = 0.95$$
$$(c) P(-k < T < k) = 0.90$$

(a)

 $\begin{array}{l} P(-1.7139 < T_{23} < k) = 0.9 \\ \leftrightarrow P(T_{23} < k) - P(T_{23} < -1.7139) = 0.9 \\ \leftrightarrow P(T_{23} < k) = 0.9 + P(T_{23} < -1.7139) \\ \leftrightarrow P(T_{23} < k) = 0.949997 \end{array}$

By Excel		By Minitab
$(using (f_x))$		$calc \rightarrow probability distribution$
fr =T.DIST(-1.7139:23:1)		
وسيطات الدالة		
	T.DIST	
1,V119- = 🔣 -1.7139 X		
۲۳ = 🐹 23 Deg	g_freedom	
TRUE = 🔣 1	mulative	
•,•£999VYEE =	الحاق تعاملا لعالم	
العرف الايمن. Cumulative قدمة منطقية: من أجل دالة التقنيع التداكمة بر استخده	ן (جוع توريع) للطالب دي ו	
الأحتمال، استخدم FALSE.		
• ,• £ ⁶	ناتج الصيغة = 33٣٧٩٩٩	
=T.INV(.949997;23)		
وسيطات الدالة		
	T.INV	
•,92999V = 💽 .949997	Probability	
۲۳ = 🎫 23	Deg_freedom	
1,VITAT9779 =		
طالب ذي الطرف الأيسر،	إرجاع عكس توزيع t للد	
Deg_freedom عدد صحیح موجب یشیر إلی عدد درجات ا		
١,٧١٣٨٢	ناتج الصيغة = ٩٣٦٩	
OIn excel you might make it in one step too		
$P(T_{23} < k) = 0.9 + P(T_{23} < -1.7139)$		
so, = T.INV(0.9 + T.DIST(-1.7139,23,1), 23	3) = 1.713839369	

(b)

 $\begin{array}{l} P(k < T_{23} < 2.807) = 0.95 \\ \leftrightarrow P(T_{23} < 2.807) - P(T_{23} < k) = 0.95 \\ \leftrightarrow P(T_{23} < k) = (T_{23} < 2.807) - 0.95 \\ \leftrightarrow P(T_{23} < k) = 0.044996 \end{array}$

By Excel	By Minitab
(using (fx))	$calc \rightarrow probability distribution$
=T.DIST(2.807;23;1)	
وسيطات الدالة	
$T.DIST$ $T, \Lambda \cdot V = 100 \text{ fm} 2.807 \text{ fm} X$ $TT = 100 \text{ fm} 2.3 \text{ fm} $	
= ۲۹۲۲۹۹۹۱۱۹ = إرجاع توزيع t للطالب ذي الطرف الأيمن. X القيمة الرقمية المراد تقييم التوزيع عندها. ناتج الصيغة = ۹۹۲۹۹۹۲۱۹۹،	
f_x =T.INV(D1;23)	
وسيطات الدالة	
T.INV ۰,۰٤٤٩٩٦/٢٩ = الله D1 Probability ۲۲ = الله 23 Deg_freedom	
= ۱٫۷۲۹۹۵۲۵۷۲ = إرجاع عكس توزيع t للطالب ذي الطرف الأيسر . Deg_freedom عدد صحيح موجب يشير إلى عدد درجات الحرية التي تم	
ناتح الصيغة = -٧١٥٢٥٧٦١ (
O In excel you might make it in one step too $P(T_{23} < k) = (T_{23} < 2.807) - 0.95$ so $= T.INV(-0.95 + T.DIST(2.807.23.1), 23) = -1.769952576$	

(*c*)

$$\begin{array}{l} (i) \ P(T_{23} < k) - P(T_{23} < -k) = .9 \\ \leftrightarrow \ P(T_{23} < k) - \{1 - P(T_{23} < k)\} = 0.9 \\ \leftrightarrow \ 2P(T_{23} < k) - 1 = 0.9 \\ \leftrightarrow \ 2P(T_{23} < k) = 1.9 \\ \leftrightarrow \ P(T_{23} < k) = 0.95 \\ so, \qquad = T. inv \ (0.95,23) = 1.71387 \\ (ii) \ P(T_{23} < k) - P(T_{23} < -k) = .9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - P(T_{23} < -k) = 0.9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - \{1 - P(T_{23} > -k)\} = 0.9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - \{1 - [1 - P(T_{23} > k)]\} = 0.9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - \{1 - 1 + P(T_{23} > k)\} = 0.9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - \{1 - 1 + P(T_{23} > k)\} = 0.9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - \{1 - 1 + P(T_{23} > k)\} = 0.9 \\ \leftrightarrow \ 1 - P(T_{23} > k) - 0.9 \\ \leftrightarrow \ 1 - 2P(T_{23} > k) = 0.1 \\ so, \qquad = T. inv. \ 2t(0.1,23) = 1.71387 \end{array}$$

By Excel	By Minitab
(using (fx))	$calc \rightarrow probability distribution$
f_x =T.INV(.95;23)	
وسيطات الدالة	
T.INV	
•,90 = .95 Probability	
۲۳ = 🐹 23 Deg_freedom	
), ٧) ٣٨٧) ٥٢٨ = الجاء عكب تعليم الله الذي الأبيان	
رجاح عصل توربع ٢ تصالب دي الطرف الإسمار. Probability الاجتمال المقترن بتوزيع t للطالب ثنائه . الطرف، رقم س. ب	
ناتج الصيغة = ٢٨٥١/١٢٨٧ الميغة	
=T.INV.2T(.1;23)	
وسيطات الدالة	
T.INV.2T	
+,) = 🐹 .1 Probability	
۲۳ = 🚺 23 Deg_freedom	
1,V1TAV10TA =	
إرجاع عكس توزيع t للطالب ثنائي الطرف،	
Deg_freedom عدد صحيح موجب يشير إلى عدد درجات الحرية	
ناتج الصيغة = ١,٧١٢٨٧١م٢٨	

4. Chi-Square Distribution

	Notes in Exc	el
1) 2) 3) 4) 5)	= CHISQ. DIST $(x, v, 0)$ = CHISQ. DIST $(x, v, 1)$ = CHISQ. DIST. RT $(x, v, 1)$ = CHISQ. INV (p, v) = CHISQ. INV. RT (p, v)	$ \begin{array}{l} \leftrightarrow f_{\chi_{\mathcal{V}}}(x) \\ \leftrightarrow P(\chi_{\mathcal{V}} \leq x) \\ \leftrightarrow P(\chi_{\mathcal{V}} \geq x) \\ \leftrightarrow P(\chi_{\mathcal{V}} \leq x_{0}) = p \\ \leftrightarrow P(\chi_{\mathcal{V}} \geq x_{0}) = p \end{array} $

By using chi- square distribution ,Find:

 $\chi^2_{0.995}$ when $\nu = 19$



5. F Distribution

\frown	Notes in Excel							
1)	= F . DIST (f , v ₁ , v ₂ , 0)	$\leftrightarrow f_{F_{v_1,v_2}}(f)$						
2)	$=$ F. DIST $(f, v_1, v_2, 1)$	$\leftrightarrow P(F_{\nu_1,\nu_2} \le f)$						
3)	$=$ F. DIST. RT($f, v_1, v_2, 1$)	$\leftrightarrow P(F_{v_1,v_2} \ge f)$						
4)	$=$ F. INV (p, v_1, v_2)	$\leftrightarrow P(F_{v_1,v_2} \leq f_0) = p$						
5)	$=$ F. INV. RT (p, v_1, v_2)	$\leftrightarrow P(F_{v_1,v_2} \ge f_0) = p$						

From the tables of F- distribution ,Find:





HYPOTHESIS TESTING STATISTICS AND CONFIDENCES INTERVAL

	By Excel (using (Data Analysis))	By Minitab
Z test one sample for mean with known variance	x	stat \rightarrow basic statistics
T test one sample for mean with unknown variance	×	\checkmark stat \rightarrow basic statistics
T test two samples for means assuming equal variance and unequal variance	\checkmark	\checkmark stat \rightarrow basic statistics
T test parried two samples for means	\checkmark	\checkmark stat \rightarrow basic statistics
One-way ANOVA (Single Factor ANOVA)	\checkmark	stat \rightarrow ANOVA \rightarrow one way \rightarrow response data are in one column for all factor levels. response data are in a separate column for each factor level.
Linear regression	\checkmark	\checkmark Stat \rightarrow regression \rightarrow regression \rightarrow fit regression model
Independent test	×	Stat \rightarrow tables \rightarrow chi-square test for association \rightarrow summarized data in a two-way table cross tabulation and chi-square \rightarrow raw data (categorical variables) + See Appendix -2-

Notes

p-value

(1) $H_1: \theta \neq \theta_0 \rightarrow p - value_{two tail} = 2P(distribution of test statistical > |test statistical|)$

(2) $H_1: \theta > \theta_0 \rightarrow p - value_{one tail(>)} = P(distribution of test statistical > test statistical)$

 $(3) H_1: \theta < \theta_0 \rightarrow p - value_{\text{one tail } (<)} = P(distribution \ of test \ statistical \ < test \ statistical)$

In the programs (Excel and Spss for symmetric distribution), how to find p-value for the one tail from p-value for two tail?

p – value one tail	test statistical > 0	Then we have $p - value_{one tail}(>)$
$= \frac{p - value_{two tail}}{p - value_{two tail}}$		and $p - value_{one tail (<)} = 1 - p - value_{one tail (>)}$
2	test statistical < 0	Then we have $p - value_{one tail}(<)$
		and $p - value_{one tail} (>) = 1 - p - value_{one tail} (<)$

- 18 -

1)

For a sample of 10 fruits from thirteen-year-old acidless orange trees, the fruit shape (determined as adiameter divided by height) was measured [Shaheen and Hamouda (1984b)]: 1.066 1.084 1.076 1.051 1.059 1.020 1.035 1.052 1.046 0.976 Assuming that fruit shapes are approximately normally distributed, find and interpret a 90% confidence interval for the average fruit shape.

(T test one sample for mean with unknown variance By Minitab)

<		One-Sample t for the Mean	One-Sample t: Options
		C1 Q1 One or more samples, each in a column	Confidence level: 90
÷	C1		
	Q1	QI	Alternative hypothesis: Mean ≠ hypothesized mean ▼
1	1.066	· · · · · · · · · · · · · · · · · · ·	
2	1.084		Help <u>OK</u> Cancel
3	1.076		
4	1.051		
5	1.059		
6	1.020	<u>Hypothesized mean</u> :	
7	1.035	Calast Crashs	
8	1.052	Opuo <u>n</u> s <u>G</u> raphs	
9	1.046		
10	0.976	Help <u>O</u> K Cancel	

One-Sample T: Q1

Variable N Mean StDev SE Mean 90% CI Q1 10 1.04650 0.03103 0.00981 (1.02851; 1.06449) 2)

[In a sample of 185 people in the Western Region who had a particular bacterial infection, the mean egg count (per gram of stool) was 141 [Ghandour et. al. (1991)]. Assume that egg counts of such people are normally distributed with a variance of 3025.]

Find and interpret a 90% confidence interval for the average egg count.

** In a sample of 185 people in the Western Region who had a particular bacterial infection, the mean egg count (per gram of stool) was 141. Assume that egg counts of such people are normally distributed with a variance of 3025. Can we conclude that the true mean egg count is different from 130. Use α =0.10.

(Z test one sample for mean with known variance By Minitab)

<u>s</u>	One-Sample Z for the Mean
	Summarized data
One-Sample Z	Sample size: 185
Test of $\mu = 130$ vs $\neq 130$	Sample mean: 141
The assumed standard deviation = 55	Known standard deviation: 55
N Mean SE Mean 90% CI Z P 185 141.00 4.04 (134.35; 147.65) 2.72 0.007	Perform hypothesis test Hypothesized mean: 130
One-Sample Z: Options	
Confidence level:	Select Options Graphs
▲ <u>A</u> lternative hypothesis: Mean ≠ hypothesized mean	Help <u>O</u> K Cancel
1 Help <u>QK</u> Cancel	

- 20 -

3)

The phosphorus content was measured for independent samples of skim and whole

Whole: 94.95 95.15 94.85 94.55 94.55 93.40 95.05 94.35 94.70 94.90 Skim: 91.25 91.80 91.50 91.65 91.15 90.25 91.90 91.25 91.65 91.00 Assuming normal populations with equal variances

- a) Test whether the average phosphorus content of skim milk is less than the average phosphorus content of whole milk. Use α =0.01
- b) Find and interpret a 99% confidence interval for the difference in average phosphorus contents of whole and skim milk

(T test two samples for means assuming equal variance By Minitab)

		Two-Sample t for the Mean	
C3	C4	C1 01 Each sample is in its own column	12 C13 C14 C15 C16
skim	whole	C3 skim	Two-Sample t: Options
91.25	94.95	C4 Whole Sample 1: skim	= (sample 1 mean) - (sample 2 mean)
91.80	95.15	Sample 2: whole	
91.50	94.85	Confidence	e level: 99
91.65	94.55	Hypothesiz	ed difference: 0.0
91.15	94.55	Altornative	bypothesist Difference a bypothesist difference
90.25	93.40		Insponesized difference
91.90	95.05	Select Options Graphs	equal variances
91.25	94.35		
91.65	94.70	Holp OK Cased Help	<u>O</u> K Cancel
91.00	94.90		

Two-Sample T-Test and CI: skim; whole

Two-sample T for skim vs whole

N Mean StDev SE Mean skim 10 91.340 0.483 0.15 whole 10 94.645 0.503 0.16 pifference = μ (skim) - μ (whole) Estimate for difference: -3.305 99% upper bound for difference: -2.742 T-Test of difference = 0 (vs <): T-Value = -14.99 P-Value = 0.000 DF = 18 Both use Pooled StDev = 0.4931

Two-Sample t: Options							
Difference = (sample 1 m	Difference = (sample 1 mean) - (sample 2 mean)						
<u>C</u> onfidence level:	99						
Hypothesized difference:	0.0						
<u>A</u> lternative hypothesis:	Difference ≠ hypothesized difference	•					
Assume equal variance	es						
Help	<u>O</u> K Cancel						

Two-Sample T-Test and CI: skim; whole

Two-sample T for skim vs whole

N Mean StDev SE Mean skim 10 91.340 0.483 0.15 whole 10 94.645 0.503 0.16

Difference = μ (skim) - μ (whole) Estimate for difference: -3.305 99% CI for difference: (-3.940; -2.670) T-Test of difference = 0 (vs \neq): T-Value = -14.99 P-Value = 0.000 DF = 18

```
Both use Pooled StDev = 0.4931
```

Or

C6	С7-Т	Two-Sample t for the Mean ×					
91.25	skim	C1 Q1 Both samples are in one column					
91.80	skim	C4 whole Samples: C6					
91.50	skim	C7					
91.65	skim	Sample IDS: 107					
91.15	skim						
90.25	skim						
91.90	skim						
91.25	skim						
91.65	skim	Select Options Graphs					
91.00	skim						
94.95	whole	Help <u>Q</u> K Cancel					
95.15	whole	Two-Sample t: Options					
94.85	whole						
94.55	whole	Difference = (sample 1 mean) - (sample 2 mean)					
94.55	whole	Confidence level: 99					
93.40	whole	Hypothesized difference: 0.0					
95.05	whole						
94.35	whole	Alternative hypothesis: Difference < hypothesized difference					
94.70	whole	Assume equal variances					
94.90	whole						
		Help <u>QK</u> Cancel					

Two-Sample T-Test and CI: C6; C7

Two-sample T for C6

C7	N	Mean	StDev	SE Mean
skim	10	91.340	0.483	0.15
whole	10	94.645	0.503	0.16

bifference = µ (skim) - µ (whole) Estimate for difference: -3.305 99% upper bound for difference: -2.742 T-Test of difference = 0 (vs <): T-Value = -14.99 P-Value = 0.000 DF = 18 Both use Pooled StDev = 0.4931

Two-Sample t: Options						
Difference = (sample 1 mean) - (sample 2 mean)						
Confidence level: 99						
Hypothesized difference: 0.0						
Alternative hypothesis: Difference = hypothesized difference						
Assume equal variances						
Help <u>O</u> K Cancel						

Two-Sample T-Test and CI: C6; C7

Two-sample T for C6

C7 N Mean StDev SE Mean skim 10 91.340 0.483 0.15 whole 10 94.645 0.503 0.16

```
Difference = \mu (skim) - \mu (whole)
Estimate for difference: -3.305
99% CI for difference: (-3.940; -2.670)
T-Test of difference = 0 (vs \neq): T-Value = -14.99 P-Value = 0.000 DF = 18
Both use Pooled StDev = 0.4931
```

(T test two samples for means assuming equal variance By Excel)

Н		G		F	E	D	С	В	А	b .
		<u>.</u> ?		Data	Analysis			whole	skim	1
				Dutu	7 that you			94.95	91.25	2
	(Ж				<u>A</u> na	alysis Tools	95.15	91.8	3
	0-		1			Moving	Histogram	94.85	91.5	4
	Ca	ncei			Random Number Generation				91.65	5
	Rank and Percentile				Percentile	94.55	91.15	6		
	t-Test: Paired Two Sample for Means		Sampling	93.4	90.25	7				
			: Paired Two Sample for Means	for Means	95.05	91.9	8			
				t-Test: Ti	vo-Sample Ass	uming Unequal	Variances	94.35	91.25	9
			~		z-Tes	t: Two Sample f	for Means	94.7	91.65	10
								94.9	91	11
										12

× ° t	Test: Two-Sample As	suming Equal Variances
ОК	5A\$1:\$A\$11	Input :Variable <u>1</u> Range
Cancel	\$B\$1:\$B\$11	:Variable <u>2</u> Range
Cuint	0	:Hypothesized Mean Difference
		0.01 : <u>Al</u> pha
	1	Output options
		:New Worksheet Ply O
		New <u>W</u> orkbook ()

	skim	whole
Mean	91.34	94.645
Variance	0.233222	0.253028
Observations	10	10
Pooled Variance	0.243125	
Hypothesized Mean Difference	0	
df	18	
t Stat	-14.9879	
P(T<=t) one-tail	6.53E-12	
t Critical one-tail	2.55238	
P(T<=t) two-tail	1.31E-11	
t Critical two-tail	2.87844	

4)

In an experiment comparing 2 feeding methods for caves, eight pairs of twins were used – one twin receiving Method A and other twin receiving Method B. At the end of a given time, the calves were slaughtered and cooked, and the meat was rated for its taste (with a higher number indicating a better taste):

Twin pair	Method A	Method B	
1	27	23	
2	37	28	
3	31	30	
4	38	32	
5	29	27	
6	35	29	
7	41	36	
8	37	31	

Assuming approximate normality, test if the average taste score for calves fed by Method B is less than the average taste foe calves fed by Method A. Use α =0.05.

(T	test	parried	two	samples	for	means	By	Minitab)
----	------	---------	-----	---------	-----	-------	----	---------	---

Worksheet 1 ***					Paire	d t f	or the	e Mean		×
C8	C9	C10	C1	Q1		Ea	ach sam	ple is in a	column	•
	Method A	Method B	C3 C4	skim whole		6	mole 1	Mother	P'	_
	27	23	C6	Mathematic		20	mple 1;	Ineutoc	ю	
	37	28	C10	Metho	d B	Sa	mple 2:	'Method	A'	
	31	30								
	38	32								
	29	27								
	35	29								
	41	36		Sele	t		Opt	ions	Graph	ns
	37	31	•							
		Paired t: O	ptions			×		<u>о</u> к	Can	cel
Differ	ence = mean of	(sample 1 - samp	le 2)							
<u>C</u> onfi	dence level:	95.0	_							
<u>H</u> ypot	thesized differer	nce: 0.0								
<u>A</u> lterr	native hypothesi	s: Difference <	< hypoth	esized dif	ference	•				
	Help		<u>о</u> к		Cancel					

- 25 -

Paired T-Test and CI: Method B; Method A

Paired T for Method B - Method A

	N	Mean	StDev	SE Mean
Method B	8	29.50	3.82	1.35
Method A	8	34.38	4.87	1.72
Difference	8	-4.875	2.532	0.895

95% upper bound for mean difference: -3.179T-Test of mean difference = 0 (vs < 0): T-Value = -5.45 P-Value = 0.000

(T test parried two samples for means By Excel)

× S	Data Analysis	Method B	Method A
	o ata 7 marjoio	23	27
ОК	<u>A</u> nalysis Tools	28	37
Cancel	 Exponential Smoothing E-Test Two-Sample for Variances 	30	31
Cancer	Fourier Analysis	32	38
تعليمات	Histogram Moving Average	27	29
	Random Number Generation	29	35
	Rank and Percentile	36	41
	Sampling	31	37
	t-Test: Paired Two Sample for Means		

× t-Test: Paired Two Sample for Means

ОК	\$F\$1:\$F\$9	Input :Variable <u>1</u> Range
Cancel	55 \$E\$1:\$E\$9	:Variable <u>2</u> Range
<u>تعليمات</u>	0	:Hypothesized Mean Difference
		Labels 🗸
		0.05 : <u>Al</u> pha
		Output options
	15	: <u>O</u> utput Range 🔘
		:New Worksheet Ply 🔘
		New Workbook ()

t-Test: Paired Two Sample for N	leans	
	Method B	Method A
Mean	29.5	34.375
Variance	14.57143	23.69643
Observations	8	8
Pearson Correlation	0.857204	
Hypothesized Mean Difference	0	
df	7	
t Stat	-5.44586	
P(T<=t) one-tail	0.00048	
t Critical one-tail	1.894579	
P(T<=t) two-tail	0.00096	
t Critical two-tail	2.364624	

5)

Two independent samples of dates were taken-one from dates in the Khalal stage and one from dates at the Tamr stage. The calcium (in mg/100g) was measured [Sawaya (1986)]:

Khalal:30,57,29,23,55,50,49,74,101,97,79,158,112,107,93,63,70,90,98,48,75,64,71,72,146,37, 82,19,115,36,34,27,38,42,18,21,75,37,80,72,73,198,107,107,35,56,25,35,26,40,75,109,27,101 Tamr:14,25,21,18,28,14,19,20,44,18,24,47,19,52,31,38,41,39,35,16,47,26,26,30,81,18,42,9,49,2 3,27,14,15,17,10,16,18,14,13,32,42,55,42,27,30,17,24,14,20,17,48,20,76

Assuming normal populations with unequal variances (α =0.05)

- a) Test whether the average calcium of dates at the khalal stage is more than this average for Tamar stage dates
- b) Find the confidence interval for the difference in the average calcium of dates at the two stage

(T test two samples for means assuming unequal variance By Minitab)

				_
	Khalal	Tamr		2
1	30	14		2
2	57	25		2
3	29	21		2
4	23	18		2
5	55	28		3
6	50	14		3
7	49	19		3
8	74	20		3
9	101	44		3
10	97	18		3
11	79	24		3
12	158	47		З
13	112	19		3
14	107	52		3
15	93	31		4
16	63	38		4
17	70	41		4
18	90	39		4
19	98	35		4
20	48	16		4
21	75	47		4
22	64	26		4
23	71	26		4
24	72	30		_

26	37	18
27	82	42
28	19	9
29	115	49
30	36	23
31	34	27
32	27	14
33	38	15
34	42	17
35	18	10
36	21	16
37	75	18
38	37	14
39	80	13
40	72	32
41	73	42
42	198	55
43	107	42
44	107	27
45	35	30
46	56	17
47	25	24
48	35	14
49	26	20
50	40	17
51	75	48
52	109	20
53	27	76
54	101	

146

81

Tv	vo-Sample t for the Mean	
C1 Q1	Each sample is in its own column	Two-Sample t: Options
C4 whole C6	Sample 1: Khalal	Difference = (sample 1 mean) - (sample 2 mean)
C9 Method A C10 Method B	Sample 2: Tamr	Confidence level: 95
C12 Khalal C13 Tamr		Hypothesized difference: 0.0
		Alternative hypothesis: Difference < hypothesized difference
]	Assume equal variances
Select	Optio <u>n</u> s <u>G</u> raphs	
	OK Const	

Two-Sample T-Test and CI: Khalal; Tamr

Two-sample T for Khalal vs Tamr

	N	Mean	StDev	SE	Mean
Khalal	54	67.7	38.0		5.2
Tamr	53	28.7	15.7		2.2

```
Difference = µ (Khalal) - µ (Tamr)
Estimate for difference: 39.02
95% upper bound for difference: 48.36
T-Test of difference = 0 (vs <): T-Value = 6.97 P-Value = 1.000 DF = 70
```

Two-Sample t for the Mean	
Each sample is in its own column	Two-Sample t: Options
Sample 1: Khalal	Difference = (sample 1 mean) - (sample 2 mean)
Sample 2: Tamr	Confidence level: 95
	Hypothesized difference: 0.0
	<u>A</u> lternative hypothesis: Difference ≠ hypothesized difference ▼
	Assume equal variances
Select Options Graphs	
Heln OK Cancel	Help <u>QK</u> Cancel

Two-Sample T-Test and CI: Khalal; Tamr

Two-sample T for Khalal vs Tamr N Mean StDev SE Mean

Khalal	54	67.7	38.0	5.2
Tamr	53	28.7	15.7	2.2

Difference = μ (Khalal) - μ (Tamr) Estimate for difference: 39.02 95% CI for difference: (27.85; 50.20) T-Test of difference = 0 (vs \neq): T-Value = 6.97 P-Value = 0.000 DF = 70

(T test two samples for means assuming unequal variance By Excel)

Р	0	Ν	М	L	К	J	I	.
×	\$	Data	Analysis			Tamr	Khalal	1
		Data	Analysis			14	30	2
0	ж			Ana	lysis Tools	25	57	3
	ncel ^			Moving	listogram Average	21	29	4
	licel		Ran	dom Number G	eneration	18	23	5
مات	ميلعت			Rank and P	Percentile	28	55	6
_					Sampling	14	50	7
_		t-Test:	t-Test: Paire Two-Sample A	d Two Sample f ssuming Equal 1	or Means Variances	19	49	8
_		t-Test: Tv	vo-Sample Assu	iming Unequal	Variances	20	74	9
-	*		z-Test	: Two Sample f	or Means	44	101	10
	1					18	97	11
						24	79	12
						4/	158	13
						19	112	14
						52	107	15
						31	93	16
						38	63	1/
						41	/0	18
						39	90	19
						35	98	20
						16	48	21
						4/	/5	22

t-Test:	ming Unequal Variances	Test: Two-Sample Assu	× ? t-
	Input		
	:Variable <u>1</u> Range	\$I\$1:\$I\$55	OK
Mean	:Variable 2 Range	\$1\$1:\$1\$54	Cancel
Varian			- 1 - 1 - 7
Obser	:Hypothesized Mean Difference	0	لعليمان
Hypoth	Labels 🖌		
df	0.05 : <u>Al</u> pha		
t Stat			
P(T<=	Output options		
t Critic	: <u>O</u> utput Range 🔘	1	
P(T<=	:New Worksheet Ply 🔘		
t Critic	New Workbook 🔘		

t-Test: Two-Sample Assuming	Unequal Varianc	es
	Khalal	Tamr
Mean	67.74074074	28.71698
Variance	1443.139064	247.2837
Observations	54	53
Hypothesized Mean Difference	0	
df	71	
t Stat	6.965139095	
P(T<=t) one-tail	6.80179E-10	
t Critical one-tail	1.666599658	
P(T<=t) two-tail	1.36036E-09	
t Critical two-tail	1.993943368	

6)

	Vi	tamin D Lev	el	
Residence type	Insufficient < 5 ng/ml	Low 5-10 ng/ml	Sufficient > 10 ng/ml	Total
Tent	6	31	97	134
Mud house	16	73	349	438
Flat	45	174	652	871
Villa	64	323	1061	1448
Brick house	51	250	886	1187
Total	182	851	3045	4078

Formation of vitamin D depends on exposure to ultraviolet radiation in sunlight. A sample of Saudis was classified by the type of residence and the level of vitamin D [Sedrani et al. (1992)]:

Test whether the Vitamin D level of Saudis is related to the type of residence. Use a level of significance of 0.05.

(Independent test By Minitab)

			C	hi-Square Test for Association
			C1 Q1 C3 skim	Summarized data in a two-way table
C16	C17	C18	C4 whole	Columns containing the table:
			C9 Method A	C16 C17 C18
6	31	97	C10 Method B C12 Khalal	×
16	73	349	C13 Tamr C16	
45	174	652	C17	Labels for the table (optional)
64	323	1061	C18	Rows: (column with row labels)
51	250	886		Col <u>u</u> mns: (name for column category)
			Select	Statistics Options
			Help	<u>O</u> K Cancel

- 31 -

Chi-Square Test for Association: Worksheet rows; Worksheet columns

Rows: Worksheet rows Columns: Worksheet columns

	C16	C17	C18	A11
1	6	31	97	134
	6.0	28.0	100.1	
2	16	73	349	438
	19.5	91.4	327.1	
3	45	174	652	871
	38.9	181.8	650.4	
4	64	323	1061	1448
	64.6	302.2	1081.2	
5	51	250	886	1187
	53.0	247.7	886.3	
A11	182	851	3045	4078
Cell (Content	s:	Count	
			Expecte	d count

Pearson Chi-Square = 9.461; DF = 8; P-Value = 0.305 Likelihood Ratio Chi-Square = 9.668; DF = 8; P-Value = 0.289

Or

C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35	(
					Cross Tabul	ation and	l Chi-Squ	are	×							
6	1	1	C1	01	Raw data	(categorical	variablec)		-							
16	2	1	C3	skim		r (categoricai	variablesj									
45	3	1	C4 C6	whole	Rows:	C21										
64	4	1	C9 C10	Method A Method B	Columns:	C22										
51	5	1	C12	Khalal	Lawara	· · · · ·				_						
31	1	2	C15 C16	Idmr	Edycis.							Cross Tabu	lation: Ch	i-Square		×
73	2	2	C17 C18		<u>F</u> requencie	es: C20		(optional)		E chi	coupro toot					
174	3	2	C20							i i cin	square test					
323	4	2	C22							Statisti	cs to display	in each cell				
250	5	2								V	Expected ce	l counts				
97	1	3			Display						Raw residua	s				
349	2	3			Co	unts					- Standardized	d residuals				
652	3	3			C Ro	w percents					Adjusted res	iduale				
1061	4	3			Col	lu <u>m</u> n percent	ts				<u>M</u> ujusteu res	nuuuna				
886	5	3				tal percents					Each ceirs co		ni-square			
				Select	Ch	i-Square	Other St	tats	Optio <u>n</u> s	н	elp			<u>o</u> ĸ	Cancel	
			н	ielp			<u></u> K		Cancel							

Tabulated Statistics: C21; C22

Using frequencies in C20

Rows:	C21	Columns	: C22	
	1	2	3	All
1	6	31	97	134
	6.0	28.0	100.1	
2	16	73	349	438
	19.5	91.4	327.1	
3	45	174	652	871
	38.9	181.8	650.4	
4	64	323	1061	1448
	64.6	302.2	1081.2	
5	51	250	886	1187
	53.0	247.7	886.3	
A11	182	851	3045	4078
Cell	Content	ts:	Count	
			Expecte	d count

Pearson Chi-Square = 9.461; DF = 8; P-Value = 0.305 Likelihood Ratio Chi-Square = 9.668; DF = 8; P-Value = 0.289

7)

A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Observation	Program 1	Program 2	Program 3	Program 4
1	9	10	12	9
2	12	6	14	8
3	14	9	11	11
4	11	9	13	7
5	13	10	11	8

(One-way ANOVA by Minitab)

				One-Way Analysis of Variance	^
C25	C26	C27	C28	C1 Q1 C3 skim	⊡
				C4 whole C6 Responses:	
9	10	12	9	C9 Method A C25 C26 C27 C28	
12	6	14	8	C10 Method B C12 Khalal	
14	9	11	11	C13 Tamr	
11	9	13	7	C17	<u> </u>
13	10	11	8	C18 C20	
				C21 C22	
				C25	
				C27	
				C28 Options Comparisons Graphs	
					-
				Select Results Storage	
Norksheet: W	orksheet 1			Help <u>O</u> K Cancel	

- 34 -

328 stat	
One-Way Analysis of Variance: C	Options ×
✓ Assume equal variances Confidence level: 95 Type of confidence interval: Two-sided	e of means and interval plot)
Help	<u>Q</u> K Cancel
One-Way Analysis of Variance: Comparisons	One-way ANOVA: C25; C26; C27; C28 Method
Comparison procedures assuming equal variances Image: Table of the second se	Null hypothesis All means are equal Alternative hypothesis At least one mean is different Significance level $\alpha = 0.05$

 $\overline{\mathbf{v}}$

-

Cancel

Dunnett

🗌 Hsu MCB

Best:

Tests

Help

Results

Control group level: C25

☑ Interval plot for differences of means

Grouping information

Largest mean is best

<u>0</u>K

Equal variances were assumed for the analysis.

Factor Information

Factor Levels Values Factor 4 C25; C26; C27; C28

Analysis of Variance

 Source
 DF
 Adj SS
 Adj MS
 F-Value
 P-Value

 Factor
 3
 54.95
 18.317
 7.04
 0.003

 Error
 16
 41.60
 2.600
 1000
 1000

 Total
 19
 96.55
 1000
 1000
 1000
 1000

Tukey Simultaneous 95% Cls



- 35 -
Or

÷	C30	C31	C32	C33	C34	C35	C36	C37	C38	C39
					On	e-Way Ar	alysis of V	ariance		×
1	12	1	C1 Q1		A Res	onse data a	re in one colur	nn for all factor l	evels	-
2	14	1	C3 skir C4 wh	n ole						
4	11	1	C6 C7		Resp	onse: C30				
5	13	1	C9 Me	thod A	Eact	r: C31				
6	10	2	C10 Me	alal						
7	6	2	C13 Ta C16	mr						
8	9	2	C17 C18							
9	9	2	C20							
10	10	2	C22							
11	12	3	C25							
12	14	3	C27 C28		~		Detions	Comparisons.	<u>G</u> rap	ohs
13	11	3								
14	13	3		Select			R <u>e</u> sults	Storage		
15	11	3		,						
16	9	4	Help	•				<u>О</u> К	Car	ncel
1/	11	4								
10	7	4								
20	, 8	4								
<u>C</u> onfid <u>T</u> ype o	ence level: of confidence ir telp	95 nterval: Two-	sided	(for tab ▼	ole of means <u>O</u> K	and interval p Cance	lot)			
One Error r	e-Way Ana ate for compa	alysis of Va risons: 5	riance: Co	mpariso	ns ×					
Compa	Comparison procedures assuming equal variances Image: I									
Result	s Interval plot Grouping info Te <u>s</u> ts Help	for differences	of means		Cancel					

- 36 -

One-way ANOVA: C30 versus C31

Method

Null hypothesis All means are equal Alternative hypothesis At least one mean is different Significance level $\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor Levels Values C31 4 1; 2; 3; 4

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
C31	3	54.95	18.317	7.04	0.003
Error	16	41.60	2.600		
Total	19	96.55			

(One-way ANOVA by Excel)

Tukey Simultaneous 95% Cls



_	V	U		Т	S	R	Q	Р	0	Ν	М	L	
	x	2		Data	Analysis								1
	_			Data	Analysis			9	12	10	9		2
		ОК				<u>A</u> na	alysis Tools	8	14	6	12		3
			^			Anova: Sinc	le Factor	11	11	9	14		4
	Ca	ancel			Anova: Two Anova: Two-Fa	-Factor With R ctor Without R	eplication	7	13	9	11		5
	ات	تعاره				C	orrelation	8	11	10	13		6
	_					Descriptive	Statistics						7
						Exponential S	moothing						8
					F-Test T	vo-Sample for	Variances						9
			× .			H	listogram						10
													11

? ×	Anova: Single Fac	tor
OK Cancel تعليمات	\$M\$2:\$P\$6 Columns () Rows ()	Input :Input Range :Grouped By Labels in first row 0.05 :Alpha
		Output options :Output Range () :New Worksheet Ply () New Workbook ()

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Column 1	5	59	11.8	3.7		
Column 2	5	44	8.8	2.7		
Column 3	5	61	12.2	1.7		
Column 4	5	43	8.6	2.3		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	54.95	3	18.31667	7.044872	0.003113	3.238872
Within Groups	41.6	16	2.6			
Total	96.55	19				

8)

Ten Corvettes between 1 and 6 years old were randomly selected from last year's sales records in Virginia Beach, Virginia. The following data were obtained, where x denotes age, in years, and y denotes sales price, in hundreds of dollars.

x	6	6	6	4	2	5	4	5	1	2
у	125	115	130	160	219	150	190	163	260	260

a) Determine the regression equation for the data.

- b) Compute and interpret the coefficient of determination, r^2 .
- c) Obtain a point estimate for the mean sales price of all 4-year-old Corvettes.

(Linear regression by Minitab)

			Regression				
Ŧ	C34	C35	C1 Q1	Responses:			
	x	у	C3 skim C4 whole	У			
1	6	125	C6				
2	6	115	C10 Method B	Continuous predictors:			
3	6	130	C12 Khalal C13 Tamr				
4	4	160	C16				
5	2	219	C18				
6	5	150	C20 C21	×			
7	4	190	C22	Categorical predictors:			
8	5	163	C26	^			
9	1	260	C27 C28				
10	2	260	C30	~			
11			C34 x				
<			C35 y	Model Options Coding Stepwise			
💷 Pr	. @ 0		Select	<u>G</u> raphs <u>R</u> esults S <u>t</u> orage			

Regression Analysis: y versus x

I .

Model Summary

S R-sq R-sq(adj) R-sq(pred) 14.2465 93.68% 92.89% 90.16%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	291.6	11.4	25.51	0.000	
x	-27.90	2.56	-10.89	0.000	1.00

Regression Equation

y = 291.6 - 27.90 x

(Linear regression by Excel)



? ×	Regression	
OK Cancel تعليمات	\$\$\$1:\$\$\$11 \$\$\$\$1:\$\$\$11 Constant is Zero	Input :Input <u>Y</u> Range :Input <u>X</u> Range Labels
	Residual Plots	Output options :Output Range :New Worksheet Ply New Workbook Residuals Residuals Standardized Residuals Normal Probability

	-	~	-	-	•	~ .		
SUMMARY OUTPU	Г							
Regression S	Statistics							
Multiple R	0.967871585							
R Square	0.936775406							
Adjusted R Square	0.928872332							
Standard Error	14.24652913							
Observations	10							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	24057.89126	24057.89	118.533	4.48427E-06			
Residual	8	1623.708738	202.9636					
Total	9	25681.6						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	291.6019417	11,43289905	25.50551	5.98E-09	265.2376293	317.9662542	265.2376293	317,9662542
x	-27.90291262	2.562889198	-10.8873	4.48E-06	-33.81294571	-21.99287953	-33.81294571	-21.99287953

328 stat Spss - 42 -

Q1)

For a sample of 10 fruits from thirteen-year-old acidless orange trees, the fruit shape (determined as adiameter divided by height) was measured [Shaheen and Hamouda (1984b)]: 1.066 1.084 1.076 1.051 1.059 1.020 1.035 1.052 1.046 0.976 Assuming that fruit shapes are approximately normally distributed, find and interpret a 90% confidence interval for the average fruit shape.

Q2)

The phosphorus content was measured for independent samples of skim and whole

Whole: 94.95 95.15 94.85 94.55 94.55 93.40 95.05 94.35 94.70 94.90 Skim: 91.25 91.80 91.50 91.65 91.15 90.25 91.90 91.25 91.65 91.00 Assuming normal populations with equal variances

- a) Test whether the average phosphorus content of skim milk is less than the average phosphorus content of whole milk. Use α =0.01
- b) Find and interpret a 99% confidence interval for the difference in average phosphorus contents of whole and skim milk

Q3)

What is the relationship between the gender of the students and the assignment of a Pass or No Pass test grade? (Pass = score 70 or above).

	Pass	No Pass	Row Totals
Males	12	3	15
Females	13	2	15
Column Totals	25	5	30

Q4)

A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Observation	Program 1	Program 2	Program 3	Program 4
1	9	10	12	9
2	12	6	14	8
3	14	9	11	11
4	11	9	13	7
5	13	10	11	8

- 43 -

Q5)

Ten Corvettes between 1 and 6 years old were randomly selected from last year's sales records in Virginia Beach, Virginia. The following data were obtained, where x denotes age, in years, and y denotes sales price, in hundreds of dollars.

x	6	6	6	4	2	5	4	5	1	2
у	125	115	130	160	219	150	190	163	260	260

a) Compute and interpret the linear correlation coefficient, r.

b) Determine the regression equation for the data.

c) Compute and interpret the coefficient of determination, r^2 .

d) Obtain a point estimate for the mean sales price of all 4-year-old Corvettes.

Q1) to use the T- test, we need to make sure that the population follows a normal distribution i.e.

 H_0 : *the population follows* a normal distribution

Vs

*H*₁: *the population does not follow* a normal distribution

However, we find the question he said that the population follows a normal distribution, so is not necessary to make this test.

Now, 90% Confidence interval of the mean can be found in two ways:

1) The first method:

ta *l	Jntitled1	[DataSet	0] - IBM	SPSS Statistics	Data Editor	-	-					
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	Direct <u>M</u> arketi	ng <u>G</u> raph	IS	<u>U</u> tilities	Add- <u>o</u> ns	Window	<u>H</u> elp
8					Re <u>p</u> o D <u>e</u> sc	rts riptive Statistics	. 1	•	1			
10 : F	FruitShap	е	.976	5	Ta <u>b</u> le	s		•				
		FruitS	hape	var	Co <u>m</u>	pare Means			M Means			
	1		1.07		<u>G</u> ene	ral Linear Mode	el l	•	Cone-S	ample T Te	est	
	2		1.08		Gene	rali <u>z</u> ed Linear N	lodels	•	Indepe	endent-San	nnles T Test	
	3		1.08		Mixed	Models		•	Doired	Complea	T Toot	
	4		1.05		<u>C</u> orre	late		•	<u>na</u> <u>r</u> aireu	-Samples	T Test	
	5		1.06		<u>R</u> egre	ession		•	one-w	ay ANOVA		
	6		1.02		L <u>o</u> glir	near		•				
	7		1.04		Neura	al Net <u>w</u> orks		•				
	8		1.05		Class	sify		•				
	9		1.05		Dime	nsion Reductio	n I	•				
	10		.98					•				
	11				Nonp	arametric Tests	; 1					
	12				Eorec	asting		•				
	13				Surviv	/al						
	14				Multin	le Response		•				
	15				Missir	na Value Analys	in					
	16				Multin	ilg value Analys	15					
	1/				Mulup	ne imputation						
	18				Com	piex Samples	'					
	19				t∰ S <u>i</u> mul	ation						
	20				Quali	ty Control						
	21				ROC	Curve						
	22								-			



+ T-Test

[DataSet0]

One-Sample Statistics

	Ν	Mean	Std. Deviation	Std. Error Mean
FruitShape	10	1.0465	.03103	.00981

One-Sample Test

			T	est Value = 0		
		9 Mean		90% Confidence Interval of the Difference		
	t	df	Sig. (2-tailed)	Difference	Lower	Upper
FruitShape	106.632	9	.000	1.04650	1.0285	1.0645
						γ
					~ ~ ~ ~	

C.I for the mean

2) The second method:

e Edit		or oo otaasacs	Data Editor						
- Eau	<u>V</u> iew <u>D</u> ata	Transform	<u>Analyze</u> Direct <u>M</u> arket	ting <u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> ir		
늘 님			Reports	*					
: EruitShon			Descriptive Statistic	s 🕨	123 <u>F</u> reque	encies			
. Fruitonap	e Facil@hana		Tables	•	<u> D</u> escri	ptives			
1	FruitSnape	var	Co <u>m</u> pare Means		<mark>.4</mark> € <u>E</u> xplor	e	-		
2	1.07		General Linear Mod	lel 🕨	E ross	tabs	-		
3	1.00		Generalized Linear	Models P	TURF	Analysis			
4	1.00		Mixed Models		Ratio				
5	1.06		<u>C</u> orrelate	P		ots			
6	1.02		<u>R</u> egression			ots			
7	1.04		Neural Networks						
8	1.05		Classify	•					
9	1.05		Dimension Reduction	on 🕨					
10	.98		Scale	•					
11			Nonparametric Test	ts 🕨					
12			Forecasting	•					
13			Survival	•					
14			Multiple Response	•					
16			Missing Value Analy	sis					
17			Multiple Imputation	•				_	
18			Complex Samples	•					It helps in
19			Simulation						a loulation of
20			Quality Control	•					calculation of
21			ROC Curve						confidence inter
22									and find
23									statistical measu
								/	
_									
	Explore				×)				
	Explore		Dependent List.	Statist Plot	23 tics S		Eplore: Statis	tics	
	Explore		Dependent List	Statist Plot Option Bootst	52 tics s rap		Eplore: Statis	stics s ce Interval 1	for Mean: 90 %
	Explore		Dependent List:	Statist Plot Option Bootst	ک۲ tics s rap		Explore: Statis Descriptive: <u>Confidenc</u> <u>M</u> -estimator <u>Outliers</u> Percentiles	stics s ce Interval 1 rs	for Mean: 90 %
	Explore		Dependent List:	Statist Plot Optio Bootst	X3 tics s ns trap		Explore: Statis Descriptive: Confident M-estimator Outliers Percentiles Continu	e) Cance	for Mean: 90 %
	Explore Display Display Display	stics © Plots	Dependent List FruitShape Factor List Label <u>Cases by:</u>	Statist Plot Optio Bootst	Σ3 tics s rap	ŧ	Eplore: Statis Confidence M-estimator Qutliers Percentiles Continu	e) Canc	for Mean: 90 %
	Explore Display @ Both @ Stati	stics O Pjots	Dependent List	Statist Plot Option Bootst	52 tics s rap		Epplore: Statis Descriptive: Confidence M-estimator Quitiers Percentiles Continu	e) Canc	for Mean: 90 %

- 47 -

Explore		8	
Image: State of the state	Factor List.	tatistics Plots 2ptions ootstrap	Boxplots © Factor levels together © Dependents together © None Normality plots with tests Pread vs Level with Levene Test © None © None © Power estimation © Transformed Power: Natural log
			Continue Cancel Help

Explore

Case Processing Summary

	Cases								
	Va	lid	Miss	sing	Total				
н. Соб	Ν	Percent	N	Percent	N	Percent			
FruitShape	10	50.0%	10	50.0%	20	100.0%			

Descriptives

Statistic Std. Error FruitShape Mean .00981 1.0465 90% Confidence Interval Lower Bound 1.0285 for Mean Upper Bound 1.0645 5% Trimmed Mean 1.0483 Median 1.0515 Variance .001 Std. Deviation .03103 Minimum .98 Maximum 1.08 Range .11 Interquartile Range .04 Skewness -1.313 .687 Kurtosis 2.276 1.334

C.I for the mean

Tests of Normality									
	Kolmo	gorov-Smir	nov ^a	Shapiro-Wilk					
	Statistic	df	Sig.	Statistic	df	Sig.			
FruitShape	.194	10	/200 [*]	.907	10	.260			
*. This is a lo a. Lilliefors (ower bound of Significance C	f the true sig correction	inificance.						
		•							
					А	s P – va			

As P - value > .1

So, we except H_0 : the population follows a normal distribution

Q2) to use the T- test for two sample, we need to make sure that

1) The independence of the two samples: It is very clear that there is no correlation between the values of the two samples.

2) The populations follow a normal distribution

i.e.

*H*₀: *the two populations follow* a normal distribution

Vs

*H*₁: *the two populations do not follow* a normal distribution

However, we find the question he said that the populations follows a normal distribution, so is not necessary to make this test.

*To make sure no more.....

)ata Editor		
<u>A</u> nalyze Direc	ct <u>M</u> arketing <u>G</u> ra	aphs <u>L</u>
× 🔛 🕌		- 11
Verichle	grouping	
Variable	grouping	var
. 94.99	Whole	
. 35.15	Whole	
. 94.05	Whole	
. 94.55	Whole	
. 94.55	Whole	
. 95.40	Whole	
. 95.05	Whole	
. 94.35	vvnole	
. 94.70	vvnole	
. 94.90	vvnole	
. 91.25	Skim	
. 91.80	Skim	
. 91.50	Skim	
. 91.65	Skim	
. 91.15	Skim	
. 90.25	Skim	
. 91.90	Skim	
. 91.25	Skim	
. 91.65	Skim	
. 91.00	Skim	

- 50 -



- 51 -

ta Explore			22	
Display- Both C	ape 01) St <u>a</u> tistics © P <u>I</u> OK Pas	Dependent List:	Statistics Plots Options Bootstrap	Explore: Plots

-

Explore

grouping

Case Processing Summary										
		Cases								
		Va	lid	Miss	sing	Total				
	grouping	N	Percent	N	Percent	N	Percent			
Variable	Skim	10	100.0%	0	0.0%	10	100.0%			
	Whole	10	100.0%	0	0.0%	10	100.0%			

		Descriptiv					
	groupin	g		Statistic	Std. Error		
Variable	Skim	Mean		91.3400	.15272		
		99% Confidence Interval	Lower Bound	90.8437		7	C.I for the mean
		for Mean	Upper Bound	91.8363		ح ا	for the skim
		5% Trimmed Mean		91.3694		_	for the skill
		Median		91.3750			
		Variance		.233			
		Std. Deviation		.48293			
		Minimum		90.25			
		Maximum		91.90			
		Range		1.65			
		Interquartile Range		.57			
		Skewness		-1.241	.687		
		Kurtosis		2.035	1.334		
	Whole	Mean		94.6450	.15907		
		99% Confidence Interval	Lower Bound	94.1281			
		for Mean	Upper Bound	95.1619			
		5% Trimmed Mean		94.6861			
		Median		94.7750			
		Variance		.253			
		OLD DE LE					

-						
	Range		1.65			
	Interquartile Range		.57		1	
	Skewness		-1.241	.687	1	
	Kurtosis		2.035	1.334	1	
Whole	Mean		94.6450	.15907	1,	
	99% Confidence Interval	Lower Bound	94.1281		ר 1	CIfe de
Double-click to	for Mean	Upper Bound	95.1619			C.I for the
activate	5% Trimmed Mean		94.6861			for the
	Median		94.7750		1	
	Variance		.253		1 '	
	Std. Deviation		.50302		1	
	Minimum		93.40		1	
	Maximum		95.15		1	
	Range		1.75		1	
	Interquartile Range		.47		1	
	Skewness		-1.864	.687	1	
	Kurtosis		4.241	1.334		



mean whole



So, we except H_0 : the two populations follow a normal distribution

Now, the goal of the question:

a)
$$H_0: \mu_{whole} - \mu_{skim} = 0$$
 Vs $H_1: \mu_{whole} - \mu_{skim} > 0$ at $\alpha = .01$

and

b) 90% Confidence interval of $\mu_{whole} - \mu_{skim}$





- 54 -





- 55 -

This for test

$$H_0: \sigma^2_{whole} = \sigma^2_{skim} ~Vs ~H_1: \sigma^2_{whole} \neq \sigma^2_{skim}$$

As P – value > .01 .So, we except H_0 . However, it is given in question.



<u>Q3)</u>

 H_0 : the gender of the students is indep. of a Pass or No Pass test grade

Vs

 H_1 : the gender of the students is not indep. of a Pass or No Pass test grade

Count	Deee OrNet	Condor	NO.
Count	PassOrivot	Gender	var
1.00	1.00	1.00	
2.00	1.00	1.00	
3.00	1.00	1.00	
4.00	1.00	1.00	
5.00	1.00	1.00	
6.00	1.00	1.00	
7.00	1.00	1.00	
8.00	1.00	1.00	
9.00	1.00	1.00	
10.00	1.00	1.00	
11.00	1.00	1.00	
12.00	1.00	1.00	
13.00	2.00	1.00	
14.00	2.00	1.00	
15.00	2.00	1.00	
16.00	1.00	2.00	
17.00	1.00	2.00	
18.00	1.00	2.00	
19.00	1.00	2.00	
20.00	1.00	2.00	
21.00	1.00	2.00	
22.00	1.00	2.00	
23.00	1.00	2.00	
24.00	1.00	2.00	
25.00	1.00	2.00	
26.00	1.00	2.00	
27.00	1.00	2.00	
28.00	1.00	2.00	
29.00	2.00	2.00	
30.00	2.00	2.00	

SPSS Statistics	Data Editor	_				
<u>T</u> ransform	<u>A</u> nalyze	Direct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	W
	Repo	orts	*	*		
	D <u>e</u> so	riptive Statistics	•	123 Frequ	encies	
	Ta <u>b</u> le	es	•	Desc	riptives	
VAR00001	Co <u>m</u>	pare Means	*	- Explo	re	P
	Gene	eral Linear Model	•		staba	
	Gene	erali <u>z</u> ed Linear Mode	ls 🕨	TUDE	Analusia	
	Mixed	d Models	*		Analysis	
	<u>C</u> orre	elate	*	<u>IZ</u> <u>R</u> atio.		
	<u>R</u> egr	ession	*	<u>Р</u> -Р Р	lots	
	L <u>o</u> gli	near	*	🛃 <u>Q</u> -Q F	lots	
	Neur	al Net <u>w</u> orks	*	•	7.00	–
	Clas	sify	*	•	8.00	
	Dime	ension Reduction	*		9.00	
	Scale	9	*	•	10.00	
	Nong	arametric Tests	*	-	12.00	
	Fore	casting	*	•	12.00	
	<u>S</u> urvi	val	*		14.00	
	Multi	ole Response	*		14.00	
	💕 Missi	ng Value Analysis			15.00	
	Multi	ole Imputation	•		17.00	
	Com	plex Samples	•		18,00	
	📰 Simu	lation			19.00	
	Qual	ity Control	•		20.00	
		Cuive	,		21.00	
					22.00	
					23.00	



- 58 -



Crosstabs

Case Processing Summary								
	Cases							
	Va	Valid Missing Total						
	Ν	N Percent N Percent N Percent						
Gender * PassOrNot	30	30 100.0% 0 0.0% 30 100.0%						

Gender * PassOrNot Crosstabulation

			Pass(DrNot		1
			1.00	2.00	Total	
Gender	1.00	Count	12	3	15	1
		Expected Count	12.5	2.5	15.0	
	2.00	Count	13	2	15	1
		Expected Count	12.5	2.5	15.0	
Total		Count	25	5	30	1
		Expected Count	25.0	5.0	30.0	



$$df = (2 - 1) * (2 - 1)$$

contain less than 5

observations. So the solution is will be Merge cells until we get the expectation greater than 5 but here it is not possible, so take a larger sample.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)	
Pearson Chi-Square	.240 ^a	1	.6⁄24			
Continuity Correction ^b	.000	1	1.000			As we can see that
Likelihood Ratio	.241	1	.623			
Fisher's Exact Test				1.000	.500	2 cells have
Linear-by-Linear	222	1	620			expected count less
Association	.232	· ·	.030			than 5 because
N of Valid Cases	30		1			these 2 cells

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 2.50.

b. Computed only for a 2x2 table

$$P - value > (\alpha = .05)$$
 so we except H_0

- 60 -

 $\mathbf{Q4}$ to use the one way ANOVA- test, we need to make sure that

1) The independence of the four samples: It is very clear that there is no correlation between the values of the four samples.

2) The populations follow a normal distribution i.e.

 H_0 : the four populations follow a normal distribution

Vs

 H_1 : the four populations do not follow a normal distribution

)4	NumberOfTask	TypesOfProgram	var
-	9.00	1.00	
-	12.00	1.00	
-	14.00	1.00	
-	11.00	1.00	
-	13.00	1.00	
-	10.00	2.00	
-	6.00	2.00	
-	9.00	2.00	
-	9.00	2.00	
-	10.00	2.00	
-	12.00	3.00	
-	14.00	3.00	
-	11.00	3.00	
-	13.00	3.00	
-	11.00	3.00	
-	9.00	4.00	
-	8.00	4.00	
-	11.00	4.00	
-	7.00	4.00	
-	8.00	4.00	
-	-	-	
-	-		
-	-		
-	-		
-	-	-	
-	-		
-	-	-	
-	-	-	
-	-	-	
	IBM SPSS	Statistics Processor is ready	/
			EN

sti	ics Data Ed	itor	_	_	-	-
	<u>A</u> nalyze	Direct <u>Marketing</u>	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indo
	Repo	orts	•	*		
	D <u>e</u> so	riptive Statistics	•	123 Frequ	iencies	
	Ta <u>b</u> le	es	•	Ha Desc	riptives	
	Co <u>m</u>	pare Means	•		re	Pass
_	Gene	eral Linear Model	•	Cros	stabs	
_	Gene	erali <u>z</u> ed Linear Mode	ls ▶	TUR	Analysis	
_	Mi <u>x</u> eo	d Models	•		Analysis	
_	<u>C</u> orre	elate	•	Ratio		
_	<u>R</u> egr	ession	•	P-P F	lots	
-	L <u>o</u> gli	near	•	Q-Q F	Plots	
-	Neur	al Net <u>w</u> orks	•	-	7.00	
-	Class	si <u>f</u> y	•	-	9.00	
-	<u>D</u> ime	ension Reduction	•	-	10.00	
-	Sc <u>a</u> le	e	•	-	11.00	
-	Nonp	arametric Tests	•		12.00	
-	Fore	casting	•		13.00	
-	Survi	val	•		14.00	
_	M <u>u</u> ltip	ole Response	•	-	15.00	
	ジ Missi	ng Value Anal <u>y</u> sis		-	16.00	
	Multip	ple Imputation	•	-	17.00	
	Com	plex Samples	►	-	18.00	
	📳 S <u>i</u> mu	lation		-	19.00	
	Quali	ity Control	•	-	20.00	
	ROC	Curve		-	21.00	
_	-	-		- ·	22.00	
	-	-		-	23.00	
_	-	-			22.00 23.00	



- 62 -

Explore

[DataSet1] E:\328\7 الدرس/Untitled1.sav

TypesOfProgram

		Case Pro	cessing Sur	nmary			
				Cas	ses		
		Va	lid	Miss	sing	To	tal
	TypesOfProgram	N	Percent	N	Percent	N	Percent
NumberOfTask	1.00	5	100.0%	0	0.0%	5	100.0%
	2.00	5	100.0%	0	0.0%	5	100.0%
	3.00	5	100.0%	0	0.0%	5	100.0%
	4.00	5	100.0%	0	0.0%	5	100.0%

Descriptives

	Types(DfProgram		Statistic	Std. Error
NumberOfTask	1.00	Mean		11.8000	.86023
		95% Confidence Interval	Lower Bound	9.4116	
		for Mean	Upper Bound	14.1884	
		5% Trimmed Mean		11.8333	
		Median		12.0000	
		Variance		3.700	
		Std. Deviation		1.92354	
		Minimum		9.00	
		Maximum		14.00	
		Range		5.00	
		Interquartile Range		3.50	
		Skewness		590	.913
		Kurtosis		022	2.000
	2.00	Mean		8.8000	.73485

_						
		2.00	Mean		8.8000	.73485
			95% Confidence Interval	Lower Bound	6.7597	
			for Mean	Upper Bound	10.8403	
			5% Trimmed Mean		8.8889	
			Median		9.0000	
			Variance		2.700	
			Std. Deviation		1.64317	
			Minimum		6.00	
			Maximum		10.00	
			Range		4.00	
			Interquartile Range		2.50	
			Skewness		-1.736	.913
			Kurtosis		3.251	2.000
		3.00	Mean		12.2000	.58310
			95% Confidence Interval	Lower Bound	10.5811	
			for Mean	Upper Bound	13.8189	
			5% Trimmed Mean		12.1667	
			Median		12.0000	
			Variance		1.700	
			Std. Deviation		1.30384	
			Minimum		11.00	
			Maximum		14.00	
			Range		3.00	
			Interquartile Range		2.50	
			Skewness		.541	.913
			Kurtosis		-1.488	2.000
		4.00	Mean		8.6000	.67823
			95% Confidence Interval	Lower Bound	6.7169	
			for Mean	Upper Bound	10.4831	
			5% Trimmed Mean		8.5556	
			Median		8.0000	
			Variance		2.300	
			Std. Deviation		1.51658	
			Minimum		7.00	
			Maximum		11.00	
			Range		4.00	
			Interquartile Range		2.50	

weuran	8.0000	
Variance	2.300	
Std. Deviation	1.51658	
Minimum	7.00	
Maximum	11.00	
Range	4.00	
Interquartile Range	2.50	
Skewness	1.118	.913
Kurtosis	1.456	2.000

Tests of Normality

		Kolmogorov-Smirnov ^a		Shapiro-Wilk			
	TypesOfProgram	Statistic	df	Sig.	Statistic	df	Sig.
NumberOfTask	1.00	.141	5	.200	.979	5	.928
	2.00	.348	5	.047	.779	5	.054
	3.00	.221	5	.200*	.902	5	.421
	4.00	.254	5	.200*	.914	5	.492

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

As P - value > .05 for the four populations.

4

So, we except H_0 : the four populations follow a normal distribution

3) Homogeneity of Variance (to get a test of the assumption of homogeneity of variance) i.e.

$$H_0: \sigma_{program 1}^2 = \sigma_{program 2}^2 = \sigma_{program 3}^2 = \sigma_{program 4}^2$$

i.e. the variances of each sample are equal

Vs

*H*₁: *The variances are not all equal*

This will be clear later.

Now, the **goal** of the question:

 $H_0: \mu_{program 1} = \mu_{program 2} = \mu_{program 3} = \mu_{program 4}$

i.e. treatments are equally effective

Vs

*H*₁: *The means are not all equal*

at $\alpha = .05$

S Statist	S Statistics Data Editor							
sform	<u>A</u> nalyze	Direct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	Window	<u>H</u> elp	
~ ~	Repo	orts	•	*				
	D <u>e</u> sc	riptive Statistics	•				1	
	Ta <u>b</u> le	s	•					
0001	Co <u>m</u>	pare Means	- F	Mean	IS		Geno	
	<u>G</u> ene	ral Linear Model	*	Cone-	Sample T Te	st		
	Gene	rali <u>z</u> ed Linear Mode	ls 🕨	lnder	🔠 Independent-Samples T Test			
	Mi <u>x</u> ed Models <u>C</u> orrelate		•	- Paire	d-Samples T	Test		
			•	Paired-Samples Trest				
	<u>R</u> egr	ession	•	<u>One-</u>	way ANOVA			
	L <u>o</u> gli	near	•		6.00	1.0	0	
	Neur	al Net <u>w</u> orks	•	-	7.00	1.0	10	
	Class	sify	•		0.00	1.0	0	
	<u>D</u> ime	nsion Reduction	•		9.00	1.0	10	
	Sc <u>a</u> le	•	•		11.00	1.0	10	
	<u>N</u> onp	arametric Tests	•		12.00	1.0	0	
	Fored	asting	•		13.00	2.0	00	
	<u>S</u> urvi	val	*		14.00	2.0	00	
	M <u>u</u> ltip	ole Response	•		15.00	2.0	00	
	ジ Missi	ng Value Analysis			16.00	1.0	00	
	Multip	ole Imputation	•		17.00	1.0	00	
	Com	p <u>l</u> ex Samples	•		18.00	1.0	00	
	🖶 Simul	ation			19.00	1.0)0	
	Quali	ty Control	•		20.00	1.0)0	
	ROC	Curve			21.00	1.0)0	
	-	-		<u> </u>	22.00	1.0)0	
					22.00	1.0	10	



Helps in the homogeneity of variance test

- 66 -



as P - value < .05, then we reject $H_0: \mu_{program 1} = \mu_{program 2} = \mu_{program 3} = \mu_{program 4}$.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: NumberOfTask LSD

		Mean 95% Confiden		ence Interval		
(I) TypesOfProgram	(J) TypesOfProgram	J)	Std. Error	Sig.	Lower Bound	Upper Bound
1.00	2.00	3.00000	1.01980	.010	.8381	5.1619
	3.00	40000	1.01980	.700	-2.5619	1.7619
	4.00	3.20000	1.01980	.006	1.0381	5.3619
2.00	1.00	-3.00000	1.01980	.010	-5.1619	8381
	3.00	-3.40000	1.01980	.004	-5.5619	-1.2381
	4.00	.20000	1.01980	.847	-1.9619	2.3619
3.00	1.00	.40000	1.01980	.700	-1.7619	2.5619
	2.00	3.40000	1.01980	.004	1.2381	5.5619
	4.00	3.60000	1.01980	.003	1.4381	5.7619
4.00	1.00	-3.20000*	1.01980	.006	-5.3619	-1.0381
	2.00	20000	1.01980	.847	-2.3619	1.9619
	3.00	-3.60000	1.01980	.003	-5.7619	-1.4381

*. The mean difference is significant at the 0.05 level.

1) $H_0: \mu_{program 1} = \mu_{program 2} \ vs \ H_1: \mu_{program 1} \neq \mu_{program 2} \ at \ \alpha = .05$

as P - value = .01 < .05, then we reject H_0 .

2) $H_0: \mu_{program 1} = \mu_{program 3}$ vs $H_1: \mu_{program 1} \neq \mu_{program 3}$ at $\alpha = .05$

as P - value = .7 > .05, then we except H_0 .

3) $H_0: \mu_{program 1} = \mu_{program 4}$ vs $H_1: \mu_{program 1} \neq \mu_{program 4}$ at $\alpha = .05$

as P - value = .006 < .05, then we reject H_0 .

4) $H_0: \mu_{program 2} = \mu_{program 3} vs H_1: \mu_{program 2} \neq \mu_{program 3} at \alpha = .05$

as P - value = .004 < .05, then we reject H_0 .

5) $H_0: \mu_{program 2} = \mu_{program 4}$ vs $H_1: \mu_{program 2} \neq \mu_{program 4}$ at $\alpha = .05$

as P - value = .847 > .05, then we except H_0 .

6) $H_0: \mu_{program 3} = \mu_{program 4} vs H_1: \mu_{program 3} \neq \mu_{program 4} at \alpha = .05$

as P - value = .003 < .05, then we reject H_0 .

<u>Q5)</u>

Enter the age values into one variable and the corresponding sales price values into another variable (see figure, below).

x	Y	var
6.00	125.00	
6.00	115.00	
6.00	130.00	
4.00	160.00	
2.00	219.00	
5.00	150.00	
4.00	190.00	
5.00	163.00	
1.00	260.00	
2.00	260.00	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
-	-	
	-	-

a) Select Analyze & Correlate & Bivariate... (see figure, below).

Statist	tics Data Ed	litor	_		_
orm	Analyze	Direct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns
	Rep	orts	*	*	
	D <u>e</u> so	criptive Statistics	►		
	Table	es	•		
Pase	Co <u>m</u>	pare Means	•	AR00004	NumberO
	Gene	eral Linear Model	►		
	Gene	erali <u>z</u> ed Linear Mode	ls 🕨	· ·	
	Mixe	d Models	•		
	<u>C</u> orr	elate	•	<u> B</u> ivaria	ate
	<u>R</u> egr	ession	*	🔚 Partia	
	L <u>o</u> gli	near	*	8 Distar	ices
	Neur	al Net <u>w</u> orks	*		
	Clas	sify	•		
	Dime	ension Reduction	•		
	Sc <u>a</u> l	e	*	-	
	<u>N</u> ong	parametric Tests	*		
	Fore	cas <u>t</u> ing	•		
	<u>S</u> urvi	val	•		
	M <u>u</u> lti	ple Response	*		
	ジ Missi	ng Value Anal <u>y</u> sis		-	
	Mulți	ple Imputation	•		
	Com	plex Samples	•	-	
	🖶 Simu	lation			
	Qual	ity Control	•	-	
	ROC	Curve			
	1.00	2.00	-	· .	
	1.00	2.00	-	-	

Select "x" and "y" as the variables, select "Pearson" as the correlation coefficient, and click " "OK" (see the left figure, below).

- 70 -

Correlations

	Correlation	ns	
		Х	Y
Х	Pearson Correlation	1	968**
	Sig. (2-tailed)		.000
	Ν	10	10
Y	Pearson Correlation	968	1
	Sig. (2-tailed)	.000	
	Ν	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

The correlation coefficient is -0.9679 which we can see that the relationship between x and y are -ve and strong.

b, c and d)

Since we eventually want to predict the price of 4-year-old Corvettes, enter the number "4" in the "x" variable column of the data window after the last row. Enter a "." for the corresponding "y" variable value (this lets SPSS know that we want a prediction for this value and not to include the value in any other computations) (see figure, below).

_			
	х	Y	
-	6.00	125.00	
-	6.00	115.00	
_	6.00	130.00	
_	4.00	160.00	
-	2.00	219.00	
_	5.00	150.00	
_	4.00	190.00	
-	5.00	163.00	
-	1.00	260.00	
-	2.00	260.00	
_	4.00	-	
-	-	-	
-	-	-	
_	_	-	
Select Analyze \Diamond Regression \Diamond Linear... (see figure).

Select "y" as the dependent variable and "x" as the independent variable. Click "Statistics", select "Estimates" and "Confidence Intervals" for the regression coefficients, select "Model fit" to obtain r², and click "Continue". Click "Save...", select "Unstandardized" predicted values and click "Continue". Click "OK".

tisti	ics Data Eo	ditor						
n	<u>A</u> nalyze	Direct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indow	<u>H</u> elp	
_	Rep	orts	•	*				4
	D <u>e</u> s	criptive Statistics	•					1
	Ta <u>b</u> l	es	•					
ass	Com	pare Means	•	AR00004	Number	OfTask	Туре	esOfPre
	<u>G</u> en	eral Linear Model	•	-		9.00		
	Gen	erali <u>z</u> ed Linear Mode	ls 🕨	-		12.00		
	Mi <u>x</u> e	d Models	•	-		14.00		
	<u>C</u> orr	elate	•	-		11.00		
	<u>R</u> eg	ression	•	Autom	atic Linear	Modeling		
	L <u>o</u> gl	inear	•	Linear	r			
	Neu	ral Net <u>w</u> orks	►	Curve	Estimation.			
	Clas	sify	•		LL east Sou	ares		
	<u>D</u> im	ension Reduction	•	Binan	Logistic			
	Sc <u>a</u> l	e	•	Dinary	Eogistic	1 -		
	Non	parametric Tests	•	Multin		uc		
	Fore	casting	►	Or <u>d</u> ina	al			
	Surv	ival	•	Probit				
	M <u>u</u> lti	ple Response	►	Nonlin Nonlin	near			
	ジ Miss	ing Value Anal <u>y</u> sis		Weigh	t Estimation	n		
	Mul <u>t</u> i	ple Imputation	•	2-Stag	ge Least Sq	uares		
	Corr	np <u>l</u> ex Samples	•	Optim	al Scaling (CATREG)		
	🖶 S <u>i</u> mu	lation		-		7.00		
	Qua	lity Control	►	-		8.00		
	ROC	Cur <u>v</u> e				-		
	1.00	2.00	-	-		-		
	1.00	2.00	-	-		-		

 Chear Regression FruitShape VAR00001 Variable grouping VAR00003 Count PassOrNot Gender VAR00002 VAR00004 NumberOfTask TypesOfProgram VAR00005 VAR00006 X 	Dependent:	Statistics Plots Save Options Style Bootstrap	Linear Regression: Statistics
---	------------	--	-------------------------------

12.0 14.0	0 1.00 0 1.00 Dependent:	Statistics Plots Save Options Style Bootstrap	Linear Regression: Save Predicted Values Unstandardized Standardized Adjusted S.E. of mean predictions Distances Mahalanobis Cook's Leverage values Prediction Intervals Mean Individual Confidence Interval: 95 % Coefficient statistics Create coefficient statistics Create a new dataset Dataset name:	Residuals Unstandardized Standardized Studentized Deleted Studentized deleted Influence Statistics DfBeta(s) Standardized DfBeta(s) DfFit Standardized DfFit Covariance ratio
ОК Р	aste <u>R</u> eset Cancel Help		Dataset name:	
	Image: state	· · ·	Export model information to XML file	Bro <u>w</u> se el Help

- 73 -

Regression

Model Summary^b

M	odel	R	R Square	Adjusted R Square	Std. Error of the Estimate
1		.968 ^a	.937	.929	14.24653

a. Predictors: (Constant), X

b. Dependent Variable: Y

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24057.891	1	24057.891	118.533	.000 ^b
	Residual	1623.709	8	202.964		
	Total	25681.600	9			

a. Dependent Variable: Y

b. Predictors: (Constant), X

Coefficients^a

		Unstandardize	d Coefficients	Standardized Coefficients			95.0% Confiden	ice Interval for B
Model		В	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	291.602	11.433		25.506	.000	265.238	317.966
	Х	-27.903	2.563	968	-10.887	.000	-33.813	-21.993

a. Dependent Variable: Y

-			
			1
Х	Y	PRE_1	
6.00	125.00	124.18447	
6.00	115.00	124.18447	
6.00	130.00	124.18447	
4.00	160.00	179.99029	
2.00	219.00	235.79612	
5.00	150.00	152.08738	
4.00	190.00	179.99029	
5.00	163.00	152.08738	
1.00	260.00	263.69903	
2.00	260.00	235.79612	
4.00		179.99029	
		-	

From above, the regression equation is: y = 29160.1942 - (2790.2913)(x).

The coefficient of determination is 0.9368; therefore, about 93.68% of the variation in y data is explained by x.



Q1)

i-
$$\binom{150}{30}$$
, $\Gamma(18)$, $\ln(14)$, $\log(17)$

ii- $P(2 < X \le 4)$ when $X \sim Poisson(3)$

iii-Write R loop and the results to calculate

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, z = -3.1, -3.0, \dots -0.1, 0, 0.1, \dots 3.0, 3.1.$$

iv- Write R code and the results to calculate:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} dt, \ z = -3, -2, -1, 0, 1, 2, 3.$$

v- Write loop structure in R for generating 5 samples each of size 100 from *Binomial* (5,0.7). Then calculate the mean, standard deviation and coefficient of variation for each sample.

Q2)

(a) Write the commends and results to calculate the following:

(i) P(-1.0 < T < 1.5), v = 10,

(ii) Find k such that P(T < k) = 0.025, $\nu = 12$,

(iii)
$$\binom{15}{9}$$
, $\log_{10}(25)$, 28!

un Ban vate=3

(b) Generate a random sample of size 12 from the exponential (3) distribution and save it to A. Next, write an \hat{k} command to create the column B such that

$$B_i = \begin{cases} 1, & A_i \le 3\\ 2, & A_i > 3 \end{cases}, \quad i = 1, 2, ..., 12.$$

Q3) (a) Find k when P(X > k) = 0.04, X - F(12, 10) $\Rightarrow 1 - P(X < k) = 0.04$ $\Rightarrow 1 - 0.04 = P(X < k)$ (b) $P(3 < X \le 7)$ when X - Poisson(3)Beta(6.5)

$$(c) \int_{0}^{1} x^{5} [1-x]^{4} dx \qquad f(x) = \underbrace{\left(\frac{6}{5}\right)_{z}^{z}}_{z} \underbrace{\left(1-x\right)_{z}^{0}}_{z} = \underbrace{\left(\frac{6}{5}\right)_{z}^{z}}_{z} \underbrace{\left(\frac{6}{5}\right)_{z}}_{z} \underbrace{\left(\frac{6}{5}\right)_{z}^{z}}_{z} \underbrace{\left(\frac{6}{5}\right)_{z}}_{z} \underbrace{\left(\frac{6}{5}\right)_{z}} \underbrace{\left(\frac{6}{5}\right)_{z}}_{z} \underbrace{\left(\frac{6}{5}\right)_{z}} \underbrace{\left(\frac{6}$$

Q4)

Γ	1	9	8	1 1	3	4	2	7
7163-1]	7	4	2		4	9	0	6
$A = \begin{bmatrix} 5 & 2 & 7 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$	5	1	5	, C =	3	8	3	2
	1	1	9		3	4	6	2

(a) A*B

(b) Determinant of C

(c) Inverse of C

(a) The following data are two independent random samples from two independent populations $A \sim (\mu_A, \sigma^2)$ and $B \sim (\mu_B, \sigma^2)$, respectively.

A: 48 39 42 52 40 48 52 52 54 48 B: 50 48 42 40 43 48 50 46 38 38

Write R command and the results to

i- Test whether $\mu_A > \mu_B$.

ii- Construct 90% confidence interval of the difference $\mu_B - \mu_A$.

Q5)

(a) Write the command and result to	calculate the following:
log(80) = 10 -	
ln(40) =	
40!=	
$\binom{50}{15} = -150$	
P(X > 2.22), where X ~ N (2,5) ⇒ 1- $P < X < 2.22$)	

(b) Write the commands and results to find the determent of the matrix and its inverse

....

	1	0	4	7
0	8	3	1	9
1,00	7	4	2	8
	LO	9	5	6

```
###Q1
##i
   choose (150, 30)
   gamma(18)
   log(14); log(14, base=exp(1))
   log(17,base=10);log10(17)
##ii
   ppois(4,lambda=3)-ppois(2,lambda=3)
##iii
   z <- seq(-3.1,3.1,by=.1)</pre>
   z
   for(i in z) {
   a=dnorm(i, mean = 0, sd = 1)
   cat(i," ",a,"\n")
   3
 #or
   for(i in z) {
   a=dnorm(i, mean = 0, sd = 1)
   print(c(i,a))
   }
##vi
   z <- seq(-3,3,by=1)</pre>
   z
   for(i in z) {
   b=pnorm(i, mean = 0, sd = 1)
   cat(i," ",b,"\n")
   }
 #or
   for(i in z) {
   b=pnorm(i, mean = 0, sd = 1)
   print(c(i,b))
   }
##v
  generating <- seq(1,5,by=1)</pre>
  generating
  generating <- c(1,2,3,4,5)
  generating
  for(i in generating) {
  c=rbinom(100, size=5, prob=.7)
  d <- mean(c)
  e <- sd(c)
  f <- e/d
  cat("sample:",c," ","mean=",d," ","sd=",e," ","cv=",f,"\n")
  }
#or
  for(i in generating) {
  c=rbinom(n=100, size=5, prob=.7)
  d <- mean(c)
  e <- sd(c)
  f <- e/d
  print(c(c,d,e,f))
  }
```

```
###Q2
##ai
  pt(1.5,df=10)-pt(-1,df=10)
##aii
   k=qt(.025, df=12)
   k.
##aiii
   choose(15,9)
   log(25,base=10);log10(25)
   factorial(28)
##b
  A <- rexp(12, rate=3)
   Α
   for(i in A) {
    if(i<=3) print(1) else print(2)
   }
 #or
   B <- vector(mode = "numeric")</pre>
   j <-0
   for(i in A) {
     j <- j+1
     if(i<=3) B[j]=1 else B[j]=2
   }
   в
###Q3
##a
   k=qf(1-.04, df1=12, df2=10)
   k
##c
   f <- function(x) \{ (x^5) * ((1-x)^4) \}
   i <- integrate(f,lower=0,upper=1)$value</pre>
   i
 #or
   f \leftarrow function(x) \{ dbeta(x, shape1=6, shape2=5) * (beta(6,5)) \}
   i <- integrate(f,lower=0,upper=1)$value</pre>
   i
I
###Q4
  a <- c(1,6,3,-1,5,2,7,4)
  A <- matrix(a,nrow = 2, ncol =4,byrow=T)
  А
  b <- c(1,9,8,7,4,2,5,1,5,1,1,9)
  B <- matrix(b,nrow = 4, ncol =3,byrow=T)</pre>
  в
   c <- c(3,4,2,7,4,9,0,6,3,8,3,2,3,4,6,2)
   C <- matrix(c,nrow = 4, ncol =4,byrow=T)
  С
##a
  A$*$B
##b
   det (C)
##c
  solve(C)
```

```
328 stat
```

```
##d
  A <- c(48,39,42,52,40,48,52,52,54,48)
  Α
  B <- c(50,48,42,40,43,48,50,46,38,38)
  В
 #or
 قبل استدعاء الملفاضع المؤشر في صفحة النتائج ثم اختار ملف ثم اختار الخيار رقم ٩ ثم احدد المجلد الذي فيه ملف البيانات#
data <- read.csv("data.csv",header=T, sep=";")
  data
  A <- data$A
  Α
  B <- data$B
   В
##di
   t.test(A,B,alternative = "greater", paired = FALSE, var.equal = T,conf.level = 0.95)
##dii
  t.test(B,A,alternative = "two.sided", paired = FALSE, var.equal = T,conf.level = 0.90)
###Q5
  1-pnorm(2.22, mean = 2, sd = sqrt(5))
```

+ See Appendix -3-