EXCEL

Write Excel commands with the results to calculate the following:

$$\binom{15}{12} =$$

By excel:
=COMBIN(15,12)=455

$$\binom{8}{4} =$$

By excel:
=COMBIN(8,4) = 70

$$\sqrt[10]{\sqrt{10!}} =$$

By excel:
=(SQRT(FACT(10))^0.1) = 2.128081
=(SQRT(FACT(10)))^0.1 = 2.128081

$Log_{10}(2.5)$

By excel:
=LOG10(2.5)=0.39794

$log_e(2.5)$

By excel:
=LN(2.5)=0.916291

$$\ln(\sqrt{7}) =$$

By excel:
=LN(SQRT(7)) = 0.972955075

$$Find \prod_{i=1}^{20} \left(\frac{i^2}{i+1}\right)$$

By excel:
=L1^2/(L1+1)
=PRODUCT(M1:M20)= 1.15852E+17

$$Let \quad A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 3 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 5 & 4 & 1 & 1 \end{bmatrix}$$

Then calculate

(i) $A\,B$

By excel: $\boxed{Ctrl + Shift + Enter}$

=MMULT(L18:N19,L21:N23)

7	2	5	4
23	17	6	5

(ii) *The determinant of* $B'B$

By excel:

=TRANSPOSE(L21:N23)

=MMULT(B1:D2,B5:E7)

=MDETERM(L6:O9) = 0

(iii) The inverse of $B\,B' = (BB')^{-1}$

By excel:

=TRANSPOSE(L21:N23)

=MMULT(V7:X9,L21:N23)=

=MINVERSE(L17:N19)

0.534435	-0.34435	-0.20937
-0.34435	0.443526	0.093664
-0.20937	0.093664	0.112948

$$if \quad A = \begin{bmatrix} 1 & 2 & 1 & 7 & 2 \\ 5 & 0 & 2 & 2 & 2 \\ 8 & 4 & 1 & 4 & 0 \\ 7 & 9 & 5 & 4 & 8 \\ 9 & 7 & 4 & 2 & 9 \end{bmatrix}, then \quad A^{-1} =$$

By excel:

=MINVERSE(B5:F9) =

-0.03	0.05	0.09	-0.10	0.08
-0.04	- 0.22	0.09	0.11	-0.04
-0.18	0.44	-0.13	0.46	-0.47
0.15	0.02	0.01	-0.03	-0.01
0.11	- 0.07	-0.11	-0.19	0.27

$$A=\begin{bmatrix}3 & 2 & 3\\ 4 & 1 & 1\end{bmatrix},\quad B=\begin{bmatrix}3 & 2 & 3\\ 2 & 7 & 0\end{bmatrix},\quad C=\begin{bmatrix}10\\ 1\\ 5\end{bmatrix}$$

Use Excel to calculate $B.C$, $|A.B'|$ *and* $(A.B')^{-1}$.

BC

By excel:
=MMULT(T23:V24,T26:T28)=
47
27

$|AB'|$

By excel:
=MMULT(T20:V21,TRANSPOSE(T23:V24))
=MDETERM(T36:U37) = -10

$(AB')^{-1}$

By excel:
=MINVERSE(T36:U37)

-1.5	2
1.7	-2.2

PDF	$P(X = k) = \boxed{?}$	= NORM.DIST (k , … ,FALSE)
CDF	$P(X \le k) = \boxed{?}$	= NORM.DIST (k , … ,TRUE)
Inverse	$P(X \le \boxed{?}) = \alpha$	= NORM.INV (α , …)

$Find\ k\ when\ P(X > k) = 0.04 \qquad X \sim F(10,11)$

$$P(X > k) = 0.04 \implies P(X \le k) = 0.96$$

By excel:
=F.INV(0.96,10,11) = 3.062037

$P(T < c) = 0.085 \qquad T \sim t\ \ with\ 6\ degrees\ of\ freedom$

By excel:
=T.INV(0.085,6) = -1.55905

$Find\ \ P(-1.5 \le X \le 1.7)\ \ X \sim t$ at 5 degrees of freedom.

$$P(-1.5 \le X \le 1.7) = P(X \le 1.7) - P(X \le -1.5)$$

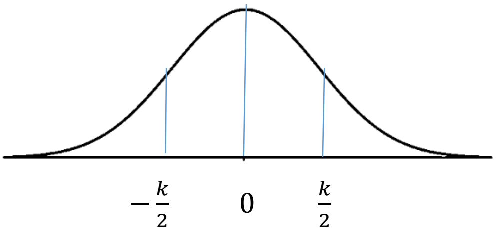
By excel:
=T.DIST(1.7,5,TRUE)-T.DIST(-1.5,5,TRUE)= 0.82811

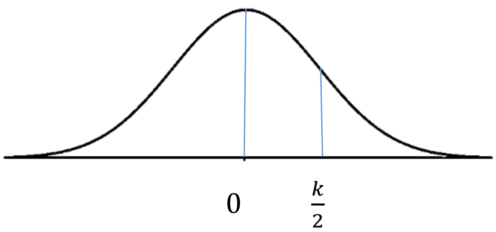
$Find\ \ k\ \ when\ \ P(X > k) = 0.42\ \ \ when\ \ X \sim \chi^2\ (12)$

$$P(X > k) = 0.42 \implies P(X \le k) = 0.58$$

By excel:
=CHISQ.INV(0.58,12)= 12.32428

$Find\ \ k\ \ when\ P\left(-\frac{k}{2} \le X \le \frac{k}{2}\right) = 0.98\ \ ,\ \ X \sim t(14)$





$$\implies 2 \times P\left(0 \le X \le \frac{k}{2}\right) = 0.98$$

$$\implies P\left(0 \le X \le \frac{k}{2}\right) = 0.49$$

$$\implies P\left(X \le \frac{k}{2}\right) - P(X \le 0) = 0.49$$

$$\implies P\left(X \le \frac{k}{2}\right) - 0.5 = 0.49$$

$$\implies P\left(X \le \frac{k}{2}\right) = 0.99$$

By excel:
=T.INV(0.99,14)=2.624494

$$\frac{k}{2} = 2.624494 \implies k = 5.248988$$

$Find\ P(2 \le X < 6)\quad and\quad P(X = 2)\quad when\quad X \sim Bino(5, 0.6)$

X	0	1	2	3	4	5
$f(x)$			*	*	*	*

$$P(2 \le X < 6) = P(X \le 5) - P(X \le 1)$$

By excel:
=BINOM.DIST(5,5,0.6,TRUE)-BINOM.DIST(1,5,0.6,TRUE)= 0.91296

$$P(X = 2)$$

By excel:
= BINOM.DIST(2,5,0.6,FALSE)= 0.2304

$Find\ P(3.5 \le X < 8)\quad when\quad X \sim Bino(10, 0.5)$

X	0	1	2	3	4	5	6	7	8	9	10
$f(x)$					*	*	*	*			

$$P(3.5 \le X < 8) = P(X \le 7) - P(X \le 3)$$

By excel:
=BINOM.DIST(7,10,0.5, TRUE)-BINOM.DIST(3,10,0.5,TRUE)= 0.773438

$P(2 \le X < 7)\qquad X \sim Poisson\ (3.5)$

X	0	1	2	3	4	5	6	7	8	9	...
$f(x)$			*	*	*	*	*				...

$$P(2 \le X < 7) = P(X \le 6) - P(X \le 1)$$

By excel:
=POISSON.DIST(6,3.5,TRUE)-POISSON.DIST(1,3.5,TRUE) = 0.798824

$P(X \ge 8.5)\qquad X \sim Poisson\ (7)$

X	0	1	2	3	4	5	6	7	8	9	10	11	...
$f(x)$										*	*	*	*

$$P(X \ge 8.5) = 1 - P(X \le 8)$$

By excel:
=1-POISSON.DIST(8,7,TRUE)= 0.270909

$$Find\ \sum_{i=7}^{10} \binom{10}{i} (0.4)^i (0.6)^{10-i} =$$

X	0	1	2	3	4	5	6	7	8	9	10
$f(x)$								*	*	*	*

$$P(7 \le X \le 10) = P(X \le 10) - P(X \le 6) \quad X \sim Bino(10,0.4)$$

By excel:
=BINOM.DIST(10,10,0.4,TRUE)-BINOM.DIST(6,10,0.4,TRUE) = 0.054762

$$P(1.5 < X < 3.5)\ when\ X\ is\ distributed\ as\ f(x) = \frac{1}{3} e^{-\frac{x}{3}}\ ,\ x > 0$$

$$P(X \le 3.5) - P(X \le 1.5) \quad X \sim Exp\left(\frac{1}{3}\right)$$

By excel:
=EXPON.DIST(3.5,0.333,TRUE)-EXPON.DIST(1.5,0.333,TRUE)=0.295

$$P(X > 4)\ when\ f(x) = \frac{\binom{8}{x}\binom{13}{5-x}}{\binom{21}{5}}\ ,\ x = 0,1,2,3,4,5$$

X	0	1	2	3	4	5
$f(x)$						*

$$P(X = 5)$$

By excel:
=HYPGEOM.DIST(5,8,5,21,FALSE) = 0.00275

$$Find \sum_{x=5}^{17} \frac{1}{2^x\, x!}$$

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad X = 0,1,2,\ldots$$

$$\Longrightarrow \sum_{i=5}^{17} \frac{1}{2^x\, x!} = \sum_{i=5}^{17} \frac{\left(\frac{1}{2}\right)^x}{x!}$$

$$\Longrightarrow \frac{1}{e^{-\left(\frac{1}{2}\right)}} \sum_{i=5}^{17} \frac{e^{-\left(\frac{1}{2}\right)}\left(\frac{1}{2}\right)^x}{x!} \quad \boxed{\lambda = \frac{1}{2}}$$

$$\Longrightarrow e^{\left(\frac{1}{2}\right)} \sum_{i=5}^{17} \frac{e^{-\left(\frac{1}{2}\right)}\left(\frac{1}{2}\right)^x}{x!}$$

$$\Longrightarrow e^{\left(\frac{1}{2}\right)} P(5 \le X \le 17) \qquad X \sim Poisson\ \left(\frac{1}{2}\right)$$

$$\Longrightarrow e^{\left(\frac{1}{2}\right)} \left[\underbrace{P(X \le 17)} - \underbrace{P(X \le 4)}\right]$$

By excel:

=EXP(0.5)*(POISSON.DIST(17,0.5,TRUE) - POISSON.DIST(4,0.5,TRUE))=0.000284

$$Find\ x\ such\ that \int_0^x x^5(1-x)^3\, dx = 0.25$$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\, \Gamma\beta} x^{\alpha-1}(1-x)^{\beta-1} \quad 0 \le X \le 1$$

$$\Longrightarrow \frac{\Gamma(6)\,\Gamma(4)}{\Gamma(10)} \int_0^x \frac{\Gamma(10)}{\Gamma(6)\,\Gamma(4)} x^{6-1}(1-x)^{4-1}\, dx = 0.25 \quad \boxed{\alpha = 6\,, \beta = 4}$$

$$\Longrightarrow \frac{\Gamma(6)\,\Gamma(4)}{\Gamma(10)} P(X \le x) = 0.25 \qquad X \sim Beta\ (6,4)$$

By excel:
= GAMMA(6)*GAMMA(4)/GAMMA(10)*BETA.INV(0.25,6,4) = 0.00099
= GAMMA(6)*GAMMA(4)/GAMMA(10)*BETA.INV(0.25,6,4,0,1)= 0.00099

$Find \int_1^2 e^{-\frac{x^2}{8}}\, dx$

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty \leq X \leq \infty}$$

$$\Longrightarrow 2\sqrt{2\pi} \int_1^2 \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}}\, dx \quad \boxed{\mu = 0\,, \sigma = 2}$$

$$\Longrightarrow 2\sqrt{2\pi}\, P(1 \leq X \leq 2) \qquad X \sim N\,(0,2)$$

$$\Longrightarrow 2\sqrt{2\pi}\, [P(X \leq 2) - P(X \leq 1)]$$

By excel:
= 2*SQRT(2*PI())*(NORM.DIST(2,0,2,TRUE)-NORM.DIST(1,0,2,TRUE))
= 0.751398

$Find \int_{-\infty}^{3} \frac{1}{\sqrt{2\pi}}\, e^{\frac{-z^2}{2}}\, dz$

$$P(Z \leq 3) \qquad Z \sim N(0,1)$$

By excel:
= NORM.DIST(3,0,1,TRUE) = 0.998565

Question:

The data given in data-1.xls represents the temperature in Riyadh city from 2005-2009. Use Excel to:

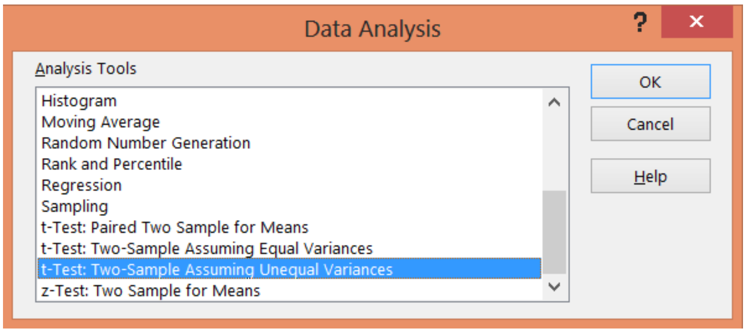
1- *Find the 20% percentile of the temperature in May, 2007.*

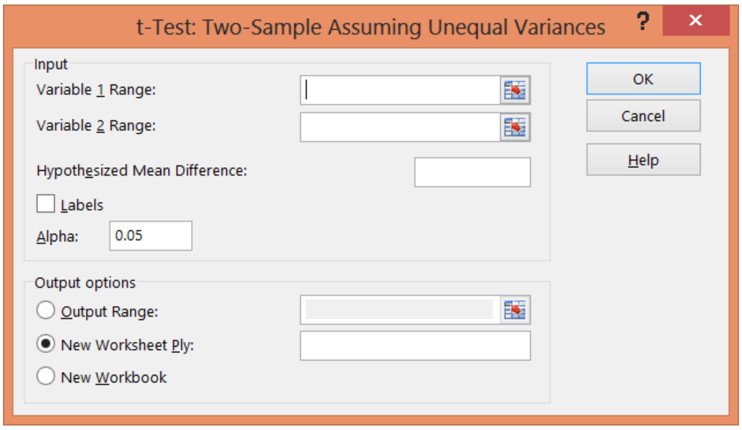
By excel: =PERCENTILE.INC(I2:I32,0.2)=30

2- *Test the hypothesis that: the mean of the temperature in 2006 is equal to the mean of the temperature in 2008. Use* $\alpha = 0.05$.

$$H_0: \mu_{2006} = \mu_{2008} \qquad \qquad H_0: (\mu_{2006} - \mu_{2008}) = 0$$
$$\Longrightarrow$$
$$H_1: \mu_{2006} \neq \mu_{2008} \qquad \qquad H_1: (\mu_{2006} - \mu_{2008}) \neq 0$$

$$\boxed{DATA} \longrightarrow \boxed{Data\ Analysis}$$





t-Test: Two-Sample Assuming Unequal Variances

	TEMPERATURE_2006	*TEMPERATURE_2008*
Mean	28.72246575	28.93278689
Variance	14.63246094	13.20593577
Observations	365	366
Hypothesized Mean Difference	0	
df	727	
t Stat	-0.762059468	
P(T<=t) one-tail	0.223135833	
t Critical one-tail	1.646952285	
P(T<=t) two-tail	**0.446271666**	
t Critical two-tail	1.963232428	

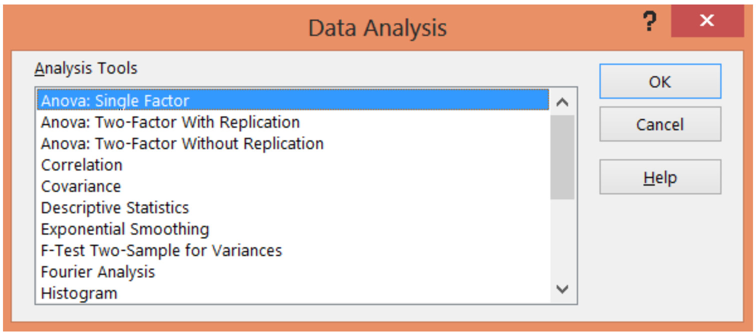
P-Value = 0.446271666 > 0.05

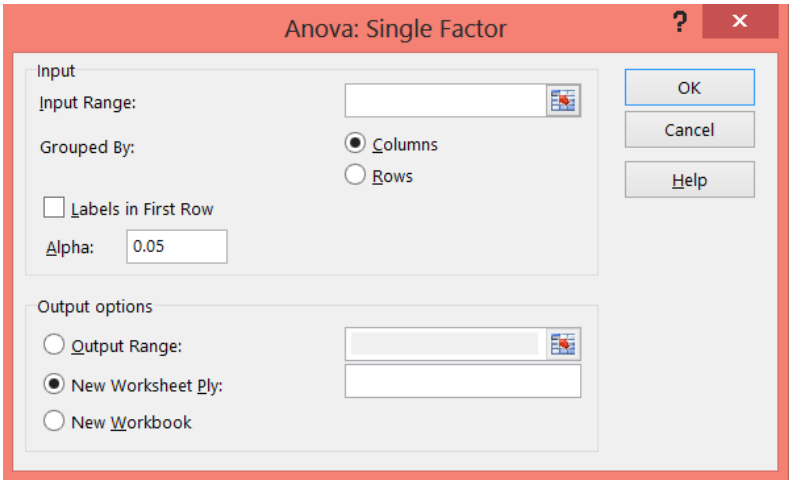
We Accpet H_0, there is no difference in Means

3- *Use ANOVA to compare between the temperatures in all given years.*

$H_0: \mu_{2005} = \mu_{2006} = \mu_{2007} = \mu_{2008} = \mu_{2009}$
$H_1: at\ least\ two\ means\ not\ equal$

$\boxed{DATA} \longrightarrow \boxed{Data\ Analysis}$





ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	16.45781	4	4.114452	0.29464	**0.881572**	2.376812
Within Groups	25443	1822	13.96432			
Total	25459.45	1826				

P-value = 0.881572> 0.05
Then we Accept H_0 *, so there no difference between the temperature means in all given years*

Question:

To study the effectiveness of wall insulation in saving energy for home heating, the energy consumption (in certain unites) for 5 houses in a certain city, was recorded for two summer; the first summer was before insulation and the second summer was after insulation:

House	*1*	*2*	*3*	*4*	*5*
Before	*12.1*	*10.6*	*13.4*	*13.8*	*15.5*
After	*12.0*	*11.0*	*14.1*	*11.2*	*15.3*

Assuming the energy consumption follows normal distribution.

(a) *Provide a 95% confidence interval for the difference between the mean energy consumption before and after the wall insulation is* installed.

$$\bar{X} \mp t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Before	After	Before-After
12.1	12	0.1
10.6	11	-0.4
13.4	14.1	-0.7
13.8	11.2	2.6
15.5	15.3	0.2

b-a	
Mean	0.36
Standard Error	0.583609458
Median	0.1
Mode	#N/A
Standard Deviation	1.304990421
Sample Variance	1.703
Kurtosis	3.541251682
Skewness	1.797562257
Range	3.3
Minimum	-0.7
Maximum	2.6
Sum	1.8
Count	5
Confidence Level(95.0%)	1.620359624

=0.36-1.62=-1.26 LB

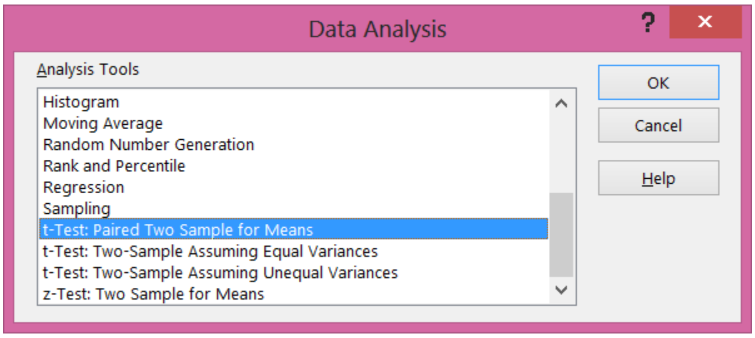
=0.36+1.62=1.98 UB

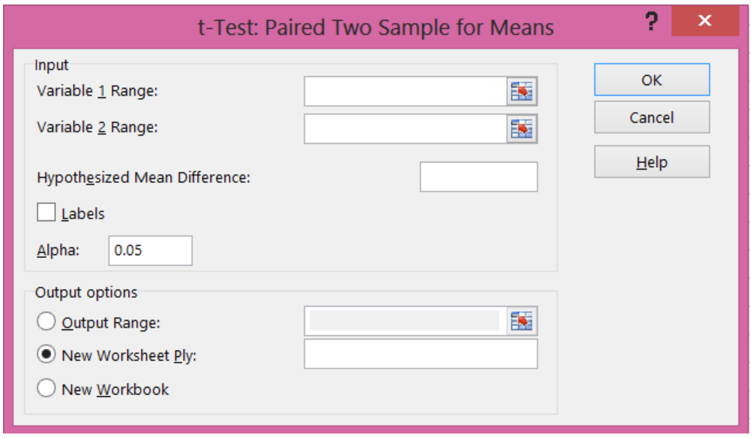
$$\boxed{-1.26 < \mu_D < 1.98}$$

(b) Is there a difference in the mean of energy consumption before and after the wall insulation is installed at the significance level 0.05?

$$\begin{matrix} H_0: \mu_{Before} = \mu_{After} \\ H_1: \mu_{Before} \neq \mu_{After} \end{matrix} \Longrightarrow \begin{matrix} H_0: \mu_{Before} - \mu_{After} = 0 \\ H_1: \mu_{Before} - \mu_{After} \neq 0 \end{matrix} \Longrightarrow \begin{matrix} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{matrix}$$

$$\boxed{DATA} \longrightarrow \boxed{Data\ Analysis}$$





t-Test: Paired Two Sample for Means

	Before	*After*
Mean	13.08	12.72
Variance	3.397	3.587
Observations	5	5
Pearson Correlation	0.756436907	
Hypothesized Mean Difference	0	
df	4	
t Stat	0.616850866	
P(T<=t) one-tail	0.285355894	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	**0.570711789**	
t Critical two-tail	2.776445105	

P-Value = 0.570711789 > 0.05, then we accept H_0 .

So, there is no difference in mean energy consumption before and after the wall insulation.

***<u>Question</u>*:**

For testing the differences between the means of two normal populations at 1% level of significance, two intendent samples are used. The excel results are given below:

t-Test: Two-Sample Assuming Unequal Variances

	x	*y*
Mean	4.83	4.29
Variance	12.57	9.57
Observations	6.00	7.00
Hypothesized Mean Difference	0.00	
Df	10.00	
t Stat	0.29	
P(T<=t) one-tail	0.39	
t Critical one-tail	1.81	
P(T<=t) two-tail	0.77	
t Critical two-tail	2.23	

(a) What are the sample sizes?

$$\text{Sample size for } X = 6$$
$$\text{Sample size for } Y = 7$$

(b) Compare between the two samples based on the C.V:

$$C.V_X = \frac{S}{\bar{X}} = \frac{\sqrt{12.57}}{4.83} = 0.734$$

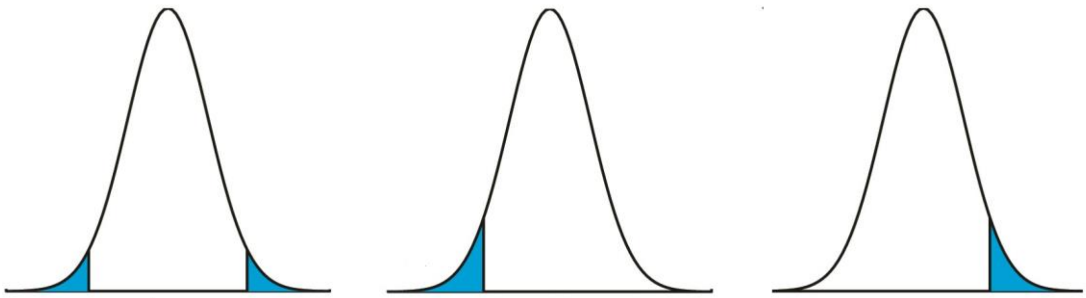
$$C.V_Y = \frac{S}{\bar{Y}} = \frac{\sqrt{9.57}}{4.29} = 0.721$$

Then X is more variation of Y.

(c) What is the test statistic?

Test statistics = 0.29

(d) What is the one-sided and two-sided critical regions?



$(-\infty, -2.23) \cup (2.23, \infty)$ $(-\infty, -1.81)$ $(1.81, \infty)$

(e) Based on the p-value, write the test steps for testing

i- $H_0: \mu_X - \mu_Y = 0 \quad vs \quad H_1: \mu_X - \mu_Y \neq 0$

P-value = 0.77 > 0.05, accept H_0, (no difference between two means).

ii- $H_0: \mu_X - \mu_Y = 0 \quad vs \quad H_1: \mu_X - \mu_Y < 0$

P-value = 0.39 > 0.05, accept H_0, (no difference between two means).

Question :

The operations manager of a company that manufactures tires wants to determine whether there is any relationship between the quality of tires and the three daily shifts. He randomly selects 1000 tires and carefully inspects them. Each tire is classified either as perfect, satisfactory, or defective, and the shift that produced it is recorded. The two categorical variables of interest are: shift and condition of the tire produced. The data can be summarized by the accompanying two-way table.

		Quality of the tires		
		Perfect	*Satisfactory*	*Defective*
shift	*Shift 1*	*105*	*100*	*4*
		106	*124*	*2*
	Shift 2	*80*	*81*	*2*
		67	*85*	*3*
	Shift 3	*42*	*80*	*5*
		37	*74*	*3*

(1) Find the mean, variance and Coefficient of variation for shift.

- *Mean:*

By excel:
=AVERAGE(C2:E7) = 55.55

- *Variance:*

By excel:
=VAR.S(C2:E7) = 1872.497

- *Coefficient of variation (C.V):*

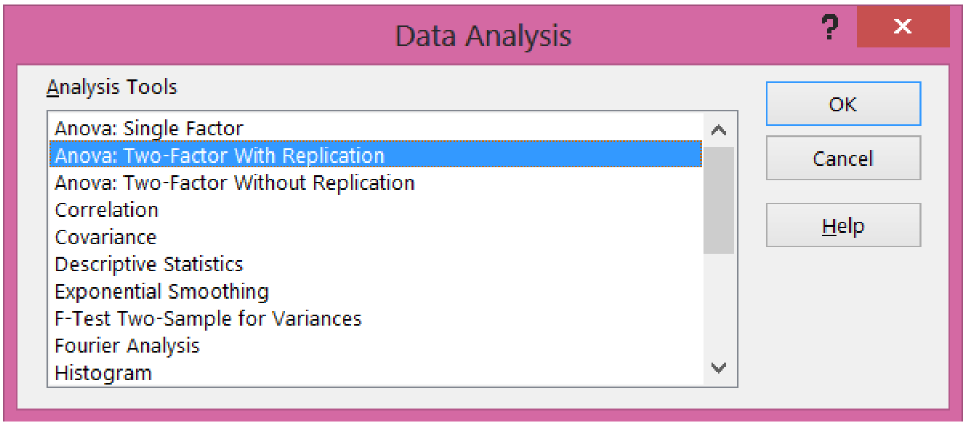
By excel:
=SQRT(B11)/B10 = 0.778902

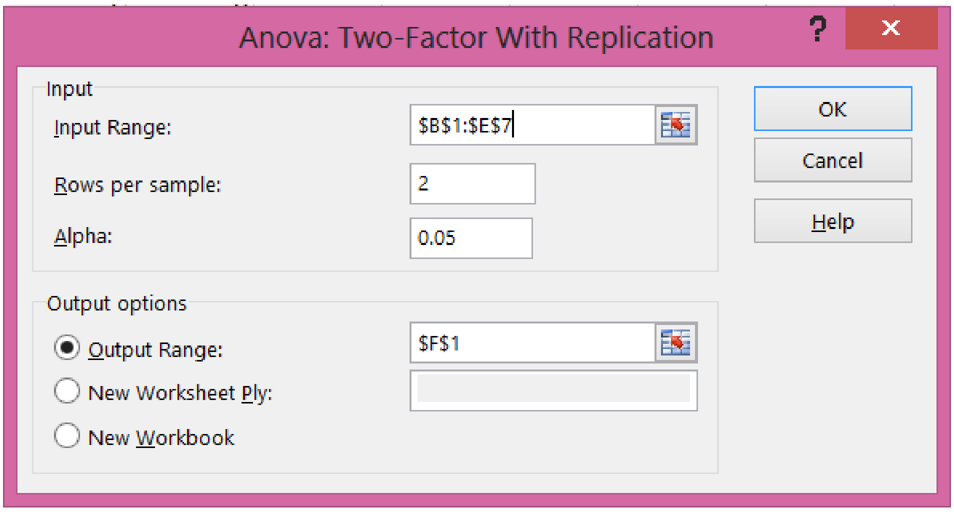
(2) Do you think there exist interaction between quality and shifts at 5% significance level? Write the name and test steps.

$$H_0: \textit{There is no interaction}$$
$$H_1: \textit{There is an interaction}$$

	Perfect	Satisfactory	Defective
Shift 1	105	100	4
	106	124	2
Shift 2	80	81	2
	67	85	3
Shift 3	42	80	5
	37	74	3

$\boxed{DATA} \longrightarrow \boxed{Data\ Analysis}$





Anova: Two-Factor With Replication				
SUMMARY	Perfect	Satisfactory	Defective	Total
Shift 1				
Count	2	2	2	6
Sum	211	224	6	441
Average	105.5	112	3	73.5
Variance	0.5	288	2	3048.7
Shift 2				
Count	2	2	2	6
Sum	147	166	5	318
Average	73.5	83	2.5	53
Variance	84.5	8	0.5	1566.8
Shift 3				
Count	2	2	2	6
Sum	79	154	8	241
Average	39.5	77	4	40.166667
Variance	12.5	18	2	1072.5667
Total				
Count	6	6	6	
Sum	437	544	19	
Average	72.833333	90.666667	3.1666667	
Variance	890.96667	343.06667	1.3666667	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	3392.1111	2	1696.0556	36.69351	4.707E-05	4.2564947
Columns	25655.444	2	12827.722	277.52284	8.188E-09	4.2564947
Interaction	2368.8889	4	592.22222	12.8125	**0.000929**	3.6330885
Within	416	9	46.222222			

P-Value = 0.000929 < 0.05, then we Reject H_0,
So,there is an interaction.

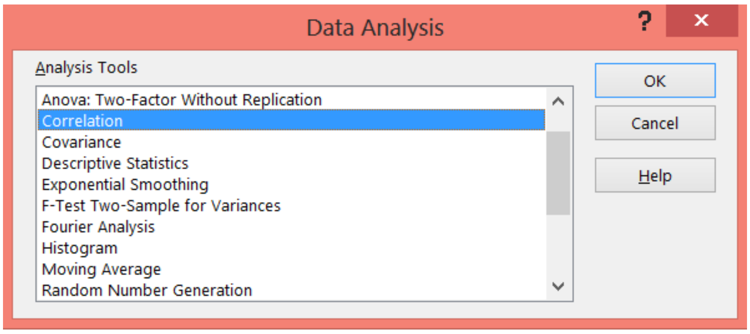
Question :

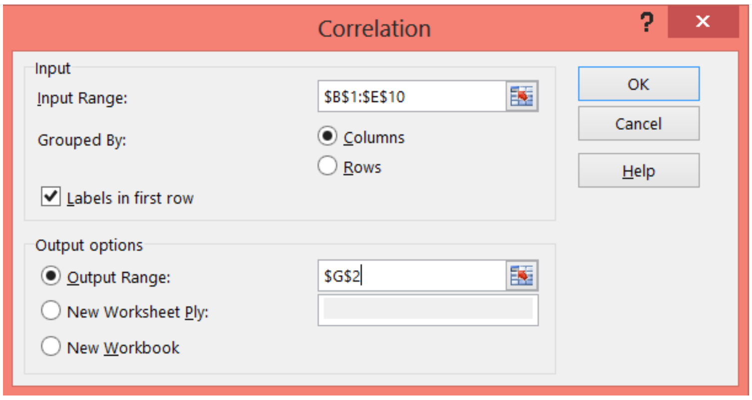
The following data is from a study on frame delay times of access point in wireless network. Frame delay times in milliseconds were recorded for 4 access point (AP1, AP2, AP3, and AP4).

- *Calculate the correlation matrix and interpret the result?*

AP1	*AP2*	*AP3*	*AP4*
93	*85*	*100*	*96*
120	*45*	*75*	*58*
65	*80*	*65*	*95*
105	*28*	*40*	*90*
115	*75*	*73*	*65*
82	*70*	*65*	*80*
99	*65*	*50*	*85*
87	*55*	*30*	*95*
100	*50*	*45*	*82*

DATA ⟶ *Data Analysis*





	AP1	*AP2*	*AP3*	*AP4*
AP1	1			
AP2	-0.4723	1		
AP3	**0.0901**	0.6337	1	
AP4	**-0.7382**	0.1884	-0.2363	1

There is a strong negative relationship between AP1 and AP4 (-0.7382).
On the other hand, the weakest relationship was between AP1 and AP3 (0.0901)

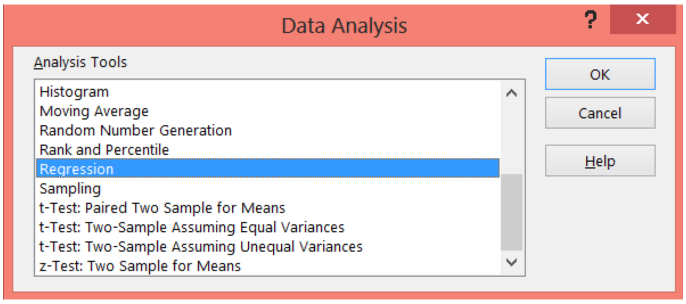
Question :

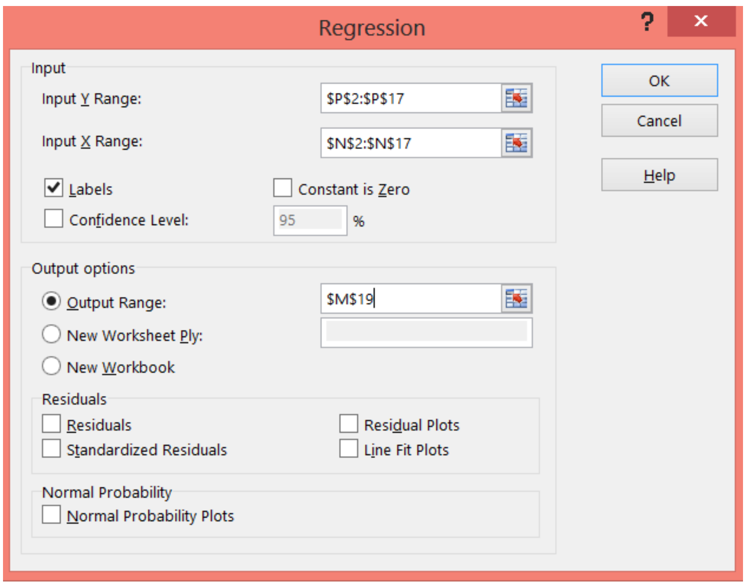
Consider the data obtained from a chemical process where the yield of the process is thought to be related to the reaction temperature (see the table below).

Observation number	*Temperature (X)*	*Yield (Y)*
1	50	122
2	53	118
3	54	128
4	55	121
5	56	125
6	50	136
7	62	144
8	65	142
9	67	149
10	71	161
11	72	167
12	74	168
13	75	162
14	76	171
15	79	175

Fit linear regression model relating Yields to Temperatures and discuss the result.

DATA ⟶ *Data Analysis*





SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.95107322
R Square	0.90454028
Adjusted R Square	0.89719722
Standard Error	6.49841701
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5201.951	5201.951	123.183	**0.00**
Residual	13	548.983	42.229		
Total	14	5750.933			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	**25.866**	10.947	2.363	0.0344	2.216	49.517	2.215	49.5172
(X)	**1.878**	0.169	11.099	0.00	1.512	2.244	1.512	2.2446

- *The regression model is* $\boxed{\hat{Y} = b_0 + b_1 X}$

$$\boxed{\hat{Y} = 25.866 + 1.878\,X}$$

- *If we test*

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

P-Value = 0.000 < 0.05, then we Reject H_0
So, there is a significant relationship between Yields and Temperature.