**Quiz 2 - Solution**

|  |  |  |
| --- | --- | --- |
| STAT 328  | Academic year 1441 H  | Send you answer before 14/6/1441 -9:00PM |
| Statistical Methods |  Second Semester | By E-mail for: wemam.c@ksu.edu.sa |

|  |  |  |
| --- | --- | --- |
| **Student’s Name** |  | **اسم الطالب** |
| **ID number** |  | **الرقم الجامعي** |
| **Section No.** |  | **رقم الشعبة** |

Write Excel commands with the results to calculate the following:

**Question 1**

Use the following data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group A | 61 | 86 | 84 | 75 | 91 | 83 | 75 | 62 | 75 | 85 |

1. Write a descriptive statistics report about the data showing, the mean, median, mode, and the sum for group A.

Mean 77.7

Median 79

Mode 75

Sum 777

1. Find 93% CI for the mean of group A.

Confidence Level(93.0%) 6.550446

93% CI for the mean of group A=(77.7-6.550446, 77.7+6.550446)

=( 71.150,84.250)

1. Test that, the mean of this data is greater than 75.

$$H\_{0}:μ\leq 75 VS H\_{1}:μ>75 $$

t-Test: Two-Sample Assuming Unequal Variances

*Group A*

Mean 77.7

Observations 10

t Stat 0.847204

P(T<=t) one-tail 0.20942

t Critical one-tail 1.833113

P(T<=t) two-tail 0.41884

t Critical two-tail 2.262157

since P(T<=t) one-tail 0.20942>0.05, then we can’t reject $H\_{0}$ and there is no different significance between the mean of group A and 75

**Question 2**

A consumer agency wanted to find if the mean time taken by each of three brands of medicine to provide relief from a headache is the same. The first drug was administered to six randomly selected patients, the second drug to four randomly selected patients, and the third to five randomly selected patients. The following gives the time (in minutes) for each patient to get relief from a headache after taking the medicine.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Drug1 | 35 | 38 | 42 | 65 | 47 | 52 |
| Drug2 Drug1 | 13 | 21 | 19 | 29 | 68 |  |
| Drug3 | 44 | 39 | 54 | 58 |  |  |

Assume the mean time to get relief follows a normal distribution for all brands of drugs and they have equal variances. At α = 0.01 , can you conclude that the mean time taken to get relief differs for the three drugs?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Anova: Single Factor |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Drug1 | 6 | 279 | 46.5 | 119.5 |  |  |
| Drug2 | 5 | 150 | 30 | 484 |  |  |
| Drug3 | 4 | 195 | 48.75 | 76.91667 |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 1021.35 | 2 | 510.675 | 2.216912 | 0.151585 | 6.926608 |
| Within Groups | 2764.25 | 12 | 230.3542 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 3785.6 | 14 |   |   |   |   |

$H\_{0}: μ\_{Drug1}=μ\_{Drug2}=μ\_{Drug3} $ vs

 $H\_{1}: there Exist difference in at least one drug$

$P\_{Value}=0.151585>0.01$, then we Accept (don’t reject $H\_{0}$, i.e. we conclude that the mean time taken to get relief differs for the three drugs

**Question 3**

Use the following data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 1 | 0 | 2 | 0 | 3 | 1 | 0 | 1 | 2 | 0 |
| *Y* | 16 | 9 | 17 | 12 | 22 | 13 | 8 | 15 | 19 | 11 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* |  |  |  |  |  |
| Multiple R | 0.949158 |  |  |  |  |  |
| R Square | 0.900901 |  |  |  |  |  |
| Adjusted R Square | 0.888514 |  |  |  |  |  |
| Standard Error | 1.48324 |  |  |  |  |  |
| Observations | 10 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 1 | 160 | 160 | 72.72727 | 2.75E-05 |  |
| Residual | 8 | 17.6 | 2.2 |  |  |  |
| Total | 9 | 177.6 |   |   |   |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 10.2 | 0.663325 | 15.37708 | 3.18E-07 | 8.67037 | 11.72963 |
| X | 4 | 0.469042 | 8.528029 | 2.75E-05 | 2.918388 | 5.081612 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

1. Estimate the simple regression model and interpret the results.

*Y* = 10.2 + 4\**X*

* There is 10$.2$ of *Y* does not depend on *X* (at X = 0)
* When *X* increase by one unit, then *Y* increase by 4 units.
1. Find the correlation coefficient between X and y.

$R=\sqrt{0.900901}=$0.949158

|  |  |  |
| --- | --- | --- |
|  | *X* | *Y* |
| X | 1 |  |
| Y | 0.949158 | 1 |

**Question 4**

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.

|  |  |  |
| --- | --- | --- |
| 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 |   |

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

|  |  |  |
| --- | --- | --- |
| 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 |  | -1.26036 | LB |
| Standard Error | 0.583609458 |  | 1.98036 | UB |
| 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 |  |  |  |
|  |  |  |  |  |



1. Can you conclude that there is a significance difference in the mean of energy consumption before and after the wall insulation is installed at the significance level 0.05? Test it and evaluate the p-value of your test.

Hypothesis:

H0: µ1 = µ2 (µD = 0) VS H0: µ1  µ2 (µD  0)

*P\_Value* = 0.570711789 > 0.05, then we accept H0, so, there is no difference in mean energy consumption before and after the wall insulation

*P\_Value* Calculation:

*P\_Value* =2\*(1-T.DIST(N17,4,TRUE)) = 0.570712

**Question 5**

The following data represents sample of size **5** student in the final exam of stat1 and stat2:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  Student | 1 | 2 | 3 | 4 | 5 |
| stat1 | 75 | 82 | 90 | 88 | 93 |
| stat2 | 90 | 84 | 92 | 81 | 75 |

1. Write a descriptive statistics report about the data showing, the mean, variance, coefficient of the variation for the sum of stat1 and stat2.

|  |
| --- |
| *sum* |
| Mean | 170.00 |
| Standard Deviation | 6.892 |
| Sample Variance | 47.50 |
| coefficient of the variation | 4.054% |
| Confidence Level(95.0%) | 8.557 |

1. Find 95% CI for the mean of sum.

95% CI= mean$\mp 8.557$=*( 161.442,178.557)*

1. Test the difference between the overall means of stat1 and stat2.

$H\_{0}: σ\_{1}^{2}=σ\_{2}^{2}$ vs $H\_{1}: σ\_{1}^{2}\ne σ\_{2}^{2}$

|  |
| --- |
| F-Test Two-Sample for Variances |
| F | 1.085 |  |
| P(F<=f) one-tail | 0.470 |  |
| F Critical one-tail | 6.388 |  |

Since F=1.085<F critical=6.388 and P(F<=f) one-tail = 0.470 > 0.05 then the variances are equal and then we must use Two-Sample Assuming Equal Variances

$H\_{0}: μ\_{1}=μ\_{2} $ vs $H\_{1}: μ\_{1}\ne μ\_{2}$

|  |
| --- |
| t-Test: Two-Sample Assuming Equal Variances |
| t Stat | 0.270 |  |
| P(T<=t) one-tail | 0.397 |  |
| t Critical one-tail | 1.860 |  |
| P(T<=t) two-tail | 0.794 |  |
| t Critical two-tail | 2.306 |  |

Since t Stat = 0.270 < 2.306 = t Critical two-tail and P(T<=t) two-tail = 0.794>0.05 then we can’t reject $H\_{0}$ and so, $ μ\_{1}=μ\_{2}$

1. Calculate the correlation coefficient between the marks of stat1 and stat2.

|  |  |  |
| --- | --- | --- |
|  | *stat1* | *stat2* |
| stat 1 | 1 |  |
| stat 2 | -0.519 | 1 |

**Question 6**

1. The following data are two independent random samples from two independent populations $A\~N\left(μ\_{1}, 16\right), B\~N\left(μ\_{2},9\right), $respectively.

A: 48 49 42 52 40 48 52 52 55 48

B: 50 48 45 40 43 48 52 46 39 38

1. Test the equality of the two means.

**H0: µ1=µ2 Vs H1: µ1≠µ2**

|  |
| --- |
| z-Test: Two Sample for Means |
|  |  |  |
|  | *A* | *B* |
| Mean | 48.6 | 44.9 |
| Known Variance | 16 | 9 |
| Observations | 10 | 10 |
| Hypothesized Mean Difference | 0 |  |
| z | 2.340085 |  |
| P(Z<=z) one-tail | 0.00964 |  |
| z Critical one-tail | 1.644854 |  |
| **P(Z<=z) two-tail** | **0.019279** |  |
| z Critical two-tail | 1.959964 |   |

**Since P\_Value = 0.019279 < 0.05, then we reject H0, so that them exist significant difference between two means.**

1. Find 90% confidence interval of the difference $μ\_{1}-μ\_{2}$

**90% confidence interval of the difference** $μ\_{1}-μ\_{2}$ **= (1.09902, 6.30097)**

**Question 7**

The following are part of the excel regression results to predict the working hours by using the lot size in a certain company

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 62.37 | 26.18 | 2.38 | 0.03 |
| Lot size  | 3.57 | 0.35 | 10.29 | 0.00 |

1. Write the estimated simple linear regression model and interpret the coefficients.

The regression equation is

Working hours = 62.37 + 3.57 \* Lot size

B0$\hat{β}\_{0}$ => The mean response Y (Working hours) is equal 62.37 hours at X (Lot size) is zero. (The part of Y does not depend on X)

B1$\hat{β}\_{1}$ => The mean of response Y (Working hours) increases by 3.57 hours when X (Lot size) increases by one unit.

1. Discuss the significance of the coefficients.

The coefficients are significant since *P\_Value* is 0.000 for B1 and 0.03 for B0.

**Question 8**

The following data is a sample of the marks of Mid-1 exam, Mid-2 exam, Homework and the final exam of a certain course

|  |  |  |  |
| --- | --- | --- | --- |
| **Mid-1** | **Mid-2** | **Homework** | **Final** |
| 16 | 18 | 8 | 34 |
| 18 | 20 | 9 | 38 |
| 20 | 21 | 8 | 36 |
| 18 | 11 | 9 | 32 |
| 23 | 23 | 9 | 38 |
| 11 | 7 | 6 | 32 |
| 16 | 17 | 7 | 34 |
| 20 | 22 | 9 | 34 |
| 21 | 22 | 10 | 34 |
| 25 | 20 | 9 | 34 |

1. Write a descriptive statistics report about the data showing, the mean, variance, coefficient of the variation of Mid-1, Mid-2 and the total marks of each students.

E=AVERAGE(A2,B2,E2)

F=VAR.S(A2,B2,E2)

G=SQRT(F2)/E2\*100

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Mid-1** | **Mid-2** | **Homework** | **Final** | **Total Marks** | **mean** | **variance** | **CV** |
| 16 | 18 | 8 | 34 | 76 | 26.22222 | 1161.333 | 129.9598 |
| 18 | 20 | 9 | 38 | 85 | 29.33333 | 1453 | 129.9485 |
| 20 | 21 | 8 | 36 | 85 | 28.66667 | 1387 | 129.9155 |
| 18 | 11 | 9 | 32 | 70 | 24.66667 | 1039 | 130.6764 |
| 23 | 23 | 9 | 38 | 93 | 31.11111 | 1633.333 | 129.9038 |
| 11 | 7 | 6 | 32 | 56 | 20.88889 | 740.3333 | 130.2562 |
| 16 | 17 | 7 | 34 | 74 | 25.55556 | 1102.333 | 129.9185 |
| 20 | 22 | 9 | 34 | 85 | 28.44444 | 1366.333 | 129.9514 |
| 21 | 22 | 10 | 34 | 87 | 29.11111 | 1430.333 | 129.9152 |
| 25 | 20 | 9 | 34 | 88 | 29.11111 | 1436.333 | 130.1874 |

1. Calculate the 3rd quartile of the final mark

|  |  |  |  |
| --- | --- | --- | --- |
| *Point* | *Final* | *Rank* | *Percent* |
| 2 | 38 | 1 | 88.80% |
| 5 | 38 | 1 | 88.80% |
| 3 | 36 | 3 | 77.70% |
| 1 | 34 | 4 | 22.20% |
| 7 | 34 | 4 | 22.20% |
| 8 | 34 | 4 | 22.20% |
| 9 | 34 | 4 | 22.20% |
| 10 | 34 | 4 | 22.20% |
| 4 | 32 | 9 | 0.00% |
| 6 | 32 | 9 | 0.00% |

1. Calculate the median and mode(s) of the homework

=MEDIAN(C2:C11)= 9 , =MODE.SNGL(C2:C11)= 9

1. Test the difference between the overall means of Mid-1 and Mid-2.

|  |
| --- |
| t-Test: Two-Sample Assuming Equal Variances |
|  |  |  |
|  | *Mid-1* | *Mid-2* |
| Mean | 18.8 | 18.1 |
| Variance | 15.73333 | 27.21111 |
| Observations | 10 | 10 |
| Pooled Variance | 21.47222 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 18 |  |
| t Stat | 0.337788 |  |
| P(T<=t) one-tail | 0.369715 |  |
| t Critical one-tail | 1.734064 |  |
| **P(T<=t) two-tail** | **0.73943** |  |
| t Critical two-tail | 2.100922 |   |

H0: µmid1= µmid2 , H1: µmid1≠ µmid2

P\_Value=0.73943 > 0.05, Accept H0.

There is no significance difference between the overall means of Mid-1 and Mid-2

1. Calculate the correlation coefficient between the marks of Mid-1 and Mid-2

|  |  |  |
| --- | --- | --- |
|  | *Mid-1* | *Mid-2* |
| Mid-1 | 1 |  |
| Mid-2 | 0.768983 | 1 |

**Question 9**

The following data are the average weekly losses of worker-hours due to accident in 10 industrial plants before and after certain safety program was put into operation:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Before | 45 | 73 | 46 | 124 | 33 | 57 | 83 | 34 | 26 | 17 |
| After | 36 | 60 | 44 | 119 | 35 | 51 | 77 | 29 | 24 | 11 |

Test whether the safety program is effective.

|  |
| --- |
| t-Test: Paired Two Sample for Means |
|  |  |  |
|  | *Before* | *After* |
| Mean | 53.8 | 48.6 |
| Variance | 1027.733 | 962.9333 |
| Observations | 10 | 10 |
| Pearson Correlation | 0.992176 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 9 |  |
| t Stat | 4.033284 |  |
| P(T<=t) one-tail | 0.001479 |  |
| t Critical one-tail | 1.833113 |  |
| P(T<=t) two-tail | 0.002958 |  |
| t Critical two-tail | 2.262157 |   |

H0: µd= 0 , H1: µd > 0

P\_Value=0.001479 < 0.05, Reject H0.

The safety program is effective.