Name ______KEY_______________ /120

Both problems are worth 60 points.

1. Liquid ethanol has approximately 65% of the energy density of petroleum based gasoline. Thus, it is to be expected that “E blend” vehicle fuels will have lower fuel efficiency than regular gasoline. However, in practice fuel efficiency is a complicated function of many other factors besides fuel energy content. To investigate the effects of different ethanol concentrations on fuel efficiency, six different cars of the same make and model, each a 2007 Toyota Camry, were driven at a nearly steady speed of 65 mph over the same 200 mile stretch of interstate highway. This trip was repeated four times for each car using a different ethanol mixture which was randomly assigned to a given car on a given trip. The fuels were E-00, regular gasoline, E-10, 10% ethanol, E-20, 20% ethanol, and E-40, 40% ethanol. The fuel efficiency results are stated below in units of mpg (miles per gallon).

E - 00	E - 10	E - 20	E- 40
33.00	32.76	31.23	31.99
34.96	32.73	33.26	31.49
34.38	35.29	33.62	31.47
34.06	31.53	32.33	32.17
34.66	33.50	33.17	30.53
33.75	33.93	33.07	31.98

Test at a 5% level of significance whether there are any differences in fuel efficiency associated with the different ethanol blends. State and explain your conclusion. Explain how you reached your conclusion. Compute and report the *P*- Value of your observed statistic.

H_0: There are no differences in fuel efficiency associated with the different ethanol blends tested.

H_1: There are differences in fuel efficiency associated with the different ethanol blends tested.

Source	Sum of Sq	DOF	Mean SQ		
				Fobs =	8.3027
Treatment	20.1463	3	6.7154	$F_{0.05}$ =	3.10
Error	16.1765	20	0.8088	P-value =	8.76E-04

Conclusion: Reject H_0. There are differences in fuel efficiency associated with the different ethanol blends tested.

Now determine by a Bonferroni procedure which, if any, of the four fuels are significantly different from each other in their mean fuel efficiency.

Bonferroni Procedure:

	E-00	E-10	E-20
E-10	1.63		
E-20	2.61	0.98	
E-40	4.87	3.25	2.26

$t_{critical}$ = 2.93

Conclusion: E-40 is significantly different from E-00 and E-10. There is insufficient evidence at a 5% level of significance to conclude any other differences in fuel efficiency associated with the different ethanol blends tested.

What assumptions are necessary for the above Single Factor ANOVA to be valid?

The populations, i.e., the mpg’s of the cars tested are assumed normally distributed with a variance that does not change with the different ethanol blends.

Do these assumptions seem to be satisfied by this particular set of data? Explain your answer. **There is not enough data to conclude whether the scores are normally distributed. This assumption is not crucial to the conclusions. The sample variances range from** 0.36 to 1.63 which is an observed F ratio of 4.51 which exceeds the critical F score $F_{.05}(6, 6) = 4.28$. So the assumption of equal variances is suspect. A second study with larger sample sizes may be warranted.

Hand in a spreadsheet with the exam. **See below:**

Problem 1

E- 00		E 10		E 20		E-40	
33.00	1089.00	32.76	1073.22	31.23	975.31	31.99	1023.36
34.96	1222.20	32.73	1071.25	33.26	1106.23	31.49	991.62
34.38	1181.98	35.29	1245.38	33.62	1130.30	31.47	990.36
34.06	1160.08	31.53	994.14	32.33	1045.23	32.17	1034.91
34.66	1201.32	33.50	1122.25	33.17	1100.25	30.53	932.08
33.75	1139.06	33.93	1151.24	33.07	1093.62	31.98	1022.72
204.81	6993.65	199.74	6657.49	196.68	6450.95	189.63	5995.05

Y bar *i*	34.14	33.29	32.78	31.61
s^2_i	0.4917	1.6292	0.7554	0.3590
A_i	6991.19	6649.34	6447.17	5993.26

A 26080.96 **SST = A-T²/n** ***s^2_i max / s^2_i min =*** 4.51

B 26097.14 **SSE = B-A**

T 790.86 $F_{.05}(6,6)$ = 4.28

ANOVA_TABLE

Source	Sum of Sq	DOF	Mean SQ
Treatment	20.1463	3	6.7154
Error	16.1765	20	0.8088
Total	36.3228	23	1.5793

Fobs = 8.3027

$F_{0.05}$ = 3.10

p-value = 8.76E-04

Reject H_0 .

There are differences in mpg associated with different ethanol mixes

Bonferroni Procedure:

	E-00	**E-10**	**E-20**
E-10	1.63		
E-20	2.61	0.98	
E-40	4.87	3.25	2.26

$t_{critical}$ = 2.93

E-40 is significantly different from E-00 and E-10

2. The data shown below gives the measured NO_X emission concentrations (in ppm) for 12 pairs of power boilers. The boilers in each pair have a pre-determined burner area liberation rate (in MBTU/ft^2 per hour). Assume the determination of burner area liberation rate is not subject to any error.

Burner Rate	NO_X Concentration
100	93
100	116
125	158
125	165
150	202
150	203
200	267
200	305
250	365
250	402
300	477
300	483
350	584
350	605
400	653
400	659
450	728
450	762
500	829
500	808
550	910
550	930
600	1037
600	1029

In Excel perform a regression analysis on this data. Your spreadsheet analysis of this data set should contain two scatter plots. The **first** should display **both the data and the predictions of the linear regression model** drawn as a continuous line. The **second** should display the **residuals** plotted versus the control variable.

Assuming that the underlying relation between burner area liberation rate and NO_X concentration is linear and that the errors in the measurements are normally distributed, calculate the lower and upper limits for a 99% confidence intervals requested in the following table.

99% Confidence Interval	Lower Limit	Upper Limit
Population Slope β_1	**1.75733**	**1.87566**
Population Intercept β_0	**-91.489**	**-47.769**
The actual NO_x Concentration when the Burner Rate = 325	**511.0**	**530.4**
A measurement of NO_x Concentration when the Burner Rate = 325	**472.3**	**569.1**
The actual NO_x Concentration when the Burner Rate = 625	**1045.8**	**1085.6**
A measurement of NO_x Concentration when the Burner Rate = 625	**1014.2**	**1117.1**

For what value of Burner Rate does the confidence interval for the NO_x Concentration have the smallest width. Explain why this is true. **At Burner Rate = 325, since this is close to the *x* average of 331.25 . The closer to the mean of *x* the smaller the spread, i.e., the more certain is the model predictions.**

Fill in the following Regression ANOVA table as shown below.

Source	Sum of Squares	Degrees of Freedom	Mean Square
Linear Model	**2121047.294**	1	**2121047.294**
Error	**6230.538973**	22	**283.2063**
Total	**2127277.833**	23	**92490.34**

Observed *F* score **7489.406724 > $F_{.01}$ = 7.9453**

At a 1% level of significance, what conclusion do you draw from this ANOVA Table?
A linear relationship between x and y explains observed pattern of the data. Randomness could only explain the observed pattern with a probability of 2.29137×10^{-29}.

Compute both the correlation coefficient and the coefficient of determination.
r = 0.998534487
r^2 = 0.997071121

Comment on how well the linear model fits this data.
It fits the data quite well! 99.7% of all of the variability in NO_x Concentration is explained by a linear dependence on burner burner area liberation rate.

What does the pattern of residuals indicate about the adequacy of the linear model?
The pattern of residuals appears random. This means that it would be pointless to go beyond a linear model. There is no pattern left over (residual) to explain. The linear model is quite adequate and explains nearly all of the observed data.

Spread Sheet Output

i	x	y	x^2	xy	y^2	y est	Residuals
1	100	93	10000	9300	8649	112.02	-19.02
2	100	116	10000	11600	13456	112.02	3.98
3	125	158	15625	19750	24964	157.43	0.57
4	125	165	15625	20625	27225	157.43	7.57
5	150	202	22500	30300	40804	202.84	-0.84
6	150	203	22500	30450	41209	202.84	0.16
7	200	267	40000	53400	71289	293.67	-26.67
8	200	305	40000	61000	93025	293.67	11.33
9	250	365	62500	91250	133225	384.49	-19.49
10	250	402	62500	100500	161604	384.49	17.51
11	300	477	90000	143100	227529	475.32	1.68
12	300	483	90000	144900	233289	475.32	7.68
13	350	584	122500	204400	341056	566.14	17.86
14	350	605	122500	211750	366025	566.14	38.86
15	400	653	160000	261200	426409	656.97	-3.97
16	400	659	160000	263600	434281	656.97	2.03
17	450	728	202500	327600	529984	747.79	-19.79
18	450	762	202500	342900	580644	747.79	14.21
19	500	829	250000	414500	687241	838.62	-9.62
20	500	808	250000	404000	652864	838.62	-30.62
21	550	910	302500	500500	828100	929.44	-19.44
22	550	930	302500	511500	864900	929.44	0.56
23	600	1037	360000	622200	1075369	1020.27	16.73
24	600	1029	360000	617400	1058841	1020.27	8.73
	7950	12770	3276250	5397725	8921982	12770	0

x-bar	331.25
y-bar	532.08
SSxx	642812.50
SSxy	1167662.5
SSyy	2127277.833
B_1	1.816490034
B_0	-69.62899044
r	0.998534487
r^2	0.997071121
Se	16.82873486
S_{B1}	0.020989849
S_{B0}	7.755185806

99% C.I.		Lower Limit		Upper Limit
Slope	1.816490034	+/-	0.059165	
		1.757325		1.875655
Intercept	-69.62899044	+/-	21.86001	
	-91.489		-47.769	
X_0 =	325			
$Y\hat{}_0$ =	520.7302706	+/-	9.689927	

Mean		**511.0403**		**530.4202**
	520.7302706	+/-	**48.41575**	
Single		**472.3145**		**569.146**
Measurement				
X_0 =	625			
$Y\hat{}_0$ =	**1065.677281**	**+/-**	**19.89513**	
Mean		**1045.782**		**1085.572**
	1065.677281	+/-	**51.43935**	
Single		**1014.238**		**1117.117**
Measurement				

ANOVA Table

	Source	Sum of SQ	DOF	Mean SQ
Linear	**Model**	2121047.294	1	2121047
	Error	6230.538973	22	283.2063
	Total	2127277.833	23	92490.34

Fobs =	**7489.406724**
$r^2(n-2)/(1-r^2)$	**7489.406724**
$F_{.01}$ =	**7.945345715**
p-value =	**2.29137E-29**