Family Name:________________________

Given Name:________________________

Student Number:________________________

MIE237S Term Test
Examination Type B; Calculator Type 2 Permitted
March 5, 2014, 12:10 P.M.
50 minutes; 40 Marks Available

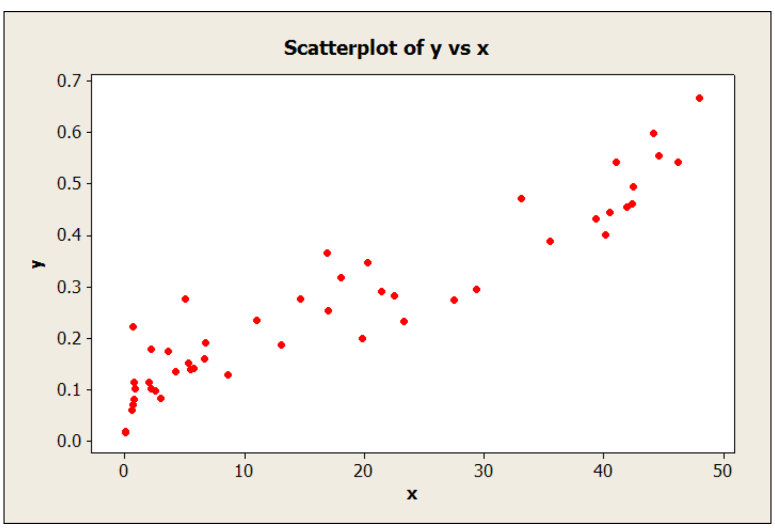
This booklet has 10 pages. The first seven pages are numbered 1–7 and contain the questions with space to answer. You may use the backs of any pages for rough work. The final three pages are not numbered and contain the aid sheet and a table of t probabilities.

If you wish to detach the last three pages you may do so *by tearing the paper along the dotted lines at the top left corner those three pages.*

1.**(20 marks total)** A paper was published in 2002 in the journal *Metallurgical and Materials Transactions B* that described some relationships among many characteristics of chromium oxides in a steel-making process.

The details of the process are not important, except to note that all of the measured variables are positive.

In this question we will focus on the two variables called "activity coefficient" (x) and "reciprocal of amount" (y). Here is a plot of the data, which consists of $n = 48$ records:



We will analyze the data using the usual simple linear regression model as specified on the aid sheet.

The sample average of the x values is $\overline{x} = 17.99$ and $S_{xx} = 12780.24$.

Here is the Minitab output with some of the entries replaced with $*$ symbols:

```
The regression equation is
y = ****** + ******* x

Predictor       Coef    SE Coef      T      P
Constant     0.09367  *********  *****  *****
x          0.0095734  0.0005080  *****  *****

S = *********   R-Sq = *****

Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  ******  ******  ******  *****
Residual Error  46  ******  ******
Total           47  1.3224

Unusual Observations

Obs     x        y      Fit   SE Fit  Residual  St Resid
  7   0.7  0.22124  0.10037  0.01207   0.12087     2.15R
 29   5.1  0.27624  0.14249  0.01056   0.13375     2.37R
 48  48.0  0.66667  0.55319  0.01735   0.11348     2.07R

R denotes an observation with a large standardized residual.
```

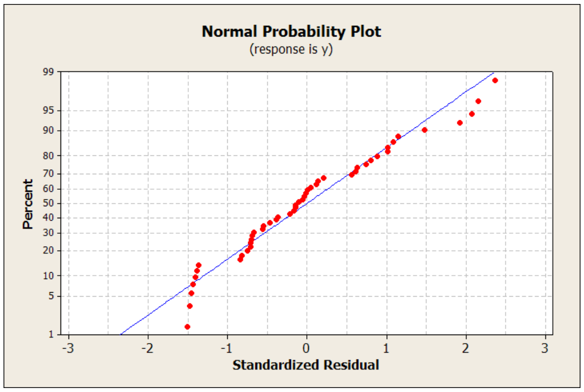
(a) **(4 marks)** Provide the missing entries for the Analysis of Variance part of the Minitab output, but do not attempt to compute the p-value.

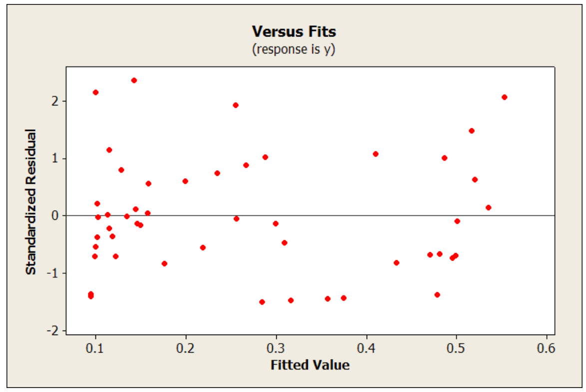
```
Analysis of Variance

Source          DF    SS              MS              F          P

Regression      __    ____________    ____________    ________   omit

Residual Error  __    ____________    ____________

Total           __    ____________
```

(b) **(4 marks)** Produce the 95% confidence interval for β_1

(c) **(4 marks)** Comment on the possible existence of unusual observations in the data.

(d) **(4 marks)** Here are a normal probability plot of the standardized residuals and a plot of the standardized residuals versus the fitted values:





Comment on the extent to which the usual lineal regression model assumptions have been satisfied.

(e) **(4 marks)** Predict the value of the response y at an x value of 10. If possible, produce a 95% prediction interval for a new response at this x value of 10. If not, state why not.

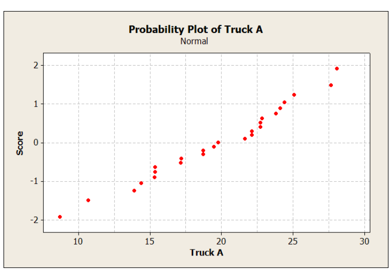
2. **(15 marks)** A mining company wants to evaluate the productivity of two brands (Brand A and Brand B) of haul truck in an open pit mining environment. They own 25 trucks of each brand and Truck B). Each truck performs the same task: moves raw material from the mine site to the processing area. Productivity is measured by the amount of raw material the truck is able to move in a given time period.

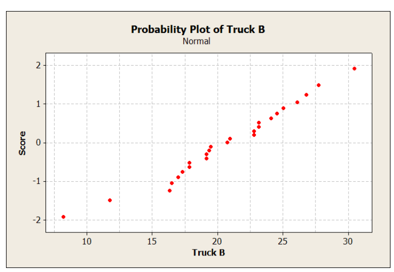
The company plans to operate all trucks in the same mine for two weeks andthey will measure the total amount of raw material moved by each truck (in hundreds of thousands of kilograms).

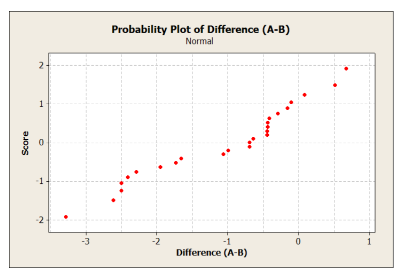
Here is an excerpt of the raw data they collected, just to show you how it was organized. Most rows of the raw data were omitted to save space. The last two rows contain the sample averages and sample standard deviations of each column.

Truck #	Brand A	Brand B	Difference (A-B)
1	15.335	16.989	-1.655
2	28.082	30.493	-2.412
3	27.653	27.755	-0.102
⋮	⋮	⋮	⋮
24	19.462	19.381	0.082
25	8.716	8.199	0.517
Average	18.51	21.02	-2.51
SD	5.38	5.91	1.32

Here are normal quantile plots of the data from the three columns:







Use the following page to perform the hypothesis test that answers the question "*Is there a difference in the average productivity between Truck A and Truck B?*". Include the following:

- perform the hypothesis test using a p-value in your conclusion;
- comment on whether or not the model assumptions you used in your calculations have been satisfied, and if they have not been satisfied, whether any violations cast doubt on the validity of your conclusions.

You won't need the entire page to answer. It's just how the question spacing worked out.

3.**(5 marks total)** Some data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is analyzed using the usual simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with slope estimator $\hat{\beta}_1 = S_{xy}/S_{xx}$.

Suppose a new variable w_i is introduced that is just a linear transformation of the y_i variable. In other words, $w_i = c + dy_i$ for each i (assume $d \neq 0$.) Consider the new simple linear regression model $w_i = \beta_0^{(w)} + \beta_1^{(w)} x_i + \varepsilon_i$.

Show that the new slope estimator $\hat{\beta}_1^{(w)}$ is equal to d multiplied by the old slope estimator $\hat{\beta}_1$.

Two Samples

Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$ with $i \in \{1,2\}$ and ε_{ij} i.i.d. $N(0,\sigma^2)$.

Pooled sample variance:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

Test Statistic:

$$T = \frac{(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$(1-\alpha)\cdot 100\%$ C.I. is

$$\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot} \pm t_{n_1+n_2-2,\alpha/2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Checking assumptions: normal plots and 3:1 ratio for equal variance assumption.

If observations are really "paired", use one-sample procedures on the paired differences using this fact:

$$\frac{\overline{Y}_d - \mu_d}{S_d/\sqrt{n}} \sim t_{n-1}$$

where $\overline{Y}_d$ is the sample average of the differences, μ_d is the mean difference between the two populations, S_d is the sample standard deviation of the differences and n is the number of paired observations.
Alternate approach for $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ uses:

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2 + \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2$$

$$SST = SSM + SSE$$

$$F = \frac{MSM}{MSE} = \frac{SSM/1}{SSE/(n_1+n_2-2)} \sim F_{1,n_1+n_2-2}$$

MSE is exactly S_p^2 and $T^2 = F$

Simple Linear Regression

Model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0,\sigma^2)$

Analysis: $\hat\beta_0 = \overline{y} - \hat\beta_1\overline{x}$

$\hat\beta_1 = S_{xy}/S_{xx}$

$S_{xy} = \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$

$S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2$

$\text{Var}\left(\hat\beta_1\right) = \frac{\sigma^2}{S_{xx}}$

Fitted value at x_i is $\hat{Y}_i = \hat\beta_0 + \hat\beta_1 x_i$

SS decomposition details:

$$\begin{array}{ccccc} \sum_{i=1}^n \left(Y_i - \overline{Y}\right)^2 & = & \sum_{i=1}^n \left(\hat{Y}_i - \overline{Y}\right)^2 & + & \sum_{i=1}^n \left(Y_i - \hat{Y}_i\right)^2 \\ \text{SST} & = & \text{SSR} & + & \text{SSE} \\ n-1 \text{ d.f.} & = & 1 \text{ d.f.} & + & n-2 \text{ d.f.} \end{array}$$

Test statistic for β_1:

$$T = \frac{\hat\beta_1 - \beta_1}{\sqrt{MSE/S_{xx}}} \sim t_{n-2}$$

The denominator is called the "standard error" of $\hat\beta_1$

$(1-\alpha)\cdot 100\%$ C.I. for β_1 is:

$$\hat\beta_1 \pm t_{n-2,\alpha/2}\sqrt{\frac{MSE}{S_{xx}}}$$

Alternate approach for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ uses (again... $T^2 = F$):

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F_{1,n-2}$$

Putting it all together as in Minitab. (Note: `#` denotes "not usually of interest".)

`The regression equation is`
`y =` $\hat\beta_0 + \hat\beta_1$`x`

Predictor	Coef	SE Coef	T	P
Constant	$\hat\beta_0$	#	#	#
x	$\hat\beta_1$	$SE(\hat\beta_1)$	$\frac{\hat\beta_1}{SE(\hat\beta_1)}$	p-value

$s = \sqrt{MSE}$ $\qquad R^2 = SSR/SST$

Source	DF	SS	MS	F	P
Regression	1	SSR	MSR	MSR/MSE	p-value
Residual Error	$n-2$	SSE	MSE		
Total	$n-1$	SST			

$(1-\alpha)\cdot 100\%$ C.I. for mean response at x_0:

$$\hat\beta_0 + \hat\beta_1 x_0 \pm t_{\alpha/2,n-2}\sqrt{MSE}\sqrt{\frac{1}{n} + \frac{(x_0-\overline{x})^2}{S_{xx}}}$$

$(1-\alpha)\cdot 100\%$ P.I. for new response at x_0:

$$\hat\beta_0 + \hat\beta_1 x_0 \pm t_{\alpha/2,n-2}\sqrt{MSE}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\overline{x})^2}{S_{xx}}}$$

Hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T}$ but mainly interested in

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{S_{xx}}$$

Studentized residuals for outlier detection:

$$\frac{\hat{\varepsilon}_i}{\sqrt{MSE}\sqrt{1-h_{ii}}}$$

Outlier heuristic: if bigger than 2 or 3, see if plots suggest true outlier.

Influential point heuristic: look for $h_{ii} > 3 \cdot \frac{k+1}{n}$ where k is the number of model terms, check plots if possible.

Sample Correlation Coefficient

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

In simple regression the square of r is equal to R^2

t distribution upper tail probabilities: $P(t_{df} \geq t)$

	Upper tail probability α												
df	.30	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
11	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.530	0.682	0.853	1.054	1.309	1.696	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.530	0.682	0.853	1.054	1.309	1.694	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.530	0.682	0.853	1.053	1.308	1.692	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.529	0.682	0.852	1.052	1.307	1.691	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.529	0.682	0.852	1.052	1.306	1.690	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.529	0.681	0.852	1.052	1.306	1.688	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.529	0.681	0.851	1.051	1.305	1.687	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.529	0.681	0.851	1.051	1.304	1.686	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.529	0.681	0.851	1.050	1.304	1.685	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
41	0.529	0.681	0.850	1.050	1.303	1.683	2.020	2.121	2.421	2.701	2.967	3.301	3.544
42	0.528	0.680	0.850	1.049	1.302	1.682	2.018	2.120	2.418	2.698	2.963	3.296	3.538
43	0.528	0.680	0.850	1.049	1.302	1.681	2.017	2.118	2.416	2.695	2.959	3.291	3.532
44	0.528	0.680	0.850	1.049	1.301	1.680	2.015	2.116	2.414	2.692	2.956	3.286	3.526
45	0.528	0.680	0.850	1.049	1.301	1.679	2.014	2.115	2.412	2.690	2.952	3.281	3.520
46	0.528	0.680	0.850	1.048	1.300	1.679	2.013	2.114	2.410	2.687	2.949	3.277	3.515
47	0.528	0.680	0.849	1.048	1.300	1.678	2.012	2.112	2.408	2.685	2.946	3.273	3.510
48	0.528	0.680	0.849	1.048	1.299	1.677	2.011	2.111	2.407	2.682	2.943	3.269	3.505
49	0.528	0.680	0.849	1.048	1.299	1.677	2.010	2.110	2.405	2.680	2.940	3.265	3.500
50	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
51	0.528	0.679	0.849	1.047	1.298	1.675	2.008	2.108	2.402	2.676	2.934	3.258	3.492
52	0.528	0.679	0.849	1.047	1.298	1.675	2.007	2.107	2.400	2.674	2.932	3.255	3.488
53	0.528	0.679	0.848	1.047	1.298	1.674	2.006	2.106	2.399	2.672	2.929	3.251	3.484
54	0.528	0.679	0.848	1.046	1.297	1.674	2.005	2.105	2.397	2.670	2.927	3.248	3.480
55	0.527	0.679	0.848	1.046	1.297	1.673	2.004	2.104	2.396	2.668	2.925	3.245	3.476
56	0.527	0.679	0.848	1.046	1.297	1.673	2.003	2.103	2.395	2.667	2.923	3.242	3.473
57	0.527	0.679	0.848	1.046	1.297	1.672	2.002	2.102	2.394	2.665	2.920	3.239	3.470
58	0.527	0.679	0.848	1.046	1.296	1.672	2.002	2.101	2.392	2.663	2.918	3.237	3.466
59	0.527	0.679	0.848	1.046	1.296	1.671	2.001	2.100	2.391	2.662	2.916	3.234	3.463
60	0.527	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
∞	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291