Topics in Biostatistics   **Homework 1**   **Due Thursday January 16th at 2.30pm**

Directions: Answer the following exercises, showing all relevant work. Students are asked to use the attached Minitab output (do not reproduce the results by hand) – but be clear which output you are using for which part below. ***Assumptions, conclusions and justifications are to be given using clear and concise English***. Please type up your solutions or write ***very*** neatly.

1. On p.146, Norman & Streiner report the medical data set reproduced below. Analyze these data by performing each of the following three analyses. In each case, list all necessary assumptions, and clearly summarize your conclusions. ***Use the attached output to answer the questions below***.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y	46	36	40	44	36	30	42	35	42	50	45	53	48	38	43	58
Treatment	A	A	B	A	A	B	A	B	B	A	A	B	B	B	B	A
X	12	14	27	35	26	21	48	51	62	64	60	77	91	84	55	74

   (a) Using the results from the Minitab output, perform the equal-variance two-sample t-test comparing the above Y averages for the two treatment groups (give the hypotheses, TS, p-value, assumptions & conclusion).
   (b) Using the results from the Minitab output, regress Y on X, obtaining and reporting parameter estimates, and testing whether X is a good linear predictor of Y. State all hypotheses in terms of model parameters.
   (c) Using the results from the Minitab output, perform the ANOCOV (analysis of covariance) analysis to determine if the Y averages differ for the two treatment groups after removing the effect of X. Note any additional assumptions/requirements here. Compare and contrast the conclusion here with that of part (a).

2. Two groups of children, one with attention deficit disorder (ADD) and a control group of children without ADD, were randomly given either a placebo or the drug Ritalin. A measure of activity was made on all the children with the results shown in the table below (higher numbers indicate more activity). Using the attached Minitab output, analyze these data (listing all necessary assumptions), including all relevant observations and implications. Clearly indicate what analysis should be performed next with these data.

Treatment	Group	Drug	Activity
1	ADD	PLACEBO	90
1	ADD	PLACEBO	88
1	ADD	PLACEBO	95
2	CONTROL	PLACEBO	60
2	CONTROL	PLACEBO	62
2	CONTROL	PLACEBO	66
3	ADD	RITALIN	72
3	ADD	RITALIN	70
3	ADD	RITALIN	64
4	CONTROL	RITALIN	86
4	CONTROL	RITALIN	86
4	CONTROL	RITALIN	82

Homework 1 Minitab Attachment

Output A

```
Two-Sample T-Test and CI: y, trt

Two-sample T for y
trt          N      Mean      StDev    SE Mean
a            8      44.63     7.23       2.6
b            8      41.13     7.22       2.6


Difference = mu (a) - mu (b)      Estimate for difference:  3.50
95% CI for difference: (-4.25, 11.25)
T-Test of difference = 0 (vs not =): T-Value = 0.97  P-Value = 0.349  DF = 14
Both use Pooled StDev = 7.22
```

Two-Sample T-Test and CI: y, trt

```
Two-sample T for y
trt          N      Mean      StDev    SE Mean
a            8      44.63     7.23       2.6
b            8      41.13     7.22       2.6


Difference = mu (a) - mu (b)      Estimate for difference:  3.50
95% CI for difference: (-4.30, 11.30)
T-Test of difference = 0 (vs not =): T-Value = 0.97  P-Value = 0.350  DF = 13
```

Output B

```
Regression Analysis: y versus x

The regression equation is
y = 35.0 + 0.158 x

Predictor         Coef      SE Coef          T        P
Constant        34.978        3.553       9.84    0.000
x              0.15774      0.06381       2.47    0.027
S = 6.227      R-Sq = 30.4%     R-Sq(adj) = 25.4%

Analysis of Variance
Source            DF           SS          MS        F        P
Regression         1       236.94      236.94     6.11    0.027
Residual Error    14       542.81       38.77
Total             15       779.75
```

Output C

```
Regression Analysis: y versus x, dum, dumx

The regression equation is
y = 35.1 + 0.228 x - 5.09 dum - 0.039 dumx

Predictor         Coef      SE Coef          T        P
Constant        35.127        4.192       8.38    0.000
x              0.22819      0.08908       2.56    0.025
dum             -5.093        6.673      -0.76    0.460
dumx           -0.0386       0.1212      -0.32    0.756


S = 5.532      R-Sq = 52.9%     R-Sq(adj) = 41.1%


Analysis of Variance
Source            DF           SS          MS        F        P
Regression         3       412.55      137.52     4.49    0.025
Residual Error    12       367.20       30.60
Total             15       779.75
```

Output D

```
Regression Analysis: y versus x, dum

The regression equation is
y = 36.0 + 0.207 x - 7.00 dum

Predictor          Coef       SE Coef            T          P
Constant         35.994         3.074        11.71      0.000
x               0.20735       0.05829         3.56      0.004
dum              -6.999         2.844        -2.46      0.029

S = 5.337      R-Sq = 52.5%      R-Sq(adj) = 45.2%

Analysis of Variance
Source            DF            SS           MS          F          P
Regression         2        409.45       204.72       7.19      0.008
Residual Error    13        370.30        28.48
Total             15        779.75
```

Output E

```
Two-way ANOVA: activity versus group, drug

Analysis of Variance for activity
Source            DF            SS           MS          F          P
group              1         114.1        114.1      10.14      0.013
drug               1           0.1          0.1       0.01      0.934
Interaction        1        1474.1       1474.1     131.03      0.000
Error              8          90.0         11.3
Total             11        1678.3
```


Interaction Plot (data means) for Activity