Spring 2019   **STAT 587C**   **Final exam**
                                                          (100 points)

Instructions:

1. Full credit will be given only if you show your work.

2. The questions are not necessarily ordered from easiest to hardest.

3. You are allowed to use any resource except aid from another individual.

4. Aid from another individual will automatically earn you a 0.

5. Feel free to tear off the last page. There is no need to turn it in.

Regression calculation

1. Suppose we have the following summary statistics for 100 pairs of response-explanatory variables with a sample correlation of -0.7.

	Explanatory (x)	Response (y)
mean	55	-221
standard deviation	9	49

Provide estimates of the following quantities. (2 points each)

(a) Intercept, $\hat{\beta}_0$

(b) Slope, $\hat{\beta}_1$

(c) Coefficient of determination, R^2

(d) Residual sum of squares, SSE

(e) Error variance, $\hat{\sigma}^2$

(f) Standard error for the intercept, $SE(\hat{\beta}_0)$

(g) Standard error for the slope, $SE(\hat{\beta}_1)$

(h) Mean response when the explanatory variable is 70, $E[Y|X = 70]$

(i) Standard error of the mean response when X is 70, $SE(E[Y|X = 70])$

(j) Standard error of prediction when X is 70, $SE(Pred\{Y|X = 70\})$

Model comparisons

2. An unreplicated completely randomized block design (CRBD) experiment is being designed to study the effect of laser etching on hardness of Boron Nitride. The experiment has 2 blocks, 3 levels of laser intensity, and 4 levels of laser speed.

   (a) Complete the degrees of freedom portion of the following ANOVA table. (12 points)

Factor	df
Block	
Intensity	
Speed	
Intensity:Speed	
Error	
Total	

   (b) The following table provides residual sums-of-squares for a sequence of nested models.

	Term included			SSE
Block	Intensity	Speed	Intensity:Speed	
X				55
X	X			43
X	X	X		30
X	X	X	X	8

   Conduct an F-test to compare the model with Block to the model that includes Block, Intensity, Speed, and the interaction between Intensity and Speed. Report the numerator and denominator degrees of freedom, estimate for $\hat{\sigma}^2$, F-statistic, p-value, and an interpretation for the test. (8 points)

Escaping hydrocarbons

3. The file `hydrocarbons.csv` is a `csv` version of the data set found at `http://people.` `sc.fsu.edu/~jburkardt/datasets/regression/x14.txt`. Use the following code to fit a main effects model:

```
d = read.csv("hydrocarbons.csv")
m = lm(Hydrocarbons.escaping..grams. ~ ., data = d)
```

Answer the following questions based on this model fit.

(a) For each of the following regression assumptions, state whether the assumption is reasonably met and state what your evidence for this conclusion. As an example, for the linearity (mean structure) assumption, you might say

> The linearity assumption is reasonably met based on no curvature observed in scatterplots of residuals vs each individual explanatory variable.

(5 points each)

   i. Normality

   ii. Constant variance

   iii. Independence

(b) Explain why the observation in line 18 has high leverage. (5 points)

Escaping hydrocarbons (continued)

(c) Provide estimates of the following quantities based on the model fit. (2 points each)

    i. Intercept

    ii. Coefficient for 'Tank temperature (F)'

    iii. Coefficient for 'Petrol temperature (F)'

    iv. Coefficient for 'Initial tank pressure (pounds/square inch)'

    v. Coefficient for 'Petrol pressure (pounds/square inch)'

    vi. Error variance

    vii. Coefficient of determination

(d) Conduct a t-test to determine whether an interaction between 'Tank temperature (F)' and 'Initial tank pressure (pounds/square inch)' is supported by the data. Report the following quantities for this test. (1 point each)

    i. Estimated coefficient

    ii. t-statistic

    iii. degrees of freedom

    iv. p-value

(e) Provide an interpretation for the result of this test. (2 points)

Story County Commercial Property Sales

4. The page entitled R Code - Story County Commercial Property Sales provides an analysis of properties sold from 2005 through February 2013 based on price ($), year built, building area (ft^2), and land area (ft^2). (5 points each)

   (a) Provide an interpretation (including units) for the following quantities:

      i. $e^{1.015018} = 3$

      ii. $e^{-0.028609} = 0.972$

      iii. $10^{0.215817} = 1.64$

   (b) Construct a two-sided equal-tail 90% confidence interval for the interaction coefficient. (5 points)

R Code - Story County Commercial Property Sales
Feel free to remove this page

```
summary(m)
##
## Call:
## lm(formula = Hydrocarbons.escaping..grams. ~ ., data = d)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.586 -1.221 -0.118  1.320  5.106
##
## Coefficients:
##                                                 Estimate Std. Error t value
## (Intercept)                                      1.01502    1.86131   0.545
## Tank.temperature..F.                            -0.02861    0.09060  -0.316
## Petrol.temperature..F.                           0.21582    0.06772   3.187
## Initial.tank.pressure..pounds.square.inch.      -4.32005    2.85097  -1.515
## Petrol.pressure..pounds.square.inch.             8.97489    2.77263   3.237
##                                                 Pr(>|t|)
## (Intercept)                                      0.59001
## Tank.temperature..F.                             0.75461
## Petrol.temperature..F.                           0.00362 **
## Initial.tank.pressure..pounds.square.inch.       0.14132
## Petrol.pressure..pounds.square.inch.             0.00319 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.73 on 27 degrees of freedom
## Multiple R-squared:  0.9261,Adjusted R-squared:  0.9151
## F-statistic: 84.54 on 4 and 27 DF,  p-value: 7.249e-15
```

(intentionally left blank)