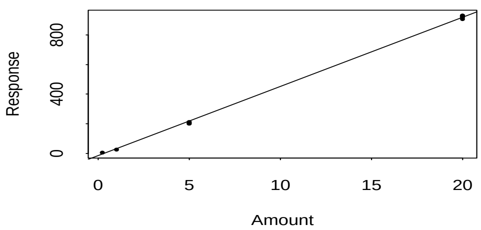
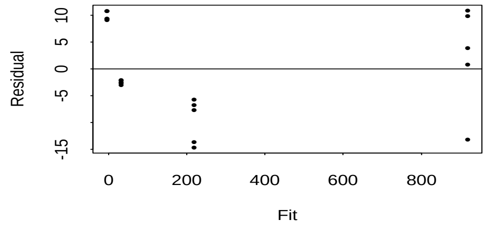
36-401 HW 10
Due 11/16/00

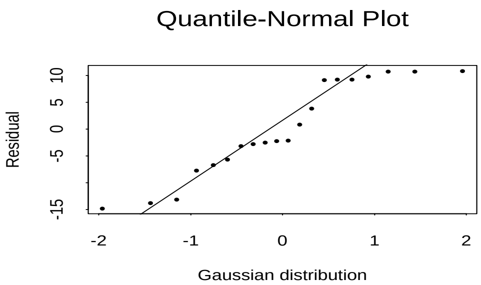
1. (**18 pts, 3 pts each**) Gas chromatography is a technique which is used to detect very small amounts of a substance. Five measurements were taken for each of four specimens (standards) containing different amounts of a particular substance. The amount of the substance in each specimen is precisely known. The response variable is the output reading from the gas chromatograph. The purpose of the study is to calibrate the chromatograph by relating the actual amount of the substance to the chromatograph reading.

   The variables are "amount" of substance in the specimen (nanograms) and "response", the output reading from the gas chromatograph.

```
chr[1:3,]                                table(chr$amount)
#   amount response                      #  0.25 1 5 20
# 1   0.25     6.55                      #     5 5 5  5
# 2   0.25     7.98
# 3   0.25     6.54
rslt_lm(response~amount, chr)
summary(rslt)
# Coefficients:   Value Std. Error   t value  Pr(>|t|)
# (Intercept)  -14.4107    2.6142    -5.5125    0.0000
#      amount   46.6287    0.2533   184.0856    0.0000
# Residual standard error: 9.023 on 18 degrees of freedom
# Multiple R-Squared: 0.9995
# F-statistic: 33890 on 1 and 18 degrees of freedom, the p-value is 0
```







(a) Write out the standard linear regression model for this problem.

(b) Assuming that the model is correct, what is the estimate of the *variance* of the error term, ϵ?

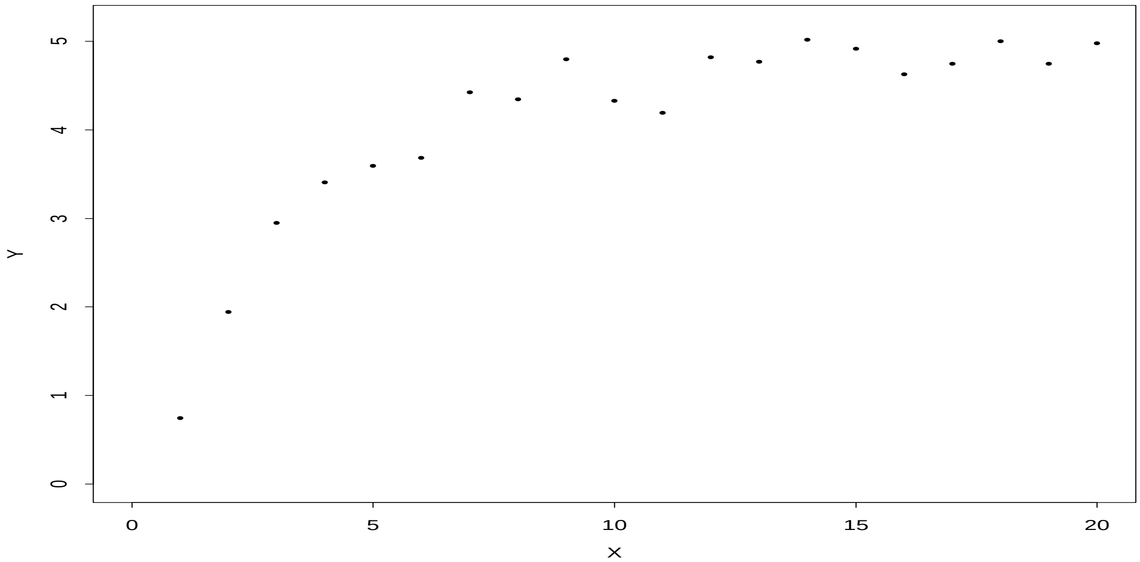
(c) Assuming that the model is correct, interpret the two coefficients.

(d) Assuming that the model is correct, what is the 95% confidence interval on b_0?

(e) To what null hypothesis does the p-value of 0 for the F-statistic refer?

(f) What assumption(s) is/are clearly violated?

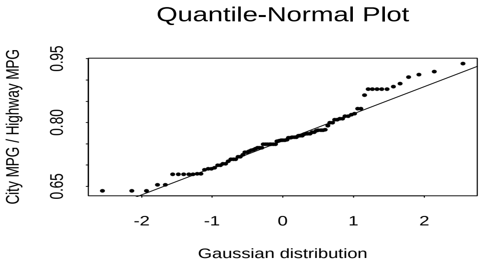
2. (**11 pts**) We want to apply the negative exponential non-linear regression model, $Y_i = \alpha(1 - e^{-\beta X_i}) + \epsilon_i$, to the data plotted below. Make estimates of the starting values. Show your work.

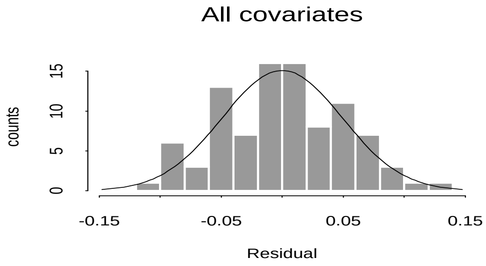


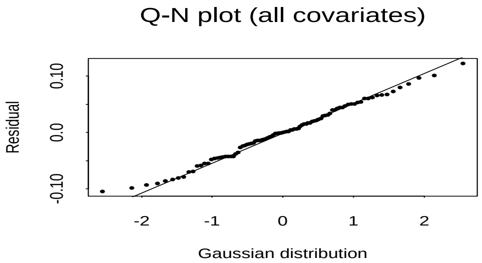
3. (**21 pts, 3 pts each**) This problem addresses the cars dataset. (A review of the variables is on the last page of the homework.) The outcome variable is the drop-off in fuel economy from highway driving to city driving, specifically log(city fuel economy / highway fuel economy).

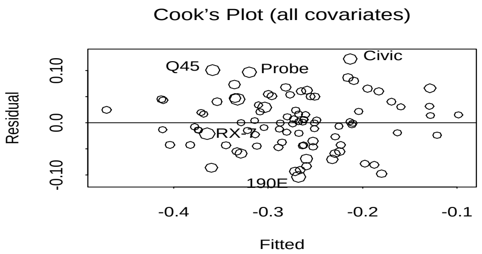
```
cars$LCHRat_log(cars$CitMPG/cars$HwyMPG)
qqn(cars$CitMPG/cars$HwyMPG, ylab="City MPG / Highway MPG")
rslt_lm(LCHRat~MinDol+OptDol+AirBag+DrvTyp+CylCnt+EngLit+HrsPwr+EngRPM+
        EngRev+ManTrx+GasGal+PasCnt+Length+WhlBas+Width+UTurn+LuggX+
        Weight+Domest+Van"), cars)  # first regress on all variables listed
ndhist(rslt$res, nclass=9, xlab="Residual", main="All covariates")
qqn(rslt$res, ylab="Residual", main="Q-N plot (all covariates)")
CookPlot(rslt, id=T, main="Cook's Plot (all covariates)", tcex=0.8)

# Influence statistics: find 5% of cases with highest Cook's distances
cd_influence(rslt)$cooksd
cutoff_4/nrow(cars) # 0.04
round(cd[cd > cutoff],3)
# Seville Stealth Probe Civic   Q45  RX-7  190E  300E Bonneville Corrado
#   0.045   0.088 0.071 0.064 0.074 0.089 0.072 0.072      0.045   0.048
```





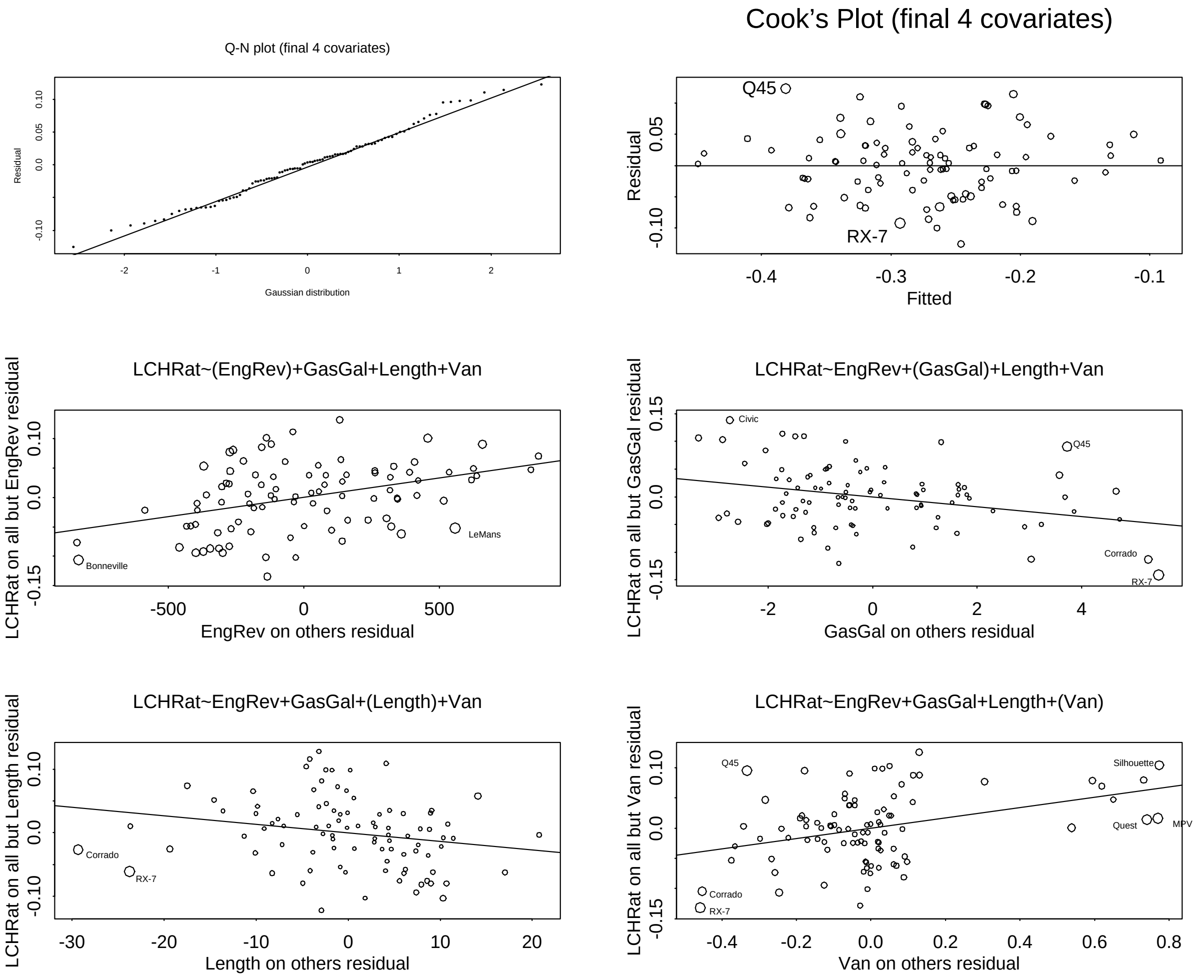





```
sw_stepwise(cars[,xlst], cars$LCHRat, nbest=5, plot=F, meth="exhaustive")
sum.step(sw, cars, "LCHRat", includeR2a=F, top=5, compact=T)
#         BIC EngRev GasGal PasCnt Length Width Van
# 4(#1) 130.4      1      1      0      1     0   1
# 3(#1) 130.4      1      1      0      0     0   1
# 4(#2) 130.3      1      1      0      0     1   1
# 4(#3) 130.2      1      1      1      1     0   0
# 3(#2) 130.2      1      0      1      1     0   0

rslt4_lm(LCHRat ~ EngRev+GasGal+Length+Van, cars)
```

```
summary(rslt4)
# Coefficients: Value Std. Error t value Pr(>|t|)
# (Intercept) -0.0412  0.1274    -0.3235  0.7470
#      EngRev  0.0001  0.0000     4.0453  0.0001
#      GasGal -0.0089  0.0029    -3.0602  0.0029
#      Length -0.0013  0.0006    -2.0933  0.0392
#         Van  0.0851  0.0226     3.7594  0.0003
#
# Residual standard error: 0.05416 on 88 degrees of freedom
# Multiple R-Squared: 0.6327
# F-statistic: 37.9 on 4 and 88 degrees of freedom, the p-value is 0
```



```
# Find 5% of cases with highest Cook's distance
inf_influence(rslt4)
CDcutoff_2*length(rslt4$coef)/nrow(cars) # 0.108
inf$cooksd[inf$cooksd > quantile(inf$cooksd,0.95)]
#       Astro      Civic        Q45       RX-7    Corrado
#  0.04700834 0.03742578 0.07926122 0.08896066 0.05900431

# Find all cases with at least one DFBETA greater than 2/sqrt(n)
DFcutoff_2/sqrt(nrow(cars))  # 0.207
DFB_inf$dfbeta[,2:5]
mx_apply(abs(DFB),1,max)
round(DFB[mx>=DFcutoff,],3)
#          EngRev GasGal Length    Van                EngRev GasGal Length    Van
#    Astro  0.018  0.286 -0.141  0.146          RX-7  0.083 -0.575  0.555  0.376
# Imperial -0.102 -0.272  0.255  0.083         Capri -0.170 -0.236 -0.052  0.073
#  Festiva -0.055 -0.006 -0.214 -0.019         Quest  0.000  0.084 -0.076 -0.317
# Aerostar  0.056 -0.062  0.115 -0.215    Silhouette -0.074 -0.107  0.058  0.264
#    Civic -0.027 -0.329  0.110  0.122        LeMans -0.294  0.129 -0.215 -0.064
#      Q45  0.099  0.500 -0.095 -0.349    Bonneville  0.263 -0.091  0.243  0.113
#    SC300  0.192  0.267 -0.019 -0.166        Passat -0.181 -0.271  0.098  0.172
#      MPV -0.051  0.096 -0.114 -0.335       Corrado  0.135 -0.404  0.500  0.271

# Details on the RX-7 model
cars[row.names(cars)=="RX-7",]
#      Manuf   Size MinDol CitMPG HwyMPG AirBag DrvTyp CylCnt EngLit HrsPwr
# RX-7 Mazda Sporty   32.5     17     25      1      0      0    1.3    255
#      EngRPM EngRev ManTrx GasGal PasCnt Length WhlBas Width UTurn Weight
# RX-7   6500   2325      1     20      2    169     96    69    37   2895
#      Domest OptDol Van LuggX      LCHRat
# RX-7      0      0   0     0  -0.3856625

# Means of 4 covariates
round(apply(cars[c("EngRev","GasGal","Length","Van")],2,mean),2)
#   EngRev   GasGal   Length    Van
#  2332.20    16.66   183.20   0.10
```

(a) Interpret the Quantile-Normal plot for “City MPG/Highway MPG”.

(b) Clearly and completely state the interpretation of the“Van” coefficient in the final model.

(c) Is the deviation of the residuals from normality in the final model mild, moderate or severe?

(d) Interpret the Cook’s plot for the “final 4 covariates” model (note the Splus code and results relating to the plot).

(e) Summarize the influence analysis for the final, 4 covariate model. Also, explain specifically what the GasGal DFBETA for the RX-7 car means.

(f) For the final model, calculate the fitted value for Highway MPG for a 200 inch long van with 2000 rev/mile, a 20 gallon fuel tank, and a city MPG rating of 15.

(g) Interpret the Van value under "Means of 4 covariates".

Data Details for Problem 2

NAME: 1993 New Car Data
DATAFRAME: cars

DESCRIPTIVE ABSTRACT:
Specifications are given for 93 new car models for the 1993 year. Several measures are given to evaluate price, mpg ratings, engine size, body size, and features.

SOURCES:

- *Consumer Reports: The 1993 Cars - Annual Auto Issue (April 1993), Yonkers, NY: Consumers Union.*
- PACE New Car & Truck 1993 Buying Guide (1993), Milwaukee, WI: Pace Publications Inc.

```
VARIABLE DESCRIPTIONS:
Rownames are Car Models
Manuf    Manufacturer
Size     Small, Sporty, Compact, Midsize, Large
MinDol   Minimum Price (in $1,000) - Price for basic version of this model
OptDol   Options Price (in $1,000) - Difference in Price for basic and premium versions
CitMPG   City MPG (miles per gallon by EPA rating)
HwyMPG   Highway MPG
AigBag   Air Bags standard: 0=none, 1=driver only, 2=driver & passenger
DrvTyp   Drive train type:  0=rear wheel drive, 1=front wheel drive, 2=all wheel drive
CylCnt   Number of cylinders  (zero for Mazda rotary engine)
EngLit   Engine size (liters)
HrsPwr   Horsepower (maximum)
EngRPM   RPM (revs per minute at maximum horsepower)
EngRev   Engine revolutions per mile (in highest gear)
ManTrx   Manual transmission available: 0=No, 1=Yes
GasGal   Fuel tank capacity (gallons)
PasCnt   Passenger capacity (persons)
Length   Length    (inches)
WhlBas   Wheelbase (inches)
Width    Width     (inches)
UTurn    U-turn space (feet)
LuggX    Luggage capacity (cu. ft.)  (zero for vans and Corvette)
Weight   Weight (pounds)
Domest   Domestic?  0=non-U.S. manufacturer, 1=U.S. manufacturer
Van      0=not a van, 1=van
```