also rob stat ... with
#11 from M595 Exam 2

needs F table

583
...n 490 Final, Fall 2011 Name___________

1) Consider the following data set: votes for preseason 1A basketball poll Nov. 22, 2011, WSIL News.

111 89 778 78 76

a) Find the sample median MED(n).

76 78 89 111 778

$MED(n) = \boxed{89}$

b) Find the sample median absolute deviation MAD(n).

$-13 \quad -11 \quad 0 \quad 22 \quad 689 = r_i = Y_i - MED(n)$

$0 \quad 11 \quad 13 \quad 22 \quad 689 \quad |Y_i - MED(n)|$

$\boxed{MAD(n) = 13}$

m480
and 18

c) Find shorth(3).

$89 - 76 = 13$

$111 - 78 = 33$

$778 - 89 = 689$

$\boxed{\text{shorth}(3) = [76, 89]}$

Parts d)-f) refer to the CI based on MED(n).

d) Find L_n. $= \lfloor \frac{n}{2} \rfloor - \lceil \sqrt{\frac{n}{4}} \rceil = \lfloor \frac{5}{2} \rfloor - \lceil \sqrt{\frac{5}{4}} \rceil$

$= \lfloor 2.5 \rfloor - \lceil 1.118 \rceil = 2 - 2 = \boxed{0}$

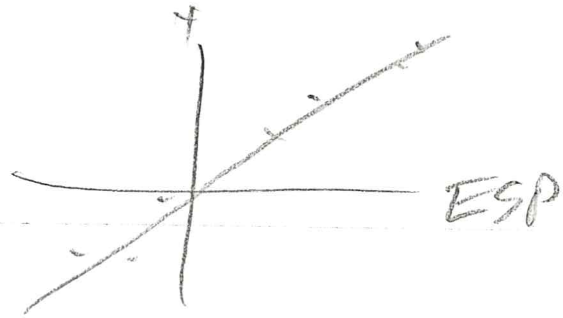
e) Find U_n. $= n - L_n = \boxed{5}$

f) Find SE(MED(n)). $0.5 \, (Y_{(U_n)} - Y_{(L_n+1)}) = .5\,(778 - 76)$

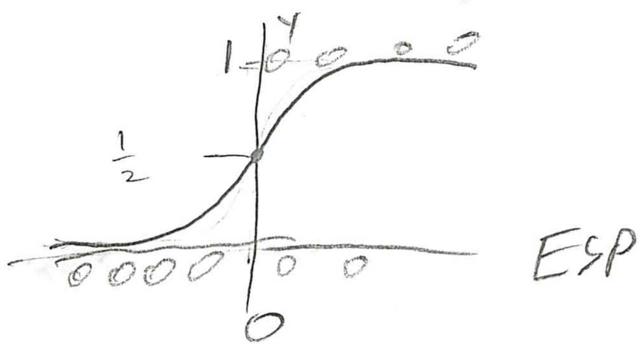
$= \frac{702}{2} = \boxed{351}$

78

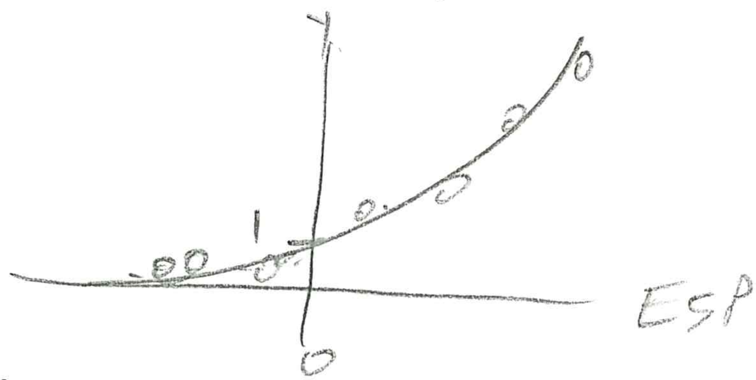
2) a) Sketch a response plot for multiple linear regression.



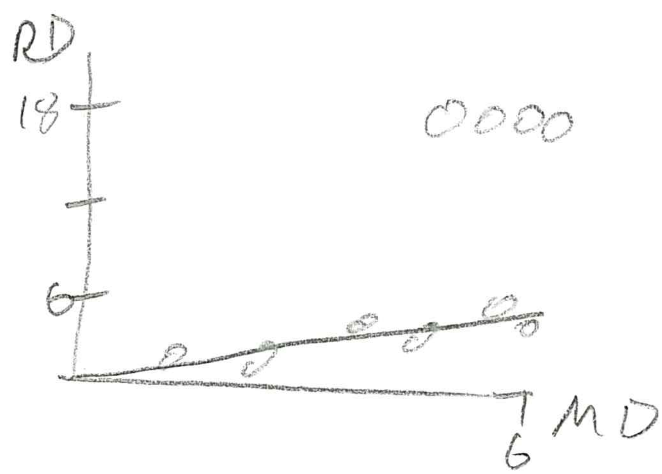
b) Sketch a response plot for binary logistic regression.



c) Sketch a response plot for Poisson regression.



d) Sketch a DD plot if outliers are present.



52

```
Base terms: ({F}LOC TYP)
            df   Deviance Pearson X2              |  k    AIC
a) Add: AGE  195    141.873  187.84               |  5  151.873

Base terms: ({F}LOC TYP AGE)
            df   Deviance  Pearson X2             |  k    AIC
b) Add: CAN  194    134.595  170.367              |  6  146.595

Base terms: ({F}LOC TYP AGE CAN)
          df  Deviance Pearson X2                 |  k    AIC
c) Add: SYS 193  128.441       179.753            |  7  142.441

Base terms: ({F}LOC TYP AGE CAN SYS)
          df   Deviance  Pearson X2               |  k    AIC
d) Add: PCO 192    126.572  186.71                |  8  142.572

Base terms: ({F}LOC TYP AGE CAN SYS PCO)
          df   Deviance      Pearson X2           |  k    AIC
e) Add: PH  191    123.285       191.264          |  9  141.285

Base terms: ({F}LOC TYP AGE CAN SYS PCO PH)
                 df   Deviance       Pearson X2   |  k    AIC
f) Add: PRE        190    121.703        203.546  | 10  141.703

Base terms: ({F}LOC TYP AGE CAN SYS PCO PH PRE)
                 df   Deviance       Pearson X2   |  k    AIC
g) Add: CPR        189    120.159        218.587  | 11  142.159

Base terms: ({F}LOC TYP AGE CAN SYS PCO PH PRE CPR)
                 df   Deviance       Pearson X2   |  k    AIC
h) Add: SEX        188    118.464         216.96  | 12  142.464

Base terms: ({F}LOC TYP AGE CAN SYS PCO PH PRE CPR SEX)
                 df   Deviance       Pearson X2   |  k    AIC
i) Add: FRA        187    117.207        215.486  | 13  143.207
```

3) Based on the output above, which model, a)–i), should be the first submodel to be examined. Explain briefly.

ignore

c) $= I_I = I_{min}$ has the fewest predictors with $AIC(I) \le AIC(I_{min}) + 2$

14 (LR but AIC means LR or PR)

Coefficient Estimates	Full Model	Poisson Regression		
Label	Estimate	Std. Error	Est/SE	p-value
Constant	-0.406023	0.877382	-0.463	0.6435
bombload	0.165425	0.0675296	2.450	0.0143
exper	-0.0135223	0.00827920	-1.633	0.1024
type	0.568773	0.504297	1.128	0.2594

Scale factor: 1 Number of cases: 30 Degrees of freedom: 26
Pearson X2: 23.928 Deviance: 25.953

Coefficient Estimates	Reduced Model	Poisson Regression		
Label	Estimate	Std. Error	Est/SE	p-value
Constant	-1.70097	0.506773	-3.356	0.0008
bombload	0.231122	0.0467598	4.943	0.0000

Scale factor: 1 Number of cases: 30 Degrees of freedom: 28
Pearson X2: 30.571 Deviance: 29.206

rastat
[illegible]

4) The response Y is the *number* of locations where aircraft was damaged. The output is from a Poisson regression. The variable *exper* = total months of aircrew experience while *type* of aircraft was coded as 0 or 1.

a) Using the reduced model, find $\hat{\mu}(x)$ if x = bombload = 8.

$ESP = \alpha + \beta x = -1.701 + .231(8) = 0.148$

$\hat{\mu} = e^{ESP} = e^{0.148} = \boxed{1.1595}$

b) There were $n = 30$ cases. Perform the 4 step change in deviance test to examine whether the reduced model can be used.

ignore

i) H_0 reduced model is good H_A use full model

ii) $G^2(R|F) = 29.206 - 25.953 = 3.253$

iii) $df = 28 - 26 = 2$

df	.25	.10
2	2.77	4.61

$.1 < pval < .25$

iv) fail to reject H_0, the reduced model is good

26

```
Response = y = low, Sequential Analysis of Deviance
All fits include an intercept.
                      Total          |  Change
Predictor       df    Deviance       |  df    Deviance
Ones            188   234.672        |
ht              187   230.650        |   1    4.02213
lwt             186   221.142        |   1    9.50777
ptl             185   215.964        |   1    5.17829
{F}race         183   210.850        |   2    5.11341
smoke           182   204.898        |   1    5.95270
```

5) Consider a study on whether the birthweight of a newborn baby is low or normal. Suppose that the response variable Y = low (0 if normal > 2500g, 1 if low birth weight < 2500 g). Predictors are ht = history of hypertension (0 no, 1 yes), lwt = weight of mother at last menstrual period, ptl = history of premature labor (0 none, 1 one, 2 two, etc), factor race (1 white, 2 black, 3 other), and smoke during pregnancy (0 no, 1 yes). Use the above output to perform a 4 step deviance test.

i) $H_0\ \underline{\beta} = \underline{0}$ $\quad H_A\ \underline{\beta} \neq \underline{0}$

ignore

ii) $G^2(O|F) = 234.672 - 204.898 = 29.774$

iii) $df = 188 - 182 = 6$ $\quad$ .001 / 22.46

pval = 0 < 0.001

iv) reject Ho there is a LR relationship between the response low and the predictors ht, ..., smoke

T3

6) The table $\boldsymbol{W}$ shown below represents 4 measurements on 5 people. Find the coordinatewise median MED($\boldsymbol{W}$)

E1d20

age	breadth	cephalic	size
39.00	149.5	81.9	3738
35.00	152.5	75.9	4261
35.00	145.5	75.4	3777
19.00	146.0	78.1	3904
0.06	88.5	77.6	933

.06	19	35	35	39
88.5	145.5	146	149.5	152.5
75.4	75.9	77.6	78.1	81.9
933	3738	3777	3904	4261

T3

Label	Estimate	Std. Error	t-value	p-value
Constant	0.311427	0.205094	1.518	0.1512
BodyWt	-0.00778306	0.0187168	-0.416	0.6838
Dose	1.48488	3.71306	0.400	0.6953
LiverWt	0.00898934	0.0186586	0.482	0.6374

Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Regression	3	0.00184396	0.000614652	0.10	0.9585
Residual	14	0.0857172	0.00612265		

7) The output above was collected in an experiment in which 18 rats were injected with a *dose* of a drug approximately proportional to *body weight*. At the end of the experiment, the animal's *liver weight* was found, and the fraction of the *drug recovered* in the liver was recorded.

a) Predict Y if $BodWt = 171$, $Dose = 0.9$ and $LiverWt = 8$.

Q6d 20

$\hat{Y} = ESP = \hat{\beta}_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 = 0.311427 - .00778306(171)$
$+ 1.48488(0.9) + .00898934(8) =$

$\boxed{0.3888}$

b) Perform a 4 step ANOVA F test.

ignore

i) Ho $\beta_2 = \beta_3 = \beta_4 = 0$ HA not Ho

ii) $F_0 = 0.10$

iii) pval = 0.9585

iv) Fail to reject Ho, There is not an MLR relationship between the response drug recovered and the predictors body weight, dose and liver weight.

26

	B1	B2	B3	B4
df	945	965	968	976
# of predictors	55	35	32	24
# with $0.01 \leq$ Wald p-value ≤ 0.05	5	4	3	0
# with Wald p-value > 0.05	8	0	0	0
G^2	892.957	922.212	929.808	959.190
AIC	1002.957	992.212	993.808	1007.190
corr(B1:ETA'U,Bi:ETA'U)	1.0	0.958	0.947	0.893
p-value for change in deviance test	1.0	0.083	0.034	0.0002

8) The above table gives summary statistics for 4 models considered as final submodels after performing variable selection. The response was binary (700 ones and 300 zeroes) and logistic regression was used. The response plot for the full model B1 was good. The minimum AIC model had AIC = 990.134. Many of the predictors were factors, and the factor was counted as a predictor with a bad pvalue if all of the predictors in the factor had bad pvalues. Which two models are the best candidates for the final submodel? Explain briefly why each of the other 2 submodels should not be used.

ignore

B2 and B3

$\frac{300}{10} = 30$

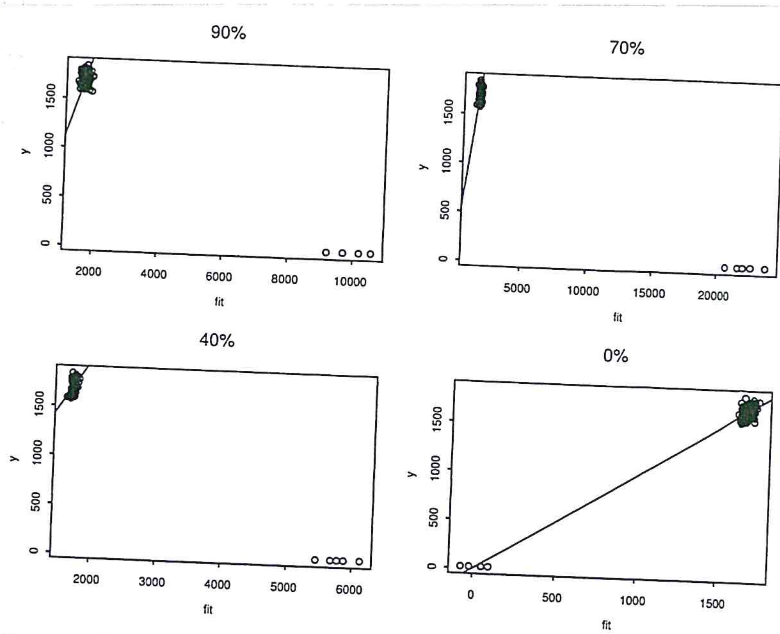
B1 has too many predictors

B4 has AIC > AIC(Imin)+7 = 997.134 and pval is too low.

T3

9) From the plot shown, which trimmed views estimator is the least outlier resistant: 90%, 70%, 40% or 0%?

R583d70
01007



$\boxed{0\%}$

identity line goes through the others

T3

10) Recall that if $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean $\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Let $\sigma_{12} = \text{Cov(Y, X)}$ and suppose Y and X follow a bivariate normal distribution

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} 15 \\ 20 \end{pmatrix}, \begin{pmatrix} 64 & \sigma_{12} \\ \sigma_{12} & 81 \end{pmatrix} \right).$$

M580
E2512
E1020

a) If $\sigma_{12} = 10$ find $E(Y|X)$.

$$15 + 10 \frac{1}{81} (X-20) = \boxed{12.531 + \frac{10}{81} X}$$

$$= \boxed{12.531 + 0.123 X}$$

b) If $\sigma_{12} = 10$, find $\text{Var}(Y|X)$.

$$64 - 10 \frac{1}{81} 10 = 64 - \frac{100}{81} = \boxed{62.765}$$

c) If $\sigma_{12} = 10$, find $\rho(Y, X)$, the correlation between Y and X.

$$\frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{10}{\sqrt{64}\sqrt{81}} = \boxed{\frac{10}{72} = 0.1389} = \frac{5}{36}$$

d) What is σ_{12} if Y and X are independent?

0

52