~~1a) included in~~ Proc Reg & proc gplot ~~"all pairs of variables"~~

&

1b) proc reg sgscatter · for the scatterplot matrix
* if you look at mpg row, it places the mpg variable on y

1b) proc corr
· computes Pearson correlation coefficients

1c) proc reg; selection cp aic msc
· reference table where
in model = 10
top ten w/ lowest cp are
at top of table

Scatter plot matrix
R pairs dataset "mtcars %>% pairs"

1d) wt qsec am : $R^2 = .85$ MSE = 6.05
cyl hp wt : $R^2 = .84$ MSE = 6.31

1e) $y = wt_{x_1} + qsec_{x_2} + am_{x_3}$

$y = wt_{x_1} + qsec_{x_2} + am_{x_3} + disp_{x_4} + disp^2_{x_5}$

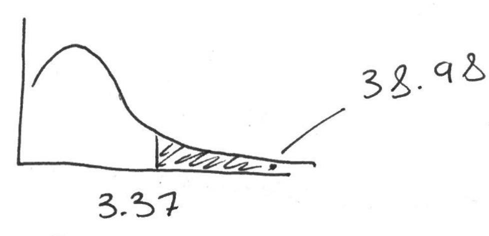
$H_0: disp = disp^2 = \emptyset$

$H_a: disp = disp^2 \neq \emptyset$

$n-p-1$
$32-5-1 = 26$

F = 38.98 qf(.95, 2, 26)

= 3.37 < 38.98



Reject H_0, at $\alpha = .05$ level

p-val: <.0001

f) MSE = 5.097 $R^2 = .88$ | $y = wt_x + qsec_{x_2} + am_{x_3} + disp_{x_4} + disp^2_{x_5}$

I believe model $y = wt_{x_1} + qsec_{x_2} + am_{x_3}$ is still superior since
TL;DR less variables and marginal variability gained.

1g) $\alpha = .10$

$y = wt x_1 + qsec x_2 + disp x_4 + disp2 x_5$

wt	.0002	< .1
qsec	.0809	< .1
~~am~~	~~.76~~	~~> .1~~
disp	.04	< .1
disp2	.01	< .1

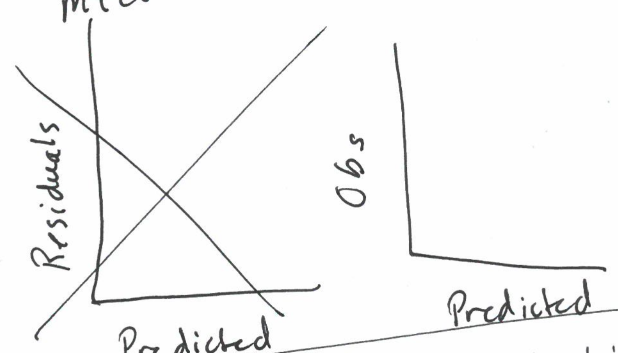
h) Mtcars.sas

i.
```
proc gplot data=residz;
  title "residuals vs fitted"
  plot r*p
run;
```

prob1-hw4-plot.sas

ii.
```
proc sort data=residz;
    by r;
run;
data residz;
    set residz;
    qnorm=quantile("norm", _n_/33);
run;
proc gplot;
    plot qnorm*r;
run;
```

iii. obs vs. fitted included in Model3 of mtcars.sas as



R^2 is multiple Coefficient of Determination = .88

1i)

		MS(E)	R^2
cyl	Model 1	6.31	~~.83~~ .84
wt am qsec	Model 2	~~5.39~~ 6.05	~~.88~~ .85
wt qsec	Model 3	4.93	.88

2) $n = 14$ trt/p = ~~#~~2 each got 7

a) $t = \dfrac{(\bar{y}_1 - \bar{y}_2) - \mu_0}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$ where $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

t dist w/ dv $v = n_1 + n_2 - 2$

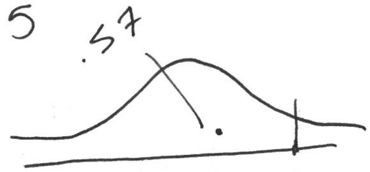
$\dfrac{(7.5 - 7.3) - 0}{\sqrt{\dfrac{.425}{7} + \dfrac{.425}{7}}}$

$s_p^2 = \dfrac{(7-1)(.6)^2 + (7-1)(.7)^2}{12}$

$s_p^2 = \dfrac{2.16 + 2.94}{12} = .425$

$\dfrac{.2}{\sqrt{.12}} = \dfrac{.2}{.35}$

$\boxed{= .57}$



qt(.95, 12)
$= 1.78$

$.57 < 1.78$, thus fail to reject H_0. There is no evidence to suggest the difference in means between treatment A & B is statistically sign.

F test $= \dfrac{SSE_f - SSE_R}{MSE}$

alternatively:

Model height = ~~treatment~~

model height = trt ....

$\sigma = \cancel{(4)}^2$ $\sigma^3 = 1.058$ or 1.06 $\sigma^2 = \sqrt{Var}^2$

~~helps on df~~

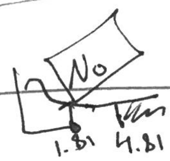
b) $y_3 = 6.3$ $s_3 = .4$ $.4 = \dfrac{\sigma_3}{\sqrt{7}}$ $Var = \sigma^2$

$Var = (1.06)^2 = 1.12$

$SSE = (Var)(n-1)$

$SSE_3 = (1.12)(6) = 6.72$ $\sqrt{s_p^2 = .425} = .65$

$SSE_p = (.425)(\cancel{6}\,14) = 5.95$



$\sigma = .65$

$\dfrac{(6.72 - 5.95)/1}{.425} = 1.81$ $= 4.84$ $MSE = \dfrac{5.95}{14 - \cancel{2}} = .425$

$14-2-1 = 11$

qf(.95, 1, 11) 14-2-1? since pooled?

2b)

Source	df	SS	MS	F
Trt	2	2.42	1.21	3.56
Err	18	6.12	.34	
Tot	20	8.54		

$\bar{y}_3 = 6.7$ $s_3 = .4$

$\bar{y}_2 = 7.3$ $s_2 = .7$

$\bar{y}_1 = 7.5$ $s_1 = .6$

$\bar{y}_{\cdot\cdot} = \frac{21.5}{3} = 7.13$

$\boxed{qf(.95, 3, 18) = 3.15 < 3.56 \text{ not sig}}$

trt df: 3-1

Err: 21-3

Tot: 21-1

$SS(tot) = SS(trt) + SS(E)$

$SS(trt) = \sum_{i=1}^{3}\sum_{j=1}^{7} (\bar{y}_i - \bar{y}_{\cdot\cdot})^2$

$MS(Trt) = 7 * \left[\frac{(\bar{y}_1 - \bar{y}_{\cdot\cdot})^2 + (\bar{y}_2 - \bar{y}_{\cdot\cdot})^2 + (\bar{y}_3 - \bar{y}_{\cdot\cdot})^2}{2}\right]$

$= 7 * \left[\frac{(7.5 - 7.13)^2 + (7.3 - 7.13)^2 + (6.7 - 7.13)^2}{2}\right]$

$= 7 * \left[\frac{(.367)^2 + (.067)^2 + (-.43)^2}{2}\right]$

$= 7[.173] = 1.21$

$MS(E) = \frac{\sum_{i=1}^{3}\sum_{j=1}^{7} (Y_{ij} - \bar{Y}_i)^2}{(21-3) \text{ or } (6 \times 3)}$

$MS(E) = \frac{.36 + .49 + .16}{3} = .34$

$MS(Trt) = \frac{SS(Trt)}{3-1} = \frac{1.21}{1}$

$SS(Trt) = 2.42$

$MS(E) = \frac{SSE}{n-t}$

$.34 = \frac{SSE}{18} = 6.12$