M48418

Outliers and for checking if a nonlinearity is monotone (so transformable to linearity, see ch 3).

Residual plots are good for visualizing $e|\underset{\sim}{x}^T\underset{\sim}{\beta}$, for checking lack of fit, nonlinearity, nonconstant variance, whether $x_j \perp\!\!\!\perp e$, if x_j^2 is needed in the model and if a predictor w_j (and w_j^2) should be added to the model. Also for checking whether the error dist is highly skewed.

81) p28 The $E(e) = 0$ assumption is free if a constant is in the model and $V(e) = \sigma^2$.

$$Y = \tilde{\beta}_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \tilde{e} \underbrace{- E\tilde{e} + E\tilde{e}}_{0}$$

$$= \beta_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e$$

where $\beta_1 = \tilde{\beta}_1 + E\tilde{e}$ and $e = \tilde{e} - E\tilde{e}$.

82) $$r_i = Y_i - \hat{Y}_i = \overbrace{Y_i - EY_i}^{e_i} + EY_i - \hat{Y}_i$$
$$\approx e_i + N(0, MSE\, \underset{\substack{\uparrow \\ \text{leverage}}}{h_i}).$$

So residuals can look normal if the 2nd term dominates. QQ plots are not very good, but the assumption of normality is not very important except for prediction intervals.

§ 2.5 83) p38 Often inference for a new or future vector of predictors $\underset{\sim}{x}_f$

is wanted.

$Y = \underline{x}^T \underline{\beta} + e$ where $E(e) = 0$. $e \perp\!\!\!\perp \underline{x}$

The regression function

$\overset{E(Y|x^T\beta)}{E(Y)} \equiv E(Y|\underline{x}) = \underline{x}^T\underline{\beta} = x_1\beta_1 + \cdots + x_p\beta_p$
is a hyperplane.

84] p38 The point estimator of Y_f given $\underline{x} = \underline{x}_f$ is $\hat{Y}_f = \underline{x}_f^T \hat{\underline{\beta}}$.

The point estimator of $E(Y_f) \equiv E(Y_f|\underline{x}_f)$ is also $\hat{Y}_f = \underline{x}_f^T \hat{\underline{\beta}}$.

85) p38 Know for final: Suppose $x_1 \equiv 1$.
A large sample $100(1-\delta)\%$ CI for $E(Y_f|\underline{x}_f) = \underline{x}_f^T\underline{\beta}$ is

$\hat{Y}_f \pm t_{n-p,1-\frac{\delta}{2}} SE(\hat{Y}_f)$ where

$SE(\hat{Y}_f)$ is given by output.

86] Usually use 90, 95, 99% CI's.
use $t_{n-p,1-\frac{\delta}{2}}$ "=" $z_{1-\frac{\delta}{2}}$ if $n-p>30$.

87) p38 If nominal coverage is $1-\delta$, a large sample CI has coverage $1-\delta_n$ where $1-\delta_n \underset{\nwarrow \text{ gets close to}}{\xrightarrow{P}} 1-\delta$ as $n \to \infty$.

know for final

88] p39 Suppose the e_i are iid $N(0,\sigma^2)$.

A large sample prediction interval (PI) for the random variable Y_f is $[\hat{L}_n, \hat{U}_n]$ where $P(\hat{L}_n \le Y_f \le \hat{U}_n) \to 1-\delta$ as $n \to \infty$. If $Y_1, \ldots, Y_n, Y_f$ are ind with $Y = \underset{\sim}{x}^T\beta + e$, a $100(1-\delta)\%$ PI for Y_f is

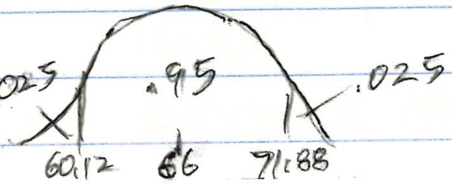
$$\hat{Y}_f \pm t_{n-p,1-\delta/2}\ se(pred) \quad \text{where}$$

se(pred) is given by output. See ex2.9 on p42. (I do an ex later)

89] A large sample $100(1-\delta)\%$ CI is for a parameter and its length $\to 0$ as $n \to \infty$. A large sample $100(1-\delta)\%$ PI is for a random variable Y_f and its length converges to $J = U - L$ where $P(L \le Y_f \le U) = 1-\delta$.

ex) Suppose heights of young women are $N(\mu=66, \sigma=3)$. Then a 95% CI for μ estimates $\mu = 66$ with $\bar{Y} \pm 1.96 \frac{S}{\sqrt{n}}$ for large n.

A 95% PI estimates the interval $[\mu - 1.96\sigma, \mu + 1.96\sigma]$.



19.5

90] The CI for $E(Y_f | \underset{\sim}{x}_f)$ is good for large n if $x_i = 1$ and $V(e) = \sigma^2$.

The PI in point 88] is only good if $e_i \overset{iid}{\sim} N(0, \sigma^2)$.

cheb The 95% PI has coverage at least 73.9% and has coverage near 0.95 for many distributions.

91] *p 37 $\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{e}$, $H = X (X^T X)^{-1} X^T$

The ith diagonal element of H is the ith leverage $h_i = H_{ii} = h_{ii} = \underset{\sim}{x}_i^T (X^T X)^{-1} \underset{\sim}{x}_i$. The leverage of a future obs $\underline{x}_f$ is

$h_f = \underset{\sim}{x}_f^T (X^T X)^{-1} \underline{x}_f$. If

$h_f > \max(h_1, \ldots, h_n)$, then extrapolation occurs: $\underline{x}_f$ is too far from the collected data $(Y_1, \underline{x}_1), \ldots, (Y_n, \underline{x}_n)$ for trustworthy prediction.

$\hat{Y}_f$ Arc: Prediction $SE(\hat{Y}_f) = \sqrt{MSE \; h_f}$, $SE(pred) = \sqrt{MSE(1 + h_f)}$

s $s(pred)$ See HW4 A

92] Other prediction intervals that

work if $x_1 = 1$ and $V(e) = \sigma^2$ are simulated in HW 4 #. 100 data sets are generated along with 100 future values x_f and Y_f. Simulated coverage is proportion of times Y_f is in PI $= \frac{x}{100}$ where $x \approx \text{bin}(100, 1-\delta_n)$ where $1-\delta_n \approx 0.95$. So want coverages near 0.95 (greater than 0.9, say).

§2.9] 93] Simple linear regression (SLR) is the special case of MLR where $p=2$ and $Y = \beta_1 + \beta_2 x + e$ (so $x_2 \equiv x$). Then $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 x$ is a line, a scatterplot of x vs Y can be made with the OLS line added, and there are formulas for finding OLS by hand.

p57-8

94] know for E2

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} = \hat{\rho} \frac{S_Y}{S_x} = \frac{\widehat{\text{cov}}(x,Y)}{S_x^2}$$

where $S_Y^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$, $S_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

are the sample variances of Y and x, and

$$\hat{\rho}(x,Y) = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(Y_i - \bar{Y})}{S_x S_Y}$$

is the sample correlation of x and Y.

Then $\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$. 20.5

9.5] Know for E2 predict Y from X given table of means, SDs and a correlation, figure out which variable is Y and which is x and find $\hat{\beta}_2, \hat{\beta}_1$ and $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$.

ex] see Hw4 B)

$Y = \log_{10}(\text{pressure})$, $X = \text{temp}$, $n = 17$

$x_f = 200$

$\hat{\beta}_1 = -0.421642$, $\hat{\beta}_2 = 0.0089562$

$se(\text{pred}) = 0.0039317$

$SE(\hat{Y}_f) = 0.00104011$

a) Find $\hat{Y}_f$.

b) Find a 95% CI for $E(Y_f / x_f) = \beta_1 + \beta_2 x_f$.

c) Find a 95% PI for Y_f.

Soln] $\hat{Y}_f = \hat{\beta}_1 + \hat{\beta}_2 x_f = -0.421642 + 0.0089562(200)$

$= 1.369598$

b) $df = n - p = 17 - 2 =$

r	
15	2.131
	95%

95% CI for $E(Y_f / x_f)$ is $\hat{Y}_f \pm t_{15, .975} \, SE(\hat{Y}_f)$

$= 1.369598 \pm 2.131 (0.00104011)$

$= [1.3674, 1.3718]$

c) 95% PI for Y_f is $\hat{Y}_f \pm t_{15, .975} \, SE(\text{pred}) =$

$1.369598 \pm 2.131(0.0039317) = [1.3612, \ 1.3780]$

ex) see HW 4D, E Given small data set, find LS line. Predict Y = bloodpressure from x = age

	age x_i	bloodpressure y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
	35	114	-20	-27	540	400
n=5	45	124	-10	-17	170	100
p=2	55	143	0	2	0	0
	65	158	10	17	170	100
	75	166	20	25	500	400
sums	275	705	0	0	1380	1000
	$=\Sigma x_i$	$=\Sigma y_i$	$=\Sigma(x_i-\bar{x})$	$=\Sigma(y_i-\bar{Y})$	$=\Sigma(x_i-\bar{x})(y_i-\bar{Y})$	$=\Sigma(x_i-\bar{x})^2$

$\bar{x} = \frac{275}{5} = 55$ $\bar{Y} = \frac{705}{5} = 141$ } always true except for rounding

$$\hat{\beta}_2 = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{Y})}{\Sigma(x_i-\bar{x})^2} = \frac{1380}{1000} = 1.380$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x} = 141 - 1.380(55) = 65.100.$$

So the $\overset{OLS}{LS}$ line (eq) is $\hat{Y} = 65.100 + 1.380\, x$.

ex) See HW 4 C Men 18-24 predict wt from ht

	mean	sd
wt	162 lb	30 lb
ht	70 in	3 in

$\hat{\rho} = 0.47$

min ht = 57, max ht = 79

soln $\bar{Y} = 162$ $s_Y = 30$

$\bar{x} = 70$ $s_x = 3$

$$\hat{\beta}_2 = \hat{\rho}\frac{s_Y}{s_X} = 0.47\,\frac{30}{3} = 4.7$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x} = 162 - 4.7(70) = -167$$

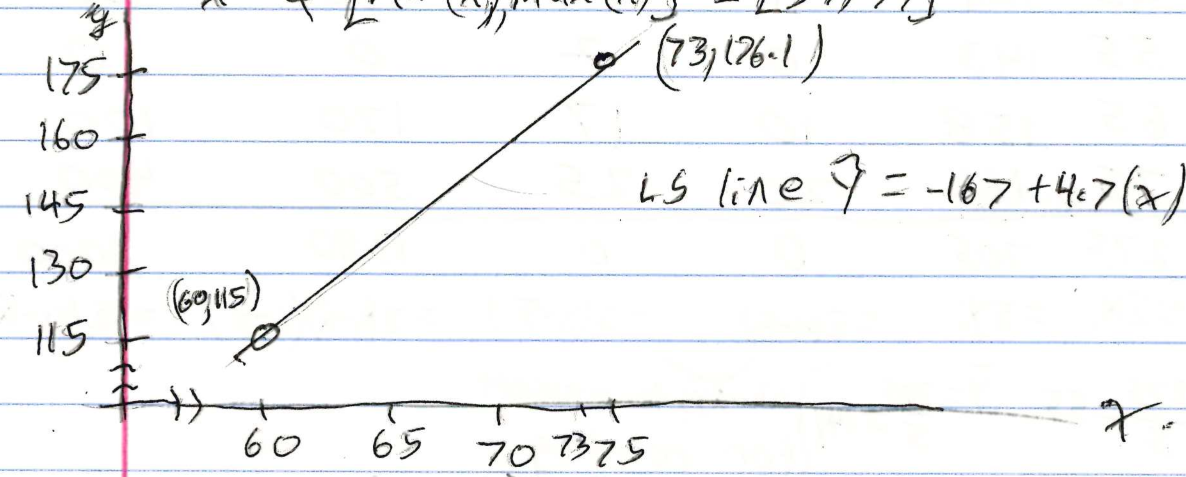
$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 x = -167 + 4.7x$ 21.5

Predict Y if $x = 60$: $\hat{Y} = -167 + 4.7(60) = 115$

$= 73$: $\hat{Y} = -167 + 4.7(73) = 176.1$

$= 1$: $\hat{Y} = -167 + 4.7(1) = -162.3$

For SLR, extrapolation occurs if $x \notin [\min(x_i), \max(x)] = [57, 79]$



§ 2.8 and 2.9 96} Know for final

$Y = \beta_1 + \beta_2 x + e$ with either β_1 or β_2 known, find the LS estimator of $\beta_i \equiv \beta$. Also, given $Q(\eta_1, \eta_2) = \sum_{i=1}^{n} (Y_i - \eta_1 - \eta_2 x_i)^2$, find $E(Y_i) \equiv E(Y_i | x_i) = \beta_1 + \beta_2 x_i$.

If β_1 is unknown, then $\eta_2 = \beta_2$ is known.
by the chain rule

$$\frac{\partial Q}{\partial \eta_1} = -2 \sum_{i=1}^{n} (Y_i - \eta_1 - \eta_2 x_i) \overset{set}{=} 0$$

$$\text{or } 2n\eta_1 - 2\sum_{i=1}^{n}(Y_i - \eta_2 x_i) = 0$$

$$\text{or } \sum_{i=1}^{n}(Y_i - \eta_2 X_i) - n\eta_1 = 0$$

$$\text{or } \underbrace{\hat{\beta}_1}_{\text{call the soln } \hat{\beta}_1} = \frac{\sum_{i=1}^{n}(Y_i - \eta_2 X_i)}{n}$$

$$\frac{\partial Q^2}{\partial^2 \eta_1} = 2n > 0$$ so $\hat{\beta}_1$ is the LS est

If β_2 is unknown, then $\eta_1 = \beta_1$ is known, and by the chain rule

$$\frac{dQ}{d\eta_2} = -2\sum_{i=1}^{n} x_i\,(y_i - \eta_1 - \eta_2 x_i) \overset{set}{=} 0$$

$$\text{or } +2\eta_2 \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i\,(y_i - \eta_1) = 0$$

$$\text{or } \underbrace{\hat{\beta}_2}_{\text{call the soln } \hat{\beta}_2} = \frac{\sum_{i=1}^{n} x_i\,(y_i - \eta_1)}{\sum_{i=1}^{n} x_i^2}$$

$$\frac{d^2Q}{d\eta_2^2} = 2\sum_{i=1}^{n} x_i^2 > 0 \quad \text{so } \hat{\beta}_2 \text{ is the LS est.}$$

ex] See HW4 F, §2.9.1; S02 Final #2

$$Y_i = 10 + 2x_{1i} + \beta_2 x_{2i} + e_i, \quad i = 1, \ldots, n.$$

$$Q(\eta_2) = \sum_{i=1}^{n} (Y_i - 10 - 2x_{1i} - \eta_2 x_{2i})^2.$$

Find LS estimator $\hat{\beta}_2$ of β_2 by setting the first derivative of $\frac{d}{d\eta_2} Q(\eta_2) \overset{set}{=} 0$.

Show $\hat{\beta}_2$ is the global minimizer of $Q(\eta_2)$ and thus the LS estimator of β_2 by showing $\frac{d^2Q}{d\eta_2^2}(\eta_2) > 0$ for all values of η_2.

soln $$\frac{dQ}{d\eta_2} = \sum 2(Y_i - 10 - 2x_{1i} - \eta_2 x_{2i})(-x_{2i}) \overset{set}{=} 0$$

$$\text{or } 2\eta_2 \sum x_{2i}^2 = 2\sum x_{2i}(Y_i - 10 - 2x_{1i})$$

$$\text{or } \hat{\beta}_2 = \frac{\sum x_{2i}(Y_i - 10 - 2x_{1i})}{\sum x_{2i}^2}$$

22.5

$$\frac{d^2 Q}{d\eta_2^2} = \frac{d}{d\eta_2} 2\, \eta_2 \sum x_{2i}^2 = 2 \sum x_{2i}^2 > 0$$

96) p55 shows that if $Q_{OLS}(\underset{\sim}{\eta}) =$

$\sum (Y_i - x_{i1}\eta_1 - \cdots - x_{ip}\eta_p)^2$, then

$$\frac{\partial}{\partial \eta_j} Q_{OLS}(\underset{\sim}{\eta}) = -2 \sum x_{ij} (Y_i - \cdots - x_{ip}\eta_p)$$

$$= -2 (\underset{\text{jth col of } \mathbf{X}}{\underset{\sim}{v}_j})^T (\underset{\sim}{Y} - \mathbf{X}\underset{\sim}{\eta}) \quad \text{for } j = 1, \ldots, p$$

Setting these p equations to 0 gives

the normal eq's: $\mathbf{X}^T \underset{\sim}{Y} - \mathbf{X}^T \mathbf{X} \hat{\underset{\sim}{\beta}} = \underset{\sim}{0}$.

so $\hat{\underset{\sim}{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \underset{\sim}{Y}$

if $\underset{n \times p}{\mathbf{X}}$ has rank p.

§2.10 97] The no intercept model = regression through the origin $Y = \underset{\sim}{x}^T \underset{\sim}{\beta} + e$ does not have a constant in the model. Each predictor, including x_1, takes on at least 2 values. The assumption $E(e) = 0$ is now important.

98] p59-60 Know $\hat{Y} = \underset{\sim}{x}^T \hat{\underset{\sim}{\beta}}$, response & residual plots, wald t and partial F test are "the same"