99] p60 Know for E2 M48423

4 step no intercept Anova F test

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

ii) $F_0 = \frac{MSM}{MSE}$ from output

iii) pval $= P(F_{p,n-p} > F_0)$ from output

iv) pval $< \delta$ reject H_0, there is an MLR relationship between Y and the predictors $x_1, \ldots, x_p$

pval $\geq \delta$ fail to reject H_0, there is not an MLR relationship between Y and the predictors $x_1, \ldots, x_p$.

100] Software default uses a constant, so need to tell software that a constant is not in the model

R: lsfit(x, y, intercept=F)

Arc: uncheck ☑ Fit Intercept box

p60

101]

Source	df	SS	MS	F	pval
model or regression	p	SSM	MSM	$F_0 = MSM/MSE$	for H_0 $\underline{\beta} = \underline{0}$
residual or error	$n-p$	SSE	MSE		

The Anova table for the no intercept Anova F test is "the same" as that for the Anova F test, but in symbols I am using

model p SSM MSM instead of

regression $p-1$ SSR MSR.

23.5

102] If $p=1$, $Y = \beta x + e$ will give a LS line through the origin: $\hat{Y} = \hat{\beta} x$.

103] p39 LS CLT Suppose $\underset{\sim}{Y} = \mathbf{X} \underset{\sim}{\beta} + \underset{\sim}{e}$

e_i iid, $E(e_i) = 0$, $V(e_i) = \sigma^2$

and $\max(h_1, \ldots, h_n) \xrightarrow{P} 0$.

Then $\hat{Y}_i = \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}} \xrightarrow{P} \underset{\sim}{x}_i^T \underset{\sim}{\beta} = E(Y_i | \underset{\sim}{x}_i)$.

$\sqrt{n}\,(\underset{\sim}{a}^T \hat{\underset{\sim}{\beta}} - \underset{\sim}{a}^T \underset{\sim}{\beta})$ is asymptotically normal for any constant $\underset{p \times 1}{\underset{\sim}{a}}$.

LS CLT: Under more conditions, $\sqrt{n}\,(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \xrightarrow{D} N_p(\underset{\sim}{0}, \sigma^2 \mathbf{W})$

where $\dfrac{\mathbf{X}^T\mathbf{X}}{n} \rightarrow \mathbf{W}^{-1}$. So $(\mathbf{X}^T\mathbf{X})^{-1} \simeq n\mathbf{W}$.

So $\hat{\beta}_i \approx N\Big(\beta_i, \underbrace{MSE\,(\mathbf{X}^T\mathbf{X})^{-1}_{ii}}_{[SE(\hat{\beta}_i)]^2}\Big)$

$\hat{Y}_i - EY_i = \underset{\sim}{x}_i^T(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \approx N\Big(0, \underbrace{MSE\ h_i}_{[se(\hat{Y}_i)]^2}\Big)$.

104] p66 Output sometimes has

lines for Anova F test $\Big\{$ Res
Resid

lack of fit $\quad F_{LF} \quad$ pval

pure error

pval $< \delta$ means there is lack of fit

pval $\geq \delta$ means no evidence of lack of fit

Ch 3 Building MLR models

1] p86 If the same data set is used to build the model and to perform inference, it may not work well. Building a model after collecting data is called "data snooping."

2) use data splitting to get valid inference. If possible, spend about 1/8 of budget to collect and build initial MLR model, 1/8 of budget to collect more data to check and change the model, resulting in a tentative MLR model. Then 3/4 of the budget to collect data assuming the tentative model will work.

3] It is ok to look at the predictors, but not the response in that $Y \mid x_2 = a_2, \ldots, x_p = a_p$

$$\stackrel{D}{=} Y \mid t_2(x_2) = t_2(a_2), \ldots, t_p(x_p) = t_p(a_p).$$

ie if $Y \mid x = a \sim N(10, 1)$,

then $Y \mid \log(x) = \log(a) \sim N(10, 1)$.
(assuming logs exist)

4] * One of the most important steps in building an MLR model is removing strong nonlinearities and outliers from the predictors. Transforming highly skewed predictors helps.

p. 92-93

5] Know A power transformation

has the form $x = t_\lambda(w) = \begin{cases} w^\lambda, & \lambda \neq 0 \\ \log(w), & \lambda = 0 \end{cases}$

Often $\lambda \in \Lambda_L = \{-1, -\frac{1}{2}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{2}, 1\}$, the "ladder of powers."

6] A rule of thumb is good advice that does not always work. Read rule of thumb 3.2 on p.102.

7] Know for final ladder rule

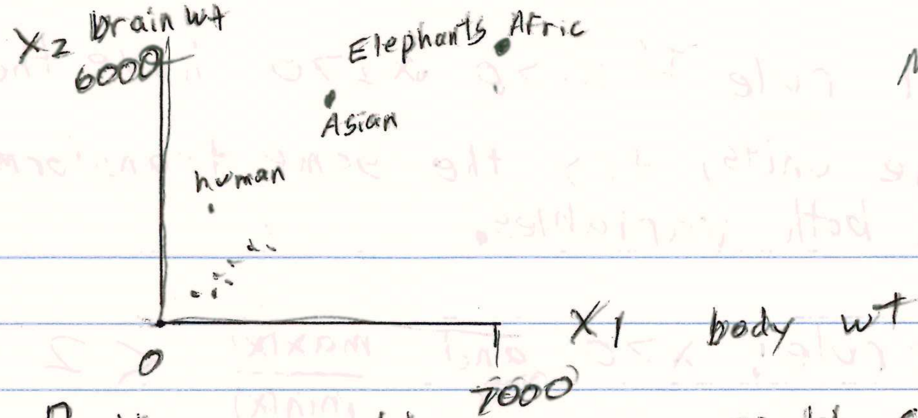
In a plot of x_1 vs x_2,

To spread small values of a variable, make λ smaller.
To spread large values of a variable, make λ larger.

8] Know for final log rule Let $x > 0$. If

$\frac{\max(x)}{\min(x)} > 10$, try $\log(x)$.

ex] Both of these rules are useful if x_1 & x_2 are highly right skewed. See brain data in HW5.

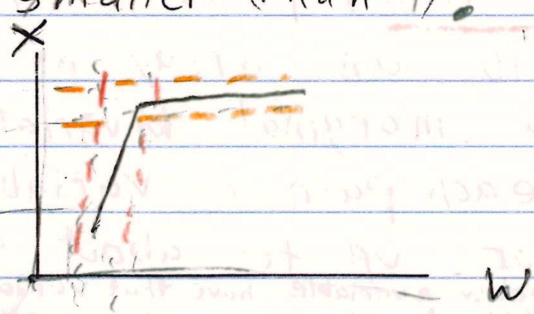


Both variables are right skewed so there are lots of small values and a few large values.

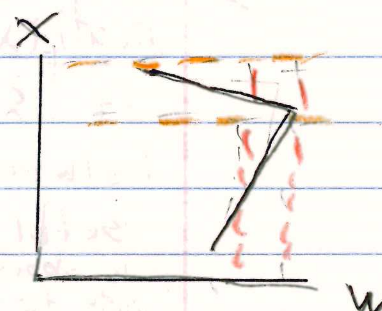
$$\left.\begin{array}{l} \dfrac{\max x_2}{\min x_2} \simeq \dfrac{6654.2}{0.005} > 10 \\[2ex] \dfrac{\max x_1}{\min x_1} \approx \dfrac{5711.9}{0.74} > 10 \end{array}\right\} \text{log rule}$$

ladder rule $\left\{ \begin{array}{l} X_1 = x_1', \quad x_2 = x_2' \quad \text{so make} \\ \lambda \text{ smaller than } 1. \end{array}\right.$

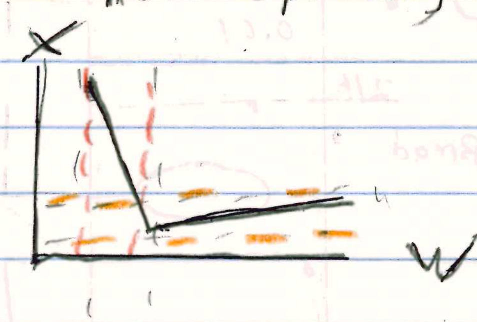
ex) Fig 3.1 p.94



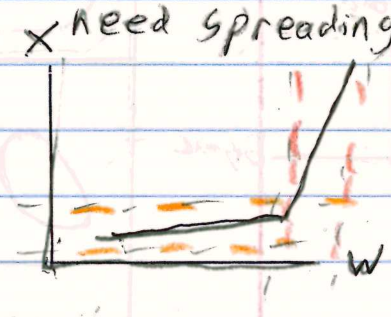
Small values of W large values of X need spreading



large values of both variables need spreading



Small values of both variables need spreading



Small values of X and large values of W need spreading

25.5

9] Unit rule If $x_1 > 0$ $x_2 > 0$ have the same units, try the same transformation on both variables.

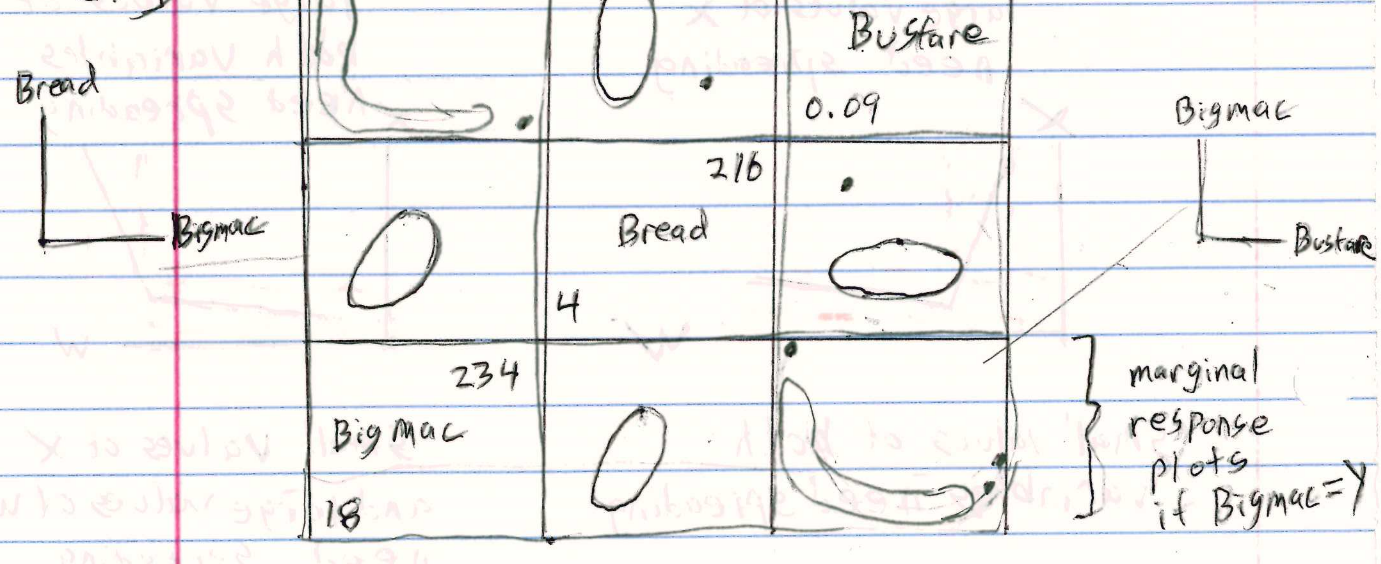
10] range rule: $x > 0$ and $\frac{\max(x)}{\min(x)} < 2$ suggests do not transform x

11] If $x_2 \approx x_1^{\lambda}$ one to one, $x_1^{\lambda} \approx x_2$ and $x_2^{1/\lambda} \approx x_1$

12] Cube root rule: If x is a volume measurement, $x^{1/3}$ may be useful.

13] Know for final: p 87 A scatterplot matrix is an array of scatterplots. It shows marginal bivariate relationships between each pair of variables and is useful for up to about 9 variables.
Plots above and below a variable have that variable on the horiz axis, plots to the left and right of a variable have that variable on the vert axis.

ex]



MH8426

14] variable names on the diagonal, label axes. The numbers in the diagonal are the min and max of the variable (Arc: log rule). The plots above the diagonal are inverses of those below. The marginal response plots correspond to the row containing the response Y: they show $Y | x_i$. Usually put the response 1st or last.

15] Nonlinearities, skewness (and outliers) can often be seen and removed with power transformations.

16] p89 No transformation ($\lambda = 1$), the log transformation ($\lambda = 0$), the square root transformation ($\lambda = \frac{1}{2}$), and the reciprocal transformation ($\lambda = -1$) are used the most.

§3.2 17] p 92 often

$$Y_i = t_{\lambda_0}(Z_i) = E(Y_i) + e_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i$$

18] Assume $Z_i > 0$ for $i = 1, \ldots, n$.

A power transformation $Y = t_\lambda(Z) = \begin{cases} Z^\lambda, & \lambda \neq 0 \\ \log(Z), & \lambda = 0 \end{cases}$

A modified power transformation is

$$Y = t_\lambda(Z) = \begin{cases} \dfrac{Z^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(Z), & \lambda = 0. \end{cases}$$

It can be shown that

$$\lim_{\lambda \to 0} \frac{Z^{\lambda}-1}{\lambda} = \log(Z).$$

Use $\lambda \in \Lambda_L = \{-1, -\frac{1}{2}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{2}, 1\}$.

19] It is a good idea to remove strong nonlinearities from the predictors before making a response transformation. Make a scatterplot matrix of the predictors and the response with the response in the top (or bottom) row. If there are $p-1 > 9$ predictors, use the 1st 9 and Y to make a scatterplot matrix

next 9 Y

⋮

last (up to) 9 and Y

(This gets tedious for more than 90 predictors.)

20] Let $W_i = t_\lambda(Z_i)$ and regress W_i on $\underset{\sim}{x}$ for each of the 7 values of λ in Λ_L. A transformation plot is a plot of $\hat{W}$ vs W with the identity line added as a visual aid.

21] p95 know for final Make the transformation plots for $\lambda \in \Lambda_L$. If the transformation plot looks like a good response plot for λ^*, take $\hat{\lambda}_o = \lambda^*$. That is, use

$Y = t_{\lambda^*}(Z)$ as the response

transformation. Given several transformation plots, you should be able to find the response transformation.

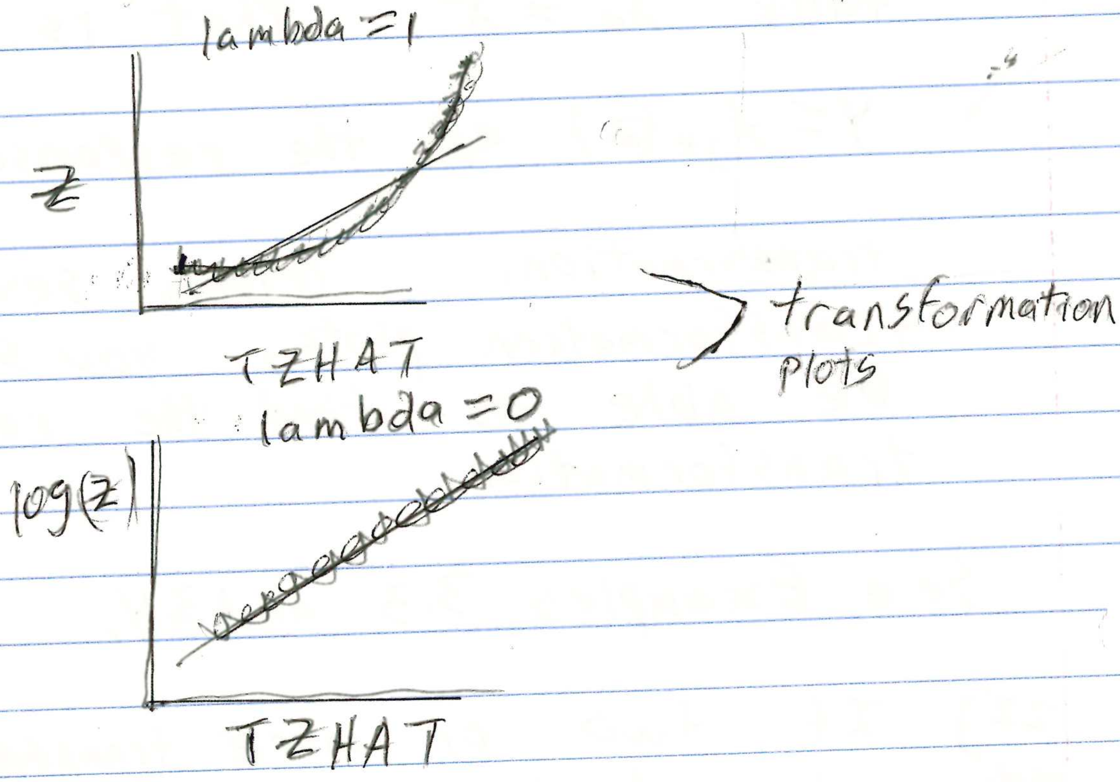
See Examples 3.3 and 3.4.

22] If two or more transformation plots work, look at the "residual plots," take the simplest or most reasonable transformation, or the transformation that makes most sense to subject matter experts.

23] The Box Cox method for finding $\hat{\lambda}$ is a numerical method available from many software packages. Often round $\hat{\lambda}$ to a value in Λ_L. Can also add

$\hat{\lambda}$ to Λ_L. Can add powers like $\pm 1/4$, $\pm 2/3$, ± 2, ± 3 to Λ_L.

ex) Horiz axis is labelled with TZHAT = "fitted values" from regressing $t_\lambda(Z)$ on $\underline{x}$.



Use $Y = \log(Z)$ because its transformation plot looks like a good response plot. (This problem is similar to giving several response plots and choosing the best one, which could appear on exam 2.)

24] p95 After choosing a response transformation to make an MLR model, do the usual checks, including making the response and residual plots.

↑ transformation plot chosen