STAT 231

Nov 23, 2016.

Tutorial: 6 pm: DC 1351.

3-30: Cynthia.

TQ: Next Wednesday

Roadmap

- 5 min recap
- Prediction Interval for Y_{new}, given $X = x_{new}$.
- How to check assumptions for the SLRM.
- Clicker Questions

Y = STAT 231 score

X = STAT 230 score.

Model: $Y_i \sim G(\alpha + \beta x_i, \sigma)$

$\Updownarrow$ $i = 1, \ldots, n$

independent.

$Y_i = \alpha + \beta x_i + R_i$, $R_i \sim G(0, \sigma)$

R_i's independe

- Find the Least Square line; and find $\hat{\sigma}$, s.

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = S_{xy}/S_{xx}.$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}[S_{yy} - \hat{\beta}S_{xy}]}$$

$$s = \sqrt{\frac{1}{n-2}[S_{yy} - \hat{\beta}S_{xy}]}$$

Notes

S = Standard error of the Regression model.

$S = S_e$ (Notation)

(i) We should be able to identify all the estimates from the R-output.

Equivalent Representations

$$\hat{\beta} = S_{xy}/S_{xx} = \frac{\sum(x_i-\bar{x})y_i}{S_{xx}} = \frac{\sum x_i(y_i-\bar{y})}{S_{xx}} \qquad \sum(x_i-\bar{x})(y_i-\bar{y})$$

$$\hat{\sigma}^2 = \frac{1}{n}\left[S_{yy} - \hat{\beta} S_{xy}\right] = \frac{1}{n}\left[\sum(y_i-\hat{\alpha}-\hat{\beta}x_i)^2\right]$$

(i) CI for β

(ii) $H_0: \beta = \beta_0$

(iii) C.I. for $\mu(x)$ when x = given.

$\mu(x) = \alpha + \beta x$ = average of the Y values when $x = x$.

C.I: $\boxed{\hat{\mu}(x) \pm t^* s\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}}$

$\hat{\alpha} + \hat{\beta} x$.

Allows from the result below.

$$\tilde{\mu}(x) \sim G\left(\mu(x), \sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right)$$

(iv) PREDICTION INTERVAL

$\mu_{new} = \mu(x_{new})$.

$x = x_{new} \qquad (x = 75)$.

$Y_{new} = \alpha + \beta x_{new}$

Objective: To construct a 95% P.I for Y_{new} given x_{new}.

$$Y_{new} \sim G(\underbrace{\alpha + \beta x_{new}}_{\mu(x_{new})}, \sigma)$$

$$\tilde{\mu}_{new} \sim G\left(\mu(x_{new}), \sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right)$$

$$Y_{new} - \tilde{\mu}_{new} \sim G\left(0, \sigma\sqrt{1+\frac{1}{n}+\frac{(x_{new}-\bar{x})^2}{S_{xx}}}\right)$$

$$\frac{Y_{new} - \tilde{\mu}_{new}}{\sigma\sqrt{1+\frac{1}{n}+\frac{(x_{new}-\bar{x})^2}{S_{xx}}}} = Z$$

P.Q P.D.

$$\boxed{\frac{Y_{new} - \tilde{\mu}_{new}}{S\sqrt{1+\frac{1}{n}+\frac{(x_{new}-\bar{x})^2}{S_{xx}}}} \sim T_{n-2}}$$

~~Confidence Interval~~

Prediction Interval for Y_{new}.

$$\boxed{(\hat{\alpha} + \hat{\beta} x_{new}) \pm t^{*} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_{new} - \bar{x})^2}{S_{xx}}}}$$

<u>Note's</u>:

(i) Can we find the CI for α ?

Yes, by plugging $x = 0$ in the CI for $\mu(x)$

(ii) Find the CI for σ^2.

$$\boxed{\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}}$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

C.I for σ^2

$$\left[\frac{(n-2)s^2}{b}, \frac{(n-2)s^2}{a}\right]$$

$$\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}$$

χ^2 table with $n-2$ df.

Checking the assumptions for the model.

$$Y_i = \alpha + \beta X_i + R_i$$

$$R_i \sim G(0, \sigma)$$

3 main assumptions

(i) Given x, Y's are Gaussian

(ii) Given x, $E(Y) = \alpha + \beta X$ = linear in X.

(iii) $V(Y) = \sigma^2$, independent of X

Pitfalls

- The tests are subjective. (by visual inspection of graphs)
- Comes with experience.

Some definitions

$\hat{r}_i$ = RESIDUAL $= y_i - \hat{\alpha} - \hat{\beta} x_i$

(thing left over after the line is fitted.)

$\hat{r}_i$s could be thought of as estimates of R_i. So if our model is correct, r_i outcomes from a Gaussian $(0, \sigma)$

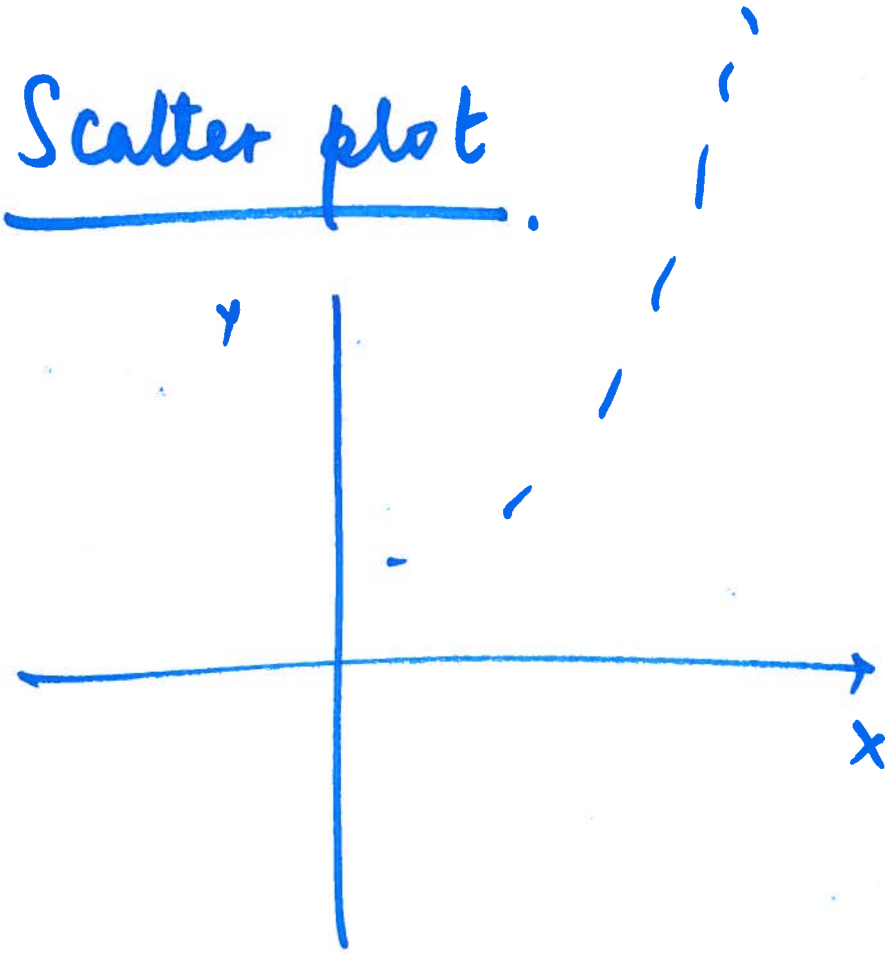
Standardized Residual

$\hat{r}_i^* = \dfrac{\hat{r}_i}{s}$ should act like $G(0,1)$ if the model is correct.

$\dfrac{r_i}{\sigma} \sim Z$

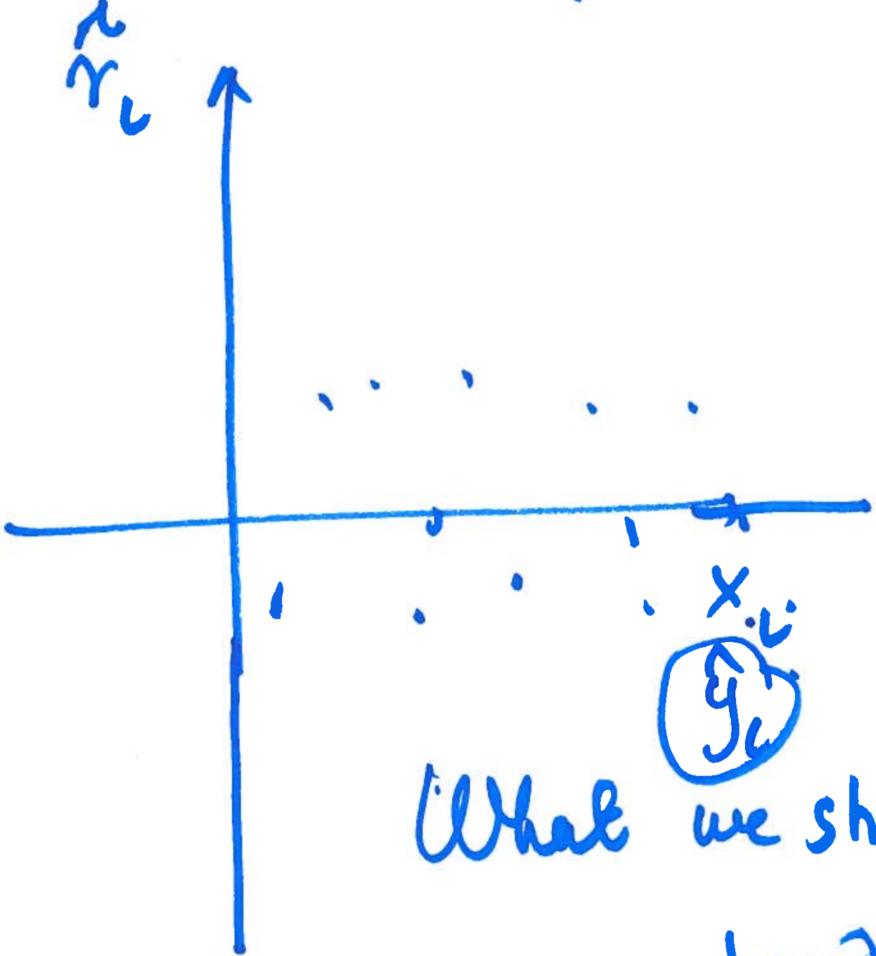
How to check for the model?

(i) Scatter plot



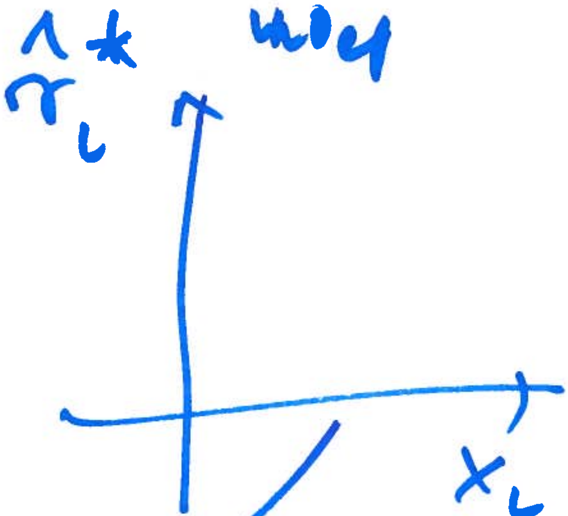
Draw the scatter plot and look for evidence of a linear relationship.

(i) Residual plots.



$\hat{r}_i \sim N(0, \sigma)$

mod



What we should expect

is a band around zero

with no obvious patterns

band between $[-3, 3]$ with

no obvious pattern

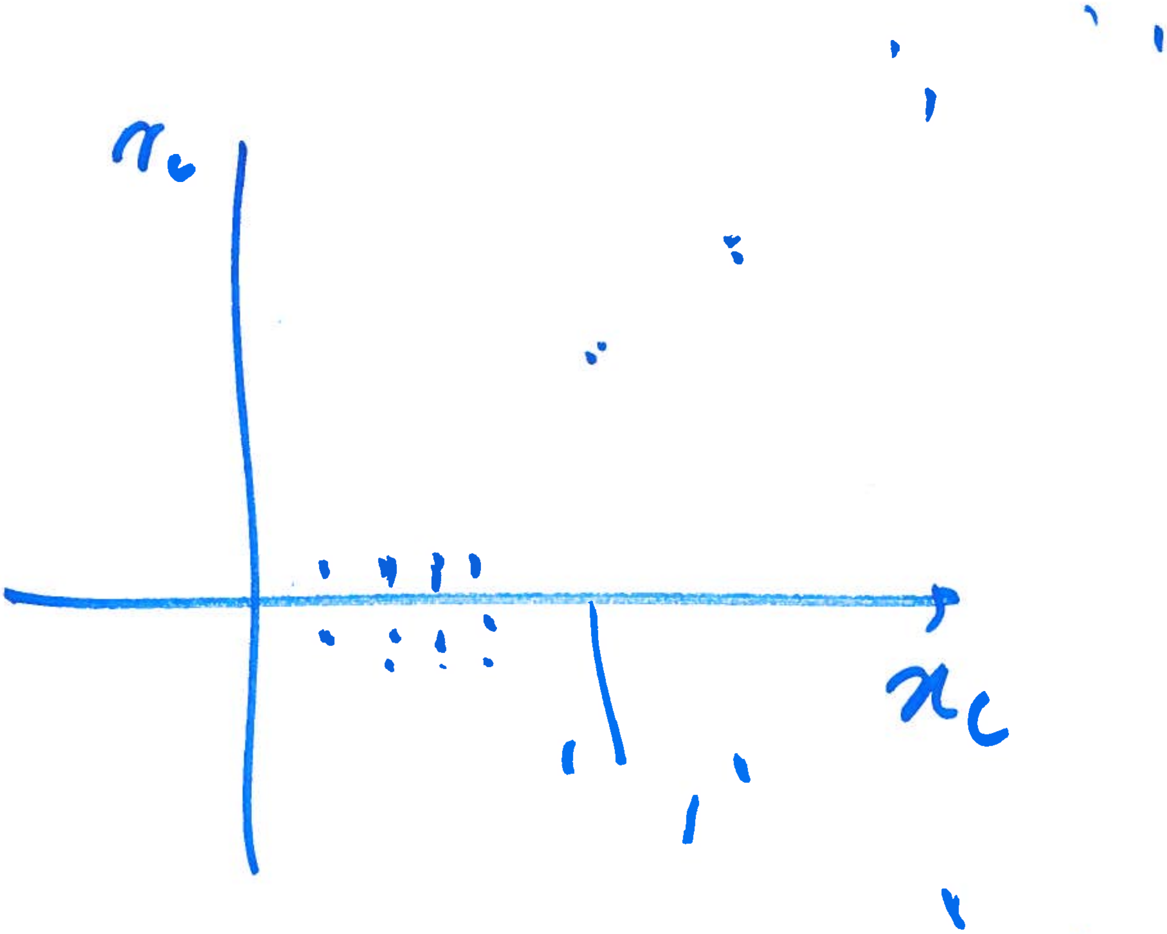
Residual plots can be of two types

Plot $\hat{r}_i$ against x_i's $(\hat{r}_i^*)$

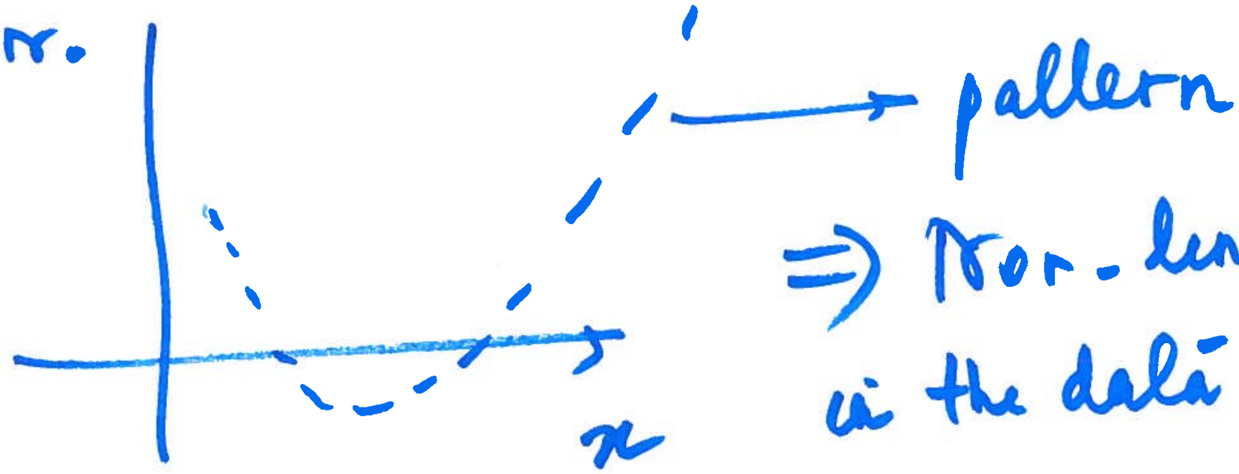
$\hat{r}_i$ against $\hat{y}_i$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i.$$

(iii) Q-Q-plot of the $\hat{r}_i^*$ against the theoretical quantiles of the Z-distribution



Fluctuation increases or decreases with x. (warning that $\sigma^2 = \sigma^2(x)$).



$\Rightarrow$ Non-linearity in the data

Clicker Questions

6. β represents the slope of the Least Square line.

(a) True — 71%

(b) False ✓

$\hat{r}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$

$\bar{r} = 0$

(a) True

(b) False.