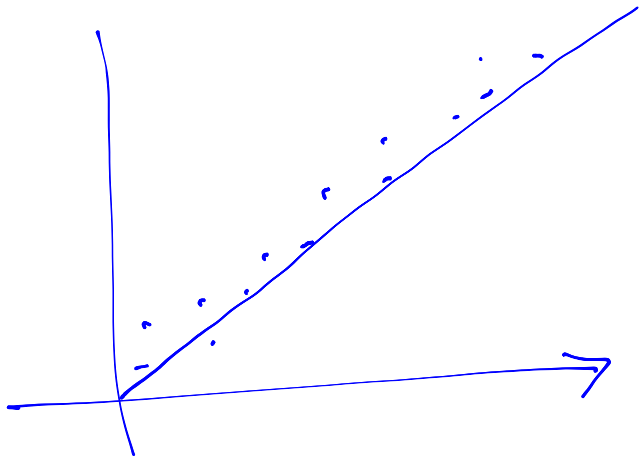
① check the normality of Residuals:



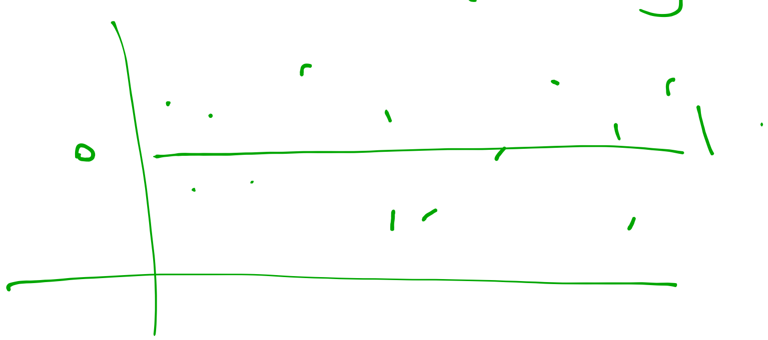
plot residuals vs. Normal quantiles ⊙ if the points are around the straight line ⇒ we conclude that the assumption of Normality of residuals is met!

② check the assumption of constant variance:

plot residuals vs. predicted values ($\hat{y}$)

or vs. Expremental variables

(x_i's, e.g wind, temp & solar. R)



If the data are scatterd with no pattern, the assumption of constant variance is met.

$p-1=3 \rightarrow p=4$

1. New York Air Quality Measurement: The daily air quality in New York was measured from May to September 1973. To see the effect of different variables on Ozone layer, linear regression was used. The JMP output shows the linear regression between wind, temprature, solar radiation and Ozone.

$$y=\beta_0+\beta_1(\text{wind})+\beta_2(\text{temp})+\beta_3(\text{Solar.R})+\varepsilon_i$$

$\varepsilon \overset{iid}{\sim} N(0,\sigma^2)$

model assumptions:

(1) residuals are normally distributed. (check by normal qq-plot)

(2) σ^2 (the variance) is constant. (residual vs. predicted plot)

Summary of Fit

RSquare	0.500653
RSquare Adj	0.490599
Root Mean Square Error *	0.063432
Mean of Response	0.119673
Observations (or Sum Wgts)	153

$\sqrt{\hat{\sigma}^2} = \sqrt{MSE} = \sqrt{S^2_{SF}}$ *

n

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.6010814	0.200360	49.7966
Error	149	0.5995125	0.004024	Prob > F
C. Total	152	1.2005939		<.0001*

$\hat{\sigma}^2 = MSE = S^2_{SF}$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.182964	0.055639	-3.29	0.0013*
Wind	-2.460579	0.548765	-4.48	<.0001*
Temp	1.4731095	0.21135	6.97	<.0001*
Solar.R	0.0518205	0.02024	2.56	0.0115*

$\hat{\beta}_0=b_0$, $\hat{\beta}_1=b_1$, $\hat{\beta}_2=b_2$, $\hat{\beta}_3=b_3$

Use the output to answer the questions.

(a) Write out the model with the appropriate estimates.

$$\hat{y}=b_0+b_1(\text{wind})+b_2(\text{temp})+b_3(\text{Solar.R})$$
$$=-0.1829-2.4605(\text{wind})+1.4731(\text{temp})+0.0518(\text{Solar.R})$$

(b) For the linear relationship, find r the sample correlation coeffecient and R^2 the coeffecient of determination and interpret R^2

Just in linear regression: $r=\sqrt{R^2}=\sqrt{0.50}$

$R^2=50\%$

(c) Provide an estimate for σ^2

$$\hat{\sigma}^2 = MSE = S_{SF}^2 = 0.0040$$

(d) Provide an estimate for the variance of the coefficient of wind. var(b_1)

$$\widehat{var}(b_1) = (SE(b_1))^2 = (0.5487)^2$$

β_1

(e) Calculate and intepret the 95% two-sided confidence interval for the coefficient of wind

$$b_1 \pm t_{(n-p,\,1-\alpha/2)} \cdot SE(b_1)$$

$$\frac{S_{SF}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$-2.46 \pm t_{(153-4,\,1-\frac{0.05}{2})} \cdot (0.5487)$$

$$= -2.46 \pm \boxed{t_{(149,\,0.975)}}(0.5487)$$

(f) Conduct a formal hypothesis test at the $\alpha = 0.05$ significance level to determine if there is significance relationship between air quality (y) and solar radiation (x_3) holding ~~depth~~ temp & wind. constant.

① $H_0: \beta_3 = 0$

$H_a: \beta_3 \neq 0$

(on formula sheet shown by Z or T)

② $\alpha = 0.05$

③ I'll use $K = \dfrac{b_3 - 0}{SE(b_3)}$ assuming that ① H_0 is true ② The MLR between y & x_1, x_2, x_3 is valid (i.e $y = \beta_0 + \beta_1 \text{wind} + \beta_2 \text{temp} + \beta_3 \text{solar.R} + \varepsilon$ is valid)

then $K \sim t_{(n-p)}$

1. New York Air Quality Measurement: The daily air quality in New York was measured from May to September 1973. To see the effect of different variables on Ozone layer, linear regression was used. The JMP output shows the linear regression between wind, temprature, solar radiation and Ozone.

Summary of Fit

RSquare	0.500653
RSquare Adj	0.490599
Root Mean Square Error	0.063432
Mean of Response	0.119673
Observations (or Sum Wgts)	153

R^2

$\sqrt{MSE} = \sqrt{S^2_{SF}}$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.6010814	0.200360	49.7966
Error	149	0.5995125	0.004024	Prob > F
C. Total	152	1.2005939		<.0001*

$P = 4 \begin{cases} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{cases}$

$MSE = S^2_{SF}$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.182964	0.055639	-3.29	0.0013*
Wind	-2.460579	0.548765	-4.48	<.0001*
Temp	1.4731095	0.21135	6.97	<.0001*
Solar.R	0.0518205	0.02024	2.56	0.0115*

$\hat\beta_0 = b_0$
$\hat\beta_1 = b_1$
$\hat\beta_2 = b_2$
$\hat\beta_3 = b_3$

$SE(b_i)$

Use the output to answer the questions.

(a) Write out the model with the appropriate estimates.

$$\hat{y} = -0.1829 - 2.4605\,\text{wind} + 1.473\,\text{temp} + 0.0518\,\text{Solar.R}$$

(b) For the linear relationship, find r, the sample correlation coeffecient and R^2, the coeffecient of determination and interpret R^2

$+ R^2 = 50\%$: 50% of the variation among air quality can be explained by a multiple regression relationship between wind, temperature, Solar Radiation & air quality.

r: sample correlation $= \sqrt{R^2} = \sqrt{0.5} = 0.7$

(c) Provide an estimate for σ^2

$MSE = (0.063432)^2 = 0.004024 = S^2_{SF}$

(d) Provide an estimate for the variance of the coefficient of wind. $\hat{\beta}_1 = b_1$

$\widehat{var}(b_1) = (SE(b_1))^2 = (0.5487)^2 = 0.00310$

(e) Calculate and intepret the 95% two-sided confidence interval for the coefficient of wind

$\hat{b}_1 \pm t_{(n-p, 1-\alpha/2)} \times SE(b_1)$ t-table

$= -2.4605 \pm t_{(149, 0.975)} \times (0.5487)$ $=\infty$ on t-table

$= -2.4605 \pm (1.96)(0.5487) = (\quad , \quad)$

(f) Conduct a formal hypothesis test at the $\alpha = 0.05$ significance level to determine if there is significance relationship between air quality (y) and solar radiation (x_3), holding depth constant.

1 $H_0: \beta_3 = 0$
$H_a: \beta_3 \neq 0$

2 $\alpha = 0.05$

③ I will use the statistic

$$k = \frac{b_3 - 0}{SE(b_3)}$$

which has $t_{(n-p)} = t_{(149)}$ distribution assuming that

① H_0 is true

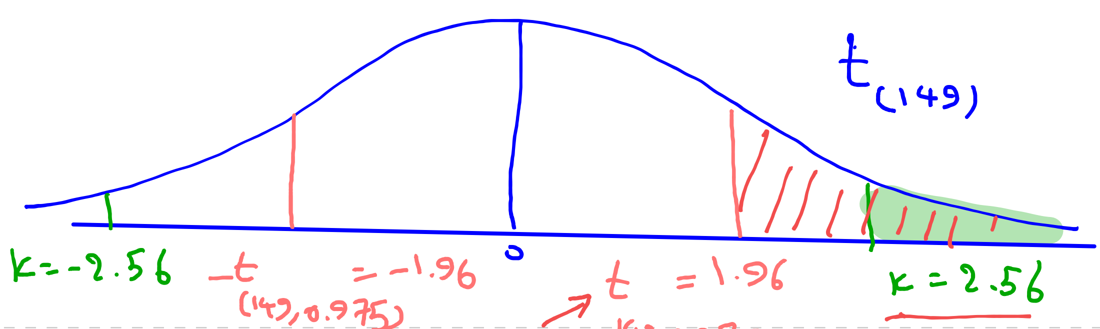
③ the MLR is valid
(i.e $y = \beta_0 + \beta_1 \text{wind} + \beta_2 \text{temp} + \beta_3 \text{solar.R} + \varepsilon$)

④ Calculate $k = \dfrac{b_3 - 0}{SE(b_3)} = \dfrac{0.05182 - 0}{0.02024}$

$= 2.56$

P-value $= P(|T| > k) = P(|T| > 2.56)$, $T \sim t_{(149)}$

$t_{(n-p, 1-\alpha/2)} = t_{(149, 0.975)} = \boxed{1.96}$



14.5, 0.975

⑤ Since p-value < α, we reject H_0.

⑥ There is enough evidence to reject H_0 concluding that there is a significant relationship between the air quality & solar radiation. holding wind and temperature constant.