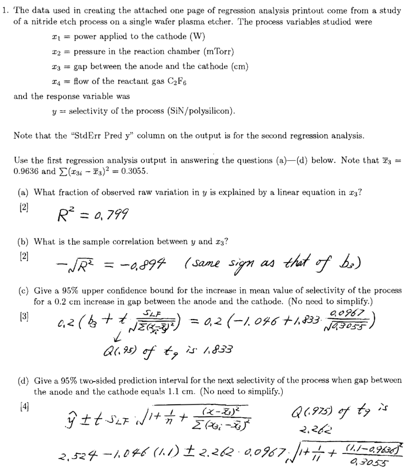
Cheatography

stats Cheat Sheet

by zippyraiden via cheatography.com/106864/cs/21439/

Old Exam



1. The data used in creating the attached one page of regression analysis printout come from a study of a nitride etch process on a single wafer plasma etcher. The process variables studied were

x_1 = power applied to the cathode (W)
x_2 = pressure in the reaction chamber (mTorr)
x_3 = gap between the anode and the cathode (cm)
x_4 = flow of the reactant gas C_2F_6

and the response variable was

y = selectivity of the process (SiN/polysilicon).

Note that the "StdErr Pred y" column on the output is for the second regression analysis.

Use the first regression analysis output in answering the questions (a)—(d) below. Note that $\bar{x}_3 = 0.9636$ and $\sum(x_{3i} - \bar{x}_3)^2 = 0.3055$.

(a) What fraction of observed raw variation in y is explained by a linear equation in x_3?

[2] $R^2 = 0.799$

(b) What is the sample correlation between y and x_3?

[2] $-\sqrt{R^2} = -0.894$ (same sign as that of b_3)

(c) Give a 95% upper confidence bound for the increase in mean value of selectivity of the process for a 0.2 cm increase in gap between the anode and the cathode. (No need to simplify.)

[3] $0.2\left(b_3 + t\,\frac{S_{LF}}{\sqrt{\sum(x_{3i}-\bar{x}_3)^2}}\right) = 0.2\left(-1.046 + 1.833\cdot\frac{0.0967}{\sqrt{0.3055}}\right)$

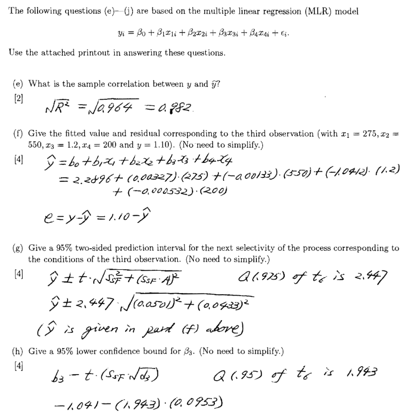
↙ Q(.95) of t_9 is 1.833

(d) Give a 95% two-sided prediction interval for the next selectivity of the process when gap between the anode and the cathode equals 1.1 cm. (No need to simplify.)

[4] $\hat{y} \pm t\cdot S_{LF}\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x}_3)^2}{\sum(x_{3i}-\bar{x}_3)^2}}$ Q(.975) of t_9 is 2.262

$2.524 - 1.046(1.1) \pm 2.262\cdot 0.0967\sqrt{1+\frac{1}{11}+\frac{(1.1-0.9636)^2}{0.3055}}$

Old Exam 2



The following questions (e)—(j) are based on the multiple linear regression (MLR) model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \epsilon_i.$$

Use the attached printout in answering these questions.

(e) What is the sample correlation between y and $\hat{y}$?

[2] $\sqrt{R^2} = \sqrt{0.964} = 0.982$

(f) Give the fitted value and residual corresponding to the third observation (with $x_1 = 275, x_2 = 550, x_3 = 1.2, x_4 = 200$ and $y = 1.10$). (No need to simplify.)

[4] $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$
$= 2.2896 + (0.00327)\cdot(275) + (-0.00133)\cdot(550) + (-1.0412)\cdot(1.2)$
$+ (-0.000532)\cdot(200)$

$e = y - \hat{y} = 1.10 - \hat{y}$

(g) Give a 95% two-sided prediction interval for the next selectivity of the process corresponding to the conditions of the third observation. (No need to simplify.)

[4] $\hat{y} \pm t\cdot\sqrt{S_{SF}^2 + (S_{SF}\cdot A)^2}$ Q(.975) of t_6 is 2.447

$\hat{y} \pm 2.447\cdot\sqrt{(0.0501)^2 + (0.0433)^2}$

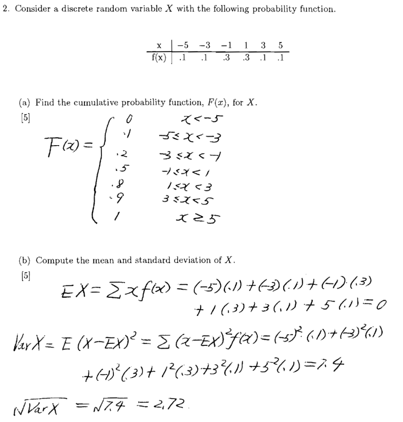
($\hat{y}$ is given in part (f) above)

(h) Give a 95% lower confidence bound for β_3. (No need to simplify.)

[4] $b_3 - t\cdot(S_{SF}\sqrt{d_3})$ Q(.95) of t_6 is 1.943

$-1.041 - (1.943)\cdot(0.0953)$

Old Exam 3



2. Consider a discrete random variable X with the following probability function.

x	−5	−3	−1	1	3	5
f(x)	.1	.1	.3	.3	.1	.1

(a) Find the cumulative probability function, $F(x)$, for X.

[5]
$$F(x) = \begin{cases} 0 & x < -5 \\ .1 & -5 \le x < -3 \\ .2 & -3 \le x < -1 \\ .5 & -1 \le x < 1 \\ .8 & 1 \le x < 3 \\ .9 & 3 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$

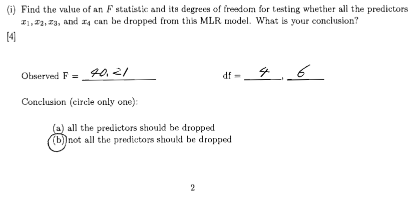
(b) Compute the mean and standard deviation of X.

[5]
$$EX = \sum x f(x) = (-5)(.1) + (-3)(.1) + (-1)\cdot(.3) + 1(.3) + 3(.1) + 5(.1) = 0$$

$$\mathrm{Var}X = E(X-EX)^2 = \sum(x-EX)^2 f(x) = (-5)^2\cdot(.1) + (-3)^2(.1) + (-1)^2(3) + 1^2(.3) + 3^2(.1) + 5^2(.1) = 7.4$$

$$\sqrt{\mathrm{Var}X} = \sqrt{7.4} = 2.72$$

Old Exam 4



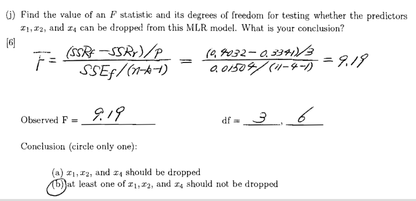
(i) Find the value of an F statistic and its degrees of freedom for testing whether all the predictors x_1, x_2, x_3, and x_4 can be dropped from this MLR model. What is your conclusion?

[4]

Observed F = 40.21 df = 4, 6

Conclusion (circle only one):

(a) all the predictors should be dropped
[✓ (b)] not all the predictors should be dropped



(j) Find the value of an F statistic and its degrees of freedom for testing whether the predictors x_1, x_2, and x_4 can be dropped from this MLR model. What is your conclusion?

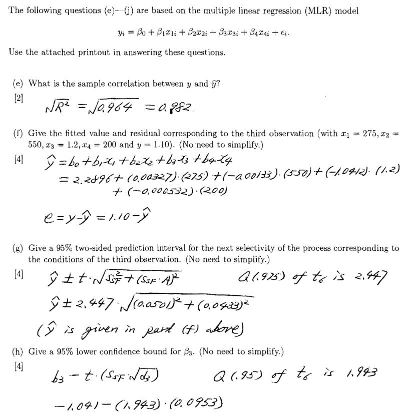
[6]
$$F = \frac{(SSR_f - SSR_r)/p}{SSE_f/(n-k-1)} = \frac{(0.4032 - 0.3341)/3}{0.01504/(11-4-1)} = 9.19$$

Observed F = 9.19 df = 3, 6

Conclusion (circle only one):

(a) x_1, x_2, and x_4 should be dropped
[✓ (b)] at least one of x_1, x_2, and x_4 should not be dropped

Old Exam 5



The following questions (e)—(j) are based on the multiple linear regression (MLR) model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \epsilon_i.$$

Use the attached printout in answering these questions.

(e) What is the sample correlation between y and $\hat{y}$?

[2] $\sqrt{R^2} = \sqrt{0.964} = 0.982$

(f) Give the fitted value and residual corresponding to the third observation (with $x_1 = 275, x_2 = 550, x_3 = 1.2, x_4 = 200$ and $y = 1.10$). (No need to simplify.)

[4] $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$
$= 2.2896 + (0.00327)\cdot(275) + (-0.00133)\cdot(550) + (-1.0412)\cdot(1.2)$
$+ (-0.000532)\cdot(200)$

$e = y - \hat{y} = 1.10 - \hat{y}$

(g) Give a 95% two-sided prediction interval for the next selectivity of the process corresponding to the conditions of the third observation. (No need to simplify.)

[4] $\hat{y} \pm t\cdot\sqrt{S_{SF}^2 + (S_{SF}\cdot A)^2}$ Q(.975) of t_6 is 2.447

$\hat{y} \pm 2.447\cdot\sqrt{(0.0501)^2 + (0.0433)^2}$

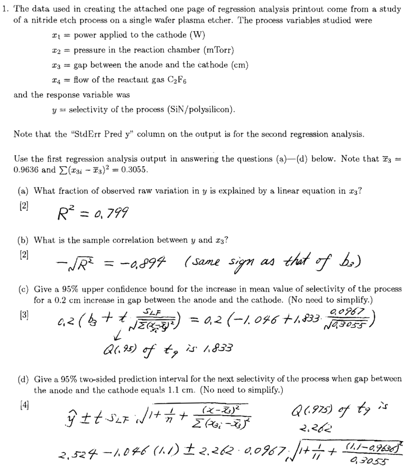
($\hat{y}$ is given in part (f) above)

(h) Give a 95% lower confidence bound for β_3. (No need to simplify.)

[4] $b_3 - t\cdot(S_{SF}\sqrt{d_3})$ Q(.95) of t_6 is 1.943

$-1.041 - (1.943)\cdot(0.0953)$

Old Exam 6



1. The data used in creating the attached one page of regression analysis printout come from a study of a nitride etch process on a single wafer plasma etcher. The process variables studied were

x_1 = power applied to the cathode (W)
x_2 = pressure in the reaction chamber (mTorr)
x_3 = gap between the anode and the cathode (cm)
x_4 = flow of the reactant gas C_2F_6

and the response variable was

y = selectivity of the process (SiN/polysilicon).

Note that the "StdErr Pred y" column on the output is for the second regression analysis.

Use the first regression analysis output in answering the questions (a)—(d) below. Note that $\bar{x}_3 = 0.9636$ and $\sum(x_{3i} - \bar{x}_3)^2 = 0.3055$.

(a) What fraction of observed raw variation in y is explained by a linear equation in x_3?

[2] $R^2 = 0.799$

(b) What is the sample correlation between y and x_3?

[2] $-\sqrt{R^2} = -0.894$ (same sign as that of b_3)

(c) Give a 95% upper confidence bound for the increase in mean value of selectivity of the process for a 0.2 cm increase in gap between the anode and the cathode. (No need to simplify.)

[3] $0.2\left(b_3 + t\,\frac{S_{LF}}{\sqrt{\sum(x_{3i}-\bar{x}_3)^2}}\right) = 0.2\left(-1.046 + 1.833\cdot\frac{0.0967}{\sqrt{0.3055}}\right)$

↙ Q(.95) of t_9 is 1.833

(d) Give a 95% two-sided prediction interval for the next selectivity of the process when gap between the anode and the cathode equals 1.1 cm. (No need to simplify.)

[4] $\hat{y} \pm t\cdot S_{LF}\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x}_3)^2}{\sum(x_{3i}-\bar{x}_3)^2}}$ Q(.975) of t_9 is 2.262

$2.524 - 1.046(1.1) \pm 2.262\cdot 0.0967\sqrt{1+\frac{1}{11}+\frac{(1.1-0.9636)^2}{0.3055}}$

By **zippyraiden**

cheatography.com/zippyraiden/