2020-15909 나승환

A.

1. if residuals follows ~~uniform~~ ~~distribution~~ uniform with mean zero ~~and~~

2. the number of variables whose value can be determined freely. it's like the dim. of vector space.

3. dividing ~~error~~ residuals by the population sd ($e_{sd,i}$) vs. by the sample sd (e_{si})

4. To penalize model whose independent variables are too many. R^2 increases even if the model include variables with less explainatory power.

5. Assuming normality is essential.

6. No. only when $\sigma^2=\sigma_1^2=\sigma_2^2$

7. No.

B.

1. chi-squared ~~d~~distribution with ~~df~~ $df=1$

2. F-distribution with $df=(2,2)$

3.

4.

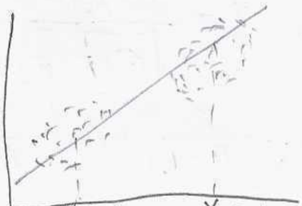
5. F-distribution with $df=(1,1)$

C. linear model should be linear about $\beta_0, \beta_1, \cdots, \beta_k$. 2, 3, 5.

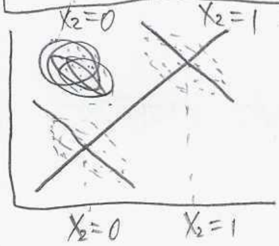
D.

1.

(consistent)



(inconsistent)



2. No. The interaction effect could made unexpected main effect. models only use X_1 ($Y_i=\beta_0+\beta_1X_{1i}+\varepsilon_{1i}$) or X_2 ($Y_i=\beta_0+\beta_1X_{2i}+\varepsilon_{2i}$) can show significant marginal effect.

E.

1. (women) $E(y_i)=\beta_0=\mu_2$

(men) $E(y_i)=\beta_0+\beta_1=\mu_1$

$\therefore \beta_0=\mu_2 \quad \beta_1=\mu_1-\mu_2$

$$\hat{\beta}_1=\frac{S_{XY}}{S_{XX}}=\frac{-\frac{n_1}{n_1+n_2}\sum_{i=1}^{n_2}Y_i+\sum_{i=1}^{n_1}Y_i}{\frac{n_1n_2}{n_1+n_2}}=\frac{n_1\mu_1-\frac{n_1}{\cancel{n_1+n_2}}(\mu_1+\mu_2)}{\frac{n_1n_2}{n_1+n_2}}$$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=\bar{y}-\frac{n_1\mu_2}{n_2}\quad \frac{-\cancel{n_1}\mu_2}{\frac{\cancel{n_1}n_2}{n_1+n_2}}$$

$$=\frac{n_1\mu_1+\mu_2n_2}{n_2+n_1}-\frac{n_1\mu_2}{n_2}$$

2. $\beta_1=0$ equals to $\mu_1-\mu_2=0\ (\Leftrightarrow \mu_1=\mu_2)$

test statistic for

$$\begin{cases}\beta_1=0 : t=\dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}=\dfrac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{XX}}}=\dfrac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\\ \mu_1=\mu_2 : t=\dfrac{(\mu_1-\mu_2)-0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\end{cases}$$

$$S_{XX}=n_1\left(1-\frac{n_1}{n_1+n_2}\right)^2+n_2\left(-\frac{n_1}{n_1+n_2}\right)^2=\frac{n_1n_2}{n_1+n_2}=\frac{1}{\frac{1}{n_1}+\frac{1}{n_2}}$$

$\therefore$ Two statistics are same.

F.

G.

H.

I. $100 \le X, Y \le 320$,

1. $Y_i = 100 + \beta_1 X_i + \varepsilon_i$

2. For model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, reduced model is $(\beta_1 | \beta_0 = 100)$:
$Y_i = 100 + \beta_1 X_i + \varepsilon_i$ but intercept is about $Y_i | X_i = 0$, which means intercept should be lower than 100 because $100 \le X_i \le 320$

3. $\sum (y_i - \beta_0 - \beta_1 x_i)^2$ should be minimized.

J. 1. Adjusted $R^2 = 1 - (1 - 0.938) \times \frac{99}{97}$
$(k=2, n=100)$
$1 - \frac{SS_E}{SS_T} \times \frac{n-1}{n-k-1}$
$= 1 - 0.062 \times \frac{99}{97}$

2. $\frac{SS_R \div 2}{0.9428/97} = 733.5$. $SS_R = \frac{2 \cdot 0.9428 \cdot 733.5}{97}$

3. From F-test with $df = (2, 97)$,
$F = \frac{SS_R/2}{SS_E/97} = 733.5$ follows F.
P-value for H_0: less than $2.2e^{-16}$.

4. Statistically, addition of x_3 isn't significant because R^2 didn't show any significant increase (to put it another way, Adjusted R^2 decreased), and F-statistic for reduced model $(\beta_0, \beta_1, \beta_2 | \beta_3)$ is greater than the full model $(\beta_0, \beta_1, \beta_2, \beta_3)$