Second midterm exam
December 15th, 2021

Time:
75 minutes

- **You are not allowed to use any documentation apart from the formula sheet you have received, and the Z(0,1) table.**
- **Use 4 decimal digits in all calculations and results.**

1. Prove the following:
   a. (1 point) If $\widehat{\theta_1}$ is an unbiased estimator for θ, and X is a random variable with mean μ=0, then $\widehat{\theta_2} = \widehat{\theta_1} + X$ is also an unbiased estimator for θ.
   b. (1 point) If $\widehat{\theta_1}$ is an unbiased estimator for θ such that $\mathrm{E}\left[\widehat{\theta_1}\right] = a^{\theta} + b$, where $a =/0$, then $\widehat{\theta_2} = \frac{\widehat{\theta_1} - b}{a}$ is also an unbiased estimator for θ.

2. To compare customer satisfaction levels of two competing cable television companies, 174 customers of Company 1 and 355 customers of Company 2 were randomly selected and were asked to rate their cable companies on a five-point scale, with 1 being least satisfied and 5 most satisfied. The survey results are summarized in the following table:

Company 1	Company 2
$n_1 = 174$	$n_2 = 355$
$\overline{x_1} = 3{,}51$	$\overline{x_2} = 3{,}24$
$S_1 = 0{,}51$	$S_2 = 0{,}52$

   a. (1,5 points) Build a 99% confident interval for the difference in average satisfaction levels of customers of the two companies as measured on this five-point scale, and explain its meaning.
   b. (1,5 points) Perform a 1% hypothesis test to decide whether customers of Company 1 are more satisfied than those of Company 2. Explain the result of the test.
   c. (1 point) Calculate the p-value of the previous test

3. We want to explain a certain varible Y by means of variables X1, X2, X3 and X4. First we obtain their dispersion matrix which is shown in figure 1, and then we try to build multiple linear regression models starting with the four variables and then removing one at a time until we try only with variables X1 and X2. The summary of each model and their corresponding residual grahps are shown in figures 2 to 4.

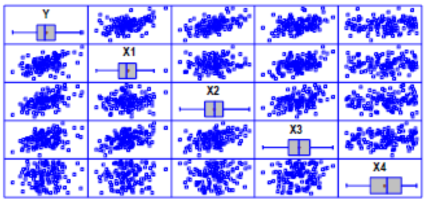

Figure 1.

		Standard	T	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	47,6932	20,2909	2,35047	0,0203
X1	25,9569	4,5788	5,66893	0,0000
X2	29,7292	3,95144	7,52365	0,0000
X3	-0,106052	0,115617	-0,917275	0,3608
X4	2,14434	13,1196	0,163446	0,8704

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	179183,	4	44795,8	27,26	0,0000
Residual	202098,	123	1643,07		
Total (Corr.)	381281,	127			

R-squared = 46,995 percent
R-squared (adjusted for d.f.) = 45,2713 percent

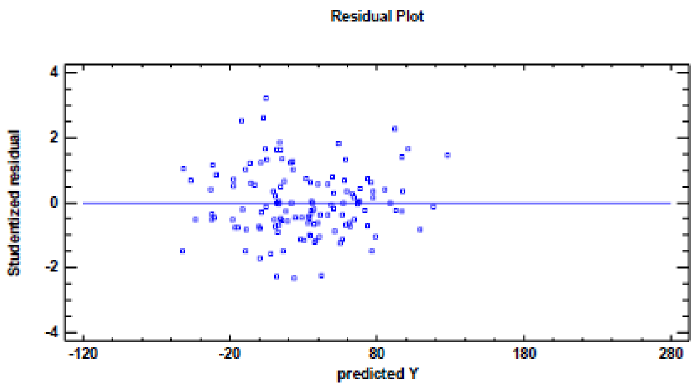

Figure 2.

		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	48,6092	19,425	2,5024	0,0136
X1	25,9502	4,56061	5,69006	0,0000
X2	29,7668	3,92923	7,57575	0,0000
X3	-0,104584	0,114814	-0,910902	0,3641

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	179139,	3	59713,1	36,63	0,0000
Residual	202142,	124	1630,18		
Total (Corr.)	381281,	127			

R-squared = 46,9835 percent
R-squared (adjusted for d.f.) = 45,7009 percent

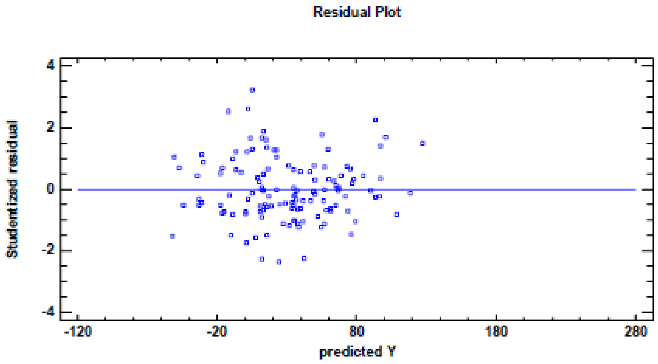

Figure 3.

			Standard	*T*	
Parameter		*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT		31,2164	3,5684	8,74801	0,0000
X1		23,9501	3,99456	5,99569	0,0000
X2		28,2857	3,57457	7,91304	0,0000

Analysis of Variance

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	177787,	2	88893,3	54,60	0,0000
Residual	203494,	125	1627,96		
Total (Corr.)	381281,	127			

R-squared = 46,6288 percent
R-squared (adjusted for d.f.) = 45,7748 percent

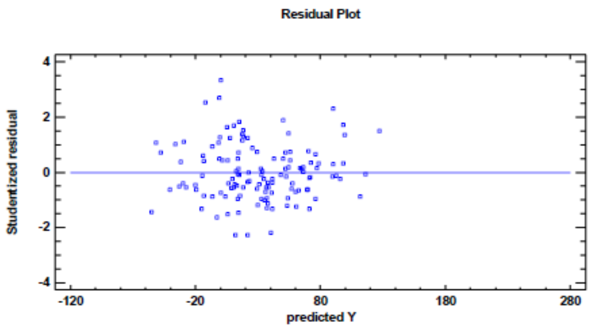

Figure 4.

a. (1 point) What conclusions can be drawn from the dispersion matrix in terms of the relation between Y and the independent variables?
b. (2 points) Write the expression for a valid and/or best model and interpret the coefficients of the regressors as well as the coefficient R^2.
c. (0,5 points) Explain what variables are removed from one model to the next one and why
d. (0,5 points) Explain what happens to the R^2 coefficients