Item 3 — 10 points

For values of $x \in \mathbb{R}$ suppose the linear model $(Y|X = x) \sim \mathcal{N}(\mu(x), \sigma^2)$ holds with

$$E(Y|X = x) = \mu(x) = 1 + 2x - x^2$$

and $\sigma = 2$. Generate $n = 50$ observations of $X \sim \mathcal{N}(0, 1)$ and calculate the values $\mu(x_i) = 1 + 2x_i - x_i^2$, $i = 1, \ldots, n$, and, having that, sample the resulting values $\boldsymbol{y} = (y_1, \ldots, y_n)$ for the response Y by using the following R code (it is important to use the specific seed):

```
set.seed(2022)
n=50
x=rnorm(n)
mu=1+2*x-x*x
y=mu+rnorm(n,sd=2)
```

(a) What is the sample standard deviation of the sample $\boldsymbol{x} = (x_1, \ldots, x_n)$? Further, give the sample mean values (i) for $\mu(\boldsymbol{x}) = (\mu(x_1), \ldots, \mu(x_n))$ and (ii) for $\boldsymbol{y} = (y_1, \ldots, y_n)$. **(requested precision: 2 digits)** — 1.5 points

sample standard deviation of $\boldsymbol{x}$ — 0.5 points

> 2

sample mean value (i) — 0.5 points

> 2

sample mean value (ii) — 0.5 points

> 2

(b) Fit the model $\mu^{(0)}(x) = \beta_1 x$. What is the resulting estimate of the regression coefficient? **(requested precision: 4 digits)** — 1 point

$\hat{\beta}_1$ — 1 point

> 2

(c) Fit the model $\mu^{(1)}(x) = \beta_0 + \beta_1 x + \beta_2 x^2$. What are the resulting estimates of the regression coefficients of x and x^2? **(requested precision: 4 digits)** — 1 point

$\hat{\beta}_1$ — 0.5 points

> 2

$\hat{\beta}_2$ — 0.5 points

> 2

(d) For both models in (b) and (c), calculate $\delta_j := \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}^{(j)}(x_i) - \mu(x_i))^2$ for $j \in \{0, 1\}$, with $j = 0$ and $j = 1$ corresponding to the models in (b) and (c), respectively. **(requested precision: 4 digits)** — 1 point

δ_0 — 0.5 points

> 2

δ_1 — 0.5 points

> 2

(e) Based on the values of δ_j, $j \in \{0, 1\}$, calculated in (d), which model do you prefer? — 0.5 points

Deselect

○ Model in (c)

○ Model in (b)

(f) For both models in (b) and (c), use the Shapiro-Wilk test to test on level $\alpha = 0.05$, if there is evidence against the assumption of normally distributed residuals. Give the resulting p-values and your decision whether there is evidence or not. **(requested precision: 4 digits)** — 4 points

p-value for model in (b) — 1 point

> 2

decision for model in (b)

1 point

reject H_0 ▾

p-value for model in (c)

1 point

2

decision for model in (c)

1 point

reject H_0 ▾

(g) In the setting of (c), test for the parameter β_2 the hypotheses $H_0 : \beta_2 = 0$ **versus** $H_1 : \beta_2 \neq 0$ on the significance level $\alpha = 0.05$. Do you reject H_0? Give the corresponding p-value. **(requested precision: 4 digits)**

1 point

p-value

0.5 points

2

decision

0.5 points

reject H_0 ▾

All answers have been saved!