Let a family of distributions be given by their pdfs (probability density functions) defined for $lpha>0,\gamma>0$ as

$$f(x;\alpha,\gamma) = \sqrt{\frac{\gamma}{2\pi x^2}} \, \exp\left(-\frac{\gamma (x-\alpha)^2}{2\alpha^2 x}\right), \quad x>0. \eqno(1)$$

For fixed (known) $\gamma>0$, $f_{\gamma}(x;\alpha)=f(x;\alpha,\gamma)$ defines a subfamily of the exponential dispersion family (EDF) of distributions with

$$c(x,\phi) = rac{1}{2} \left(\ln(\gamma) - \ln(2\pi x^2) - rac{\gamma}{x}
ight).$$

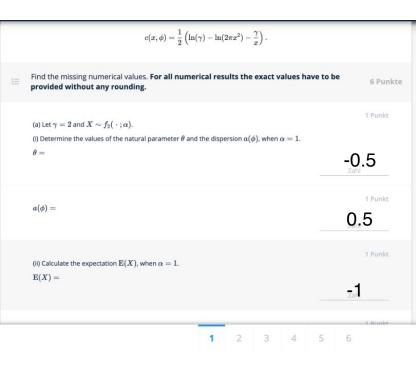
Find the missing numerical values. For all numerical results the exact values have to be provided without any rounding.

6 Punkte

(a) Let $\gamma=2$ and $X\sim f_2(\;\cdot\;;lpha).$

(i) Determine the values of the natural parameter heta and the dispersion $a(\phi)$, when lpha=1.

1 2 3 4 5 6



1 Punkt

(iii) Calculate the variance $\mathrm{Var}(X)$, when lpha=1.

$$Var(X) =$$

0.5

(b) Further assume that Y is a binary response variable and let

$$\pi(x) = P(Y = 1 \mid X = x).$$

Suppose that

$$(X|Y=j) \sim f(\,\cdot\,|lpha_j,\gamma_j),$$

that is, conditionally on Y=j the explanatory variable X has pdf (1) with parameters $lpha_j,\gamma_j>0, j\in\{0,1\}$, and consider the model

$$\operatorname{logit}(\pi(x)) = \operatorname{log}\left(rac{\pi(x)}{1-\pi(x)}
ight) = eta_0 + eta_1 x^{-1} + eta_2 x.$$

Assume that $\alpha_0=\gamma_0=2$ and $\alpha_1=\gamma_1=1$. Calculate the values of β_1 and β_2 .

Hint: Use Bayes' Theorem applied to probability distributions: For random variables X_1 , X_2 with pdfs or pmfs f^{X_1}, f^{X_2} , respectively, it holds

$$f^{X_1|X_2=x_2}(x_1) = \frac{f^{X_2|X_1=x_1}(x_2)f^{X_1}(x_1)}{f^{X_2}(x_2)} \ I_{\operatorname{supp}(X_2)}(x_2), \quad x_1 \in \mathbb{R},$$

(b) Further assume that $oldsymbol{Y}$ is a binary response variable and let

$$\pi(x) = P(Y = 1 \mid X = x).$$

Suppose that

$$(X|Y=j) \sim f(\,\cdot\,|lpha_j,\gamma_j),$$

that is, conditionally on Y=j the explanatory variable X has pdf (1) with parameters $\alpha_j, \gamma_j>0, j\in\{0,1\}$, and consider the model

$$\operatorname{logit}(\pi(x)) = \operatorname{log}\left(rac{\pi(x)}{1-\pi(x)}
ight) = eta_0 + eta_1 x^{-1} + eta_2 x.$$

Assume that $lpha_0=\gamma_0=2$ and $lpha_1=\gamma_1=1$. Calculate the values of eta_1 and eta_2 .

Hint: Use Bayes' Theorem applied to probability distributions: For random variables X_1 , X_2 with pdfs or pmfs f^{X_1} , f^{X_2} , respectively, it holds

$$f^{X_1|X_2=x_2}(x_1) = \frac{f^{X_2|X_1=x_1}(x_2)f^{X_1}(x_1)}{f^{X_2}(x_2)} \ I_{\operatorname{supp}(X_2)}(x_2), \quad x_1 \in \mathbb{R},$$

where $f^{X_i|X_j}$, $i \neq j, i, j \in \{1,2\}$ is the conditional probability density or mass function of X_i given X_j and I_A is the indicator function on a set A.

$$\beta_1 =$$

Zahl

 f^{X_1}, f^{X_2} , respectively, it holds

$$f^{X_1|X_2=x_2}(x_1)=rac{f^{X_2|X_1=x_1}(x_2)f^{X_1}(x_1)}{f^{X_2}(x_2)}\,I_{\mathrm{supp}(X_2)}(x_2),\quad x_1\in\mathbb{R},$$

where $f^{X_i|X_j}$, $i \neq j$, $i,j \in \{1,2\}$ is the conditional probability density or mass function of X_i given X_j and I_A is the indicator function on a set A.

$$\beta_1 =$$

 $\beta_2 =$

Nächste Aufgabe → Consider a GLM $\mathrm{E}(m{Y})=g(m{\mu})=m{X}m{eta}$, where g is the **canonical** link, $m{eta}=(eta_1,eta_2)'$ and

$$m{X} = egin{pmatrix} 1 & 0 \ 0 & 1 \ 1 & -1 \end{pmatrix}.$$

Further assume that the responses Y_1,Y_2,Y_3 are jointly independent with $Y_i\sim \mathcal{P}(\mu_i)$, where $\mathcal{P}(\mu_i)$ denotes the Poisson distribution with parameters $\mu_i=2i^2$, $i\in\{1,2,3\}$.

Determine the expected Fisher information matrix

$${\cal I}_F = egin{pmatrix} i_1 & i \ & \ i & i_2 \end{pmatrix}$$

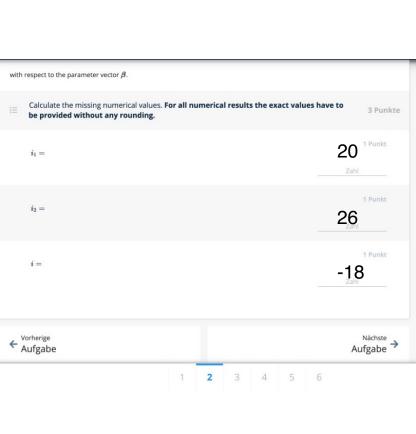
with respect to the parameter vector $\boldsymbol{\beta}$.



Calculate the missing numerical values. For all numerical results the exact values have to

- be provided without any rounding.
- $i_1 =$

3 Punkte



Aufgabe 3 4 Punkte

Let $(X_i)_i \stackrel{iid}{\sim} \mathcal{P}(\mu_1), (Y_i)_i \stackrel{iid}{\sim} \mathcal{P}(\mu_2)$ be two jointly independent sequences of Poisson random counts with parameter $\mu_i, i \in \{1, 2\}$, respectively. Further, define

$$\hat{\mu}_{1(n)}=rac{1}{n}\sum_{i=1}^n X_i, \qquad \hat{\mu}_{2(n)}=rac{1}{n}\sum_{i=1}^n Y_i, \quad n\in\mathbb{N}.$$

Derive the asymptotic variance $\sigma^2>0$, say, of

$$Z_n = \sqrt{n}(\hat{\mu}_{1(n)} + \hat{\mu}_{2(n)}^2)$$

as $n o \infty$ for $\mu_1 = 1$ and $\mu_2 = 2$.

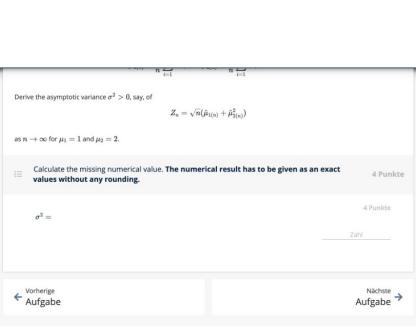
as
$$n o \infty$$
 for $\mu_1 = 1$ and $\mu_2 = 2$

Calculate the missing numerical value. The numerical result has to be given as an exact values without any rounding.

 $\sigma^2 =$

$$\sigma^z =$$

4 Punkte



Alle Antworten wurden gespeichert!

1 2 3 4 5 6

Aufgabe 4 The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong. Task 24 3 Punkte Which of the three models of task 24(b) and 24(c) has the best fit in terms of AIC? Type in "1" for the logistic regression model, "2" for the probit and "3" for the cloglog regression model (without quotation marks). Provide the corresponding AIC value for this model (requested precision: 1 digit) best model AIC value 1 Punkte 1 Transform the variable sex into a factor variable. Fit a cloglog regression model that predicts transport using time, sex and their interaction. What is the percentage of correct classified observations for this model? (requested

6

precision: 4 digits)

model, "2" for the probit corresponding AIC value best model	and "3" for	the cloglo	g regress	ion mode	l (without		" for the logistic regression s). Provide the
							Zahl
AIC value							0.5 Punkt
							Zahl
Transform the variable s							
time, sex and their int							
	eraction. W	hat is the	percenta	ge of corr	ect classi	fied observations	ts transport using for this model? (requested
time, sex and their int precision: 4 digits)	eraction. W	hat is the	percenta	ge of corr	ect classi	fied observations	ts transport using for this model? (requested 0.7479
time, sex and their int precision: 4 digits) What are the values for A	eraction. W	hat is the	percenta	ge of corr	ect classi	fied observations	ts transport using for this model? (requested 0.7475)

time, sex and their i	sex into a factor variable. Fit a cloglog regression model that predict interaction. What is the percentage of correct classified observations	
precision: 4 digits)		Zahl
What are the values for	r AIC and BIC of the model above? (requested precision: 1 digit)	0.5 Punkte
value AIC		Zahl
Value BIC		0.5 Punkte
		Zahi
← ^{Vorherige} Aufgabe		Nächste Aufgabe →
	Alle Antworten wurden gespeichert!	
	1 2 3 4 5 6	

5.5 Punkte Aufgabe 5

The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong.

Task 25

At the beginning of your code, set the seed to 2021. Load the dataset credits.csv from RWTHmoodle into your workspace. Make sure that your working directory contains an

version of credits.csv from the RWTHmoodle space.

unmodified version of credits.csv. In particular, we recommend downloading a fresh

5.5 Punkte

In the *credits* dataset, the variable **guaran** stands for wheather the possible credits recipient is "No further guarantor of debtor" (guaran = 1), "co applicant" (guaran = 2) or "guarantor" (guaran = 3). Introduce a new variable ${\tt guaran2} \ \hbox{with two categories where} \ {\tt guaran2} = 1 \ \hbox{for} \ {\tt guaran} = 1 \ \hbox{and} \ {\tt guaran} = 2 \ \hbox{and} \ {\tt guaran2} = 2 \ \hbox{otherwise}.$ For how many cases in the data set credits it holds that guaran2=1? (requested precision: whole number)

Split randomly the data in training (65 % of the dat) and testing data (the remaining 35 % of the data). Transform account, behavior, rate, finance, furthered, home, job, pers into factor variables. Fit the following logistic regression model on the training data:

In the credits dataset, the variable guaran stands for wheather the possible credits recipient is "No further guarantor of debtor" (guaran = 1), "co applicant" (guaran = 2) or "guarantor" (guaran = 3). Introduce a new variable guaran = 2 and guaran = 2 and guaran = 2 and guaran = 2 otherwise. For how many cases in the data set credits it holds that guaran = 2? (requested precision: whole number)

Zahl

0.5 Punkte

Split randomly the data in training (65% of the dat) and testing data (the remaining 35% of the data). Transform account, behavior, rate, finance, furthered, home, job, pers into factor variables. Fit the following logistic regression model on the training data:

 ${\tt repayment} \sim$

 ${\tt time+age+account+behavior+savings+rate+guaran2+finance+furthcred+home+job+pers} \\ In the following, we will refer to this model as $model b$.}$

What are the AIC and BIC values for this model? (requested precision: 1 digit)

3

4 5

6

AIC value

663.1

BIC value

0.5 Punkte

779.5

Zahī

1 Puni

BIC value						0.5 Punkt
						Zahl
What is the AUC (area under th	ne curve) of m	odel la trans	uested precision	o: 3 digite\	1 Puni
What is the AOC (area under u	ie curve) or m	ouer or (req	uesteu precisioi	i. 3 digits)	Zahl
Fit the null model	correspondi	ng to <i>model b</i>	. What is the	null deviance for	r this model? (requ	1 Punl ested precision: 4 digits)
Fit the null model	correspondi	ng to <i>model b</i>	. What is the	null deviance for	r this model? (requ	ested precision: 4 digits)
Select the best m following, we will	odel nested in	n <i>model b</i> in t	erms of AIC o	using a backward 2. What is the med	l stepwise selection dian of the estimate	797.5 0.5 Punkt algorithm. In the ed probabilities for on time
Select the best m following, we will	odel nested in	n <i>model b</i> in t	erms of AIC o	using a backward 2. What is the med	l stepwise selection	797.5 0.5 Punkt algorithm. In the ed probabilities for on time
Select the best m following, we will credit repayment Select the best m following, we will	odel nested in refer to this s for the traini odel nested in refer to this s	n model b in t selected mode ng sample, ba n model b in t selected mode	erms of AIC of all as model eased on this reference of BIC of all and model of all and model of all all and model of all all and model of all all all and model of all all all and model of all all all all all all all all all al	using a backward be write the med condel? (requeste condels a backward f. Test model f ar	I stepwise selection dian of the estimate d precision: 4 digi	797.5 0.5 Punkt algorithm. In the ed probabilities for on timets) 2ahl 0.5 Punkt algorithm. In the ed probabilities for on timets)
Select the best m following, we will credit repayment Select the best m following, we will	odel nested ir refer to this s for the traini odel nested ir refer to this s ed residual su	n model b in t selected mode ng sample, ba n model b in t selected mode	erms of AIC of all as model eased on this reference of BIC of all and model of all and model of all all and model of all all and model of all all all and model of all all all and model of all all all all all all all all all al	using a backward be write the med condel? (requeste condels a backward f. Test model f ar	I stepwise selection dian of the estimat ad precision: 4 digi	797.5 0.5 Punkt algorithm. In the ed probabilities for on timets) 2ahl 0.5 Punkt algorithm. In the ed probabilities for on timets)

	orresponding to mod	<i>lel b</i> . What is the n	ull deviance for this n	model? (requested precision: 4 digits)
				Zahl
				0.5 Punkto
following, we will r		odel as model e. \	What is the median of	vise selection algorithm. In the f the estimated probabilities for on time cision: 4 digits)
				Zahl
PRESS for model e				Zahi
PRESS for model f				0.5 Punkt
•				Zahl

Aufgabe 6 The questions of the third E-Test are based on the tasks of R-Laboratory 8 and 9. Please provide numbers in the requested precision within each question. The use of different precision is evaluated as wrong. Task 26 3.5 Punkte

0.5 Punkts
What is the value of the predicted probability for a good kick when the distance is 47 yards for *model 1*? (requested precision: 4 digits)

Fit a probit regression model that predicts Good. using Dist as explanatory variable. In the following, we will refer to

this model as model 1. What is the deviance of this model? (requested precision: 4 digits)

zahl

What is the 90 % profile confidence interval for the intercept of model~1? (requested precision: 4 digits) lower bound of CI

Fit a probit regression model that predicts Good. using Dist as explanatory variable. In the following, we this model as <i>model 1</i> . What is the deviance of this model? (requested precision: 4 digits)	0.5 Punkte e will refer to Zahl
What is the value of the predicted probability for a good kick when the distance is 47 yards for <i>model 1?</i> (precision: 4 digits)	0.5 Punkte requested
What is the 90 % profile confidence interval for the intercept of <i>model 1?</i> (requested precision: 4 digits) lower bound of CI	0.5 Punkte
upper bound of CI	0.5 Punkte

Fit a probit model that predicts <code>Good.</code> using the attributes <code>Dist, Blk., Pressure, Roof.type, Altitude</code> and <code>Field</code>. Select the model with smallest AIC using a backward stepwise selection algorithm. In the following, we will refer to this model as <code>model 2</code>. What is the number of parameters included in the selected model beyond the intercept?

	CI	
		Zahl
Field. Select to to this model as	el that predicts Good. using the attributes Dist, B. the model with smallest AIC using a backward stepwise model 2. What is the number of parameters include cision: whole number)	selection algorithm. In the following, we will refer
		Zahl
0.5? (requested	rcentages of correct classified observations for $mode$ $oldsymbol{I}$ precision: $oldsymbol{4}$ digits) ect classified observations $model~1$	0.5 Punkti I I and model 2 when the threshold probability is Zahl
0.5? (requested percentage corr	precision: 4 digits)	2l l and $model$ 2 when the threshold probability l :
0.5? (requested percentage corr	I precision: 4 digits) ect classified observations <i>model 1</i>	l l and $model 2$ when the threshold probability is