

Aufgabe 3 (9 + 5 + 3 + 7 + 7 + 8 + 3 + 9 = 51 Punkte)

Hinweis: Beachten Sie die Tabellen mit Quantilen am Ende der Aufgabenstellung!

Im Rahmen einer wissenschaftlichen Studie wurden Probanden gebeten, Schätzfragen über mehrere Runden wiederholt zu beantworten und zwischen den Antwortrunden nach vorgegebenen Regeln miteinander zu kommunizieren. Die Kommunikationsregeln gab es dabei in drei mit T1, T2 und T3 bezeichneten Varianten, bei der Datenerhebung wurden diese kodiert mittels Dummy-Variablen `dummy_T1`, `dummy_T2` und `dummy_T3`, welche jeweils den Wert 1 annehmen, wenn die namensgebenden Kommunikationsregeln in Kraft waren, und sonst den Wert 0. Die folgenden Analysen sind damit befasst, den Schätzfehler nach Beendigung aller Kommunikation (`error_6`) mit Hilfe des anfänglichen Schätzfehlers (`error_1`) sowie den unterschiedlichen Kommunikationsregeln zu erklären.

- (a) Zunächst wurde ein lineares Modell geschätzt, das folgenden Output ergab:

Call:

```
lm(formula = error_6 ~ error_1 + dummy_T1 + dummy_T2)
```

Residuals:

Min	1Q	Median	3Q	Max
-35.953	-4.987	-0.987	4.416	43.321

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.59627	0.43761	10.503	<2e-16 ***
error_1	0.54140	0.01624	33.339	<2e-16 ***
dummy_T1	???	0.58572	-1.322	0.186
dummy_T2	0.23803	0.58562	???	0.684

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.97 on 1404 degrees of freedom

Multiple R-squared: ???, Adjusted R-squared: 0.4424

F-statistic: 373.1 on 3 and 1404 DF, p-value: < 2.2e-16

Stellen Sie das zugrunde liegende Modell in formaler Schreibweise dar und geben Sie an, welche Variable hier mit Hilfe welcher Regressoren erklärt wird.

Bestimmen Sie ferner die mit „???“ markierten Größen (mit den Bezeichnungen $\hat{\beta}_{\text{dummy_T1}}$, $t_{\text{dummy_T2}}$ und R^2).

- (b) Geben Sie an, wie viele Daten in die obige Schätzung eingingen und welche Regressionskoeffizienten positiv bzw. negativ signifikant sind zum Signifikanzniveau 10%. Begründen Sie jeweils Ihre Antwort.
- (c) Warum wäre es problematisch, `dummy_T3` als zusätzlichen Regressor aufzunehmen?
- (d) Testen Sie zum Signifikanzniveau 5%, ob eine Erhöhung des Anfangs-Schätzfehlers um eine Einheit den End-Schätzfehler um mehr als eine halbe Einheit anwachsen lässt.

Geben Sie hierzu die Hypothesen, die Teststatistik mit ihrer Verteilung unter H_0 , den kritischen Bereich, die realisierte Teststatistik sowie die Testentscheidung an. Beantworten Sie auch explizit die oben formulierte Fragestellung.

- (e) Es wird vermutet, dass die unterschiedlichen Kommunikationsregeln sich auch dahingehend auswirken könnten, dass der Einfluss des Anfangs- auf den End-Schätzfehler mit den Kommunikationsregeln variiert. Daher wird folgendes Modell geschätzt:

Call:

```
lm(formula = error_6 ~ error_1 + dummy_T1 + dummy_T2
    + I(error_1 * dummy_T1)
    + I(error_1 * dummy_T2))
```

Residuals:

Min	1Q	Median	3Q	Max
-40.848	-4.884	-0.982	4.473	44.747

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.98195	0.51412	5.800	8.18e-09	***
error_1	0.63574	0.02284	27.836	< 2e-16	***
dummy_T1	1.68801	0.89110	1.894	0.05839	.
dummy_T2	4.02523	0.87790	4.585	4.94e-06	***
I(error_1 * dummy_T1)	-0.14666	0.04108	-3.570	0.00037	***
I(error_1 * dummy_T2)	-0.21691	0.03772	-5.750	1.09e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.863 on 1402 degrees of freedom

Multiple R-squared: 0.4576, Adjusted R-squared: 0.4556

F-statistic: 236.5 on 5 and 1402 DF, p-value: < 2.2e-16

Testen Sie unter Zuhilfenahme geeigneter Bestimmtheitsmaße zum Signifikanzniveau 5%, ob wenigstens einer der zusätzlichen Regressoren tatsächlich relevant ist.

Geben Sie hierzu die Hypothesen, die Teststatistik mit ihrer Verteilung unter H_0 , den kritischen Bereich, die realisierte Teststatistik sowie die Testentscheidung an. Beantworten Sie auch explizit die oben formulierte Fragestellung.

- (f) Interpretieren Sie die in der vorigen Teilaufgabe neu hinzugekommenen Regressionskoeffizienten. Geben Sie ferner den geschätzten Zusammenhang zwischen Anfangs- und End-Schätzfehlern unter allen drei Kommunikationsregeln an.
- (g) Zu welchem Ergebnis gelangt man vermutlich, wenn man das einfache Regressionsmodell, welches den End-Schätzfehler ausschließlich mit Hilfe des Anfangs-Schätzfehlers erklärt, bezüglich der unterschiedlichen Kommunikationsregeln auf Strukturbrüche untersucht? Geben Sie eine kurze Begründung für Ihre Vermutung! *Hinweis:* Zur Beantwortung dieser Teilaufgabe sind keinerlei Rechnungen vonnöten.
- (h) Alle Kommunikationsregeln sahen vor, dass ein Proband im Zentrum der Kommunikation war, dies wurde im Datensatz kodiert durch eine Dummy-Variable namens **zentrum**, die entsprechend den Wert 1 annimmt für Daten von Probanden, die im Zentrum der Kommunikation waren, und sonst den Wert 0. Im Raum steht die Vermutung, dass für das erweiterte Modell aus Teilaufgabe (e) ein Strukturbruch vorliegen könnte im Hinblick auf die Frage, ob ein Proband im Zentrum der Kommunikation stand oder nicht. Um dies zu überprüfen, wurde das Modell separat geschätzt für die Probanden im Zentrum bzw. außerhalb des Zentrums:

Call:

```
lm(formula = error_6 ~ error_1 + dummy_T1 + dummy_T2
    + I(error_1 * dummy_T1)
    + I(error_1 * dummy_T2),
    subset = which(zentrum == 1))
```

Residuals:

Min	1Q	Median	3Q	Max
-30.996	-4.337	-0.742	3.633	44.201

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.62642	0.91915	3.945	9.65e-05 ***
error_1	0.58607	0.04175	14.037	< 2e-16 ***
dummy_T1	-0.62615	1.64325	-0.381	0.703
dummy_T2	1.05048	1.65685	0.634	0.526
I(error_1 * dummy_T1)	0.13870	0.09385	1.478	0.140
I(error_1 * dummy_T2)	-0.05335	0.07623	-0.700	0.485

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.122 on 346 degrees of freedom

Multiple R-squared: 0.4978, Adjusted R-squared: 0.4906

F-statistic: 68.6 on 5 and 346 DF, p-value: < 2.2e-16

Call:

```
lm(formula = error_6 ~ error_1 + dummy_T1 + dummy_T2
    + I(error_1 * dummy_T1)
    + I(error_1 * dummy_T2),
    subset = which(zentrum == 0))
```

Residuals:

Min	1Q	Median	3Q	Max
-41.754	-4.984	-0.892	4.682	35.811

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.75386	0.61189	4.501	7.53e-06 ***
error_1	0.65217	0.02699	24.166	< 2e-16 ***
dummy_T1	1.99866	1.06199	1.882	0.0601 .
dummy_T2	4.87288	1.03032	4.729	2.56e-06 ***
I(error_1 * dummy_T1)	-0.19227	0.04672	-4.115	4.17e-05 ***
I(error_1 * dummy_T2)	-0.26004	0.04343	-5.988	2.92e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.057 on 1050 degrees of freedom

Multiple R-squared: 0.4527, Adjusted R-squared: 0.4501

F-statistic: 173.7 on 5 and 1050 DF, p-value: < 2.2e-16

Testen Sie zum Signifikanzniveau 5%, ob ein Strukturbruch vorliegt hinsichtlich der Eigenschaft, im Zentrum der Kommunikation zu sein.

Hinweis: Verwenden Sie die folgende Tabelle mit Quantilen einiger $t(n)$ -Verteilungen:

$n \backslash p$	0.85	0.9	0.95	0.975	0.99	0.995	0.9995
1	1.962611	3.077684	6.313752	12.706205	31.820516	63.656741	636.619249
2	1.386207	1.885618	2.919986	4.302653	6.964557	9.924843	31.599055
3	1.249778	1.637744	2.353363	3.182446	4.540703	5.840909	12.923979
4	1.189567	1.533206	2.131847	2.776445	3.746947	4.604095	8.610302
5	1.155767	1.475884	2.015048	2.570582	3.364930	4.032143	6.868827
6	1.134157	1.439756	1.943180	2.446912	3.142668	3.707428	5.958816
7	1.119159	1.414924	1.894579	2.364624	2.997952	3.499483	5.407883
8	1.108145	1.396815	1.859548	2.306004	2.896459	3.355387	5.041305
1400	1.036817	1.282157	1.645943	1.961660	2.329014	2.579346	3.297490
1401	1.036817	1.282156	1.645942	1.961659	2.329013	2.579343	3.297485
1402	1.036817	1.282156	1.645941	1.961657	2.329011	2.579341	3.297480
1403	1.036817	1.282155	1.645940	1.961656	2.329009	2.579338	3.297475
1404	1.036816	1.282155	1.645940	1.961655	2.329007	2.579336	3.297470
1405	1.036816	1.282154	1.645939	1.961654	2.329005	2.579333	3.297465
1406	1.036816	1.282154	1.645938	1.961653	2.329003	2.579331	3.297460
1407	1.036816	1.282154	1.645937	1.961651	2.329001	2.579328	3.297455
1408	1.036815	1.282153	1.645937	1.961650	2.328999	2.579326	3.297450

sowie folgende Tabellen mit 0.95-Quantilen einiger $F(m, n)$ -Verteilungen:

$n \backslash m$	2	3	4	5	6	8	10	12	1396	1397
2	19.000	19.164	19.247	19.296	19.330	19.371	19.396	19.413	19.495	19.495
3	9.552	9.277	9.117	9.013	8.941	8.845	8.786	8.745	8.528	8.528
4	6.944	6.591	6.388	6.256	6.163	6.041	5.964	5.912	5.631	5.631
5	5.786	5.409	5.192	5.050	4.950	4.818	4.735	4.678	4.368	4.368
6	5.143	4.757	4.534	4.387	4.284	4.147	4.060	4.000	3.672	3.672
8	4.459	4.066	3.838	3.687	3.581	3.438	3.347	3.284	2.931	2.931
10	4.103	3.708	3.478	3.326	3.217	3.072	2.978	2.913	2.542	2.542
12	3.885	3.490	3.259	3.106	2.996	2.849	2.753	2.687	2.300	2.300
1396	3.002	2.611	2.378	2.221	2.105	1.945	1.837	1.759	1.092	1.092
1397	3.002	2.611	2.378	2.221	2.105	1.945	1.837	1.759	1.092	1.092
1398	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1399	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1400	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1401	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1402	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1403	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1404	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1405	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1406	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092
1408	3.002	2.611	2.378	2.220	2.105	1.945	1.837	1.759	1.092	1.092

$n \backslash m$	1398	1399	1400	1401	1402	1403	1404	1405	1406	1408
2	19.495	19.495	19.495	19.495	19.495	19.495	19.495	19.495	19.495	19.495
3	8.528	8.528	8.528	8.528	8.528	8.528	8.528	8.528	8.528	8.528
4	5.631	5.631	5.631	5.631	5.631	5.631	5.631	5.631	5.631	5.631
5	4.368	4.368	4.368	4.368	4.368	4.368	4.368	4.368	4.368	4.368
6	3.672	3.672	3.672	3.672	3.672	3.672	3.672	3.672	3.672	3.672
8	2.931	2.931	2.931	2.931	2.931	2.931	2.931	2.931	2.931	2.931
10	2.542	2.542	2.542	2.542	2.542	2.542	2.542	2.542	2.542	2.542
12	2.300	2.300	2.300	2.300	2.300	2.300	2.300	2.300	2.300	2.300
1396	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1397	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1398	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1399	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1400	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1401	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1402	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1403	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1404	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1405	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1406	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092
1408	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092	1.092

Aufgabe 4 (5 + 4 + 3 + 2 + 3 + 14 = 31 Punkte)

In der Situation von Aufgabe 3 werden nun ausschließlich Daten betrachtet, die unter den Kommunikationsregeln T1 erhoben wurden und von Probanden stammen, die im Zentrum der Kommunikation standen.

- (a) Die Schätzung des einfachen linearen Zusammenhangs zwischen Anfangs- und End-Schätzfehler ergab folgenden Output:

```
Call:
lm(formula = error_6 ~ error_1,
    subset = which(zentrum == 1 & dummy_T1 == 1))

Residuals:
    Min       1Q   Median       3Q      Max
-23.643  -3.998  -1.624   3.970  19.055

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)         ?     1.17402    ???   0.0124 *
error_1              ?     0.07244    ??? 4.45e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7 on 86 degrees of freedom
Multiple R-squared:  0.5379,    Adjusted R-squared:  0.5325
F-statistic: 100.1 on 1 and 86 DF,  p-value: 4.451e-16
```

Bestimmen Sie zunächst mit Hilfe des geeigneten Outputs der letzten Teilaufgabe von Aufgabe 3 die mit „?“ markierten fehlenden Schätzwerte der Regressionskoeffizienten. Ermitteln Sie anschließend die mit „???“ markierten fehlenden t -Werte der Schätzer.

- (b) Eine Untersuchung des Modells aus der vorhergehenden Teilaufgabe mit Hilfe des Breusch-Pagan-Tests nach Koenker ergab folgenden Output:

```
studentized Breusch-Pagan test

data:  lm(error_6 ~ error_1,
          subset = which(zentrum == 1 & dummy_T1 == 1))
BP = 5.7151, df = 1, p-value = 0.01682
```

Erläutern Sie kurz, welche Fragestellung mit diesem Test untersucht wird. Wie lautet das Ergebnis der Untersuchung, wenn man ein Signifikanzniveau von $\alpha = 0.05$ zu Grunde legt?

- (c) Zur Behebung eines potentiellen Problems mit heteroskedastischen Störgrößen wird das Modell um einen quadratischen Term erweitert und erneut geschätzt. Als Ergebnis erhält man:

Call:

```
lm(formula = error_6 ~ error_1 + I(error_1^2),
    subset = which(zentrum == 1 & dummy_T1 == 1))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-25.075	-4.120	-1.416	3.815	19.817

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.240143	1.644136	2.579	0.0116 *
error_1	0.456619	0.259475	1.760	0.0820 .
I(error_1^2)	0.008053	0.007483	1.076	0.2849

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.994 on 85 degrees of freedom

Multiple R-squared: 0.5441, Adjusted R-squared: 0.5334

F-statistic: 50.72 on 2 and 85 DF, p-value: 3.18e-15

Lassen diese Ergebnisse auf ein Problem imperfekter Multikollinearität schließen? Begründen Sie Ihre Antwort!

- (d) Ein Breusch-Pagan-Test nach Koenker ergibt für das Modell aus Teilaufgabe (c) einen p-Wert von 0.07226. Zu welchem Ergebnis gelangt dieser Test daher zum Signifikanzniveau 5%?
- (e) Als weitere Alternative wird schließlich einem möglichen Einfluss der insgesamt acht verschiedenen Fragen (bezeichnet mit Q1 bis Q8) Rechnung getragen. Dabei ergab sich folgender Output, bei dem „Frage Q n “ mit $n = 2, \dots, 8$ Dummy-Variablen bezeichnen, die für Frage n den Wert 1 annehmen und sonst den Wert 0:

Call:

```
lm(formula = error_6 ~ error_1 + Frage,
    subset = which(zentrum == 1 & dummy_T1 == 1))
```

Residuals:

	Min	1Q	Median	3Q	Max
	-18.4711	-3.5200	-0.1308	2.9810	15.6034

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.15823	1.90801	1.131	0.2614
error_1	0.61451	0.08232	7.465	9.52e-11 ***
Frage Q 2	6.49915	3.12877	2.077	0.0410 *
Frage Q 3	3.47581	2.91400	1.193	0.2365
Frage Q 4	-2.19675	2.69551	-0.815	0.4175
Frage Q 5	6.71917	3.02429	2.222	0.0292 *
Frage Q 6	-0.58041	2.85301	-0.203	0.8393
Frage Q 7	2.34107	2.78571	0.840	0.4032
Frage Q 8	3.52536	2.93692	1.200	0.2336

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.717 on 79 degrees of freedom

Multiple R-squared: 0.6091, Adjusted R-squared: 0.5695

F-statistic: 15.39 on 8 and 79 DF, p-value: 2.173e-13

Verwenden Sie ein geeignetes Güte-Kriterium zur Messung der Modellgüte, um zu entscheiden, ob dieses Modell das beste unter allen in dieser Aufgabe betrachteten Modellen ist.

- (f) Benutzen Sie das Modell aus der vorhergehenden Teilaufgabe, um für einen Probanden, der unter den Kommunikationsregeln T1 im Zentrum der Kommunikation war und bei der Beantwortung von Frage Q1 einen Anfangs-Schätzfehler von 10 begangen hat, Prognoseintervalle zur Vertrauenswahrscheinlichkeit 95% sowohl für den End-Schätzfehler als auch für dessen Erwartungswert zu berechnen. Benutzen Sie hierfür die folgende Tabelle mit Quantilen der t -Verteilung sowie die Information, dass die Kovarianz zwischen den Regressionsschätzern für das Absolutglied und `error_1` auf -0.03388 geschätzt wird.

$n \backslash p$	0.90	0.95	0.975	0.99
75	1.292941	1.665425	1.992102	2.377102
76	1.292790	1.665151	1.991673	2.376420
77	1.292643	1.664885	1.991254	2.375757
78	1.292500	1.664625	1.990847	2.375111
79	1.292360	1.664371	1.990450	2.374482
80	1.292224	1.664125	1.990063	2.373868
81	1.292091	1.663884	1.989686	2.373270
82	1.291961	1.663649	1.989319	2.372687
83	1.291835	1.663420	1.988960	2.372119
84	1.291711	1.663197	1.988610	2.371564
85	1.291591	1.662978	1.988268	2.371022
86	1.291473	1.662765	1.987934	2.370493