

11. weitere Übungsaufgaben Statistik II WiSe 2019/2020

1. **Aufgabe:** Bei 100 Fahrzeugen des gleichen Typs sind neben dem Preis (PREIS) auch die gefahrene Strecke (MEILEN) und die Anzahl der Werkstattbesuche (SERVICE) bekannt. Aus diesen Daten wurde das folgende Modell geschätzt.

1. **Modell:** Simple Regression - PREIS vs. MEILEN

Dependent variable: PREIS
 Independent variable: MEILEN
 Linear model: $Y = a + b \cdot X$

Coefficients

	<i>Least Squares</i>	<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
Intercept	6533,38	84,5123	77,3069	0,0000
Slope	-0,0311577	0,0023089	-13,4947	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	4,18353E6	1	4,18353E6	182,11	0,0000
Residual	2,25136E6	98	22973,1		
Total (Corr.)	6,43489E6	99			

Correlation Coefficient = -0,806308
 R-squared = 65,0132 percent
 Standard Error of Est. = 151,569

- a) Wie lautet die Modellgleichung?
 b) Welche Hypothese wurde im 1. Modell in der Tabelle „Coefficients“ in der Zeile „Slope“ getestet und wie lautet die Testentscheidung bei $\alpha = 0,05$?

Im folgenden Modell wurde die Anzahl der Werkstattbesuche als weitere Einflussgröße mit aufgenommen.

2. **Modell:** Multiple Regression - PREIS

Dependent variable: PREIS
 Independent variables:
 MEILEN
 SERVICE

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	6206,13	24,9662	248,581	0,0000
MEILEN	-0,0314627	0,000631928	-49,7884	0,0000
SERVICE	135,837	3,90259	34,807	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	6,268E6	2	3,134E6	1821,53	0,0000
Residual	166892,	97	1720,53		
Total (Corr.)	6,43489E6	99			

R-squared = 97,4065 percent
 Standard Error of Est. = 41,4793

- c) Wie lautet im 2. Modell die geschätzte Regressionsfunktion?
 d) Es wird behauptet, dass das 2. Modell dem 1. Modell vorzuziehen ist. Geben Sie dafür zwei Gründe an.
 (Ein weiterer Test soll dabei aus Zeitgründen nicht durchgeführt werden.)

2. **Aufgabe:** Es soll die Abhängigkeit der Bevölkerung (in Millionen) der USA von der Jahreszahl Jahr= 1790, 1800, ..., 2000, 2010 untersucht werden. Es liegen 23 Stichprobenpaare vor. Ein erstes Modell (Modell A) liefert das folgende Ergebnis:

Simple Regression - Bevölkerung vs. Jahr

Dependent variable: Bevölkerung

Independent variable: Jahr

Linear model: $Y = a + b \cdot X$

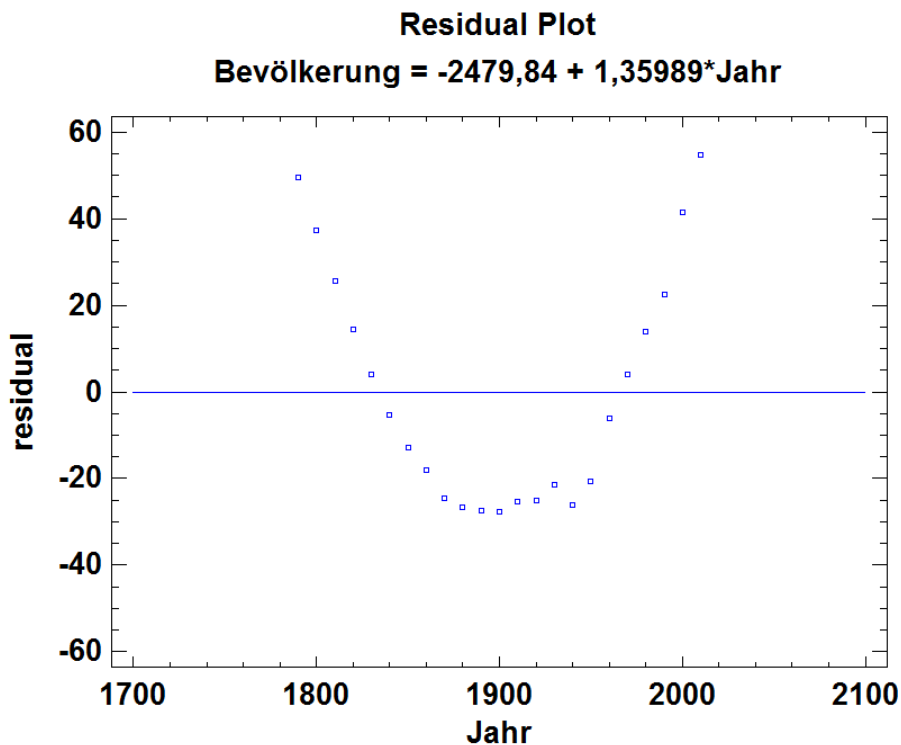
Coefficients

	<i>Least Squares</i>	<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
Intercept	-2479,84	166,974	-14,8517	0,0000
Slope	1,35989	0,0878274	15,4837	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	187150	1	187150	239,74	0,0000
Residual	16393,1	21	780,622		
Total (Corr.)	203543	22			

- Wie lautet die geschätzte Regressionsfunktion?
- Bestimmen Sie das Bestimmtheitsmaß.
- Was können Sie aus der folgenden Grafik ablesen?



Ein zweites Modell (Modell B) liefert als Ergebnis:

Multiple Regression - Bevölkerung

Dependent variable: Bevölkerung

Independent variables:

Jahr
Jahr²

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	21913,8	576,209	38,0309	0,0000
Jahr	-24,3489	0,607054	-40,1099	0,0000
Jahr ²	0,00676548	0,000159732	42,3552	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	203362	2	101681	11251,47	0,0000
Residual	180,743	20	9,03715		
Total (Corr.)	203543	22			

R-squared = 99,9112 percent
Standard Error of Est. = 3,00618

- d) Testen Sie zum Signifikanzniveau $\alpha = 0,01$, ob das kleinere Modell A gegenüber dem größeren Modell B genügt, oder ob das größere Modell B zu einer signifikanten Verbesserung führt.

3. **Aufgabe:** Für 29 PKWs wurden die Merkmale **Preis** Y (in €), gefahrene **Kilometer** X_1 , **Leistung** X_2 (in Ps) und **Alter** X_3 (in Jahren) erfasst. Aus den Daten erhält man das folgende Ergebnis.

Multiple Regression - Preis

Dependent variable: Preis

Independent variables:

Kilometer
Leistung
Alter

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	16646,2	771,963	21,5635	0,0000
Kilometer	-0,0157827	0,00253198	-6,23334	0,0000
Leistung	5,92975	5,19015	1,1425	0,2641
Alter	-907,708	44,4003	-20,4437	0,0000

- a) Wie lautet die geschätzte Regressionsfunktion?
b) Welche der drei Einflussgrößen würden Sie am ehesten aus dem Modell entfernen? Begründen Sie Ihre Entscheidung kurz.

- c) Im Folgenden wurden für zwei weitere Regressionsmodelle die Ergebnisse mit Statgraphics erstellt.

Modell 1:

Simple Regression - Preis vs. Kilometer

Dependent variable: Preis

Independent variable: Kilometer

Linear model: $Y = a + b \cdot X$

Coefficients

	<i>Least Squares</i>	<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
Intercept	15966,7	1626,16	9,81863	0,0000
Slope	-0,0460363	0,00863454	-5,33165	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	1,93448E8	1	1,93448E8	28,43	0,0000
Residual	1,8374E8	27	6,8052E6		
Total (Corr.)	3,77188E8	28			

R-squared = 51,2868 percent

Standard Error of Est. = 2608,68

Modell 2:

Simple Regression - Preis vs. Alter

Dependent variable: Preis

Independent variable: Alter

Linear model: $Y = a + b \cdot X$

Coefficients

	<i>Least Squares</i>	<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
Intercept	15775,4	450,764	34,9971	0,0000
Slope	-1071,45	54,8556	-19,5323	0,0000

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	3,52258E8	1	3,52258E8	381,51	0,0000
Residual	2,49299E7	27	923328,		
Total (Corr.)	3,77188E8	28			

- i) Bestimmen Sie für das Modell 2 das Bestimmtheitsmaß.
- ii) Bestimmen Sie für das Modell 2 die Schätzung für die Standardabweichung des Fehlers.
- iii) Welches der beiden Modelle würden Sie dem anderen Modell vorziehen? Begründen Sie Ihre Wahl kurz.

4. **Aufgabe:** Für 20 Fahrzeuge eines Typs wurden das Alter, die gefahrenen Kilometer und der aktuelle Preis erfasst. Mit diesen Daten wurde folgendes Statgraphics-Ergebnis erstellt.

Multiple Regression - Preis

Dependent variable: Preis

Independent variables:

Alter

Kilometer

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	12526,3	574,898	21,7888	0,0000
Alter	-827,245	42,2323	-19,588	0,0000
Kilometer	-0,00635206	0,00918315	-0,691708	0,4990

Analysis of Variance

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	4,94745E7	2	2,47372E7	192,35	0,0000
Residual	2,05767E6	16	128604,		
Total (Corr.)	5,15322E7	18			

- Wie lautet die geschätzte Regressionsfunktion?
- Bestimmen Sie das Bestimmtheitsmaß B .
- Welchen Verkaufspreis würden Sie mit obigem Modell für ein 3 Jahre altes Auto, welches nur 15000 Kilometer gefahren ist, prognostizieren?