5.2.3 Der globale F-Test

Mit der in diesem Unterkapitel berechneten Prüfgröße f wird der globale F-Test durchgeführt. Mit ihm kann untersucht werden, ob mindestens ein Steigungsparameter (also ein Parameter β_i mit $i \geq 1$) signifikant von Null verschieden ist. Wir testen also die

Nullhypothese: $\beta_i = 0$ für alle $i \geq 1$

gegen die

Alternativhypothese: $\beta_i \neq 0$ für mindestens ein $i \geq 1$.

Bei der einfachen linearen Regressionsanalyse ist dieser Test identisch mit dem t-Test zur Regressionsanalyse und der

Nullhypothese: $\beta_1 = 0$

gegen die

Alternativhypothese: $\beta_1 \neq 0$,

da es nur einen Steigungsparameter gibt. Dieser Test wird deshalb eigentlich erst bei der multiplen linearen Regression benötigt, da man hier zeigen kann, dass mindestens eine der unabhängigen Variablen im Modell einen Einfluss auf die abhängige Variable hat.

Für diesen Test verwendet man die Quadratsummen der Varianzzerlegung. Es sei wobei r wieder die Anzahl der Spalten von X ist ($r = k + 1$), dann gilt (wie immer unter H_0):

1) SSE/σ^2 ist Realisierung einer Chi-Quadrat-verteilten Zufallsvariablen mit n - r Freiheitsgraden.

2) SSR/σ^2 ist Realisierung einer Chi-Quadrat-verteilten Zufallsvariablen mit r - 1 Freiheitsgraden.
3) SST/σ^2 ist Realisierung einer Chi-Quadrat-verteilten Zufallsvariablen mit n - 1 Freiheitsgraden.

Es folgt damit:

$$f = \frac{SSR/(r-1)}{SSE/(n-r)}$$

ist Realisierung einer F-verteilten Zufallsvariablen mit r - 1 und n - r Freiheitsgraden und somit Prüfgröße des F-Tests. Die Nullhypothese wird dann verworfen, wenn $f \geq F^{-1}_{F_{r-1,n-r}}(1-\alpha)$, bzw. wenn

$$\text{p-Wert} = 1 - F_{F_{r-1,n-r}}(f) \leq \alpha \ .$$

Bei der F-Verteilung wird bei allen Verfahren vom System der p-Wert nur dann exakt berechnet, wenn für die beiden Freiheitsgrade der F-Verteilung m_1 und m_2 die Bedingung $m_1 + m_2 \leq 240$ erfüllt ist, oder wenn $m_1 = 1$ und m_2 beliebig ist, oder wenn $m_2 = 1$ und m_1 beliebig ist. Sonst wird für $m_1 + m_2 > 240$ die F-Verteilung über die Chi-Quadrat-Verteilung approximiert.

Im Online-Modul ist die Bedingung $m_1 + m_2 \leq 240$ bei diesem Test immer erfüllt, da der Stichprobenumfang n höchsten den Wert 100 haben kann.

Im Beispiel ist f = 8,4093… und p-Wert ≈ 0,0199. β_1 ist also, wie bereits gezeigt, signifikant von Null verschieden (p-Wert ≤ 0,05).

Umsetzung mit SAS:

```
data dat1;
input x y;
datalines;
170     54
170     67
167     60
177     63
182     75
167     63
164     56
170     65
169     61
176     80
run;

proc reg data=dat1;
model y=x /dw clm cli;
/* plot residual.*x residual.*predicted./symbol ='*'; */
run;
```

SAS-Output zur Prozedur REG:

Die Prozedur REG
Model: MODEL1
Dependent Variable: y

Number of Observations Read	10
Number of Observations Used	10

Varianzanalyse					
Quelle	**DF**	**Summe der Quadrate**	**Mittleres Quadrat**	**F-Statistik**	**Pr > F**
Model	1	295.38887	295.38887	8.41	0.0199
Error	8	281.01113	35.12639		
Corrected Total	9	576.40000			

Root MSE	5.92675	**R-Square**	0.5125
Dependent Mean	64.40000	**Adj R-Sq**	0.4515
Coeff Var	9.20303		

Parameterschätzer

Variable	DF	Parameter-schätzer	Standard-fehler	t-Wert	Pr > |t|
Intercept	1	-114.80119	61.82445	-1.86	0.1004
x	1	1.04674	0.36096	2.90	0.0199