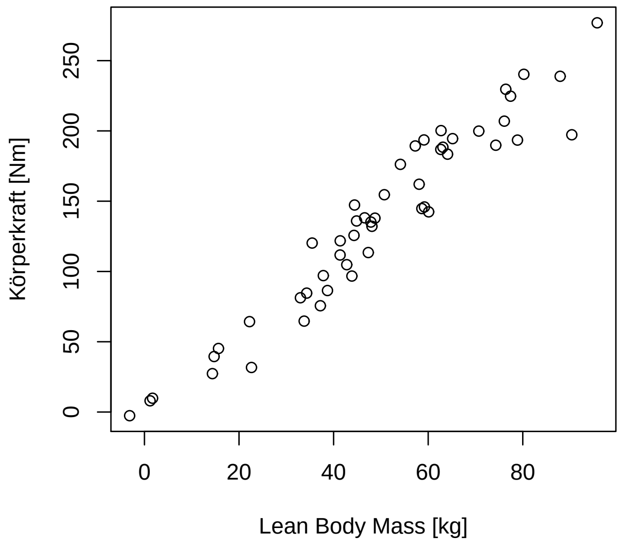

Dieses Quiz soll Ihnen helfen, den `R` Output einer einfachen linearen Regression besser zu verstehen (s. Kapitel 5.4.1)

Es wurden 50 Personen untersucht. Für jede Person wurde die "Lean Body Mass" (Variable `lbm`; LBM = Körpermasse ohne Fett; Einheit: kg) und die Körperkraft (Variable `strength`; maximales Drehmoment am rechten Knie, wenn Oberschenkelstrecker maximal angespannt wird; Einheit: Nm) gemessen. In einem Streudiagramm sieht man, dass Personen mit grosser LBM auch eine grosse Körperkraft aufweisen. Um diesen Zusammenhang genauer zu untersuchen, versuchen wir folgendes Modell anzupassen: $\text{strength}_i = \beta_0 + \beta_1 \text{lbm}_i + E_i$; $E_i \sim \mathcal{N}(0, \sigma^2)$ *i.i.d.* Mit `R` und dem Befehl "summary(lm(strength ∼ lbm))" berechnen wir eine lineare Regression. Wir nehmen an, dass die Modellvorraussetzungen gut erfüllt sind. `R` liefert folgenden Output:

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

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.5221	6.1138	???	0.682
lbm	2.8080	0.1126	24.941	<2e-16 ***

Residual standard error: 18.15 on ?? degrees of freedom

Auswertung und Lösung

Abgaben: 56 / 265

Maximal erreichte Punktzahl: 8

Minimal erreichte Punktzahl: 1

Durchschnitt: 5.91

Frage 1

Genau die korrekten Antworten: ca. 96% - Keine Antwort: ca. 0%.

Was ist gemäss R Output die Schätzung für den Parameter β_0?

√ **Ca. 96%** ◯ -2.5221

Richtig!

Ca. 2% ◯ 2.8080

Leider nicht.

Ca. 2% ◯ 6.1138

Leider nicht.

Ca. 0% ◯ 0.1126

Leider nicht.

Ca. 2% ◯ Weiss nicht.

Danke für Ihr Feedback!

Frage 2

`Genau die korrekten Antworten: ca. 95% - Keine Antwort: ca. 0%.`

Was ist gemäss `R` Output die Schätzung für den Parameter β_1?

Ca. 0% ◯ -2.5221

Leider nicht.

√ **Ca. 96%** ◯ 2.8080

Richtig!

Ca. 0% ◯ 6.1138

Leider nicht.

Ca. 4% ◯ 0.1126

Leider nicht.

Ca. 2% ◯ Weiss nicht.

Danke für Ihr Feedback!

Frage 3

`Genau die korrekten Antworten: ca. 55% - Keine Antwort: ca. 2%.`

Die erwartete Kraft (genauer: das Drehmoment) für eine Person mit 50 kg Lean Body Mass ist gemäss dem geschätzten Modell:

Ca. 2% ◯ 129.45

Leider nicht.

Ca. 7% ◯ 133.49

Leider nicht.

√ **Ca. 55%** ◯ 137.88

Richtig!

Ca. 27% ◯ Kann man mit dem Output nicht berechnen.

Leider nicht.

Ca. 7% ◯ Weiss nicht.

Danke für Ihr Feedback!

Das Modell sagt folgenden Zusammenhang zwischen erwarteter Kraft y und Lean Body Mass x vorher: $y = -2.5221 + 2.8080 \cdot x$. Wenn $x = 50$ ist also $y = 137.88$.

Frage 4

`Genau die korrekten Antworten: ca. 82% - Keine Antwort: ca. 0%.`

Hat LBM einen signifikanten (5% Niveau) Einfluss auf die Körperkraft?

√ **Ca. 82%** ◯ Ja

Richtig!

Ca. 16% ◯ Nein

Leider nicht.

Ca. 2% ◯ Keine Aussage möglich

Leider nicht.

Ca. 0% ◯ Weiss nicht.

Danke für Ihr Feedback!

Der p-Wert in der Zeile `lbm` ist sehr klein (kleiner als 5%). Also kann die Nullhypothese $H_0 : \beta_1 = 0$ auf dem 5% Signifikanzniveau verworfen werden. LBM hat also einen signifikanten Effekt auf die Körperkraft.

Frage 5

`Genau die korrekten Antworten: ca. 57% - Keine Antwort: ca. 0%.`

Was ist ein approximatives zweiseitiges 95%-Vertrauensintervall für β_1? (Ist die Null enthalten? Passt diese Beobachtung zu dem p-Wert im `R` Output?)

Ca. 2% ◯ $-2.5221 \pm 2 * 2.8080$

Leider nicht.

Ca. 29% ◯ $-2.5221 \pm 2 * 6.1138$

Leider nicht.

Ca. 7% ◯ $6.1138 \pm 2 * 0.1126$

Leider nicht.

√ **Ca. 57%** ◯ $2.8080 \pm 2 * 0.1126$

Richtig!

Ca. 5% ◯ Weiss nicht.

Danke für Ihr Feedback!

Ein approximatives 95%-Vertrauensintervall erhält man, indem man "Estimate" ± 2*"Std. Error" rechnet. Für β_1 ergibt das also $2.8080 \pm 2 * 0.1126$. (Die Null ist im 95%-Vertrauensintervall nicht enthalten. D.h., selbst wenn der p-Wert nicht im Output angegeben wäre, wüssten wir, dass die Nullhypothese $H_0 : \beta_0 = 0$ zu Gunsten von $H_A : \beta_0 \neq 0$ auf dem 5% Signifikanzniveau verworfen werden würde. Der p-Wert wäre also sicher kleiner als 5%.)

Frage 6

`Genau die korrekten Antworten: ca. 66% - Keine Antwort: ca. 0%.`

Wie gross ist der beobachtete Wert der Teststatistik in einem Test $H_0 : \beta_0 = 0$ gegen $H_A : \beta_0 \neq 0$ (das ist der t-Wert / "t value" in der Zeile, die zu β_0 gehört)?

Ca. 4% ◯ 6.1138

Leider nicht.

Ca. 5% ◯ -0.682

Leider nicht.

Ca. 18% ◯ 24.941

Leider nicht.

√ **Ca. 68%** ◯ -0.413

Richtig!

Ca. 7% ◯ Weiss nicht.

Danke für Ihr Feedback!

Der beobachtete Wert der Teststatistik berechnet sich aus "Estimate"/"Std. Error". In unserem Fall ist das also $\frac{-2.5221}{6.1138} = -0.413$.

Frage 7

`Genau die korrekten Antworten: ca. 77% - Keine Antwort: ca. 0%.`

Welche Schätzung wird für σ^2 ausgegeben?

Ca. 20% ◯ 18.15

Leider nicht.

√ **Ca. 77%** ◯ 18.15^2

Richtig!

Ca. 2% ◯ Kann man nicht aus dem Output ablesen.

Leider nicht.

Ca. 2% ◯ Weiss nicht.

Danke für Ihr Feedback!

Der "Residual Standard Error" (hier 18.15 ist der Schätzwert für σ. Also ist 18.15^2 eine Schätzung von σ^2.

Frage 8

`Genau die korrekten Antworten: ca. 63% - Keine Antwort: ca. 0%.`

Angenommen, die "degrees of freedom" wären 10. Was wäre dann ein exaktes zweiseitiges 95%-Vertrauensintervall für β_1?

Ca. 14% ◯ $2.8080 \pm 2 * 0.1126$

Leider nicht.

√ **Ca. 63%** ◯ $2.8080 \pm 2.228 * 0.1126$

Richtig!

Ca. 11% ◯ $2.8080 \pm 1.96 * 0.1126$

Leider nicht.

Ca. 13% ◯ Weiss nicht.

Danke für Ihr Feedback!

Ein exaktes zweiseitiges 95%-Vertrauensintervall für β_1 lässt sich mit der Formel Estimate $\pm\, t_{df;0.975} \cdot$ Std.Error berechnen. Dabei sind *df* die "degrees of freedom", also die Anzahl Beobachtungen minus die Anzahl im Modell verwendeter βs. Da wir die "degrees of freedom" als 10 angenommen haben (eigentilch sind es 50-2 = 48), suchen wir in der Tabelle $t_{10;0.975} = 2.228$. Damit ergibt sich für das exakte zweiseitige 95% Vertrauensintervall $2.8080 \pm 2.228 * 0.1126$.