1. Je 1 Punkt.
1) c
2) a
3) a
4) a
5) b
6) d
7) c
8) d

2. **a)** Der Effekt von x auf y ist verschieden in den beiden Gruppen (0.5 P): Der Achsenabschnitt und die Steigung unterscheiden sich in den beiden Gruppen. (0.5 P) **(1 Punkt)**

b) Die Variablen x und z sind stark positiv korreliert. Für $z = 0$ nimmt x kleinere Werte an als für $z = 1$. **(1 Punkt)**

c) Model mit interaktion zwischen x und z. (0.5 P) $y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + \epsilon$ (0.5 P) **(1 Punkt)**

d) Für die Gruppe $z = 0$:

$$y = 1.312 + 1.09x.$$

(0.5 P) Für die Gruppe $z = 1$:

$$y = (1.312 - 1.253) + (1.09 - 0.352)x = 0.059 + 0.738x.$$

(0.5 P) **(1 Punkt)**

e) Nein ist es nicht, denn die Interaktion $x : z$ im Model ist nicht signifikant. **(1 Punkt)**

f) bei adjusted R-squared (0.5 P), bei F-statistic (0.5 P), bei Residual standard error (0.5 P), wenn multiple R-squared genannt wird (-0.5 P) maximal(**(1 Punkt)**) minimal(**(0 Punkte)**)

3. **a)** i. Falsch. Der zugehörige 5%-Test akzeptiert die Nullhypothese $\beta_1 = 0$, da 0 im 95%-Vertrauensbereich liegt; dies gilt daher erst Recht für den 1%-Test. **(1 Punkt)**

ii. Richtig. R^2 hängt von der Streuung der Fehler in der Zielvariable ab. Bei verschiedenen Datensätzen könnte diese unterschiedlich sein. **(1 Punkt)**

iii. Falsch. Wenn mehr Parameter geschätzt werden müssen, ist i.A. auch die Varianz der Prognose grösser. Bei einfacheren Modellen ist der Bias i.A. natürlich grösser; es muss also eine geeignete Balance zwischen Bias und Varianz der Prognose gefunden werden. **(1 Punkt)**

iv. Richtig. `x*y = x + y + x:y`. Da `x^2 = x:x = x` gilt `(x + y)^2 = x + x:y + y`. **(1 Punkt)**

v. Richtig. Das kann passieren wenn die erklrenden Variablen stark korreliert sind.

b) $\hat{Y}|X = x \sim \text{Poisson}(\hat{\beta}x)$, daher ist die geschätzte Varianz einer neuen Beobachtung an der Stelle $x = 3$ gegeben durch $e^{\hat{\beta}x} = e^{1.8 \cdot 3} = e^{5.4} = 221.4064 \approx 221.41$. **(1 Punkt)**

c) In einem gesättigten Modell sind alle Residuen = 0. Der Residual Standard Error wird daher auf 0 geschätzt. **(1 Punkt)**

d) Das adjusted R^2 lsst sich berechnen mit der Formel $adjR^2 = 1 - \frac{n-1}{p} \cdot \frac{R^2}{F} = 1 - \frac{9}{5} \cdot \frac{0.86}{20.5} = 0.9244878 \approx 0.924$

4. **a)** **(1 Punkt)**

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

b) **(2 Punkt)** Das relevante Quantil für einen zweiseitigen z-Test mit Niveau 5% ist 1.96. Daher ist der Koeffizient von x_3 (1P) und der Intercept (1P) signifikant von 0 verschieden.

c) **(1 Punkt)** $n = df_{Resid} + p = 20$

d) **(1 Punkt)** Die gesuchte Odds Ratio beträgt $\exp(\hat{\beta}_2) = 1.758267$.

e) **(2 Punkte)** $\text{logit}(\hat{p}) = -15.5543 - 0.5859 \cdot 3 + 0.5643 \cdot 25 + 1.9639 \cdot 2 = 0.7239306$
$\hat{p} = \frac{\exp(0.7239306)}{1+\exp(0.7239306)} = 0.673472 \approx 0.673$ (1P)
Daher prognostiziert man $\hat{y} = 1$ (1P).

f) **(2 Punkt)** Damit die Wahrscheinlichkeit, dass $y = 1$, genau 50% (also $p = 0.5$) beträgt, müssen die Log-Odds $\eta = \log(p/(1-p) = \log(1) = 0$ sein. Bei $x_1 = 5$ und $x_2 = 25$ erhalten wir aus der Modellgleichung

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

die Gleichung

$$0 = -15.5543 - 0.5859 \cdot 5 + 0.5643 \cdot 25 + 1.9639 \cdot x_3 \ ,$$

(1P) was sich auflösen lässt zu $x_3 = 2.228076 \approx 2.23$. (1P)

g) **(2 Punkt)** Das grössere Model, da der AIC für dieses Model kleiner ist. $AIC =$ Residual deviance $+ 2 \cdot$ number of parameters

grosses Model $AIC = 14.177 + 2 \cdot 4 = 22.177$ (1P)
kleines Model $AIC = 18.177 + 2 \cdot 3 = 24.158$ (1P)

5. **a)** Die Prävalenz von Lungenkrebs Toten nimmt mit dem Alter zu.(0.5P) Die Prävalenz ist höher für rauchende Personen.(0.5P) **(1 Punkt)**

b) Wenn wir die Interaktion `smoke:age` noch ins Modell aufnehmen, hat das Modell mehr Parameter als Beobachtungen. Das heisst, dieses Modell wäre nicht identifizierbar. **(1 Punkt)**

c) Die erwartete Anzahl von Todesfällen ist proportional zur Populationsgrösse. **(1 Punkt)**

d) $656 \cdot e^{-3.70919} = 16.07027 \approx 16.07$ **(1 Punkt)**

e) $e^{0.41044} = 1.507481 \approx 1.51$ **(1 Punkt)**

f) $436 \cdot e^{-3.70919} \cdot e^{2.59447} \cdot e^{0.41044} = 215.5865 \approx 215.59$ **(1 Punkt)**

g) Nein es ist nicht plausibel. Ein 95% Vorhersageintervall für eine $Pois(\lambda)$-verteilte Zufallsvariable ist approximativ gegeben durch

$$[\lambda - 1.96 \cdot \sqrt{\lambda}, \lambda + 1.96 \cdot \sqrt{\lambda}]$$

für grosses λ. Wenn wir nun unsere Schätzung für λ, $\widehat{\lambda} \approx 215.59$, einsetzen erhalten wir das Intervall $[186.8114, 244.3686] \approx [186, 244]$. (1P) Ein 95% Vorhersageintervall für die Zahl in **f)** muss mindestens so gross sein da wir auch noch die Unsicherheit der Parameterschätzung beachten müssen. (1P) **(2 Punkte)**