

Statistik I Übungsblatt 9

Abgabe: Donnerstag, 13.07.06, 12:00 Uhr, Briefkasten: Statistik I.
Insgesamt können maximal 25 Punkte erreicht werden.

1. Nach dem dramatischen Spiel vom letzten Mittwoch streiten sich Jürgen und Marcello darum, ob in Spielen der italienischen oder deutschen Nationalmannschaft mehr Tore geschossen werden. Aus den je letzten 100 Spielen ergaben sich folgende Zahlen:

Anzahl Tore:	0	1	2	3	4	5	6	7	8	≥ 9
Italien	11	25	28	20	9	3	3	0	1	0
Deutschland	13	27	28	17	11	1	2	1	0	0

- (a) Bestimmen Sie die Tor-Raten der einzelnen Mannschaften, sowie die gemeinsame Rate.
(b) Als der französische Fußballnationaltrainer Raymond Domenech davon hört, sagt er nur: "Die spielen doch beide gleich schlecht!". Kann er diese Aussage auf einem Niveau von $\alpha=5\%$ aufrechterhalten?
(c) (Bonus: Wer kann uns die echten Zahlen besorgen?)
2. Beim Verlassen des Dortmunder Westfalen-Stadions wurden zufällig 277 Fans der deutschen Fußball-Nationalmannschaft nach der Rückennummer des besten Spielers gefragt. (Es wird nur die Startelf berücksichtigt)

Nummer	1	3	5	11	13	16	17	18	19	20	21
Anzahl	8	6	12	32	25	23	28	70	41	17	15

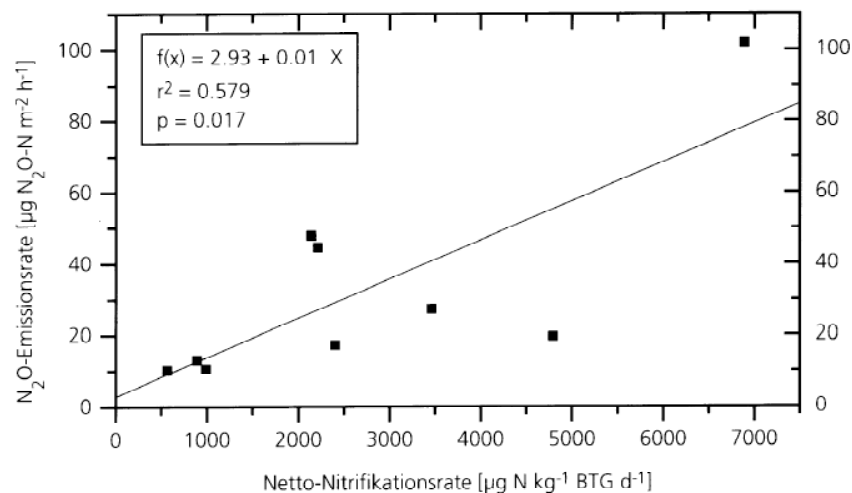
Testen Sie mit einem χ^2 -Test, ob die Fans rein zufällig geantwortet haben, oder nicht. Welche Annahmen treffen Sie?

3. Aus dem Genlabor der Iowa State University stammen folgende, replizierte Gen-Expressionswerte von Arabidopsis ohne Stressfaktor (A) und Arabidopsis unter UV-Bestrahlung (B).

A:	27	34	20,5	29,5	20	28	20	26,5	22	24,5	34	35,5	19
B:	18	14,5	13,5	12,5	23	24	21	17	18,5	9,5	14	-	-

- (a) Berechnen Sie ausführlich(!) entweder die Wilcoxon Rangsummen- oder Mann-Whitney-U-Teststatistik, und geben Sie den Wert für die andere an.
(b) Bestimmen Sie den p -Wert
- exakt (kritische Werte sind zahlreich vertafelt, und Google hilft gerne)
 - über die Normalapproximation
- für die einseitige Hypothese.

4. Modelliert werden soll der Verbrauch: MPG City (Datensatz siehe Übungsblatt 1).
- Führen Sie eine Regression durch für MPG vs. Weight und für MPG vs. Horsepower. Stellen Sie beide Zusammenhänge in Scatterplots dar, in die Sie die Regressionslinie einzeichnen. Sind Sie mit dem Modell zufrieden? Gibt es Ausreißer, sind Verletzungen der Annahmen sichtbar?
 - Erzeugen Sie Scatterplots von den Residuen vs. vorhergesagten Werten. Beschreiben und interpretieren Sie den Unterschied in den beiden Plots.
 - Transformieren Sie MPG zu $1/\text{MPG}$ und führen Sie die Regression $1/\text{MPG}$ vs. Weight durch. Erzeugen Sie einen Scatterplot Residuen vs. vorhergesagte Werte. Wie interpretieren Sie das Resultat? Welche Probleme erwarten Sie aufgrund der Transformation, und wie kann man sie handhaben?
5. Unter der abgebildeten Graphik war der folgende Text abgedruckt:
Korrelationsanalyse [=Regressionsanalyse] zwischen den unter Einsatz der neuen Methode bestimmten Netto-Nitrifikationsraten und der [...] N_2O -Emissionsrate. Ergebnis der Korrelationsanalyse zwischen N_2O -Emissionsrate und Netto-Nitrifikationsrate des Buchenbestandes im Höglwald.



- Beurteilen Sie die Güte der Regression. Hätten Sie hier eine Regression empfohlen? Gehen Sie dabei auch auf die Annahmen, die der linearen Regression zugrunde liegen, ein.
- Betrachten Sie den Punkt oben rechts. Wie würden Sie den Einfluss dieses Punktes auf die Regression abschätzen? Ist der Einfluss dieses Punktes anders als der eines anderen Punktes?
- Die Steigung der Geraden ist fast Null, aber statistisch (einigermaßen) signifikant. Erläutern Sie einem Laien den Unterschied zwischen der statistischen Signifikanz eines Parameters und der Frage, ob ein Parameter verschwindet.