

Testatfragen zur Vorlesung “Biostatistik” (I)

1. Nenne den Gegenstand der Statistik in den Humanwissenschaften.
2. “In Leipzig kommt 2020 im Durchschnitt auf 92 Einwohner ein Neugeborenes. Am 31.12.2020 werden 100 Einwohner von Leipzig zufällig ausgewählt. Dann ist mindestens eine der ausgewählten Personen im Jahr 2020 geboren.” Ist dieser Schluß korrekt? Begründe die Antwort.
3. Was versteht man unter dem primären Endpunkt einer Interventionsstudie?
4. Nenne die vier wichtigsten Grundsätze, die bei Konzeption und Durchführung einer randomisierten Studie zu beachten sind.
5. Nenne drei Beispiele für kategoriale, nominale Variablen, die nicht ordinal sind.
6. Richtig oder falsch? “In einer Studie mit $N = 50$ Teilnehmern kann das Geburtsjahr als kategoriale, ordinale Variable aufgefaßt werden.” Begründe die Antwort.
7. Erstelle für die Buchstaben im eigenen Vor- und Familiennamen eine Häufigkeitstabelle.
8. Wie ist die Verteilungsfunktion F für eine metrische Variable mit endlich vielen Messungen definiert?
9. Richtig oder falsch? “ m ist genau dann Median einer metrischen Variable, wenn mindestens die Hälfte der Messungen nicht größer als m ist und mindestens die Hälfte der Messungen nicht kleiner als m ist.” Begründe die Antwort.
10. Wo kann der Interquartil-Range im Boxplot abgelesen werden?
11. Nenne die Definition der empirischen Standardabweichung.
12. Warum ist der Mittelwert empfindlich gegen Ausreißer, der Median aber nicht?
13. Wie sind die Stichproben-Mittelwerte $\hat{\mu}$ einer metrischen Variablen für unabhängige, gleichgroße Stichproben verteilt?
14. Was versteht man unter dem 95-%-Konfidenzintervall für den Stichproben-Mittelwert?
15. Unter welchen Bedingungen darf ein Konfidenzintervall mit der empirischen Standardabweichung gebildet werden?
16. In 500 unabhängigen, gleichgroßen Stichproben aus \mathcal{P} wurde jeweils das 90-%-Konfidenzintervall für den Stichproben-Mittelwert der metrischen Variable x berechnet. Wieviele dieser Intervalle enthalten den wahren Mittelwert μ von x in \mathcal{P} ?
17. Sei x eine binäre Variable. Wie lautet das 95-%-Konfidenzintervall für die relative Häufigkeit \hat{p} von $x = TRUE$ in einer Stichprobe \mathcal{S} ?
18. Was versteht man unter einer Vierfeldertafel?
19. Wie lautet die Indifferenztafel zur Vierfeldertafel $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$?
20. Unter welchen Voraussetzungen darf der Chi-Quadrat-Test für eine Vierfeldertafel durchgeführt werden?

Testatfragen zur Vorlesung “Biostatistik” (II)

21. Wie wird die Prüfgröße beim Chi-Quadrat-Test für eine Vierfeldertafel gebildet?
22. Beschreibe den Ablauf eines Signifikanztests.
23. Bilde die Alternativhypothese zu \mathcal{H}_0 : “Das Bestehen der Fahrprüfung und die sofortige Erteilung des Führerscheins sind unabhängig davon, ob es am Vorabend geregnet hat.”
24. Erkläre: Fehler I. und II. Art beim Hypothesentest.
25. Wieviele Freiheitsgrade besitzt eine $(k \times m)$ -Kontingenztafel?
26. Unter welchen Voraussetzungen darf der klassische t -Test durchgeführt werden?
27. Gegeben seien zwei Stichproben-Mittelwerte, deren 90%-Konfidenzintervalle sich nicht überschneiden. Wie groß ist die Wahrscheinlichkeit, daß die wahren Mittelwerte verschieden sind? Begründe die Antwort.
28. Definiere: gepoolte empirische Standardabweichung $\hat{\sigma}_p$.
29. Wie lautet die Prüfgröße im klassischen t -Test, und welcher Verteilung unterliegt sie?
30. Wie lautet die Alternativhypothese beim zweiseitigen t -Test?
31. Wo liegt beim zweiseitigen t -Test der Akzeptanzbereich für \mathcal{H}_0 ?
32. Richtig oder falsch? “Wenn der zweiseitige t -Test zum 5%-Niveau nicht signifikant ausfällt, stimmen die Mittelwerte μ' und μ'' in \mathcal{P}' und \mathcal{P}'' mit einer Wahrscheinlichkeit von 95 % überein.” Begründe die Antwort.
33. Wo liegt der Ablehnbereich beim einseitigen t -Test mit der Nullhypothese “ $\mu' \leq \mu''$ ”?
34. Wie verhält sich der Fehler II. Art, wenn der Stichprobenumfang anwächst?
35. Wann sollte anstatt des klassischen t -Tests der Welch-Test durchgeführt werden?
36. Erläutere: Cohen’s d und die zugehörigen Effektstärken.
37. Nenne ein Beispiel für zwei verbundene Stichproben.
38. Welche Prüfgröße wird beim t -Test für Paardifferenzen benutzt, und wie ist sie verteilt?
39. Wie lautet die Mosteller-Formel zur Berechnung der Körperoberfläche?
40. Wie ändern sich Ablehnbereich und Prüfgröße beim einseitigen t -Test, wenn eine Verschiebung v zu berücksichtigen ist?

Testatfragen zur Vorlesung

"Biostatistik" (III)

41. Beschreibe den Ablauf einer einfaktoriellen ANOVA.
42. Warum vermeidet man beim Mittelwertvergleich für drei oder mehr unabhängige Stichproben das mehrfache Testen?
43. Wie lautet die Alternativhypothese beim Globaltest?
44. Wie ist die Prüfgröße beim Globaltest verteilt?
45. Erläutere: Bonferroni-Korrektur.
46. Wie werden die Rangsummen R' , R'' beim U-Test gebildet?
47. Erläutere beim U-Test die Umrechnung von R' in U' und T' .
48. Was besagen die Shift-Hypothesen beim U-Test, und welche Folgerungen kann man aus ihnen ziehen?
49. Wie werden die Rangsummen R^- , R^+ beim Wilcoxon-Rangsummentest gebildet?
50. Nenne die Voraussetzungen für den Kruskal-Wallis-Test.
51. Was versteht man unter einer Ausgleichsgerade, und wie lauten ihre Koeffizienten?
52. Definiere: Pearsonscher Korrelationskoeffizient für zwei metrische Variablen Z und K .
53. Zwei Merkmale Z und K sind korreliert. Welche Ursachen kann es dafür geben?
54. Skizziere jeweils einen Scatterplot für Paare von überhaupt nicht, schwach positiv und perfekt negativ korrelierten Größen.
55. Nenne vier Modellvoraussetzungen für das einfache lineare Regressionsmodell für Stichproben.
56. Nenne vier Modellvoraussetzungen für das GLM für Stichproben.
57. Wie wird eine kategoriale Variable mit den vier Stufen A, B, C und D ins GLM eingebunden?
58. Welche Prüfgröße wird beim globalen Signifikanztest beim GLM für Stichproben benutzt, und wie ist sie verteilt?
59. Wie lauten die 95-%-Konfidenzintervalle für die Koeffizienten \hat{c}_0 und \hat{c}_1 im GLM für Stichproben?
60. Richtig oder falsch? "Der Koeffizient vor dem Interaktionsterm $K_{j_1} \cdot K_{j_2}$ ist das Produkt der Koeffizienten c_{j_1} und c_{j_2} vor K_{j_1} bzw. K_{j_2} ."

Testatfragen zur Vorlesung “Biostatistik” (IV)

61. Nenne das Akaike-Informationskriterium.
62. Erkläre den Aufbau eines GLM durch Rückwärtsselektion.
63. Wieviele Kovariaten sollten in ein GLM zu einer Stichprobe vom Umfang N höchstens einbezogen werden?
64. Richtig oder falsch? “Läßt man in einem GLM eine Kovariate weg, so muß der AIC-Wert zunehmen, weil damit Information verlorengeht.”
65. Was versteht man unter einem Innersubjektfaktor? Nenne zwei Beispiele.
66. Worauf werden die Korrekturfaktoren $\hat{\epsilon}_{G-G}$ bzw. $\hat{\epsilon}_{H-F}$ angewendet?
67. Beschreibe die Zerlegung der Gesamtvarianz bei der zweifaktoriellen ANOVA mit Meßwiederholung.
68. In einer zweifaktoriellen ANOVA mit Meßwiederholung ist $N = 24$, $R = 3$ und $T = 5$. Wieviele Freiheitsgrade hat $SS_{RestInnSubj}$?
69. Wie lauten die einfachen Innersubjekt Kontraste bei der zweifaktoriellen ANOVA mit 5 Meßzeitpunkten?
70. Wie lautet bei einer zweifaktoriellen ANOVA mit $N = 24$, $R = 3$ und $T = 5$ das lineare Querschnittsmodell zum Zeitpunkt t_2 ?
71. Erkläre: Prävalenz eines Merkmals in einer Population.
72. Auf welches Risikomaß bezieht sich folgende Aussage? “Das Risiko, seinen Regenschirm zu vergessen, ist für Autofahrer doppelt so groß wie für Fußgänger.”
73. Gegeben sind die Daten aus einer Vierfeldertafel für $E+/E-$ vs. $D+/D-$. Welcher Unterschied besteht zwischen den Risikomaßen RR und RRR?
74. In einer Vierfeldertafel für $E+/E-$ vs. $D+/D-$ sind nur die Zeilen- und Spaltensummen bekannt. Kann daraus eindeutig die Odds Ratio O_{E+}/O_{E-} bestimmt werden?
75. Wie lautet das 95%-Konfidenzintervall für die Odds Ratio?
76. Welche Größe wird in einem logistischen Regressionsmodell approximiert?
77. Was ist eine Sigmoidfunktion? Nenne drei Haupteigenschaften.
78. Wie werden die Koeffizienten im Modell $\text{logit}(p) = \hat{c}_0 + \hat{c}_1 K$ für eine metrische Variable K interpretiert?
79. Wie werden die Koeffizienten im Modell $\text{logit}(p) = \hat{c}_0 + \hat{c}_1 K$ für eine binäre Variable K interpretiert?
80. Wie erhält man zum Modell $\text{logit}(p) = \hat{c}_0 + \hat{c}_1 K$ die ROC-Kurve?