

Tentamen Inleiding Statistiek (wi2615)
21 januari 2011, 8.30-11.30u.

Een antwoord alleen is *niet* voldoende: er dient een berekening, toelichting en/of motivatie aanwezig te zijn. Dit alles goed leesbaar en in goed Nederlands. Succes gewenst met dit tentamen!

1. Beschouw een populatie met getallen $x_1, x_2, \dots, x_N$ en schrijf, als gebruikelijk, $\mu = \frac{1}{N} \sum_{i=1}^N x_i$ en $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$. We trekken vervolgens *met teruglegging* n keer aselekt een waarde uit de populatie en schrijven $X_1, X_2, \dots, X_n$ voor de bijbehorende stochastische variabelen.

- a. Laat zien dat

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \text{ met } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

een zuivere schatter voor σ^2 .

- b. Is $S = \sqrt{S^2}$ een zuivere schatter voor $\sigma = \sqrt{\sigma^2}$?

2. Zij X een discrete stochastische variabele met $P(X=1) = \lambda$ en $P(X=2) = 1 - \lambda$, voor een onbekende $\lambda \in [0, 1]$. Drie onafhankelijke trekkingen worden gedaan uit de verdeling van X . Deze worden genoteerd als x_1, x_2 en x_3 .

- a. Leid de momentenschatting voor λ af.
b. Leid de maximum likelihood schatter voor λ af.
c. Leid de a posteriori dichtheid voor Λ af, op basis van een uniforme a priori verdeling voor Λ

Hint: de Beta verdeling met parameters a en b heeft kansdichtheid

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{[0,1]}(x).$$

3. We willen beoordelen of data $x_1, \dots, x_n$ redelijkerwijs gemodelleerd kunnen worden als een realisatie uit een exponentiële verdeling. Geef aan hoe dit met behulp van een probability plot kan. Geef in het bijzonder aan welke punten getekend worden in de probability plot. N.B. de dichtheid van een exponentiële verdeling met parameter λ wordt gegeven door $f(x) = \lambda e^{-\lambda x}$, als $x \geq 0$.
4. In een dataset bestaande uit 25 metingen $x_1, \dots, x_{25}$ zitten 5 uitschieters. We noteren met N het aantal uitschieters dat voorkomt in een bootstrap steekproef van de 25 metingen. We gaan uit van de empirische bootstrap.
- a. Leg uit waarom de verdeling van N binomiaal is.
b. Als we in totaal 1000 bootstrap steekproeven trekken, hoe vaak verwacht je dan dat een steekproef 2 of meer uitschieters bevat?
5. Er wordt een experiment gepland om het gemiddelde van een controlegroep te vergelijken met dat van een behandelde groep. Stel dat er 25 personen zijn in iedere groep. Stel verder dat de observaties binnen een groep onafhankelijk en normaal verdeeld zijn en dat de standaardafwijking van een individuele waarneming in iedere groep gelijk is aan 5. De observaties in de controlegroep noteren we met Y_i ($1 \leq i \leq 25$), die in de behandelde groep met X_i ($1 \leq i \leq 25$). We veronderstellen verder dat observaties onafhankelijk zijn.

Z.O.Z

- a. Bereken de standaard-error van $\bar{Y} - \bar{X}$.
 - b. We toetsen de nulhypothese dat de $\mu_X = E X_i$ gelijk is aan $\mu_Y = E Y_i$ tegen het alternatief $\mu_Y > \mu_X$. Als toetsingsgrootte kiezen we $T = \bar{Y} - \bar{X}$. Wat is het kritieke gebied, bij significantieniveau 0.05? Uw antwoord mag een niet verder uitgewerkt kwantiel van de standaard normale verdeling bevatten.
 - c. Wat is het onderscheidend vermogen van de toets als $\mu_Y = \mu_X + 1$? U mag het antwoord uitdrukken in termen van de cumulatieve verdelingsfunctie Φ van de standaard normale verdeling.
 - d. Stel dat de p -waarde van de toets gelijk is aan 0.04. Verwerpen we de test bij significantieniveau 0.05?
 - e. Stel dat de p -waarde van de test gelijk is aan 0.04. Stel verder dat we als alternatieve hypothese nu $\mu_X \neq \mu_Y$ nemen. Wat wordt de p -waarde met deze alternatieve hypothese? Of is daar niets over te zeggen?
6. Stel dat we eenzelfde meting doen op 3 verschillende machines. We veronderstellen het 1-factor ANOVA-model. De j -de meting op de i -de machine noteren we met Y_{ij} . Per machine verrichten we 2 metingen. Het 1-factor model is te schrijven als

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i \in \{1, 2, 3\} \quad j \in \{1, 2\}$$

waarbij $\{\varepsilon_{ij}\}$ onafhankelijke stochastische variabelen zijn met een $N(0, \sigma^2)$ -verdeling. Noteer de vector met metingen als

$$\mathbf{Y} = (Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}, Y_{3,1}, Y_{3,2})^T.$$

- a. Schrijf het bovenstaande model als een lineair model. Dwz. in de vorm $\mathbf{Y} = X\beta + \varepsilon$.
- b. Is er een unieke kleinste-kwadraten oplossing?
- c. Met welk R-commando kun je een lineair model fitten? Met welk commando verkrijgt je de ANOVA-tabel?

Puntenverdeling:

1a:3	2a:2	2c:3	4a:3	5a:2	5c:2	5e:1	6b:3
1b:3	2b:3	3 :3	4b:3	5b:2	5d:1	6a:3	6c:2

Het tentamencijfer wordt als volgt bepaald: $\frac{\text{aantal punten} + 4}{4.3}$.