Examen statistiek I januari 2017

July 23, 2017

1. Beschouw een gerandomiseerde, placebo-gecontroleerde studie naar de cardiovasculaire veiligheid van nieuwe diabetestherapieën in patienten met type 2-diabetes.

   (a) (1.5pt) De daling in lichaamsgewicht over het verloop van de studie was gemiddeld 4.35 kg (SE 0.3 kg) groter in de semaglutide groep (1.0 mg, n=822) dan in de placebogroep (1.0 mg, n=825). Bereken een bijhorend 95%-betrouwbaarheidsinterval voor het gemiddeld verschil in de daling in lichaamsgewicht in beide groepen.

   (b) (1.5pt) Bereken een bijhorend 95%-referentieinterval voor de daling in lichaamsgewicht in de semaglutide groep (1.0 mg, n=822) wetende dat de gemiddelde daling in lichaamsgewicht in de semaglutide groep 4.0 kg bedraagt. Reken daartoe eerst uit wat de variantie is op de daling in lichaamsgewicht. U mag er daarbij van uitgaan dat de daling in lichaamsgewicht even variabel is in alle behandelingsgroepen.

   (c) (1pt) Geef een benadering voor de p-waarde (of toon ten minste hoe men die berekent) van de toets van de nulhypothese dat de gemiddelde daling in lichaamsgewicht gelijk is in beide groepen.

2. (1pt) Onderzoekers voeren 2 onafhankelijke gerandomiseerde studies uit. In beide studies gebruiken ze een ongepaarde t-test of het behandelingseffect θ_j, waarbij $j = 1, 2$ het nummer van de studie aanduidt, gelijk is aan nul. Om de nulhypothese te testen ($H_0 = \{\theta_1 = 0 \vee \theta_2 = 0\}$) tegenover de alternatieve hypothsese dat ze beiden verschillen van nul, kiezen ze als p-waarde het maximum van de p-waarden van de afzonderlijke testen van $H_0 : \theta_1 = 0$ en $H_0 : \theta_2 = 0$. Ze gebruiken deze p-waarde zoals gebruikelijk is bij toetsen. Wat is de kans dat deze test type I fouten maakt als $\theta_1 = \theta_2 = 0$. Is deze kans even groot/kleiner/groter als $\theta_1 = 0$ en $\theta_2 \neq 0$?

3. (2pt) Een gerandomiseerde behandeling is onafhankelijk van geslacht (door randomisatie), maar afhankelijk van sterfte omwille van een behandelingseffect op sterfte. Doe een gepaste berekening om te onderzoeken of behandeling en geslacht ook gegarandeerd onafhankelijk zijn in de subgroep van mensen die op het einde van de studie nog in leven zijn.

4. Bij het nemen van een steekproef van stukken land voor registratie van het voorkomen van een diersoort bekomt men in de tellingen $Y_1, ..., Y_n$ waarbij elke Y_i het aantal geobserveerde dieren van de beschouwde soort voorstelt. Onderstel dat deze Poisson is verdeeld met parameter λ.

   (a) (0.5pt) De maximum kansschatter $\hat{\lambda}_1$ van λ is gelijk aan het steekproefgemiddelde. Bepaal de exacte verdeling van deze maximum kansschatter.

   (b) (0.5pt) Stel op basis van $\hat{\lambda}_1$ een uitdrukking op voor een asymptotisch 95% betrouwbaarheidsinterval voor λ.

   (c) (1pt) Voor de eenvoud registreert men soms enkel of de diersoort al dan niet voorkomt op het beschouwde stuk land, maar niet hoe frequent. Zij n_0 het aantal Y_i's dat 0 bedraagt. Toon aan dat de maximum kans schatter van λ op basis van enkel n_0 gelijk is aan $\hat{\lambda}_2 = -\log \frac{n_0}{n}$.

   (d) (1pt) Hoe zou men de maximum kans schatter van λ kunnen bekomen als men enkel registreert of $Y_i = 0$, $Y_i = 1$, $Y_i = 2$ of $Y_i > 2$, maar men om werk te sparen niet telt hoeveel groter dan 2 de waarde van Y_i eventueel is? Het volstaat om de likelihood neer te schrijven.

   (e) (1pt) Toon aan dat $\hat{\lambda}_2$ een consistente schatter van λ is.

   (f) (1pt) Toon aan hoe de bias van $\hat{\lambda}_2$ kan berekend woden als functie van n en λ.

   (g) (1pt) Bepaal de asymptotische relatieve efficiëntie van $\hat{\lambda}_1$ en $\hat{\lambda}_2$ in functie van λ.

   (h) (1pt) Stel dat we de nulhypothese wensen te toetsen dat $H_0 : \lambda = 1$ versus $H_1 : \lambda > 1$. Toon dan aan hoe op basis van n_0 een exacte p-waarde kan worden berekend, en toon aan hoe men deze kan berekenen voor $n_0 = 3$.

   (i) (1pt) Op basis van de p-waarde berekening uit de vorige vraag, toon aan hoe de power van de bijhorende toets berekend kan worden voor een gegeven waarde van λ, bijvoorbeeld $\lambda = 1.5$. U mag hierbij vanuit gaan dat $p(n_0, n)$ een functie is die voor gegeven waarden van n_0 en n de p-waarde geeft van de test in de vorige vraag. Het is niet nodig de power uit te rekenen, wel aan te tonen hoe ze kan worden berekend.

   (j) (1pt) Veronderstel dat voor elk van de n bekomen stukken land additionele tellingen $Z_1, ..., Z_n$ worden gemaakt van een andere diersoort. Met behulp van welke toets zou u dan, mits n voldoende groot is, de gemiddelde tellingen van beide diersoorten kunnen vergelijken? Leg uit.

5. (a) (1pt) Toon aan dat als $X_n \to^d X$ en $Y_n \to^d 0$, dat dan $X_n Y_n \to^d 0$.

   (b) (1pt) Gebruik makend van bovenstaand resultaat, toon aan dat als $X_n \to^d X$ en $Y_n \to^d c$, dat dan $X_n Y_n \to^d cX$, c constant.