

Øving 9, ST1301

Oppgave 1 *Regresjon. Estimering av arvbarhet.*

a) Legg inn din egen høyde, din mors høyde, din fars høyde, og ditt kjønn via linken på fagets hjemmeside¹. Last så ned dataene fra samme link og les de inn i R ved bruk av `read.table` (bruk tilleggsargumentet `header=T` slik at kolonnene får navn). Lagre resultatet i en `data.frame` med navn `data`. Skriv så `attach(data)` for å gjøre variablene i `data.frame`n direkte tilgjengelig tilgjengelige uten bruk av `$` notasjon.

Beregn såkalt “midparental values”, d.v.s., gjennomsnittlig høyde til hvert foreldrepar og lagre resultatet i en variabel med navn `midparent`.

Lag et spredningsplot av variabelen `avkom` mot variabelen `midparent`. Velg forskjellige symboler for hvert kjønn i spredningsplottet ved å oppgi en vektor med elementer lik 1 og 2 tilsvarende hvert kjønn som tilleggsargumentet `pch`. Denne vektoren kan lages ved å konverte variabelen `kjonn` (en faktor) til en vektor av heltall med funksjonen `as.integer`. Argumentet `pch` er beskrevet på hjelpesiden til `par`.

Tilpass så en lineær modell ved hjelp av `lm` hvor du bruker høyde til `avkom` som responsvariabel og `midparent` og `kjonn` som forklaringsvariable. Undersøk den tilpassede modellen med `summary`, `confint`, og eventuelt `dummy.coef`.

Hva blir estimatet av stigningstallet i modellen? Er regresjonskoeffisienten (stigningstallet) for variabelen `midparent` signifikant forskjellig fra 0?

Det kan virke rimelig at forventet høyde til `avkom` av et gitt foreldrepar er lik snittet av høyden til hver forelder? Dette vil i tilfelle innebære at regresjonskoeffisienten er lik 1. Utfør en test av denne hypotesen. (Hint: ta utgangspunktet i estimatet av koeffisienten og standardfeilen i utskriften du får når du bruker `summary` og finne p -verdien for testen med `pt`.)

Hva blir estimatet av og konfidensintervallet for gjennomsnittlig høydeforskjell mellom kjønnene? Er høydeforskjellen statistisk signifikant? Inneholder konfidensintervallet tallet 0?

Legg til to regresjonslinjer i spredningsplottet over som svarer til forventet høyde som funksjon av `midparental value` for hvert kjønn ved hjelp to kall til `abline`.

b) Hent ut residualene fra den tilpassede modellen og undersøk om disse er normalfordelte ved å lage et histogram av disse. Lag også et Q-Q-plot med `qqnorm`. c) I populasjonsgenetisk teori antas det ofte såkalt tilfeldig parring (“random mating”) som betyr at egenskapene til et tilfeldig valgt foreldrepar er uavhengige tilfeldige variable. Dersom individene i en populasjon foretrek-

¹<http://www.math.ntnu.no/~jarlet/ST1301-2004v/hoyde.imf?forste=1>

ker partnere som er lik seg selv vil dette føre til at egenskapene blir positivt korrelerte. Dette kalles “assortative mating” og vil kunne ha evolusjonære konsekvenser, bl.a. vil det føre til en økning i den genetiske variansen i egenskapen som studeres (her høyde).

Undersøk om det er noen korrelasjon mellom høyden til foreldrene ved å bruke funksjonen `cor.test`. Lag også et spredningsplot av høyde til mor versus høyde til far.