

## 1. Estimering og hypotesetesting

La  $Y$  være en normalfordelt stokastisk variabel med forventning  $\beta x$  og varians  $\sigma^2$ . Her er  $\beta$  en ukjent parameter, og  $x$  er et valgt nivå av en forklaringsvariabel. Variansen  $\sigma^2$  antas å være kjent. Fra et forsøk har vi samlet inn  $n$  (uavhengige) parvise observasjoner  $(x_i, y_i)$  for  $i = 1, \dots, n$ .

- (a) Vis at rimelighetsfunksjonen for  $\beta$  er proporsjonal med  $e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2}$ .
- (b) Hva blir sannsynlighetsmaksimeringsestimatoren  $\hat{\beta}$  for  $\beta$ ? Vis utregning. Argumenter for at  $\hat{\beta}$  er normalfordelt og forventningsrett. Hva blir variansen?

En testobservator for  $H_0 : \beta = b$  mot  $H_1 : \beta \neq b$  er

$$Z = \frac{\hat{\beta} - b}{\sqrt{\sigma^2 / \sum_i x_i^2}}.$$

Når nullhypotesen er sann så er  $Z \sim N(0, 1)$ .

- (c) Vi forkaster  $H_0$  dersom vi observerer  $Z < -1.44$  eller  $Z > 1.44$ . Hva blir testens signifikansnivå?

Sannsynlighetskvoteprinsippet ('likelihood ratio') er en metode vi kan bruke for å konstruere en test. For denne oppgaven vil vi da ta utgangspunkt i sannsynlighetskvoten

$$\lambda = \frac{L(b)}{L(\hat{\beta})},$$

der  $L()$  er rimelighetsfunksjonen for  $\beta$ ,  $b$  er verdien for  $\beta$  under  $H_0$  og  $\hat{\beta}$  er sannsynlighetsmaksimeringsestimatoren for  $\beta$ . Sannsynlighetskvoten vil ta verdier mellom 0 og 1.

- (d) Er det verdier av  $\lambda$  nære 0 eller nære 1 som tyder på at null-hypotesen stemmer? Forklar kort. Vis at du med sannsynlighetskvoteprinsippet kan utlede samme test som i oppgave (c).

## 2. Enkel lineær regresjon

Vi har  $n = 35$  parvise observasjoner  $(x_i, y_i)$  der  $x_i = 0$  for  $i = 1, \dots, 15$  og  $x_i = 1$  for  $i = 15, \dots, 35$ . Observasjonene er illustrert i kryssplottet i Figur 1.

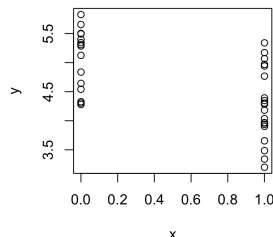


Figure 1: Kryssplott av  $n = 30$  parvise observasjoner

En enkel lineær regresjonsmodell er tilpasset ved hjelp av R og resultatet (modellsammenheng) er presentert her:

Call:

```
lm(formula = y ~ x)
```

Residuals:

|  | Min      | 1Q       | Median  | 3Q      | Max     |
|--|----------|----------|---------|---------|---------|
|  | -1.06263 | -0.38814 | 0.04095 | 0.44000 | 1.07685 |

Coefficients:

|             | Estimate | Std. Error | t value | Pr(> t )     |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 5.0590   | 0.1501     | 33.708  | < 2e-16 ***  |
| x           | -0.7959  | 0.1985     | -4.008  | 0.000328 *** |

---

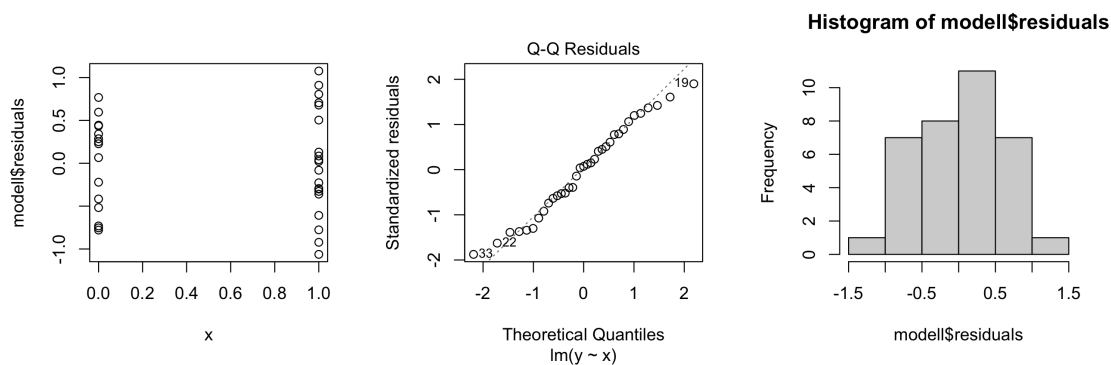
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5813 on 33 degrees of freedom

Multiple R-squared: 0.3275, Adjusted R-squared: 0.3071

F-statistic: 16.07 on 1 and 33 DF, p-value: 0.0003285

- (a) Skriv ned en formel for den enkle lineære regresjonsmodellen og hvilke antagelser som gjelder for denne. Figur 2 viser tre ulike plott av residualene  $\hat{\epsilon}_i = y_i - \hat{y}_i$  fra modelltilpasningen. Bruk en eller flere av disse, samt eventuelt Figur 1, til å kommentere modellantagelsene.

(a) Residualer mot  $x$ 

(b) Kvantil-kvantil plott

(c) Histogram

Figure 2: Residualplott

- (b) Hva blir predikert verdi for  $y$  dersom vi gjør et nytt forsøk med  $x = 1$ ? Regn ut et 95% prediksjonsintervall.

### 3. ANOVA

En forsker undersøker effekten av tre ulike gjødselsorter (A, B og C) på plantevekst (i cm). Det brukes uavhengige stikkprøver på 10 planter i hver gruppe.

Resultatene av eksperimentet er gitt i følgende tabell:

| Gjødsel | $n_j$ | $\bar{Y}_j$ | $s_j^2$ |
|---------|-------|-------------|---------|
| A       | 10    | 15.38       | 2.38    |
| B       | 10    | 12.82       | 2.59    |
| C       | 10    | 12.93       | 2.32    |

- (a) Følgende tabell er en delvis utfylt variansanalysetabell (ANOVA-tabell) hvor noe informasjon mangler (stjerner). Konstruer den komplette variansanalysetabellen og foreta en test på nullhypotesen om at forventet vekst er den samme innenfor hver kategori (gjødselsort). Skriv ned null- og alternativ-hypotese. Hva blir konklusjonen ved signifikansnivå 0.05?

| Source    | df | SS     | MS | F    |
|-----------|----|--------|----|------|
| Treatment | *  | *      | *  | 3.52 |
| Error     | *  | 160.52 | *  |      |
| Total     | *  | *      |    |      |

- (b) Forskeren ønsker å teste om den gjennomsnittlige effekten av gjødsel A og B er større enn effekten av C. Uttrykk dette som en kontrast og forklar kort når det er rimelig å utføre en slik test. Til slutt, utfør testen og rapporter  $p$ -verdien. Hva er din konklusjon?

#### 4. Ikke-parametrisk test

En klinisk forsker måler systolisk blodtrykk (mmHg) på 10 pasienter før og etter et nytt kosttilskudd. Målet er å undersøke om kosttilskuddet bidrar til å redusere blodtrykket. Målingene er:

| Pasient | Før | Etter |
|---------|-----|-------|
| 1       | 150 | 145   |
| 2       | 160 | 158   |
| 3       | 155 | 152   |
| 4       | 148 | 150   |
| 5       | 162 | 160   |
| 6       | 158 | 156   |
| 7       | 151 | 150   |
| 8       | 149 | 147   |
| 9       | 157 | 154   |
| 10      | 153 | 152   |

Forskeren mener at målingene ikke er normalfordelte og bestemmer seg derfor for å utføre ikke-parametriske tester.

- (a) To ikke-parametriske tester er tilgjengelige: fortegnstesten og Wilcoxons rank test. Angi forutsetningene for hver av de to testene og formuler nullhypotesen og den alternative hypotesen.
- (b) Utfør fortegnstesten og Wilcoxons rank test. Hva er konklusjonen hvis  $\alpha = 0.05$ ?