Skisseløsning eksamen STK1000 4. desember 2006

Oppgave 1

a). Median 10.5, gjennomsnitt 11.4. Med median som stoppetid, vil 50% av fotgjengerne ikke komme over i tide.

b). $n = 10$ uavhengige enkeltforsøk, registrerer om hver enkelt rekker over eller ikke, sannsynligheten for å ikke rekke over er lik p i hvert enkeltforsøk, X er antall som ikke rekker over blant de 10 $\Rightarrow X \sim B(10, p)$.

Estimator for p: $\hat{p} = \frac{X}{10}$

Denne estimatoren er forventningsrett fordi $\mu_{\hat{p}} = \frac{1}{10} \cdot 10 \cdot p = p$.

Estimat: $\hat{p} = \frac{2}{10} = 0.2$

c). Gitt at $p = 0.08$. Da er $X \sim B(10, 0.08)$ og

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 0.1879.$$

Oppgave 2

a). Estimator for forventning δ: $\hat{\delta} = \bar{D}$

$\mu_{\hat{\delta}} = \mu_{\bar{D}} = \delta$

$\sigma^2_{\hat{\delta}} = \sigma^2_{\bar{D}} = \frac{\sigma^2_D}{10}$

$\sigma_{\hat{\delta}} = \frac{\sigma_D}{\sqrt{10}}$

$\hat{\delta} \sim N(\delta, \frac{\sigma_D}{\sqrt{10}})$ fordi $\hat{\delta}$ er en lineærkombinasjon av D_i-ene, som er uavhengige og normalfordelte.

b).

$$H_0 : \delta = 0 \text{ mot } H_a : \delta > 0.$$

$$t = \frac{\bar{D}}{s_D/\sqrt{10}} = \frac{5.5}{2.21736} = 2.48$$

c). Antall frihetsgrader blir $10 - 1 = 9$. Fra tabellen får vi at $P = P(T \geq 2.48)$ må ligge mellom 0.01 og 0.02. Da P-verdien er såpass liten, forkaster vi nullhypotesen og påstår at bloddoping forbedrer farten ($P < 0.02$). NB! Om noen velger å ikke forkaste (nivå 0.01), er det også godtatt.

d). 1. er en parret t-test, mens 2. er en to-utvalgs-t-test. 1. brukes for å sammenligne behandlinger utført på samme individer, mens 2. brukes til å sammenligne behandlinger på to uavhengige grupper. Her vil 1. være riktig fordi løperne er de samme. Ved å bruke differansene reduserer vi effekten av individuell variasjon.

$P = 0.017$ fra utskriften.

Oppgave 3

a). Observasjonspar (x_i, Y_i) av temperatur og frekvens, $i = 1, ..., 15$. Enkel lineær regresjons-modell:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

der β_0 er skjæringspunktet med y-aksen og β_1 er stigningskoeffisienten. Vi antar at ϵ_i er uavhengige og normalfordelte $N(0, \sigma)$ for alle i.

Parametre i modellen er β_0, β_1 og σ, estimatene for disse finner vi fra utskriften som $b_0 = 6.47$, $b_1 = 0.381$ og $S = 0.971518$.

b). Plottet av data og tilpasset regresjonslinje viser at en lineær modell kan passe rimelig bra. R-Sq er andelen av variasjonen i Y som forklares av den lineære sammenhengen, og en slik andel på 69.7% er ok. Residualplottet kan derimot gi mistanke om at sammenhengen ikke er lineær! Normalantakelsen virker OK fra normalfordelingsplottet for residualene (en tilnærmet rett linje).

c). Ønsker et 95% konfidensintervall for forventet frekvens ved 31 grader, dvs. for

$$\mu_Y = \beta_0 + \beta_1 \cdot 31.$$

Fra utskriften finner vi intervallet (17.452, 19.143). Antall frihetsgrader er $15-2 = 13$. t^* for 13 frihetsgrader og C=0.95 er 2.16.