
Oblig 3c - Lineær regresjon med usikkerhet

Levering: 1 PDF, i rett mappe på Canvas. Lever eventuell [R](#)/[MatLab](#)/[Wolfram](#)-kode som kildefil i tillegg.

Førstefrist: 24. apr., 15:00 **Merk tiden!**

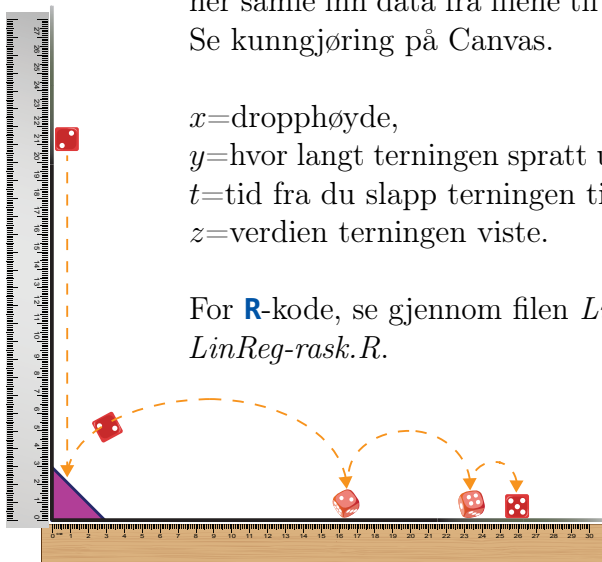
Sistefrist: 1. mai, 15:00

1. kap. 17: oppgave 1.c
2. kap. 17: oppgave 1.d
3. **Terningdropp-oppgaven:** (Totalt 50%)

Som med seigmennene i oblig 3a, skal dere her samle inn data fra filene til alle grupper. Se kunngjøring på Canvas.

x =dropphøyde,
 y =hvor langt terningen spratt ut fra veggen
 t =tid fra du slapp terningen til den lå i ro
 z =verdien terningen viste.

For [R](#)-kode, se gjennom filen *LinReg.R* eller *LinReg-rask.R*.



Kreditering: Terninger fra [all-free-download.com](#). Linjaler fra [fuzzimo.com](#).

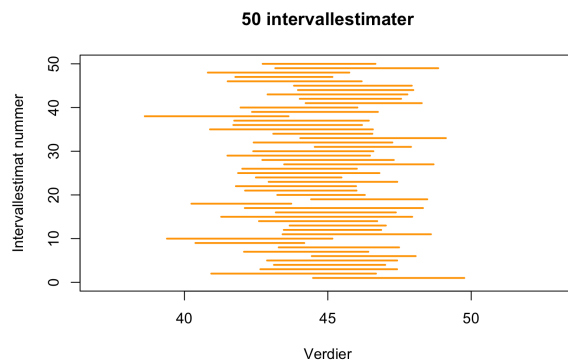
- Tegn et diagram med samtlige datapunkter, og legg på den lineære regresjonslinjen.
- Bruk nøytrale prior hyperparametre, og finn *posterior* og *prediktive* sannsynlighetsfordelinger, det vil si, sannsynlighetsfordelinger for τ , σ , b , $y(x)$ og $Y_+(x)$.
- Finn et 80% kredibilitetsintervall (intervallestimat) for stigningstallet b .
- Finn et 80% kredibilitetsintervall (intervallestimat) for standardavviket σ .
- Finn et 80% kredibilitetsintervall (intervallestimat) for $y(x)$.
- 80% intervallestimatet for $y(x)$ er funksjoner av x , og en kurve over, og en under regresjonslinjen. Plott disse kurvene inn sammen med regresjonslinjen.

-
- (g) Finn verdien $R^2 = \frac{SS_y - SS_e}{SS_y}$. Dette tallet forteller hvor stor del av variasjonen i y som kan forklares av linja $y = a + bx$. Det finnes (minst) 2 måter å regne ut dette på.
- (h) Finn R^2 for regresjonen mellom z (utfall på terningen) og x (dropphøyde), og også for regresjonen mellom t (tid fra du slapp terningen til den lå i ro) og x (dropphøyde). Kommenter hva de forskjellige R^2 sier oss, og hvorfor du tror de har de verdiene de har.

4. Følgende R-kode vil plukke ut et utvalg av 20 av observasjonene.

```
antall_rader = dim(dropp_df)[1]
N = 20
utvalg = sort(sample(1:antall_rader,N)) # Sortering er ikke nødvendig
utvalg                                     # men du får da se hvilke rader som er plukket
ny_dropp_df = dropp_df[utvalg,]          # Dette er kjernen; plukker ut radene
rownames(ny_dropp_df)=1:N                # Lurt hvis du skal kjøre for-løkke.
ny_dropp_df                             # Ikke nødvendig, men du får se den nye data
```

- (a) Kjør 50 runder med $N = 5$, for å finne regresjonslinjen for de 5 punktene. Tegn de 50 regresjonslinjene sammen, i samme graf. Hva ser du?
- (b) Som forrige oppgave, men med N henholdsvis lik 15, 50 og 200. Hva ser du?
- (c) Gjør oppgave 3c for 50 runder med $N = 5$, og tegn de 50 intervallestimatene på samme måte som i grafen under.



- (d) Gjør nå det samme med N henholdsvis lik 15, 50 og 200. Hva ser du?
- (e) Gjør illustrasjonene fra oppgave 3f for lineær regresjon laget ut fra utvalg med $N = 5, 15, 50$ og 200. Hva ser du?