

Oblig 3c

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

Førstefrist: 9. april., 18:00

Sistefrist: 16. april., 18:00

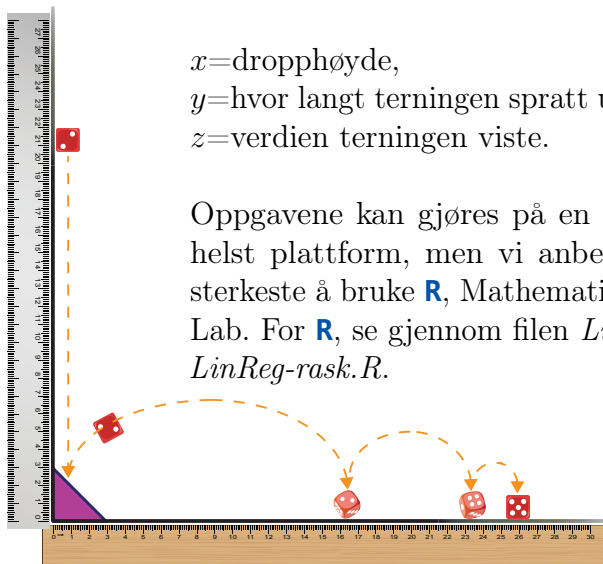
Godkjent: 40% + (antall i gruppa) * 10%

1. (15%) kap. 17: oppgave 1.c
2. (15%) 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
 z =verdien terningen viste.

Opgavene kan gjøres på en hvilken som helst plattform, men vi anbefaler på det sterkeste å bruke [R](#), Mathematica eller MatLab. For [R](#), se gjennom filen *LinReg.R* eller *LinReg-rask.R*.



Kreditering: Terninger fra all-free-download.com. Linjaler fra fuzzimo.com.

- (a) (5%) Tegn et diagram med samtlige datapunkter, og legg på den lineære regresjonslinjen.
- (b) (15%) Bruk nøytrale prior hyperparametre, og finn *posterior* og *prediktive* sannsynlighetsfordelinger, det vil si, sannsynlighetsfordelinger for τ , b , $y(x)$ og $Y_+(x)$.
- (c) (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for stigningstallet b .
- (d) (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for standardavviket σ . (Hint: Bruk verdiene fra τ og regn om ved å bruke at $\tau = \frac{1}{\sigma^2}$)
- (e) (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for $y(x)$.

-
- (f) (5%) 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) (5%) 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 $y = a + bx$. For de av dere som bruker dataverktøy for å finne dette: angi hvordan dere fant det.
- (h) (5%) Finn R^2 for regresjonen mellom z (utfall på terningen) og x (dropphøyde). Kommenter hva forskjellen mellom R^2 for y og R^2 for z sier oss.

4. (Totalt 20%) Følgende **R**-kode vil plukke ut et utvalg av observasjonene.

```
antall_rader = dim(dropp_df)[1]
N = 20 # (Eksempel; se oppgavene)
utvalg = sort(sample(1:antall_rader,N)) # Sortering er ikke nødvendig
utvalg # men du får da se hvilke rader som er plukket ut
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-rammen.
```

- (a) (5%) Kjør 50 runder, og bruk $N = 15$. For hver runde, gjør oppgave 3a, men tegn regresjonslinjene sammen, i samme graf. Hva ser du?
- (b) (5%) Kjør en runde med N henholdsvis lik 5, 15, 50 og 200. For hver runde, gjør oppgavene 3c og 3d. Hva ser du?
- (c) (10%) Kjør en runde med N henholdsvis lik 5, 15, 50 og 200. For hver runde, gjør oppgaven 3f. Tegnes i hvert sitt diagram. Hva ser du?