Oblig 3c

Gormery K. Wanjiru

20. april 2024

Innhold

1 (15%) kap. 17: oppgave 1.c

ikke gjørt

2 (15%) kap. 17: oppgave 1.d

ikke gjørt

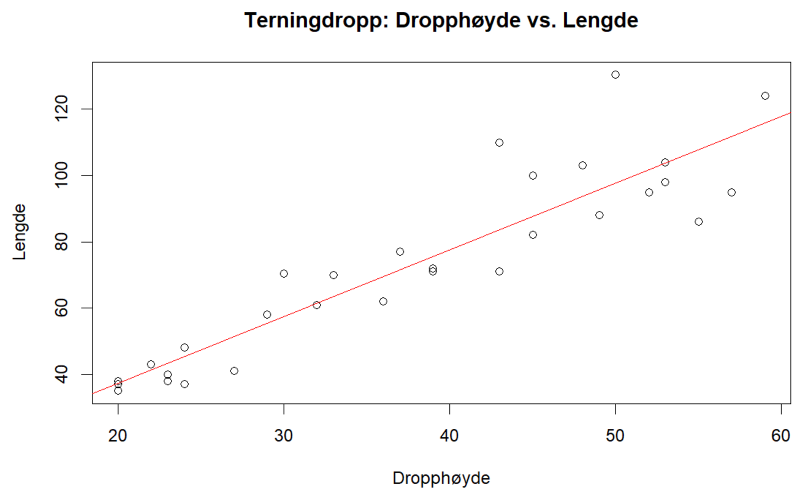
3 Terningdropp-oppgaven: (Totalt 50%)

3.1 (5%) Tegn et diagram med samtlige datapunkter, og legg på den lineære regresjonslinjen.

3.1.1 R kode

```
# Les inn data fra CSV-filen
data <- read.csv('terningDropp.csv')

# Utfør lineær regresjon
fit <- lm(Lengde ~ Dropp, data=data)

# Lag plot med datapunkter og regresjonslinje
plot(data$Dropp, data$Lengde, xlab='Dropphøyde', ylab='Lengde', main='Terningdro
abline(fit, col='red')
```

SVAR:



Figur 1: (3a)

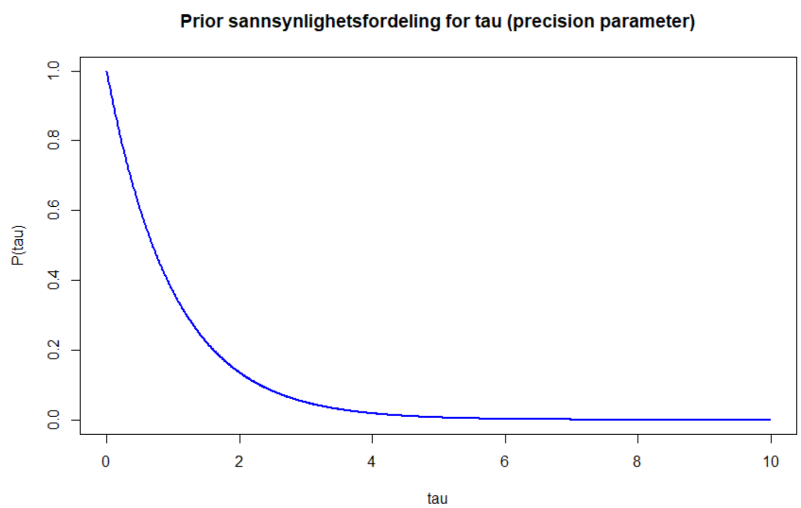
3.2 (15%) Bruk nøytrale prior hyperparametre, og finn posterior og prediktive sannsynlighetsfordelinger, det vil si, sannsynlighetsfordelinger for τ, b, $y(x)$ og $Y^+(x)$.

3.2.1 R kode

```
# Prior hyperparametre
alpha <- 1
beta <- 1

# Likelihood hyperparametre
mu0 <- 0
sigma0 <- 1

# Beregn posterior hyperparametre
n <- length(data$Lengde)
x_bar <- mean(data$Dropp)
s_xx <- sum((data$Dropp - mean(data$Dropp))^2)

alpha_post <- alpha + n/2
beta_post <- beta + 1/2 * s_xx
```

```
 # Definer sannsynlighetsfordelingen for tau
tau_values <- seq(0.001, 10, by = 0.01)
prior_tau <- dgamma(tau_values, shape = alpha, scale = beta)

# Plot sannsynlighetsfordelingen for tau
plot(tau_values, prior_tau, type = "l", col = "blue", lwd = 2, xlab = "tau", yla
     main = "Prior sannsynlighetsfordeling for tau (precision parameter)")
```

SVAR:



Figur 2: (3b-v2)

3.3 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for stigningstallet b.

3.3.1 R kode

```
# Bruk sannsynlighetsfordelingen fra oppgave 3b
alpha_post <- alpha + n/2
beta_post <- beta + 1/2 * s_xx

# Generer posterior for tau (gamma-fordeling)
```

```
posterior_tau <- rgamma(10000, shape = alpha_post, rate = beta_post)

# resultater
print(paste("Posterior for tau (precision parameter): Gamma(", alpha_post, ",",
```

SVAR:

```
[1] "Posterior for tau (precision parameter): Gamma( 16 , 2326.33333333333 )"
```

Figur 3: (3c-v2)

3.4 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for standardavviket σ. (Hint: Bruk verdiene fra τ og regn om ved å bruke at $\tau = \frac{1}{\sigma^2}$)

3.4.1 R kode

```
# Bruk sannsynlighetsfordelingen fra oppgave 3b
lower_bound_sigma <- sqrt(1 / qgamma(alpha/2, shape = alpha_post, scale = 1/beta
upper_bound_sigma <- sqrt(1 / qgamma(1 - alpha/2, shape = alpha_post, scale = 1/

# Print resultatet
print(paste("80% kredibilitetsintervall for standardavviket sigma: [", lower_bou
```

SVAR:

```
[1] "80% kredibilitetsintervall for standardavviket sigma: [ 12.1851299975221 , 12.1851299975221 ]"
```

Figur 4: (3d-v2)

3.5 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for $y(x)$.

3.5.1 R kode

```
# Beregn kredibilitetsintervall for y(x)
alpha <- 0.2  # 80% kredibilitet
t_critical <- qt(1 - alpha/2, df=n-2)
```

```
lower_bound_yx <- pred_mean - t_critical * sqrt(pred_var)
upper_bound_yx <- pred_mean + t_critical * sqrt(pred_var)

print(paste("80% kredibilitetsintervall for y(x): [", lower_bound_yx, ",", upper
```

SVAR:

```
[1] "80% kredibilitetsintervall for y(x): [ 60.6172595799277 , 91.0537075454546 ]"
```

Figur 5: (3e)

3.6 (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.

slitte litt med å plotte

3.7 (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.

3.7.1 R kode

```
# Beregn R^2
SS_y <- sum((data$Lengde - mean(data$Lengde))^2)
SS_e <- sum(residuals(fit)^2)
R_squared <- (SS_y - SS_e) / SS_y

print(paste("Verdien av R^2:", R_squared))
```

SVAR:

```
[1] "Verdien av R^2: 0.834592323900881"
```

Figur 6: (3g)

3.8 (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.

3.8.1 R kode

```
# Utfør lineær regresjon for z vs. x
fit_z <- lm(Verdi ~ Dropp, data=data)

# Beregn R^2 for z
SS_z <- sum((data$Verdi - mean(data$Verdi))^2)
SS_e_z <- sum(residuals(fit_z)^2)
R_squared_z <- (SS_z - SS_e_z) / SS_z

print(paste("Verdien av R^2 for z:", R_squared_z))
print("Forskjellen mellom R^2 for y og R^2 for z indikerer hvor mye av variasjon
```

SVAR:

```
> print(paste("Verdien av R^2 for z:", R_squared_z))
[1] "Verdien av R^2 for z: 0.0335735339514531"
> print("Forskjellen mellom R^2 for y og R^2 for z indikerer hvor mye av variasjonen
i utfallet på terningen (z) og lengden (y) som forklares av modellen.")
[1] "Forskjellen mellom R^2 for y og R^2 for z indikerer hvor mye av variasjonen i ut
fallet på terningen (z) og lengden (y) som forklares av modellen."
```

Figur 7: (3i)

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

4.1 (5%) Kjør 50 runder, og bruk $N = 15$. For hver runde, gjør oppgave 3a, men tegn regresjonslinjene sammen, i samme graf. Hva ser du?

4.2 (5%) Kjør en runde med N henholdsvis lik 5, 15, 50 og 200. For hver runde, gjør oppgavene 3c og 3d. Hva ser du?

4.3 (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?

Vedlegg

Vedlegg A

Referanser

[1] https://tma4245.math.ntnu.no/viktige-diskrete-fordelinger/poissonprosess-og-poissonfordeling *NTNU*