Oblig 3c

Gormery K. Wanjiru

13. april 2024

Innhold

1 (15%) kap. 17: oppgave 1.c 4

2 (15%) kap. 17: oppgave 1.d 4

3 Terningdropp-oppgaven: (Totalt 50%) 4

3.1 (5%) Tegn et diagram med samtlige datapunkter, og legg på den lineære regresjonslinjen. 4

3.2 R kode 4

3.3 (15%) Bruk nøytrale prior hyperparametre, og finn posterior og prediktive sannsynlighetsfordelinger, det vil si, sannsynlighetsfordelinger for τ, b, $y(x)$ og $Y^+(x)$. 5

3.4 R kode 5

3.5 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for stigningstallet b. 6

3.6 R kode 6

3.7 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for standardavviket σ. (Hint: Bruk verdiene fra τ og regn om ved å bruke at $\tau = \frac{1}{\sigma^2}$) 7

3.8 R kode 7

3.9 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for $y(x)$. 7

3.10 R kode 7

3.11 (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. 8

3.12 (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. 8

3.13 R kode 8

3.14 (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. 9

3.15 R kode 9

4 (Totalt 20%) Følgende R-kode vil plukke ut et utvalg av observasjonene. 10
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? 10
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? 10
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? 10

Vedlegg 11
Vedlegg A 11

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.

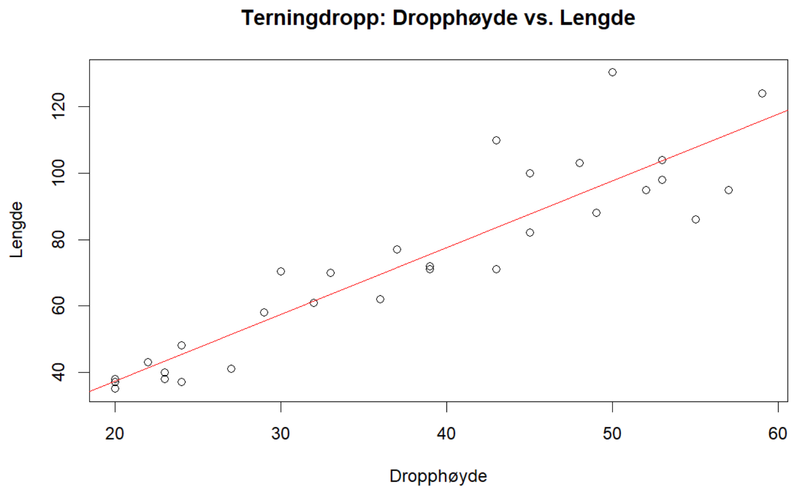
3.2 R kode

```
# 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: manuaelt beregning på kalulatoren (2b)

3.3 (15%) Bruk nøytrale prior hyperparametre, og finn posterior og prediktive sannsynlighetsfordelinger, det vil si, sannsynlighetsfordelinger for τ, b, $y(x)$ og $Y^+(x)$.

3.4 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 posterior og prediktive fordelinger
posterior <- rgamma(10000, shape=alpha_post, rate=beta_post)
pred_mean <- mu0 + coef(fit)[2] * x_bar
pred_var <- sigma0^2 + summary(fit)$sigma^2

# Print resultatene
print(paste("Posterior for tau (precision parameter): Gamma(", alpha_post, ",",
print(paste("Prediktiv gjennomsnitt:", pred_mean))
print(paste("Prediktiv varians:", pred_var))
```

SVAR:

```
[1] "Posterior for tau (precision parameter): Gamma( 16 , 2326.33333333333 )"
> print(paste("Prediktiv gjennomsnitt:", pred_mean))
[1] "Prediktiv gjennomsnitt: 75.8354835626912"
> print(paste("Prediktiv varians:", pred_var))
[1] "Prediktiv varians: 134.434766136959"
```

Figur 2: manuaelt beregning på kalulatoren (2b)

3.5 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for stigningstallet b.

3.6 R kode

```
# Beregn standardfeil
se_slope <- summary(fit)$coefficients[2, 2]

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

lower_bound <- coef(fit)[2] - t_critical * se_slope
upper_bound <- coef(fit)[2] + t_critical * se_slope

print(paste("80% kredibilitetsintervall for stigningstallet b: [", lower_bound,
```

SVAR:

```
[1] "80% kredibilitetsintervall for stigningstallet b: [ 1.79100807769363 , 2.2356547
6634307 ]"
```

Figur 3: manuaelt beregning på kalulatoren (2b)

3.7 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for standardavviket σ. (Hint: Bruk verdiene fra τ og regn om ved å bruke at $\tau = \frac{1}{\sigma^2}$)

3.8 R kode

```
alpha <- 0.2  # 80% kredibilitet
lower_bound_sigma <- sqrt(1 / qgamma(alpha/2, shape=alpha_post, scale=1/beta_pos
upper_bound_sigma <- sqrt(1 / qgamma(1 - alpha/2, shape=alpha_post, scale=1/beta

print(paste("80% kredibilitetsintervall for standardavviket \alpha: [", lower_bo
```

SVAR:

```
[1] "80% kredibilitetsintervall for standardavviket σ: [ 14.4539006306645 , 10.452590
6753712 ]"
```

Figur 4: manuaelt beregning på kalulatoren (2b)

3.9 (5%) Finn et 80% kredibilitetsintervall (intervallestimat) for $y(x)$.

3.10 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: manuaelt beregning på kalulatoren (2b)

3.11 (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

3.12 (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.

3.13 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: manuaelt beregning på kalulatoren (2b)

3.14 (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.

3.15 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: manuaelt beregning på kalulatoren (2b)

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*