

Oblig 1b: Analyse av Terningdropp

Gormery Kombo

1. februar 2024

1 Introduksjon

I denne rapporten analyserer jeg data samlet fra et eksperiment hvor en terning ble droppet fra ulike høyder. Jeg undersøker sammenhengen mellom høyden terningen ble sluppet fra (Dropp) og lengden den rullet (Lengde) etter å ha truffet bakken. Analysen inkluderer regresjonsanalyse og residualanalyse for å forstå forholdet mellom disse variablene.

2 Oppgave 2

2.1 2a: Datainnsamling og Analyse

Dataene som ble samlet inn er tilgjengelige i filen `terningDropp.csv`. Dataene ble analysert ved hjelp av R for å utføre statistiske beregninger og visuelle representasjoner.

2.2 2b: Regresjonsanalyse for de Første 5 Målingene

Først ble det utført en regresjonsanalyse på de første fem målingene for å estimere sammenhengen mellom 'Dropp' og 'Lengde'. Resultatene gir oss en innsikt i hvordan lengden varierer med høyden i de første forsøkene.

```
1 lm_first5 <- lm(Lengde ~ Dropp, data=df[1:5, ])  
2 summary(lm_first5)
```

Manuell beregning av regresjonslinjen: bruker følgende formel for manuell beregning:

$$\beta = \frac{n \sum(xy) - \sum x \sum y}{n \sum(x^2) - (\sum x)^2}$$
$$\alpha = \bar{y} - \beta \bar{x}$$

Beregner summene:

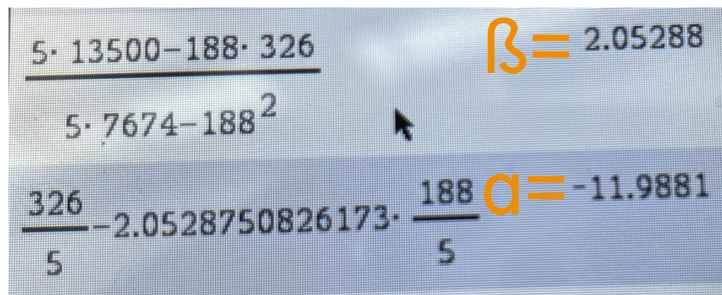
$$\begin{aligned}\sum x &= 188 \\ \sum y &= 326 \\ \sum xy &= 13500 \\ \sum x^2 &= 7674\end{aligned}$$

Deretter beregner vi β og α :

$$\begin{aligned}\beta &= \frac{5 \times 13500 - 188 \times 326}{5 \times 7674 - 188^2} \\ \alpha &= \frac{326}{5} - \beta \times \frac{188}{5}\end{aligned}$$

Sammenligning: Sammenlign de manuelt beregnede verdiene av α og β med de fra R-utdataen for å sikre nøyaktighet. R-koden gir oss $\beta = 2.0529$ og $\alpha = -11.9881$.

og manuelt beregnet gir oss: $\beta = 2.0529$ og $\alpha = -11.9881$.



Figur 1: manuelt beregning på kalkulatoren (2b)

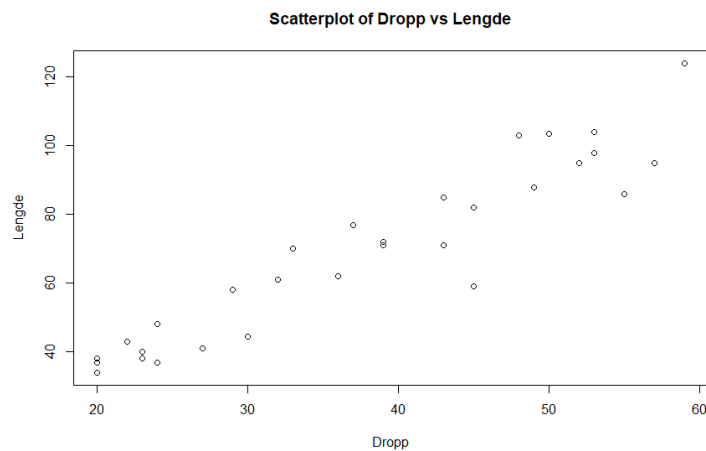
2.3 2c: Regresjonsanalyse for Hele Datasettet

Deretter ble en lignende regresjonsanalyse utført på hele datasettet for å se om det samme mønsteret holder for en større mengde data.

```
1 lm_full <- lm(Lengde ~ Dropp, data=df)
```

2.4 2d: Spreddiagram av Data

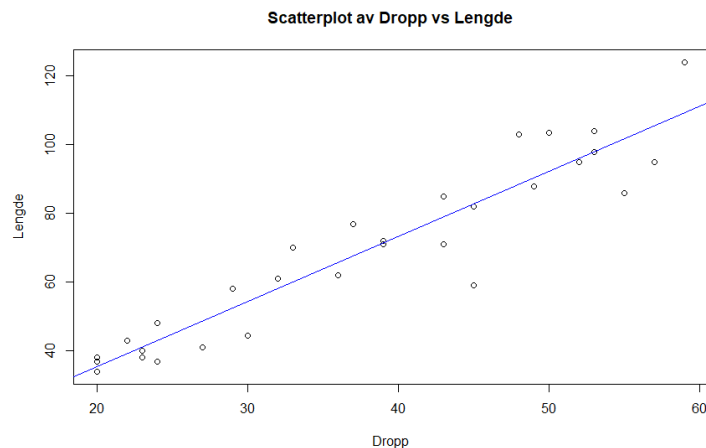
Følgende spreiddiagram viser alle datapunktene (Dropp mot Lengde) fra eksperimentet. Dette gir oss en visuell representasjon av datadistribusjonen.



Figur 2: Spreddiagram av Dropp mot Lengde (2d)

2.5 2e: Spreddiagram med Regresjonslinje

Dette spreiddiagrammet inkluderer regresjonslinjen som viser den estimerte lineære sammenhengen mellom 'Dropp' og 'Lengde' for hele datasettet.



Figur 3: Spreddiagram med regresjonslinje (2e)

2.6 2g: Sum av Kvadrerte Residualer (SSe)

Summen av de kvadrerte residualene (SSe) gir oss en kvantitativ måling av hvor godt regresjonsmodellen passer dataene. Lavere verdier indikerer en bedre

passform.

```
1 ssr_first5 <- sum(residuals(lm_first5)^2)
2 ssr_full <- sum(residuals(lm_full)^2)
```

2.7 2g: Sum av Kvadrerte Residualer (SSe)

```
1 ssr_first5 <- sum(residuals(lm_first5)^2)
2 ssr_full <- sum(residuals(lm_full)^2)
```

SSR (2g): SSR gir en kvantitativ måling av hvor godt regresjonsmodellen passer dataene. Lavere verdier indikerer en bedre passform.

Beregning:

$$\text{SSR for de første 5 målingene: } \sum_{i=1}^5 (y_i - \hat{y}_i)^2 = 748.808$$

$$\text{SSR for hele datasettet: } \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 2114.794$$

Her representerer y_i de faktiske observasjonene, \hat{y}_i er de estimerte verdiene fra regresjonsmodellen, og n er antall observasjoner.

2.8 2h: Standardfeil (se)

Standardfeilen (SE) gir oss en idé om usikkerheten i estimatet av regresjonskoeffisientene. Den hjelper oss å forstå hvor nøyaktig vår lineære modell estimerer den avhengige variabelen.

```
1 se_first5 <- sqrt(ssr_first5 / lm_first5$df.residual)
2 se_full <- sqrt(ssr_full / lm_full$df.residual)
```

Standardfeil (2h): SE gir en indikasjon på nøyaktigheten i estimatene av regresjonskoeffisientene.

Beregning:

$$\text{SE for de første 5 målingene: } SE = \sqrt{\frac{\text{SSR}}{df}} = \sqrt{\frac{748.808}{df.residual_first5}} = 15.79882$$

$$\text{SE for hele datasettet: } SE = \sqrt{\frac{\text{SSR}}{df}} = \sqrt{\frac{2114.794}{df.residual_full}} = 8.690704$$

Her representerer `df.residual_first5` og `df.residual_full` frihetsgradene for residualene for henholdsvis de første 5 målingene og hele datasettet.

2.9 2i: Lagring av Arbeid (for meg selv)

Jeg Lagret det utførte arbeidet som en r prosjekt på github:

<https://github.com/valiantlynx/statistics-course/tree/main/statistikk-oblig-1b>