

## Skilaverkefni 3

*Brynjólfur Gauti Jónsson*  
*Pórarinn Jónmundsson*

### Dæmi 1

Ný aðferð til að mæla fosfór í jarðvegi er kynnt til sögunnar. Úrtak af 11 jarðvegssýnum úr jarðvegi með raunverugeu fosfórmagni  $548 \text{ mg/kg}$  er greint með nýju aðferðinni. Úrtaksmeðaltalið reynist vera 587 og úrtaksstaðalfrávik 10.

(1) Gefa gögnin til kynna að nýja aðferðin mæli fosfórmagnið rétt? Nota skal núlltilgátupróf með marktektarkröfu  $\alpha = 0.05$ .

Við ætlum að prófa  $\mu$  sem er hið sanna meðaltal þýðisins og gerum ráð fyrir því að meðaltalið komi úr normaldreifingunni (skv T3\_Nulltilgatu.pdf). Tilgátuprófið er

$$H_0 : \mu = 548 \quad \text{v.s.} \quad H_1 : \mu \neq 548.$$

Þar sem hið sanna staðalfrávik er ekki gefið þurfum við að styðjast við  $t$ -próf. Prófstærðin er:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

Hér er  $\bar{x}$  úrtaksmeðaltalið,  $s$  úrtaksstaðalfrávik,  $\mu_0$  núlltilgátugildið sem við viljum prófa og  $n$  fjöldi mælinga. Í okkar tilfelli verður því jafnan að ofan:

$$t = \frac{587 - 548}{10/\sqrt{11}} = \frac{39}{10/\sqrt{11}} \approx 12.935$$

Þar sem gagntilgátuprófið okkar er tvíhliða þá höfnum við  $H_0$  ef  $|t| > t_{\alpha/2, n-1}$ . Þar sem  $\alpha = 0.05$  er  $t_{0.025, 10} = 2.228$  og því ljóst að

$$|t| = |12.935| > 2.228 = t_{0.025, 10}.$$

Við höfnum því núlltilgátunni  $H_0$  undir marktektarkröfunni  $\alpha = 0.05$  og samþykkjum gagntilgátuna undir sömu marktektarkröfu.

(2) Hverju þurfum við að gera ráð fyrir til að núlltilgátuprófið í (1) verði viðeigandi?

Að gögnin komi úr normaldreifingunni og fyrst úrtakið er lítið að gögnin séu án útlaga eða séu ekki skeif.

## Dæmi 2

Framleiðslufyrirtæki þarf að velja á milli tveggja birgja sem selja örflögur í tölvur. Slembiúrtak af 200 örflögum frá birgi  $A$  eru kannaðar og 8 þeira reyndust vera gallaðar, á meðan 13 í slembiúrtaki af 250 örflögum frá birgi  $B$  voru gallaðar. Notið núlltilgátu til að meta hvort þessi niðurstaða ætti að hafa áhrif á val fyrirtækisins á birgi með marktektarkröfu  $\alpha = 0.01$ .

### Lausn:

Við leysum dæmið út frá tvíkosta (*e. binomial*) dreifingunni. Þar sem úrtakið er nægilega stórt notum við normaldreifinguna til að nálgast prófstærð miðað við núlltilgátuna  $H_0 : p_A = p_B$ .

Höfum að  $\hat{p}_A = \frac{8}{200}$ ,  $\hat{p}_B = \frac{13}{250}$ ,  $\hat{p} = \frac{8+13}{200+250} = \frac{21}{450}$ . Prófstærðin okkar verður

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}} = -0.6$$

Flettum upp  $p$ -gildi prófstærðar með R skipuninni `2 * dnorm(-0.6)` og fáum  $p = 0.67$ . Við getum því ekki sagt að munur sé á birgjum og því ætti niðurstaðan ekki að hafa áhrif á val fyrirtækisins.

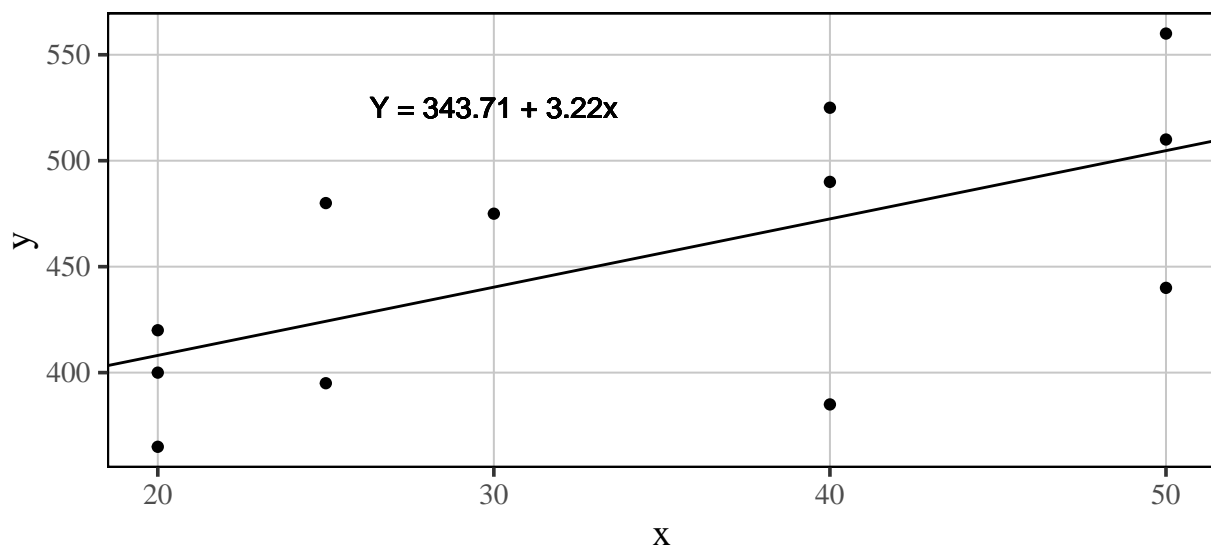
p # Dæmi 3

Í þessu dæmi á að skila R kóða ásamt svörum. Rannsókn var gerð til að ákvarða samband milli vikulegra útgjalda til auglýsinga og söluhagnaðs. Niðurstaðan var eftirfarandi:

Útgjöld til auglýsinga [mISK]	Söluhagnaður [mISK]
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

(a) Notið R til að gera línulega aðhvarfsgreiningu á gögnunum með því að nota formúlurnar í bókinni beint og ákvarðið þannig jöfnu matlínunnar fyrir  $Y = \beta_0 + \beta_1 x$ . Teiknið gögnin.

```
x <- data %>% mutate(x0 = 1) %>% select(x0, x) %>% as.matrix
y <- as.matrix(data$y)
hat_matrix <-
betas <- solve(t(x) %*% x) %*% t(x) %*% y
fmla <- paste0("Y = ", round(betas[1], 2), " + ", round(betas[2], 2), "x")
ggplot(data, aes(x, y)) + geom_point() +
  geom_abline(intercept = betas[1], slope = betas[2]) +
  geom_text(aes(x = 30, y = 525, label = fmla))
```



(b) Ákvarðið 95% öryggisbil fyrir  $\beta_1$  með því að reikna sjálf í R.

```
e <- y - (x %>% betas)
sigma2 <- ((t(e) %>% e) / (nrow(y) - nrow(betas))) %>% as.vector
var_beta <- solve(t(x) %>% x) %>% diag %>% as.vector %>% (function(x) x * sigma2)
se_beta <- sqrt(var_beta)

data_frame(term = c("Skurðpunktur", "x"),
            beta = as.vector(betas),
            se = se_beta) %>%
  mutate(lower = beta + qt(0.025, nrow(y)) * se,
         upper = beta + qt(0.975, nrow(y)) * se) %>%
  set_names(c("Breyta", "Stuðull", "Staðalvilla", "Neðri", "Efri")) %>%
  kable(format = "latex", booktabs = TRUE, digits = 3, align = c("l", rep("c", 4))) %>%
  kable_styling() %>%
  add_header_above(c("", "", "", "95% Öryggisbil" = 2))
```

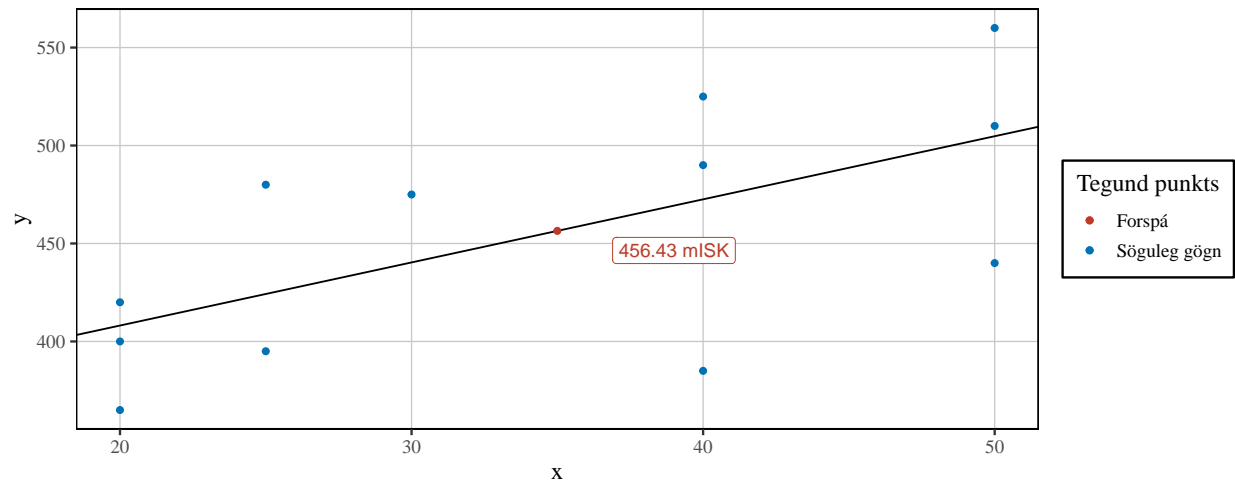
Breyta	Stuðull	Staðalvilla	95% Öryggisbil	
			Neðri	Efri
Skurðpunktur	343.706	44.766	246.168	441.243
x	3.221	1.240	0.520	5.922

(c) Notið fallið  $\text{lm}(y \sim x)$  og staðfestið reikningana í (a) og (b).

```
mod <- lm(y ~ x, data = data)
mod %>%
  tidy %>%
  select(term, estimate, std.error) %>%
  mutate(lower = estimate + qt(0.025, nrow(y)) * std.error,
         upper = estimate + qt(0.975, nrow(y)) * std.error) %>%
  set_names(c("Breyta", "Stuðull", "Staðalvilla", "Neðri", "Efri")) %>%
  kable(format = "latex", booktabs = TRUE, digits = 3) %>%
  kable_styling(full_width = TRUE) %>%
  add_header_above(c("", "", "", "95% Öryggisbil" = 2))
```

Breyta	Stuðull	Staðalvilla	95% Öryggisbil	
			Neðri	Efri
(Intercept)	343.706	44.766	246.168	441.243
x	3.221	1.240	0.520	5.922

(d) Spáið fyrir vikulegum söluhagnaði þegar útgjöld til auglýsinga eru 35 milljónir ISK.



(e) Á 5% prófsstigi, getum við dregið þá ályktun að auglýsingar hafi áhrif á sölur?

Þar sem öryggisbil fyrir hallastuðul við auglýsingar inniheldur ekki núll getum við ályktan með 95% vissu að áhrif auglýsinga á sölu séu ekki núll.

(Auka) Hvernig skal túlka  $\beta_0$  og  $\beta_1$  fyrir þessi gögn?

$\beta_0$  er vænt sölumagn þegar engu fjármagni er varið í auglýsingar. Hafa ber varann á þar sem úrtak inniheldur ekki mælingar á sölu fyrir auglýsingaútgjöld minni en 20 mISK eða meiri en 50 mISK. Mælt er með að heimfæra ekki niðurstöður greiningar á auglýsingafjármagn langt utan þess bils.

$\beta_1$  er vænt aukning í sölumagni þegar vikulegum útgjöldum til auglýsinga er aukið um 1 mISK.