

# Lösningar

## Tentamen i Statistisk analys, 17 augusti 2026

---

### Uppgift 1

- a) Falskt
- b) Falskt
- c) Sant
- d) Sant
- e) Sant

### Uppgift 2

- a) Vi får  $\bar{x} = 4.0$  och  $s^2 = (1/(6-1)) \sum_i (x_i - 4.0)^2 = 0.2((-0.6)^2 + (-0.3)^2 + 0.3^2 + 0.7^2 + (-0.1)^2 + 0) = 0.208$ , så  $s = \sqrt{0.208} \approx 0.456$ .
- b) Väntevärdet  $\mu$  skattas med  $\bar{x} = 4.0$ . Ett 95% konfidensintervall ges av  $\bar{x} \pm \lambda_{0.025} * \frac{\sigma}{\sqrt{n}} = 4.0 \pm 1.96 * 0.5/\sqrt{6} = 4.0 \pm 0.40$ .
- c) Om  $\sigma$  är okänd så skattas den med  $s = 0.456$  och motsvarande konfidensintervall ges av  $\bar{x} \pm t_{0.025}(6-1) \frac{s}{\sqrt{n}} = 4.0 \pm 2.571 \frac{0.456}{\sqrt{6}} = 4.0 \pm 0.479$ . Eftersom nollhypotesens värde ( $\mu = 3.5kg$ ) ligger långt utanför intervallet så förkastas denna hypotes.

### Uppgift 3

- a) Vi använder enkel linjär regression. Dvs löptiderna är oberoende av varanadra med en linjär trend:  $Y_i = \alpha + \beta x_i + \epsilon_i$ , där  $\epsilon_i$  är normalfördelad med väntevärde 0 och varians  $\sigma^2$ .

Vi får följande summer och medelvärden:  $\bar{x} = 6.5$  (om vi kallar april månad  $X_1 = 4$ , annars 3.5),  $S_{xx} = 17.5$ ,  $\bar{y} = 6.5$ ,  $S_{yy} = 0.0304$ ,  $S_{xy} = 0.7$ .

Från detta erhåller man skattningar enligt formelsamlingen:  $\hat{\beta} = S_{xy}/S_{yy} = 0.04$ ,  $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 4.19$  (4.31 om  $x_1$  sätts till 1) och  $s = \sqrt{(S_{yy} - (S_{xy}^2/S_{xx}))/n} = 0.0283$ .

b) Att se om det finns en trend eller inte är detsamma som att testa nollhypotesen  $H_0 : \beta = 0$  mot  $H_1 : \beta \neq 0$ . Vi bildar ett 95% konfidensintervall för  $\beta$ :

$$\hat{\beta} \pm t_{0.025}(4) \frac{s}{\sqrt{S_{xx}}} = 0.04 \pm 2.78 \frac{0.0283}{\sqrt{17.5}} = 0.04 \pm 0.019$$

Konfidensintervallet täcker inte in nollhypotesens värde (0). Vi kan därför (dessvärre) förkasta nollhypotesen. Toms löptider har blivit signifikant sämre (dvs högre) med tiden ...

## Uppgift 4

a) Kontingenstabell. Vi beräknar först förväntat antal observationer i cell  $(i, j)$ :  $e_{ij} = n_i * O_{.j}/n$ . Det blir  $e_{11} = 50 * 24/150 = 8$  vilket även gäller för övriga i den kolumnen eftersom  $n_1 = n_2 = n_3 = 50$ . Man får:  $e_{11} = e_{21} = e_{31} = 8$ ,  $e_{12} = e_{22} = e_{32} = 13$ ,  $e_{13} = e_{23} = e_{33} = 16$ ,  $e_{14} = e_{24} = e_{34} = 13$ . Man får då att  $Q = \sum_{i,j} (O_{ij} - e_{ij})^2 / e_{ij} = 12.16$ . Antal frihetsgrader är  $(3 - 1) * (4 - 1) = 6$ . Från tabellen har vi att  $\chi_{0.05}^2(6) = 12.59$ . Således går det inte att hävda att färdssätten mellan städerna skiljer sig åt signifikant, men det är ju ganska nära, så en viss tendens åt att det skiljer sig åt.

b) Andelen som tar bil i Xstad är  $\hat{p}_X = 13/50 = 0.26$  och i Sveastad  $\hat{p}_S = 20/50 = 0.40$ . Ett 95% konfidensintervall för skillnaden  $p_X - p_S$  ges av

$$\hat{p}_X - \hat{p}_S \pm 1.96 * \sqrt{\frac{\hat{p}_X(1 - \hat{p}_X)}{n_X} + \frac{\hat{p}_S(1 - \hat{p}_S)}{n_S}} = -0.14 \pm 0.18.$$

Skattingen är således att Sveastad använder bil nominellt 14% mer än Xstad, men skillnaden är inte statistiskt säkerställd.

## Uppgift 5

a) Vi väljer 2-sidig hypotes och  $\alpha = 5\%$  felrisk, och testar således om socialdemokraterna ligger oförändrat ( $p = 0,303$ ) mot alternativet att en ändring har skett. Vi bildar ett 95% konfidensintervall för  $p$  genom att först skatta  $p$  med  $\hat{p} = X/n = 0,339$ , där  $X$  = antal som sagt sig vilja rösta på socialdemokraterna och  $n = 4000$  antal svarande i undersökningen.  $X$  är således  $\text{Bin}(n = 4000, p)$  som har varians  $np(1-p)$ .  $\hat{p} = X/n$  har således varians  $p(1-p)/n$  och standardavvikelse  $\sqrt{p(1-p)/n}$ , vilket skattas genom att ersätta  $p$  med dess skattning  $\hat{p} = 0,339$  instoppat för  $p$ . Ett 95% konfidensintervall ges därför av

$$\hat{p} \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0,339 \pm 1,96 * 0,015 = [0,324, 0,354].$$

Eftersom 30.3% inte ingår i intervallet så kan nollhypotesen förkastas. Dvs ökningen sedan valet på 3.6% är statistiskt säkerställd.

b) Skillnaden mellan de två tidpunkterna för opinionsmätningarna  $p_{2026} - p_{2025}$  skattas med  $\hat{p}_{2026} - \hat{p}_{2025} = 0,339 - 0,362 = -0,023$ , alltså en nedgång på 2.3 procentenheter. Denna skattning har varians  $p_{2026}(1 - p_{2026})/n + p_{2025}(1 - p_{2025})/n$ . Ett 95% konfidensintervall ges således av

$$\hat{p}_{2026} - \hat{p}_{2025} \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}_{2026}(1 - \hat{p}_{2026})}{n} + \frac{\hat{p}_{2025}(1 - \hat{p}_{2025})}{n}} = -0,023 \pm 0,021.$$

Eftersom intervallet inte täcker in 0 (som svarar mot ingen förändring) så är socialdemokraternas nedgång gentemot förra året statistiskt säkerställd på 5%-nivån (men inte på 1%-nivån).

## Uppgift 6

a) Likelihooden ges av  $L(p) = \prod_i p(1 - p)^{z_i - 1} = p^n (1 - p)^{\sum_i z_i - n}$ . Loglikelihooden blir  $\ell(p) = n \log(p) + (\sum_i z_i - n) \log(1 - p)$ . Derivering ger  $\ell'(p) = \frac{n}{p} - \frac{\sum_i z_i - n}{1 - p}$ . Om man sätter derivatan till 0 och löser ut  $p$  får man  $\hat{p} = \frac{n}{\sum_i z_i} = \frac{1}{\bar{z}}$ .

b) ML-skattningen av en funktion av  $p$  ges av funktionen evaluerad i ML-skattningens värde. ML-skattningen för  $1/p$  blir således  $1/\hat{p} = \bar{z}$ .

c) Från formelsamligen har vi att  $E(Z) = 1/p$  och  $V(Z) = (1 - p)/p^2$ . Medelvärde av  $n$  oberoende sådana variabler har således väntevärde  $E(\bar{Z}) = E(Z) = 1/p$  och variansen blir  $V(\bar{Z}) = V(Z)/n = (1 - p)/np^2$ . Centrala gränsvärdessatsen gör att  $\bar{Z}$  är approximativt normalfördelad med detta väntevärde och varians. Ett 95% konfidensintervall för  $1/p$  ges således av

$$1/\hat{p} \pm 1.96 \sqrt{\frac{1 - \hat{p}}{n\hat{p}^2}} = \bar{z} \pm 1.96 \sqrt{\frac{1 - 1/\bar{z}}{n(1/\bar{z})^2}} = \bar{z} \pm 1.96 \sqrt{\frac{\bar{z}^2 - \bar{z}}{n}}$$