STOCKHOLMS UNIVERSITET
MATEMATISKA INSTITUTIONEN
Avd. Matematisk statistik, TB

MS 4001
LÖSNINGAR
21 augusti 2020

# Lösningar

## Tentamen i Statistisk analys, 21 augusti 2020

---

### Uppgift 1

**a)** Falskt

**b)** Sant

**c)** Sant

**d)** Sant

**e)** Sant

### Uppgift 2

**a)** Antal smittade i kommun 1 antas vara $Bin(n_1 = 18352,\ p_1)$ och antal smittade i kommun 2 antas vara $Bin(n_2 = 33428,\ p_2)$. Samma dödlighet svarar mot att $p_1 = p_2$ vilket blir vår nollhypotes. Detta testas genom att beräkna

$$z = \frac{p_1^* - p_2^*}{\sqrt{\frac{p_1^*(1-p_1^*)}{n_1} + \frac{p_2^*(1-p_2^*)}{n_2}}} = 5.57,$$

där $p_1^* = 352/18352 = 0.0192$ och $p_2^* = 421/33428 = 0.0126$ Detta skall jämföras med normalfördelningen och det blir ju ett mycket signifikant resultat. På 95% nivån ska man t ex förkasta $H_0$ om $|z| > 1.96 = z_{0.025}$. Man kan därför utan tvekan påstå att kommun 2 hade högre dödlighet.

**b)** Testar utgår ifrån att de enskilda observationerna är oberoende (och likafördelade) vilket gör att antal med en viss egenskap blir $Bin(n, p)$. Men eftersom individer smittar varandra så är antagande inte helt uppfyllt (observationerna är positivt beroende).

## Uppgift 3

**a)** Vi ansätter enkel linjär regression, dvs $Y_i = \alpha + \beta x_i + \epsilon_i$, där $Y_i$ är andra komponenten, $x_i$ första komponententen, och $\epsilon_1, \ldots, \epsilon_n$ är oberoende och normalfördelade med väntevärde 0 och standardavvikelse $\sigma$. Med skattningsformler från formelsamlingen erhåller man skattningarna $\hat{\alpha} = 49$, $\hat{\beta} = -0.72$ och $\hat{\sigma} = s = 0.94$.

**b)** Vi vill testa hypotesen att $\beta = -1$. För detta gör vi ett 95% konfidensintervall med formeln $\beta \pm t_{0.025}(11-2)s/\sqrt{S_{xx}}$ och får intervallet $(-0.908,\ -0.538)$. Efter hypotesens värde -1 *inte* ingår i intervallet så förkastas nollhypotesen. Andra komponenten avtar *inte* exakt 1 enhet i genomsnitt när första komponenten ökar med en enhet.

**c)** Ett prediktionsintervall för $Y_0 = \alpha + \beta * 60.0 + \epsilon$ ges av $I_{Y_0} = \hat{\alpha} + \hat{\beta} * 60.0 \pm t_{0.025}(11-2)s\sqrt{1 + 1/11 + \frac{(60-\bar{x})^2}{S_{xx}}} = 5.70 \pm 2.23$.

## Uppgift 4

**a)** $x$ och $y$ variablerna rangordnas. De två $y$-observationerna med samma värde får båda rang 6.5. Låt $d_i$ vara skillnaden mellan $i$:e observationens ranger. Vi beräknar $D = \sum_i d_i^2 = 406.5$. Spearmeans rangkorrelation ges av

$$r_S = 1 - \frac{6D}{n(n^2-1)} = -0.85.$$

**b)** Vi testar $H_0$: $X$ och $Y$ är oberoende, mot $H_1$: ej oberoende. Testvariablen ges av $z = \sqrt{n-1}r_S = -2.69$. Under $H_0$ är $z \sim N(0,1)$ approximativt. Eftersom $|z| > 2.57 = \lambda_{0.005}$ så förkastas $H_0$. Det finns således ett signifikant negativt beroende mellan variablerna.

## Uppgift 5

**a)** Observationerna är parvisa. Vi bildar därför differenserna $z_i = y_i - x_i$ som vi antar är approximativt oberoende och normalfördelade med väntevärde $\Delta$

och varians $\sigma^2$. Vi får $\bar{z} = 1.96$ och $s = 1.09$. Ett 95% konfidensintervall för $\Delta$ (=Coronaeffekten på arbetslösheten) blir

$$\bar{z} \pm t_0.025(8-1)s/\sqrt{8} = (1.05,\ 2.87).$$

Således en tydlig positiv effekt.

**b)** Frågan här handlar alltså om vilken signifikansnivå $\alpha$ som gör att

$$\bar{z} \pm t_0.025(7)s/\sqrt{8}$$

exakt inkluderar talet 0 (dvs ingen effekt). För detta krävs att $t_{\alpha/2}(7)s/\sqrt{8} = 1.96$, dvs att $t_{\alpha/2}(7) = 5.09$. Från tabellen ser man då att $\alpha/2$ bör ligga ungefär mitt emellan 0.001 och 0.0005, dvs att $\alpha$ är mitt emellan 0.002 och 0.001, så $p$-värdet är ungefär 0.0015.

## Uppgift 6

**a)** Likelihooden för $\beta$ ($\alpha = 10$ är känt) ges av

$$L(\beta) = \prod_{i=1}^{n} \left( \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \right).$$

Loglikelihooden för $\beta$ blir således

$$\ell(\beta) = n\alpha \log(\beta) - 10 \log(\Gamma(\alpha)) + (\alpha - 1) \sum_i \log(x_i) - \beta \sum_i x_i.$$

Derivatan blir

$$\ell'(\beta) = n\alpha/\beta - \sum_i x_i.$$

Om vi sätter derivatan till 0 och $\alpha = 10$ och löser ekvationen erhålls ML-skattningen

$$\hat{\beta} = 10n/\sum_i x_i.$$

**b)**

Ekvationerna för momentmetoden blir

$$\frac{\alpha}{\beta} = \bar{x}$$
$$\frac{\alpha}{\beta^2} = \frac{\sum_i (x_i - \bar{x})^2}{n}.$$

Om man löser ut $\alpha$ och $\beta$ ur dessa ekvationer så får vi

$$\alpha^* = \frac{\bar{x}^2}{n^{-1}\sum_i (x_i - \bar{x})^2} \quad \text{och} \quad \beta^* = \frac{\bar{x}}{n^{-1}\sum_i (x_i - \bar{x})^2}.$$

Modifierade momentmetoden blir densamma som ovan fast nämnaren ersatts med $s^2$.