

---

## SDAI (ST1101), Tentamen 2, 6 hp, första tentamenstillfället

**Kurs: Statistik och dataanalys I, 15 hp**

**Tentamensdatum: 2023-04-19**

Skrivtid:	kl. 14 - 19 (5 timmar)
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Formel- och tabellsamling för Statistik och dataanalys I, 15 hp

---

Tentamen består av 5 uppgifter, uppdelade i deluppgifter.

Maximalt antal poäng anges per deluppgift.

Svar med fullständiga redovisningar ska lämnas.

- Använd endast skrivpapper som tillhandahålls i skrivsalen.
- För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
- Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
- Om du inte lyckas lösa en deluppgift och behöver det svaret för en senare deluppgift så kan du hitta på värdet för att kunna göra beräkningarna i de efterföljande uppgifterna.
- I beräkningar från R-utskrifter får du utgå från det som är givet.

Tentamen kan maximalt ge 100 poäng och för godkänt resultat krävs minst 50.

Betygsgränser:

- A: 90–100p
- B: 80–89p
- C: 70–79p
- D: 60–69p
- E: 50–59p
- Fx: 40–49p
- F: 0 – 40p

OBS! Fx och F är underkända betyg som kräver omexamination.

Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.

Lösningsförslag läggs ut på Athena efter tentamen i samband med rättningen.

**Lycka till!**

---

**UPPGIFT 1 (16 POÄNG)**

Låt  $A$  och  $B$  vara två händelser med  $P(A) = 0.4$  och  $P(B) = 0.2$ . Sannolikheten att både  $A$  och  $B$  inträffar är 0.1.

- (a) Vad är sannolikheten att åtminstone någon av de två händelserna inträffar? (4 p)
- (b) Är  $A$  och  $B$  oberoende? (4 p)
- (c) Vad är sannolikheten för  $A$  givet att  $B$  har inträffat? (4 p)
- (d) Vad är sannolikheten att **exakt en** av händelserna  $A$  och  $B$  inträffar? (4 p)

**UPPGIFT 2 (16 POÄNG)**

En stödlinje tar dagligen emot ett stort antal telefonsamtal. Tyvärr händer det med sannolikheten 0.1 att samtalet görs på skämt och inte av en person som verkligen behöver stöd. Utöver det är också 5% av samtalen felringningar.

- (a) Vad är sannolikheten att det första riktiga stödsamtalet (dvs ej skämt eller felringning) kommer först vid dagens tredje samtal? Ange eventuella antaganden du måste göra för dina beräkningar. (4 p)
- (b) Vad är sannolikheten att ett av de tre första samtalen för dagen är ett riktigt stödsamtal? (6 p)
- (c) Vad är sannolikheten att bland de tre första samtalen få:
  - ett riktigt samtal,
  - ett skämtsamtal och
  - en felringning?(6 p)

**UPPGIFT 3** (20 POÄNG)

- (a) En slumpvariabel  $X$  har sannolikhetsfördelningen som anges i tabellen nedan. Beräkna väntevärdet och standardavvikelsen för  $X$ . (6 p)

$x$	1	2	3	4
$P(X = x)$	0.2	0.4	0.3	0.1

- (b) Låt  $Y = 3 + 2 \cdot X$ , där  $X$  är slumpvariabeln i uppgift 3a). Beräkna väntevärde och standardavvikelse för  $Y$ . (6 p)
- (c) Låt  $Z$  vara en Binomialfördelad slumpvariabel med parametrar  $n = 10$  och  $p = 0.3$ . Beräkna väntevärde och varians för  $X + Z$ , där  $X$  återigen är slumpvariabeln i Uppgift 3a). Ange eventuella antaganden du måste göra. (8 p)

**UPPGIFT 4** (25 POÄNG)

Ett företag massproducerar en komplicerad delkomponent till datorer där 22% av produkterna blir defekta eftersom en metall som ingår i produktionen är känslig och blir lätt förstörd i produktionsprocessen. Ett nytt material har lanserats på marknaden som kan ersätta det känsliga materialet. Företaget köper in materialet och provar att producera 80 produkter med det, och fann att 12 produkter blev defekta.

- (a) Låt  $p$  vara andelen i defekta produkter i populationen av produkter som produceras med det nya materialet. Beräkna ett approximativt 95% konfidensintervall för  $p$ . Tolka intervallet på ett sätt som visar att du *förstått* dess innebörd. (7 p)
- (b) Testa på 5% signifikansnivå om andelen defekta för det nya materialet skiljer från andelen defekta med det gamla materialet. Ställ upp hypotester, utför testet och dra korrekt slutsats. (7 p)
- (c) Upprepa hypotestestet i Uppgift 4b), men denna gång genom att beräkna och tolka  $p$ -värdet. (7 p)
- (d) Förklara vad som menas med påståendet: 'Estimatorn  $\hat{p}$  är en väntevärderiktig estimator av  $p$ '. (4 p)

**UPPGIFT 5 (23 POÄNG)**

I den här uppgiften analyseras priset i US dollar (**price**) för 100 diamanter. För varje diamanter finns även data på diamanterens vikt (**carat**), djup (**depth**), bredden på toppen relativt diamanterens bredaste punkt (**table**) och en binär variabel (**clear**) som antar värdet 1 om diamanter är klar i sin färg, och värdet 0 om diamanter är mer grumlig. Nedan ger en sammanfattning av data och en utskrift på anpassningen av den linjära regressionsmodellen

$$\text{price} = \beta_0 + \beta_1 \cdot \text{carat} + \beta_2 \cdot \text{depth} + \beta_3 \cdot \text{table} + \beta_4 \cdot \text{clear} + \varepsilon, \quad \varepsilon \stackrel{iid}{\sim} N(0, \sigma_\varepsilon).$$

Vissa värden har dock tagits bort från utskriften av den anpassade modellen.

```
diamond.price  diamond.carat    diamond.depth    diamond.table    diamond.clear
Min.   : 558    Min.   :0.2500    Min.   :57.70    Min.   :53.00    Min.   :0.00
1st Qu.:2960    1st Qu.:0.7100    1st Qu.:61.20    1st Qu.:56.00    1st Qu.:0.00
Median :3078    Median :0.7500    Median :62.05    Median :57.00    Median :1.00
Mean   :2846    Mean   :0.7594    Mean   :62.08    Mean   :57.57    Mean   :0.54
3rd Qu.:3193    3rd Qu.:0.9000    3rd Qu.:62.90    3rd Qu.:59.00    3rd Qu.:1.00
Max.   :3303    Max.   :1.5000    Max.   :66.10    Max.   :67.00    Max.   :1.00
```

Call:

```
lm(formula = price ~ carat + depth + table + clear, data = diamond)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-2018.6 -157.7  144.4   301.3  829.1
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5164.04    2427.28   2.128 0.035968 *
carat        3368.77     255.25  13.198 < 2e-16 ***
depth        -63.52      33.62   -1.888 0.061812 .
table        -19.77      102.37   -0.193 0.849658
clear         381.32     102.37   3.726 0.000258 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Multiple R-squared:  0.6493,    Adjusted R-squared:  0.6346
```

- Testa om **depth** är en signifikant förklarande variabel på 5% signifikansnivå. Ställ upp hypoteser, beräkna teststatistikan med fördelningen under  $H_0$  och utför testet. Dra slutsats. (7 p)
- Testa om **table** är en signifikant förklarande variabel på 10% signifikansnivå. (5 p)
- Tolka skattningen av  $\beta_4$  och skapa ett 90%-igt konfidensintervall för denna parameter. (7 p)
- Förklara utförligt betydelsen av värdet 102.37 för variabeln **clear** under kolumnen **Std.Error** i utskriften ovan. (4 p)

**Lycka till!**