

# Tentamen i Statistik och mätmetoder

## Uppgift 1

- a) Beräkna medelvärdet för variabeln `unique_pageviews`

Jag gjorde beräkning av medelvärdet för variabeln i R med hjälp av detta kommando:

```
mean(data_tent_1$unique_pageviews, na.rm = TRUE)
```

Resultatet visar att medelvärdet är 439.3859

- b) Skapa en frekvenstabell av variabeln `design`.

```
> crossTable(data_tent_1$design)
```

```
Cell Contents
-----|
          N |
N / Table Total |
-----|
```

```
Total observations in Table: 640
```

```
-----|-----|
          A |          B |
-----|-----|
          329 |          311 |
          0.514 |          0.486 |
-----|-----|
```

- c) Beräkna medelvärdet för `unique_pageviews` fördelat på design (A eller B) och redovisa enligt följande.

	Medelvärde <code>unique_pageviews</code>
Design A	425
Design B	453

## Uppgift 2

a) Ange vilken metod du vill använda

Jag väljer metoden konfidensintervall för jämförelse av populationsandelar

b)  $H_0: \pi_1 - \pi_2 = d_0$

$H_a: \pi_1 - \pi_2 \neq d_0$

c) Visa resultatet av analysen. Välj 5 % signifikansnivå.

$$z = \frac{(p_1 - p_2) - d_0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{(0.514 - 0.486) - 0}{\sqrt{\frac{0.514(1-0.514)}{329} + \frac{0.486(1-0.486)}{311}}}$$

$$\Rightarrow \frac{0.028}{\sqrt{0.000759 + 0.0008}} = \frac{0.028}{0.0395} = 0.708$$

p-värdena är punktskattningar från svaret i uppgift 1 b).

Testvariabeln är 0.708

Jag har valt dubbelsidig mothypotes och de kritiska värdena bestäms till

$$Z_{\alpha/2} = Z_{0.05/2} = 0.025 = -1.96$$

och

$$Z_{1-\alpha/2} = Z_{1-0.05/2} = 0.975 = 1.96$$

Testvariabeln 0.708 faller inte i de kritiska värdena varför  $H_0$  inte kan förkastas.

d) Med 95 % säkerhet kan vi konstatera att det inte föreligger statistiskt säkerställda skillnader i frekvensen mellan design A och design B.

- e) Metoden förutsätter att vi har att göra med 2 obundna och slumpmässigt dragna urval och att  $np(1-p) > 5$  för båda stickproven.

Jag förutsätter att stickproven uppfyller kraven när det gäller obundenhet och slumpmässighet (det är svårt för mig att kontrollera). När det gäller andra kravet går det lätt att beräkna:

$$329 * 0,514 (1 - 0,514) = 82,18$$

$$311 * 0,486 (1 - 0,486) = 77,68$$

Det framgår med all tydlighet att det andra kravet är uppfyllt med råge.

## Uppgift 3.

Pröva påståendet att genomsnittligt antal unique\_pageviews är statistiskt säkerställt högre för design B än för design A genom att redovisa följande:

- a) Ange vilken metod du väljer att använda.

Eftersom vi här har att göra med medelvärden väljer jag metoden Konfidensintervall för jämförelse av populationsmedelvärden.

- b)  $H_0: \mu_1 = \mu_2$   
 $H_a: \mu_1 > \mu_2$

Här har jag valt att låta  $\mu_1$  stå för design B och  $\mu_2$  stå för design A.

- c) Visa resultatet. Välj 5 % signifikansnivå.

Först beräknas testvariabeln,  $t =$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(453 - 425) - 0}{\sqrt{\frac{214^2}{311} + \frac{217^2}{329}}} =$$

$$= \frac{28}{\sqrt{147,254 - 143,127}} = \frac{28}{\sqrt{290,3817}} =$$

$$= 1,643$$

Standardavvikelseerna räknade jag fram med excel.

Kritiskt värde  $t_{n^*-1, 1-0.05} = t_{310, 0.95} = 1.660$

Egentligen så är  $t_{100, 0.95} = 1.660$  och

$$t_{1000, 0.95} = 1.646$$

så  $t$ -värdet hamnar mittemellan.

Hur som helst är  $t$ -värdet 1.643 lägre än det kritiska värdet,  $H_0$  kan ej förkastas.

- d) Med 95 % säkerhet kan vi konstatera att det inte går att statistiskt säkerställa att genomsnittligt antal unique\_pageviews skulle vara högre för design b än för design a.
- e) Ange vilka förutsättningar metoden baseras på.  
Metoden förutsätter att vi har dragit två obundna slumpmässigt utvalda stickprov och att samplingfördelningarna för de två stickprovsmedelvärdena är normalfördelade.

Jag har svårt att avgöra om det första kravet är uppfyllt men det andra kravet bör vara det med tanke på att vi har att göra med ett stickprov som är relativt stort (långt högre än 30 enheter se även uppgift 4d).

## Uppgift 4.

Beräkna ett 95 %-igt konfidensintervall för skillnaden i genomsnittligt antal unique\_pageviews mellan design A och design B genom att redovisa följande:

- a) Ange vilken metod du väljer att använda.  
Liksom i föregående uppgift väljer jag ett 95 %-igt konfidensintervall för jämförelse av populationsmedelvärden.
- b)

$$\begin{aligned}
 & (\bar{X}_1 - \bar{X}_2) \pm t_{n^*-1; 1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \\
 & = (453 - 425) \pm t_{310; 0,975} \sqrt{\frac{214^2}{311} + \frac{217^2}{329}} = \\
 & = 28 \pm 1,984 \cdot \sqrt{290,3817} = \\
 & = 28 \pm 33,80 \\
 & -5,8 \leq \mu_1 - \mu_2 \leq 61,80
 \end{aligned}$$

- c) Tolka konfidensintervallet med ord.  
Eftersom konfidensintervallet täcker noll kan vi konstatera att det på 95 % konfidensnivå inte föreligger några statistiskt säkerställda skillnader mellan design A och design B när det gäller genomsnittligt antal unique\_pageviews.

(Precis som i tidigare uppgift har jag låtit design B stå för x-streck och även  $\mu_1$ )

- d) Ange vilka förutsättningar metoden baseras på och redogör för om de är uppfyllda. Metoden förutsätter att stickproven har dragits slumpmässigt och att samplingfördelningarna är normalfördelade.

Eftersom stickprovet överstiger 30 enheter kan vi tillämpa oss av centrala gränsvärdessatsen som säger att om stickprovet är tillräckligt stort får fördelningen anses vara approximativt normalfördelad. Så ja det andra kravet bör vara uppfyllt. Kan inte uttala mig om stickproven är slumpmässigt eller obundet utvalda men förutsätter att så givetvis är fallet.

## Uppgift 5.

Andelen bounce för en webbplats är 34,6 %. Vad är sannolikhet att av 3 besökare ingen bouncar? Besvara detta genom följande steg:

- a) Vilken sannolikhetsfördelning följer slumpförsöket?  
Försöket följer en binomialfördelning eftersom varje delförsök kan endast anta ett av två möjliga värden: bounce eller inte bounce.
- b) Visa resultatet av sannolikhetsberäkningen i R.

```
> dbinom(0, size = 3, prob = 0.36)
[1] 0.262144
```

Sannolikheten för att 0 personer av 3 bouncar för denna webbplats är 0.26214 eller 26,2 %.