STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)

VT2020

# Tentamen i Undersökningsmetodik (4.5 hp)

## Kurs: Regressionsanalys och undersökningsmetodik

**2020-08-17**

| | |
|---|---|
| **Skrivtid:** | kl. 15.00 - 21.00 (6 timmar, inklusive en timme för digital överföring) |
| **Godkända hjälpmedel:** | Miniräknare, dator, kurslitteratur och föreläsningsanteckningar |
| **Vidhäftade hjälpmedel:** | Formelsamling och Statistiska tabeller (endast de tabeller som krävs) |

- Tentamen består av 5 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.

- Svar med fullständiga redovisningar ska lämnas.
  - För full poäng krävs tydliga, utförliga och väl motiverade lösningar.
  - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
  - Använd minst fem värdesiffror i dina beräkningar (1,2345 och 1234,5 är exempel på tal med fem värdesiffror). I förekommande fall är det inte möjligt pga. avrundning i t.ex. SAS-utskrifter men utgå då ifrån det som är givet. Du kan dock avrunda ditt slutliga svar.

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

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

  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 kort efter tentamen.

## KONTAKT MED EXAMINATOR UNDER SKRIVNINGEN

- Om det är något som är oklart eller något som verkar konstigt kan du kontakta examinatorn under pågående skrivning via e-post: michael.carlson@stat.su.se eller telefon: 08-16 29 82.
- OBS! Undvik att använda Athena för att kontakta examinator.
- Bevaka dock din e-post och Athena under skrivningen för eventuella meddelanden som rör provet.

**LYCKA TILL!**

**Uppgift 1. (30p)**

I en avlägsen del av landet går en busslinje mellan en liten ort och en större tätort. Man ville undersöka om det finns ett behov av att utöka antalet avgångar på kvällstid från den mindre orten (och tillbaka lite senare på kvällen förstås, det finns en biograf och andra nöjen i den större orten).

Man drog ett obundet slumpmässigt urval, utan återläggning, bestående av $n = 50$ personer från populationen i den mindre orten som vid undersökningstillfället bestod av $N = 200$ invånare. Man frågade via telefonintervjuer de utvalda hur många gånger de hade åkt med busslinjen under kvällstid en given referensvecka (samma för alla) och fick följande resultat:

| Antal gånger man åkt | 0 | 1 | 2 | 3 | $\leq 4$ |
|---|---|---|---|---|---|
| Antal personer | 14 | 18 | 10 | 8 | 0 |

I dina beräkningar ska ändlighetskorrektion användas.

a) Beräkna ett 90% konfidensintervall för andelen personer i den mindre orten som åkt med busslinjen minst en gång under referensveckan. (8p)

b) Anta att den totala längden på ett 99% konfidensintervall för andelskattningen i a) högst får vara 2 procentenheter. Hur många personer måste man minst undersöka då? Anta att antalet personer i den mindre orten som åker med bussen är lika med antalet som inte åker. (8p)

c) Beräkna 95% konfidensintervall för det genomsnittliga antalet gånger per person man åkt med bussen under referensveckan. (10p)

   TIPS: Låt $y_k$ beteckna antalet resor som person $k$ gjort. Tänk igenom hur du på ett enkelt sätt tar fram $\sum_{k \in S} y_k$ och $\sum_{k \in S} y_k^2$ alt. $\sum_{k \in S}(y_k - \bar{y})^2$ för att sedan beräkna stickprovsmedelvärde och stickprovsvarians.

d) Diskutera kortfattat för- och nackdelar med att fråga enskilda om deras resvanor enligt beskrivningen ovan. Föreslå en alternativ metod för att samla in data som kan vara lämplig för att analysera trafikbehovet. Max ½ sida räcker, det behövs ingen lång utredning för att få full poäng. (4p)

**Uppgift 2. (25p)**

En populär dansk restaurangkedja specialiserade på kötträtter har utöver $N_1 = 75$ restauranger i Danmark, även $N_2 = 30$ i Sverige och $N_3 = 28$ i Norge. Man har precis introducerat en ny vegetarisk rätt på menyn och för att analysera hur populär den verkar vara samlade man in data från första veckan efter introduktionen för en delmängd av kedjans restauranger i de tre länderna. Man hade nämligen inte fått in siffrorna från alla restauranger än. Man noterade antalet beställningar per restaurang och beräknade stickprovsmedelvärde och -standardavvikelse per land enligt följande tabell:

| Land | Stickprovsstorlek | Medelvärde | Standardavvikelse |
|---|---|---|---|
| Danmark | 15 | 21.2 | 12.8 |
| Sverige | 8 | 14.5 | 11.4 |
| Norge | 5 | 25.8 | 9.2 |

I dina beräkningar ska ändlighetskorrektion användas.

a) Skatta $\tau =$ totala antalet beställningar för samtliga restauranger i kedjan och beräkna ett 95% konfidensintervall för skattningen med formlerna för stratifierat OSU. (10p)

b) Nu var det ju ingen riktig, i förväg bestämd, allokering av stickprovsstorlekarna för de tre länderna, man fick vad man fick i antal per land. Men du är en nyfiken statistiker och bestämmer dig för att jämföra den observerade allokeringen (stickprovets fördelning över de tre länderna) med vad man hade fått om man hade planerat för och använt (i) proportionell och (ii) optimal allokering. Utgå ifrån de observerade standardavvikelserna och avgör frågan. OBS! du ska alltså redovisa två olika allokeringar och jämföra mot den observerade. (10p)

c) Går det att lita på resultatet i a)? Resonera kortfattat vad det är som är fel med undersökningen om man vill skatta populationsparametrar och beräkna precisionen för dessa med formlerna för stratifierat urval. Det finns minst två viktiga aspekter som är problematiska i detta fall. Skriv kortfattat, ca ½ sida max bör räcka. (5p)

**Uppgift 3. (20p)**

Ett stort internationellt företag i klädesbranschen vill analysera hur intäkterna kan komma att förändras jämfört med förra året. Från de $N = 350$ butiker företaget runt om i Europa drogs ett OSU utan återläggning av storlek $n = 11$. Man hade redan uppgift om $x_k$ = intäkterna i mkr för förra året för samtliga butiker. För innevarande år beräknades på bästa sätt prognoser $y_i$ = intäkter i mkr för innevarande år för de butiker som drogs till urvalet. Man fick följande data:

| $k$ | $x_k$ | $y_k$ | $x_k^2$ | $y_k^2$ | $x_k y_k$ |
|---|---|---|---|---|---|
| 1 | 6 | 8 | 36 | 64 | 48 |
| 2 | 12 | 14 | 144 | 196 | 168 |
| 3 | 16 | 25 | 256 | 625 | 400 |
| 4 | 18 | 18 | 324 | 324 | 324 |
| 5 | 26 | 36 | 676 | 1 296 | 936 |
| 6 | 30 | 26 | 900 | 676 | 780 |
| 7 | 36 | 48 | 1 296 | 2 304 | 1 728 |
| 8 | 39 | 78 | 1 521 | 6 084 | 3 042 |
| 9 | 45 | 56 | 2 025 | 3 136 | 2 520 |
| 10 | 50 | 57 | 2 500 | 3 249 | 2 850 |
| 11 | 52 | 74 | 2 704 | 5 476 | 3 848 |
| SUMMA | 330 | 440 | 12 382 | 23 430 | 16 644 |

Utöver detta visste man förstås att intäkterna totalt under förra året summerade till $\tau_x = 7\,680$ mkr.

I dina beräkningar ska ändlighetskorrektion användas.

a) Åskådliggör de observerade $x$- och $y$-värdena från stickprovet i ett lämpligt diagram som visar hur sambandet mellan variablerna ser ut. Om du skulle använda $x$ som en hjälpvariabel för att skatta totalvärdet $\tau_y$, skulle du välja en kvot- eller en regressionsestimator? Motivera! (5p)

b) Skatta $\tau_y$ med antingen en kvotestimator eller en regressionsestimator, med $x$ som hjälpvariabel, i enlighet med ditt val i a) ovan. Beräkna sedan standardfelet för skattningen. (10p)

OBS! Du ska inte göra både och, endast <u>en</u> skattning ska beräknas. Om du gör båda kommer endast den första att beaktas vi rättningen. Notera också att du kan få full poäng för uppgiften även om du väljer ”fel” estimator i a) ovan.

c) Skatta nu både $\tau_y$ och $\tau_x$ med de vanliga OSU-formlerna. Du behöver inte beräkna standardfelen, bara punktskattningarna. Jämför sedan $\hat{\tau}_{y,OSU}$ med punktskattningen i b) ovan. Kan du förklara skillnaden mellan dem? TIPS: Vad är värdet på $\tau_x$? (5p)

**Uppgift 4. (15p)**

För var och en av följande två deluppgifter ska du svara kortfattat. Hela uppgiften bör kunna redovisas på ca 1½, max 2 A4-sidor. Du får gärna komplettera med bilder och skisser.

a) Förklara begreppet *inklusionssannolikhet*. Ange två saker som måste vara uppfyllda för inklusionssannolikheter och vilka konsekvenserna är om de inte är uppfyllda. (5p)

b) En mängd olika datainsamlingsmetoder finns beskrivna i litteraturen och många används i praktiken. Förklara vad som menas med begreppet *mixed mode*. Ange två fördelar och två nackdelar med mixed mode. (5p)

c) Standarder för bortfallsredovisning har diskuterats på föreläsningarna och i kurslitteraturen. En formel definieras enligt

$$BA(1) = 1 - \frac{n_{\mathrm{S}}}{n_{\mathrm{S}} + n_{\mathrm{B}} + n_{\mathrm{O}}}$$

Förklara vad detta mått anger och förklara vad de olika termerna i högerledet står för. (5p)

**Uppgift 5. (10p)**

Man har en population $U$ bestående av $N = 4$ element med följande värden på en variabel $Y$:

$$U = \{2, 3, 5, 8\}$$

a) Beräkna medelvärde och varians för $Y$ i denna ändliga population. (2p)

b) Man drar ett obundet slumpmässigt stickprov av storlek $n = 2$ från $U$ utan återläggning. Beräkna med formler väntevärde och varians för stickprovsmedelvärdet $\bar{Y}$. (2p)

c) Man drar ett obundet slumpmässigt stickprov av storlek $n = 2$ från $U$ med återläggning. Beräkna med formler väntevärde och varians för stickprovsmedelvärdet $\bar{Y}$. (2p)

d) Ange utfallsrummen för $\bar{Y}$ när man drar utan respektive med återläggning. Du behöver inte ange sannolikhetsfördelningarna, endast de möjliga värden $\bar{Y}$ kan anta med respektive design behöver anges. (2p)

e) Utgå utifrån ditt svar i d) och ge en enkel förklaring till varför det verkar logiskt att variansen i c) är större än variansen i b). (2p)

# Formel- och tabellsamling

## DESKRIPTIV STATISTIK

Notation: $U$ = populationen
$S$ = stickprov (stort $S$); $\subseteq U$

| | | |
|---|---|---|
| Medelvärde: | $\mu = \frac{1}{N}\sum_{k\in U} y_k$ | Varians: $\sigma^2 = \dfrac{\sum_{k\in U}(y_k-\mu_y)^2}{N} = \dfrac{\sum_{k\in U} y_k^2 - N\mu_y^2}{N}$ |
| | $\bar{y} = \frac{1}{n}\sum_{k\in S} y_k$ | $s^2 = \dfrac{\sum_{k\in S}(y_k-\bar{y})^2}{n-1} = \dfrac{\sum_{k\in S} y_k^2 - n\bar{y}^2}{n-1}$ |
| Andel: ($y_k = 0$ eller 1) | $P = \frac{1}{N}\sum_{k\in U} y_k$ <br> $\hat{p} = \frac{1}{n}\sum_{k\in S} y_k$ | $\sigma^2 = P(1-P)$ <br> $s^2 = \frac{n}{n-1}\hat{p}(1-\hat{p})$ |
| Kovarians: | $\sigma_{xy} = Cov(x,y) = \dfrac{\sum_{k\in U}(x_k-\mu_x)(y_k-\mu_y)}{n-1} = \dfrac{\sum_{k\in U} x_i y_i - n\bar{x}\bar{y}}{n-1}$ <br> $s_{xy} = Cov(x,y) = \dfrac{\sum_{k\in U}(x_k-\bar{x})(y_k-\bar{y})}{n-1} = \dfrac{\sum_{k\in U} x_k y_k - n\bar{x}\bar{y}}{n-1}$ | |
| Korrelation: | $r_{xy} = Corr(x,y) = \dfrac{s_{xy}}{s_x \cdot s_y} = \dfrac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}$ | |

## Beräkningsformler för VARIANSER och REGRESSIONSKOEFFICIENT

$$s^2 = \frac{n\sum y_k^2 - (\sum y_k)^2}{n(n-1)} = \frac{\sum y_k^2 - \frac{(\sum y_k)^2}{n}}{n-1} = \frac{\sum y_k^2 - n\bar{y}^2}{n-1} = \frac{\sum (y_k-\bar{y})^2}{n-1}$$

$$b = \frac{n\sum x_k y_k - (\sum x_k)(\sum y_k)}{n\sum x_k^2 - (\sum x_k)^2} = \frac{\sum x_k y_k - \frac{(\sum x_k)(\sum y_k)}{n}}{\sum x_k^2 - \frac{(\sum x_k)^2}{n}} = \frac{\sum x_k y_k - n\bar{x}\bar{y}}{\sum x_k^2 - n\bar{x}^2}$$

$$= \frac{\sum(x_k-\bar{x})(y_k-\bar{y})}{\sum(x_k-\bar{x})^2} = \frac{\sum(x_k-\bar{x})(y_k-\bar{y})\,/(n-1)}{\sum(x_k-\bar{x})^2\,/(n-1)}$$

$$= \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x^2}\cdot\frac{s_x s_y}{s_x s_y} = \frac{s_{xy}}{s_x s_y}\cdot\frac{s_y}{s_x} = r_{xy}\cdot\frac{s_y}{s_x}$$

OBS! Notationen har förenklats ovan, summationsindex är alltid $k$: ex. $\sum y_k = \sum_{k\in S} y_k$

## OBUNDET SLUMPMÄSSIGT URVAL u.å.

| Parameter: | Estimator: | Varians $V(\cdot)$: | Variansskattning $\hat{V}(\cdot)$: |
|---|---|---|---|
| $\mu$ | $\bar{y} = \frac{1}{n}\sum_{k\in S} y_k$ | $V(\bar{y}) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$ | $\hat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$ |
| $\tau$ | $\hat{\tau} = N\bar{y}$ | $V(\hat{\tau}) = N^2 V(\bar{y})$ | $\hat{V}(\hat{\tau}) = N^2 \cdot \hat{V}(\bar{y})$ |
| $P$ | $\hat{p} = \frac{1}{n}\sum_{k\in S} y_k$ | $V(\hat{p}) = \left(\frac{N-n}{N-1}\right)\frac{P(1-P)}{n}$ | $\hat{V}(\hat{p}) = \left(1 - \frac{n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}$ |
| $A$ | $\hat{A} = N\hat{p}$ | $V(\hat{A}) = N^2 V(\hat{p})$ | $\hat{V}(\hat{A}) = N^2 \cdot \hat{V}(\hat{p})$ |

Stickprovsstorlek: $$n \geq \frac{N\sigma^2}{D^2(N-1) + \sigma^2}$$

## STRATIFIERAT URVAL u.å.

Notation: $L =$ antal strata

$N_k =$ populationsstorleken för stratum $k = 1, \dots, L$

$n_k =$ stickprovets storlek i stratum $k = 1, \dots, L$

$W_k = N_k / N$

$\bar{y}_k =$ stickprovsmedelvärde i stratum $k = 1, \dots, L$

$s_k^2 =$ stickprovsvarians i stratum $k = 1, \dots, L$

| Parameter | Estimator | Varians $V(\cdot)$ | Variansskattning $\hat{V}(\cdot)$ |
|---|---|---|---|
| $\mu$ | $\bar{y}_{\text{str}} = \sum_{k=1}^{L} W_k \bar{y}_k$ | $\sum_{k=1}^{L} W_k^2 \left(\frac{N_k-n_k}{N_k-1}\right)\frac{\sigma_k^2}{n_k}$ | $\sum_{k=1}^{L} W_k^2 \left(1 - \frac{n_k}{N_k}\right)\frac{s_k^2}{n_k}$ |
| $\tau$ | $\hat{\tau}_{\text{str}} = N\bar{y}_{\text{str}}$ | $\sum_{k=1}^{L} N_k^2 \left(\frac{N_k-n_k}{N_k-1}\right)\frac{\sigma_k^2}{n_k}$ | $\sum_{k=1}^{L} N_k^2 \left(1 - \frac{n_k}{N_k}\right)\frac{s_k^2}{n_k}$ |
| $P$ | $\hat{p}_{\text{str}} = \sum_{k=1}^{L} W_k \hat{p}_k$ | $\sum_{k=1}^{L} W_k^2 \left(\frac{N_k-n_k}{N_k-1}\right)\frac{P_k(1-P_k)}{n_k}$ | $\sum_{k=1}^{L} W_k^2 \left(1 - \frac{n_k}{N_k}\right)\frac{\hat{p}_k(1-\hat{p}_k)}{n_k-1}$ |
| $A$ | $\hat{A}_{\text{str}} = N\hat{p}_{\text{str}}$ | $\sum_{k=1}^{L} N_k^2 \left(\frac{N_k-n_k}{N_k-1}\right)\frac{P_k(1-P_k)}{n_k}$ | $\sum_{k=1}^{L} N_k^2 \left(1 - \frac{n_k}{N_k}\right)\frac{\hat{p}_k(1-\hat{p}_k)}{n_k-1}$ |

Optimal allokering: $$n_k = n \cdot \frac{N_k\sigma_k}{\sum_{j=1}^{L} N_j\sigma_j}$$

**KLUSTERURVAL – OSU u.å.**

Notation: $U$ = population av kluster
$S$ = stickprov av kluster
$N$ = antal kluster totalt
$n$ = antal kluster i stickprovet

$M$ = totalt antal element
$m_i$ = antal element i kluster nr $i = 1,2 \ldots, N$
$\bar{m}$ = stickprovsmedelvärde av klusterstorlekarna $m_i$
$s^2_{m_i}$ = stickprovsvariansen av klusterstorlekarna $m_i$

$\tau = \sum_{k \in U} y_k$ = totalvärdet för $y$ i hela populationen
$\mu = \tau / M$ = populationsmedelvärde av $y$
$\tau_i = \sum_{k \in C_i} y_k$ = totalvärdet för kluster nr $i = 1,2 \ldots, N$
$\bar{\tau}$ = stickprovsmedelvärdet av totalvärdena $\tau_i$
$s^2_{\tau_i}$ = stickprovsvariansen av totalvärdena $\tau_i$

$A = \sum_{k \in U} y_k$ = antalet ettor i hela populationen; ($y_k = 0$ eller 1)
$P = A/M$ = andelen ettor i hela populationen; ($y_k = 0$ eller 1)

| Parameter | Estimator | Variansskattning |
|---|---|---|
| $M$ | $\widehat{M}_{\text{vvr}} = N \cdot \bar{m}$ | $\widehat{V}(\widehat{M}_{\text{vvr}}) = N^2 \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{s^2_{m_i}}{n}$ |
| $\mu$ | $\bar{y}_{\text{vvr}} = \frac{\hat{\tau}_{\text{vvr}}}{M} = \frac{N\bar{\tau}}{M}$<br>$\bar{y}_{\text{kvot}} = \frac{\hat{\tau}_{\text{vvr}}}{\widehat{M}} = \frac{\sum_{i \in S} \tau_i}{\sum_{i \in S} m_i}$ | $\widehat{V}(\bar{y}_{\text{vvr}}) = \frac{N^2}{M^2} \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{s^2_{\tau_i}}{n}$<br>$\widehat{V}(\bar{y}_{\text{kvot}}) = \left(\frac{1}{\bar{m}}\right)^2 \left(1 - \frac{n}{N}\right) \frac{\sum_{i \in S}(\tau_i - \bar{y}_{kvot} m_i)^2}{n(n-1)}$ |
| $\tau$ | $\hat{\tau}_{\text{vvr}} = N\bar{\tau}$<br>$\hat{\tau}_{\text{kvot}} = M\bar{y}_{\text{kvot}} = \frac{M}{\widehat{M}} \hat{\tau}_{vvr}$ | $\widehat{V}(\hat{\tau}_{\text{vvr}}) = N^2 \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{s^2_{\tau_i}}{n}$<br>$\widehat{V}(\hat{\tau}_{\text{kvot}}) = \left(\frac{M}{\bar{m}}\right)^2 \left(1 - \frac{n}{N}\right) \frac{\sum_{i \in S}(\tau_i - \bar{y}_{kvot} m_i)^2}{n(n-1)}$ |
| $P$ | *formler utgår* | |
| $A$ | *formler utgår* | |

## SKATTNINGSMETODER

Notation: $\tau_y$ = totalvärdet för variabeln $y$ för hela populationen
$\hat{\tau}_y$ = skattningen av $\tau_y$ under OSU
$\mu_y$ = populationsmedelvärdet av för variabeln $y$

**Kvotskattning under OSU u.å.:**

| Parameter | Punkt- resp. variansskattning |
|---|---|
| $\tau_y$ | $\hat{\tau}_{\text{kvot}} = \hat{R} \cdot \tau_x = \dfrac{\sum_{k \in S} y_k}{\sum_{k \in S} x_k} \cdot \tau_x = \dfrac{\tau_x}{\hat{\tau}_x} \cdot \hat{\tau}_y$ där $\hat{R} = \dfrac{\sum_{k \in S} y_k}{\sum_{k \in S} x_k} = \dfrac{\hat{\tau}_y}{\hat{\tau}_x}$ <br> $\hat{V}(\hat{\tau}_{\text{kvot}}) = N^2 \cdot \left(1 - \dfrac{n}{N}\right) \cdot \dfrac{1}{n} \cdot \left(\dfrac{\sum_{k \in S}\left(y_k - \hat{R}x_k\right)^2}{n-1}\right)$ <br> där $\sum_{k \in S}\left(y_k - \hat{R}x_k\right)^2 = \sum_{k \in S} y_k^2 - 2\hat{R}\sum_{k \in S} x_k y_k + \hat{R}^2 \sum_{k \in S} x_k^2$ |
| $\mu_y$ | $\hat{\mu}_{\text{kvot}} = \hat{R} \cdot \mu_x = \dfrac{\sum_{k \in S} y_k}{\sum_{k \in S} x_k} \cdot \mu_x = \dfrac{\mu_x}{\bar{x}} \cdot \bar{y}$ <br> $\hat{V}(\hat{\mu}_{\text{kvot}}) = \left(1 - \dfrac{n}{N}\right) \cdot \dfrac{1}{n} \cdot \left(\dfrac{\sum_{k \in S}\left(y_k - \hat{R}x_k\right)^2}{n-1}\right)$ |

**Regressionsskattning under OSU u.å.:**

| Parameter | Punkt- och variansskattning |
|---|---|
| $\mu_y$ | $\hat{\mu}_{\text{reg}} = \bar{y} + b(\mu_x - \bar{x})$ där $b = \dfrac{\sum_{k \in S}(y_k - \bar{y})(x_k - \bar{x})}{\sum_{k \in S}(x_k - \bar{x})^2}$ <br> $\hat{V}\left(\hat{\mu}_{\text{reg}}\right) = \left(1 - \dfrac{n}{N}\right) \cdot \dfrac{1}{n} \cdot \left(\dfrac{\sum_{k \in S}(y_k - \bar{y})^2 - b^2 \sum_{k \in S}(x_k - \bar{x})^2}{n-2}\right)$ <br> där $\sum_{k \in S}(y_k - \bar{y})^2 = \sum_{k \in S} y_k^2 - n\bar{y}^2$ |
| $\tau_y$ | $\hat{\tau}_{\text{reg}} = N \cdot \hat{\mu}_{\text{reg}}$ <br> $\hat{V}\left(\hat{\tau}_{\text{reg}}\right) = N^2 \cdot \hat{V}\left(\hat{\mu}_{\text{reg}}\right)$ |

**Poststratifiering under OSU u.å.:**

Parametrar och estimatorer - se under **Stratifierat urval** ovan

OBS! Populationsvikterna $W_k$ måste vara kända.

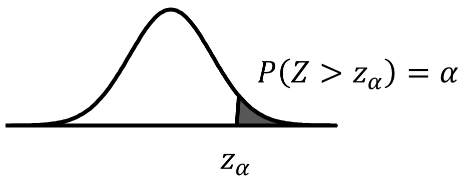
Variansskattning – *formler utgår*

## Från tabellsamlingen

**TABELL 2.** Normalfördelningens kvantiler, standardiserad

$Z \in N(0, 1)$. Vilket värde har $z_\alpha$ om $P(Z > z_\alpha) = \alpha$ där $\alpha$ är en given sannolikhet.

Utnyttja även $\Phi(-z) = 1 - \Phi(z)$ för $P(Z \leq -z_\alpha)$.




| $\alpha$ | $z_\alpha$ |
|---:|---:|
| 0,25 | 0,6745 |
| 0,10 | 1,2816 |
| 0,05 | 1,6449 |
| 0,025 | 1,9600 |
| 0,010 | 2,3263 |
| 0,005 | 2,5758 |
| 0,0025 | 2,8070 |
| 0,0010 | 3,0902 |
| 0,0005 | 3,2905 |
| 0,00025 | 3,4808 |
| 0,00010 | 3,7190 |
| 0,00005 | 3,8906 |
| 0,000025 | 4,0556 |
| 0,000010 | 4,2649 |
| 0,000005 | 4,4172 |