

## **Tentamen i Undersökningsmetodik (4,5 hp)**

### **Kurs: Regressionsanalys och undersökningsmetodik**

**2023-02-16**

---

<b>Skrivtid:</b>	kl. 8.00 - 13.00 (5 timmar)
<b>Godkända hjälpmedel:</b>	Miniräknare utan lagrade formler och lagrad text
<b>Vidhäftade hjälpmedel:</b>	Formelsamling och Statistiska tabeller (endast de tabeller som krävs)

- Tentamen består av 4 uppgifter uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.
- Svar med fullständiga redovisningar och motiveringar ska lämnas.
  - Använd endast skrivpapper som tillhandahålls i skrivsalen.
  - Max en uppgift per blad, t.ex. uppgift 1a-d på ett eller flera blad men börja på nytt blad för nästa uppgift 2a-d osv.
  - För full poäng på en uppgift 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. tabellerna men utgå då ifrån det som är givet. Du kan dock med fördel avrunda det slutliga svaret.
- Maxpoäng är 100 och för godkänt resultat krävs minst 50 poäng. 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.

- Lösningförslag läggs ut på Athena kort efter tentamen.

**LYCKA TILL!**

### Uppgift 1. (25p)

Ett konsultföretag genomförde en undersökning kring sexuella trakasserier på arbetsplatsen bland jurister. Ett enkelt slumpmässigt urval av 200 jurister drogs från en population bestående av 1 500 jurister. Svaren fördelade sig över kön (juridisk definitionen av kön, kvinnor och män) och svaret på frågan (Ja/Nej) om det förekommer sexuella trakasserier på deras arbetsplats enligt följande tabell:

	Ja	Nej
Kvinnor	34	36
Män	12	118

Du ska i det följande inkludera ändlighetskorrektionen ( $fpc$ ) i dina beräkningar.

- Skatta  $P$  = andelen jurister i populationen som anser att det förekommer sexuella trakasserier. Beräkna sedan ett 90 % konfidensintervall för  $P$ . (6p)
- Skatta  $A$  = antalet jurister i populationen som anser att det förekommer sexuella trakasserier. Beräkna sedan ett 90 % konfidensintervall för  $A$ . (2p)

En kollega till dig vill gå ut med ett pressmeddelande att det förekommer sexuella trakasserier på  $100 \cdot \hat{p}$  % av advokatkontoren i Sverige där  $\hat{p}$  är din skattning från a) ovan.

- Förklara för din kollega som saknar utbildning i statistik vad det är för fel med den slutsatsen. TIPS: Det finns minst tre fel. ( $3 \cdot 2 = 6p$ )

Det visar sig senare att urvalsstorleken egentligen var  $n = 250$ . Bortfallet i undersökningen uppgick alltså till  $n_b = 50$  jurister som av olika skäl inte deltog i undersökningen och antalet svarande var  $n_s = 200$ . Man misstänker starkt att det inte går att ignorera bortfallet, dvs. att svaret på frågan är korrelerad med viljan (eller sannolikheten) att svara på frågan.

- Vilka möjliga värden hade din skattning  $\hat{p}$  kunnat bli om du hade fått svar från samtliga 250? Vilken skattningsmetod kan du använda i detta fall för att justera för bortfallet? Du behöver inte göra någon beräkning men ett svar på ett ord räcker inte, däremot 1-2 meningar. (5p)

Vid ett projektmöte med dina kollegor bestämde man sig för att börja planera för en ny undersökning. Ni är överens om att utvidga populationen till samtliga jurister i Sverige (ni kan få tillgång till Advokatsamfundets medlemsregister som ni kan använda som ram). Ni vill ha ett stratifierat urval med kön som stratifieringsvariabel för att bättre kunna skatta varje kön för sig. För kvinnorna vill ni dessutom att konfidensintervallets längd inte ska överskrida 3.92 %-enheter vilket är samma sak som att säga att felmarginalen högst får vara 1.96 %-enheter.

- Hur stort måste  $n_k$  = stickprovsstorleken i kvinnogruppen vara för att *garantera* att felmarginalen för  $P_k$  i detta stratum som mest kan bli 1.96 %-enheter? Utgå ifrån 95 % konfidensgrad. Eftersom  $N_k$  = antalet kvinnliga jurister i dagsläget inte är känt så kan du bortse ifrån ändlighetskorrektionen  $\frac{N_k - n_k}{N_k - 1}$  i dina beräkningar. TIPS: Utgå ifrån den teoretiska variansen i formelsamlingen. (6p)

## Uppgift 2. (25p)

Ett stort företag har gett dig i uppdrag att analysera resultatet av företagets utvecklingsprojekt som genomfördes och infördes i verksamheten under år 2021. Totalt hade  $N = 121$  projekt slutförts.

Du får lätt reda den totala kostnaden för varje enskilt projekt. Däremot är vinsten i termer av ökade intäkter och/eller besparingar av varje enskilt projekt svårare och mer tidskrävande att mäta. Du har efter några dagars hårt arbete lyckats beräkna total vinst under tolv månader efter att resultaten införts i verksamheten för  $n = 10$  av projekten. Du använder följande beteckningar:

$$x_k = \text{total kostnad i mkr för projekt } k = 1, \dots, 121$$

$$y_k = \text{total vinst i mkr för projekt } k = 1, \dots, 121$$

Du vet också att den totala kostnaden för samtliga 121 projekt var  $\tau_x = 965$  mkr. Du kan med gott samvete anta att de  $n = 10$  granskade projekten utgör ett OSU u.å. Dina data kan sammanställas enligt följande:

Projekt $k$	1	2	3	4	5	6	7	8	9	10	Summa
$x_k$	2	4	4	6	8	8	10	12	12	14	80
$y_k$	8	12	10	16	21	16	18	20	19	17	157
$x_k^2$	4	16	16	36	64	64	100	144	144	196	784
$y_k^2$	64	144	100	256	441	256	324	400	361	289	2 635
$x_k y_k$	16	48	40	96	168	128	180	240	228	238	1 382

Du ska i det följande inkludera ändlighetskorrektionen (*fpc*) i dina beräkningar.

- Resonera kortfattat utifrån sakproblemet och variabelernas definitioner om en kvotskattning eller en regressionsskattning är bäst i detta fall och välj en av metoderna. Motivera ditt svar. (5p)
- Skatta den totala vinsten med den skattningsmetod som du i a) ovan kom fram till är bäst och beräkna standardfelet för skattningen. (10p)
- Skatta den totala vinsten med den väntevärdesriktiga HT-estimatorn under OSU u.å. och beräkna standardfelet för skattningen. (5p)
- I diagrammet på sidan 5 åskådliggörs de tio observationsparen  $(x_k, y_k)$ . Resonera kring vad du kan utläsa ur diagrammet med avseende på de villkor som ska vara uppfyllda för skattningsmetoden du valde i a) och använde i b). Kommentera även dina resultat i b) och c) ovan. Förklara kortfattat varför det blev bättre eller sämre med skattningsmetoden i b) jämfört med den i c). NOT: Du kan resonera utifrån diagrammet även om du inte har gjort samtliga uppgifterna innan. (5p)

### Uppgift 3. (30p)

Du har blivit ombedd att planera en statistisk undersökning gällande sockerbetsodling. Du finner i Jordbruksverkets statistik att det år 2020 producerades drygt 2 miljoner ton betor i Sverige. Antal hektar odlingsmark som användes uppgick till ca 29 800 och genomsnittlig skörd var 68 ton per hektar. Antalet sockerbetsodlare var (nästan) 1 000.

I tabellen nedan har de  $N = 1000$  jordbruken delats in i fyra storleksklasser 1-4 dvs. fyra olika strata efter storlek = odlingsmark. Vidare redovisas total och genomsnittlig skörd per jordbruk i ton (1000 kg) samt standardavvikelse för skörden per jordbruk (dessa siffror är påhittade!).

Jordbruk		Odlingsmark år 2020		Skörd år 2020		
Storlek: $h$	Antal: $N_h$	Totalt	Medel	Total: $\tau_h$	Medel: $\mu_h$	St.avv: $\sigma_h$
1: små	200	1 200	6	81 600	408	450
2: medium	420	9 240	22	628 320	1 496	700
3: stora	250	10 000	40	680 000	2 720	1400
4: jättestora	130	9 360	72	636 480	4 896	2500
Totalt	1 000	29 800		2 026 400		

Medelskörd per jordbruk och standardavvikelsen för hela populationen kan beräknas till  $\mu_y = 2026.4$  ton respektive  $\sigma_y = 1 838.9$

- Argumentera kortfattat och utifrån givna data varför det bör löna sig att dra ett stratifierat OSU u.å. med odlingsmark som stratifieringsvariabel jämfört med ett enkelt OSU u.å. (5p)
- Du har blivit ombedd att dra ett urval av  $n = 100$  jordbruk för att analysera skörden för år 2022. Använd all tillgänglig information och ta fram en optimal allokering av stickprovet för din kommande undersökning. (5p)

Man drar sedan ett stratifierat OSU u.å. av  $n = 100$  slumpmässigt valda gårdar fördelade enligt din allokering som du fick i b) ovan. Resultatet för 2022 års skörd blev följande (observera att  $N_h =$  antal gårdar per stratum har förändrats något jämfört med år 2020):

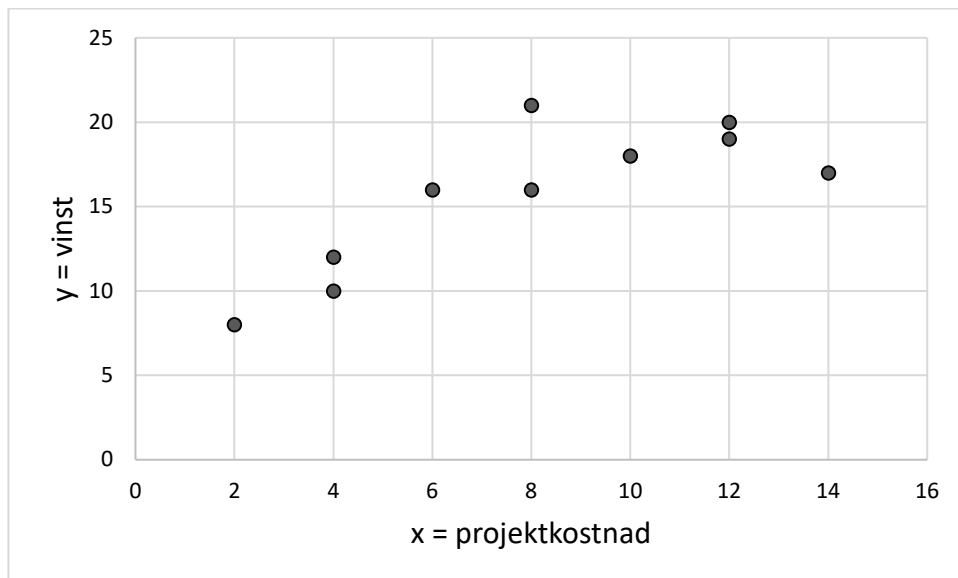
Storlek: $h$	$N_h$	$n_h$	$W_h$	$N_h^2$	$\bar{y}_h$	$s_h^2$
1: små	190			36 100	425	176 400
2: medium	410			168 100	1 620	656 100
3: stora	260			67 600	3 010	2 016 400
4: jättestora	140			19 600	5 040	6 760 000
Totalt	1 000	100	1			

- Beräkna en punktskattning och ge ett 90% konfidensintervall för  $\tau_y =$  den totala skörden för hela populationen. Ändlighetskorrektions ska användas genomgående i din lösning. Utgå ifrån din allokering som du fick i b) ovan; om du inte har gjort uppgift b) använd istället  $n_1 = 19$ ,  $n_2 = 41$ ,  $n_3 = 26$  och  $n_4 = 14$ . (10p)

(forts. nästa sida)

- d) Förklara vad en inklusionssannolikhet är för något och beräkna sedan dessa sannolikheter för samtliga jordbruk i populationen. TIPS: Det är ett OSU u.å. inom varje enskilt stratum och oberoende dragningar mellan strata. (5p)
- e) Förklara vad som menas med att en skattning är väntevärdesriktig (*unbiased*). Om man drar ett stratifierat OSU u.å. och använder de skattningsformler som finns i formelsamlingen, kommer detta upplägg (design) ge väntevärdesriktiga skattningar? Motivera ditt svar. (5p)

**OBS! Detta diagram hör till uppgift 2**



#### Uppgift 4. (20p)

För var och en av följande deluppgifter ska du svara kortfattat. Hela uppgiften bör kunna redovisas på maximalt ca två A4-sidor. Inga beräkningar behövs men du får gärna illustrera med bilder och skisser om det underlättar, även formler dvs. matematisk framställning går bra om du tycker att det är relevant.

- a) Om man utgår från de olika potentiella felkällorna i en statistisk undersökning kan det **totala felet** (TSE) definieras som summan av fem olika fel som samtliga har tagits upp under kursen. Lista dessa feltyper och förklara kortfattat med en mening per typ hur de definieras. (5p)
- b) Välj en av feltyperna som du har angett i uppgift a) ovan och beskriv denna mer utförligt. Beskriv hur felet uppstår, vilka effekterna av felet kan bli, hur man kan planera sin undersökning för att minimera risken för stora fel och vad man kan göra efter genomförd datainsamling för minska effekterna av felet. (5p)
- c) Det finns ett antal s.k. kvalitetskriterier som enligt lag ska beaktas och deklarerats i Sveriges officiella statistik. Välj ut två av dessa kriterier och beskriv dem kortfattat. Finns det någon målkonflikt mellan de två kriterierna? Motivera ditt svar. (5p)
- d) Förklara vad CASM-modellen handlar om och beskriv de olika stegen som definierar modellen. Resonera sedan utifrån modellen skillnaden mellan faktafrågor och attitydfrågor. Det räcker om du kan peka på någon väsentlig skillnad. (5p)

**Obs! Om du är osäker på hur mycket som krävs för varje delfråga så reflektera över hur många poäng du kan få för varje delfråga. Du behöver alltså inte skriva långt och mycket, bara det viktigaste!**

# Formel- och tabellsamling

## DESKRIPTIV STATISTIK

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

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

## 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  $k$ , ex.  $\sum y_k = \sum_{k \in S} y_k$

---

**OBUNDET SLUMPMÄSSIGT URVAL u.å. (HT)**

Parameter	Punktskattning	Teoretisk 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. (HT)**

Notation:  $L =$  antal strata

$N_h =$  populationsstorleken för stratum  $h = 1, \dots, L$

$n_h =$  stickprovets storlek i stratum  $h = 1, \dots, L$

$W_h = N_h/N$

$\bar{y}_h =$  stickprovsmedelvärde i stratum  $h = 1, \dots, L$

$s_h^2 =$  stickprovsvarians i stratum  $h = 1, \dots, L$

Parameter	Punktskattning	Variansskattning $\hat{V}(\cdot)$
$\mu$	$\bar{y}_{\text{str}} = \sum_{h=1}^L W_h \bar{y}_h$	$\hat{V}(\bar{y}_{\text{str}}) = \sum_{h=1}^L W_h^2 \left( 1 - \frac{n_h}{N_h} \right) \frac{s_h^2}{n_h}$
$\tau$	$\hat{\tau}_{\text{str}} = N\bar{y}_{\text{str}}$	$\hat{V}(\hat{\tau}_{\text{str}}) = \sum_{h=1}^L N_h^2 \left( 1 - \frac{n_h}{N_h} \right) \frac{s_h^2}{n_h}$
$P$	$\hat{p}_{\text{str}} = \sum_{h=1}^L W_h \hat{p}_h$	$\hat{V}(\hat{p}_{\text{str}}) = \sum_{h=1}^L W_h^2 \left( 1 - \frac{n_h}{N_h} \right) \frac{\hat{p}_h(1-\hat{p}_h)}{n_h-1}$
$A$	$\hat{A}_{\text{str}} = N\hat{p}_{\text{str}}$	$\hat{V}(\hat{A}_{\text{str}}) = \sum_{h=1}^L N_h^2 \left( 1 - \frac{n_h}{N_h} \right) \frac{\hat{p}_h(1-\hat{p}_h)}{n_h-1}$

Optimal allokering: 
$$n_h = n \cdot \frac{N_h \sigma_h}{\sum_{j=1}^L N_j \sigma_j}$$



## 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  $i = 1, 2, \dots, N$

$\bar{m}$  = stickprovsmedelvärde av klusterstorlekarna  $m_i$

$s_m^2$  = 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  $i = 1, 2, \dots, N$

$\bar{\tau}$  = stickprovsmedelvärde av totalvärdena  $\tau_i$

$s_\tau^2$  = 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  $1$ )

vvr står för den vanliga HT-estimatorn under OSU u.å.

Parameter	Punktskattning	Variansskattning
$M$	$\hat{M}_{\text{vvr}} = N \cdot \bar{m}$	$\hat{V}(\hat{M}_{\text{vvr}}) = N^2 \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{s_m^2}{n}$
$\mu$	$\bar{y}_{\text{vvr}} = \frac{\hat{t}_{\text{vvr}}}{M} = \frac{N\bar{\tau}}{M}$	$\hat{V}(\bar{y}_{\text{vvr}}) = \frac{N^2}{M^2} \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{s_\tau^2}{n}$
	$\bar{y}_{\text{kvot}} = \frac{\hat{t}_{\text{vvr}}}{\hat{M}_{\text{vvr}}} = \frac{\sum_{i \in S} \tau_i}{\sum_{i \in S} m_i}$	$\hat{V}(\bar{y}_{\text{kvot}}) = \left(1 - \frac{n}{N}\right) \cdot \frac{1}{\bar{m}^2} \cdot \frac{\sum_{i \in S} (\tau_i - \bar{y}_{\text{kvot}} m_i)^2}{n(n-1)}$
		där $\sum_{i \in S} (\tau_i - \bar{y}_{\text{kvot}} m_i)^2 = [\text{jmf}r \text{ nästa sida}]$ $= \sum_{i \in S} \tau_i^2 - 2\bar{y}_{\text{kvot}} \sum_{i \in S} \tau_i m_i + \bar{y}_{\text{kvot}}^2 \sum_{i \in S} m_i^2$
$\tau$	$\hat{t}_{\text{vvr}} = N\bar{\tau}$	$\hat{V}(\hat{t}_{\text{vvr}}) = N^2 \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{s_\tau^2}{n}$
	$\hat{t}_{\text{kvot}} = M\bar{y}_{\text{kvot}} = \frac{M}{\hat{M}_{\text{vvr}}} \hat{t}_{\text{vvr}}$	$\hat{V}(\hat{t}_{\text{kvot}}) = \left(\frac{M}{\bar{m}}\right)^2 \left(1 - \frac{n}{N}\right) \frac{\sum_{i \in S} (\tau_i - \bar{y}_{\text{kvot}} m_i)^2}{n(n-1)}$
$P$	<i>formler utgår</i>	
$A$	<i>formler utgår</i>	

## SKATTNINGSMETODER

Notation:  $\tau_y$  = totalvärdet för variabeln  $y$  för hela populationen  
 $\hat{t}_y$  = HT-skattningen av  $\tau_y$  under OSU  
 $\mu_y$  = populationsmedelvärdet av för variabeln  $y$   
 Motsvarande beteckningar gäller för variabeln  $x$

### Kvotskattning under OSU u.å.:

Parameter	Punkt- och variansskattning
$\mu_y$	$\hat{\mu}_{\text{kvot}} = \hat{R} \cdot \mu_x = \frac{\sum_{k \in S} y_k}{\sum_{k \in S} x_k} \cdot \mu_x = \frac{\mu_x}{\bar{x}} \cdot \bar{y} \quad \text{där} \quad \hat{R} = \frac{\sum_{k \in S} y_k}{\sum_{k \in S} x_k} = \frac{\hat{t}_y}{\hat{t}_x} = \frac{\bar{y}}{\bar{x}}$ $\hat{V}(\hat{\mu}_{\text{kvot}}) = \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \left(\frac{\sum_{k \in S} (y_k - \hat{R}x_k)^2}{n-1}\right)$ <p style="text-align: center;">där <math>\sum_{k \in S} (y_k - \hat{R}x_k)^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</math></p>
$\tau_y$	$\hat{t}_{\text{kvot}} = N \cdot \hat{\mu}_{\text{kvot}} = \hat{R} \cdot \tau_x = \frac{\sum_{k \in S} y_k}{\sum_{k \in S} x_k} \cdot \tau_x = \frac{\tau_x}{\hat{t}_x} \cdot \hat{t}_y$ $\hat{V}(\hat{t}_{\text{kvot}}) = N^2 \cdot \hat{V}(\hat{\mu}_{\text{kvot}})$

### Regressionsskattning under OSU u.å.:

Parameter	Punkt- och variansskattning
$\mu_y$	$\hat{\mu}_{\text{reg}} = \bar{y} + b(\mu_x - \bar{x}) \quad \text{där} \quad b = \frac{\sum_{k \in S} (y_k - \bar{y})(x_k - \bar{x})}{\sum_{k \in S} (x_k - \bar{x})^2} \quad (\text{se sid 7})$ $\hat{V}(\hat{\mu}_{\text{reg}}) = \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \left(\frac{\sum_{k \in S} (y_k - \bar{y})^2 - b^2 \sum_{k \in S} (x_k - \bar{x})^2}{n-2}\right)$ <p style="text-align: center;">där <math>\sum_{k \in S} (y_k - \bar{y})^2 = \sum_{k \in S} y_k^2 - n\bar{y}^2 = (n-1)s_y^2</math></p>
$\tau_y$	$\hat{t}_{\text{reg}} = N \cdot \hat{\mu}_{\text{reg}} = N\bar{y} + Nb(\mu_x - \bar{x})$ $\hat{V}(\hat{t}_{\text{reg}}) = N^2 \cdot \hat{V}(\hat{\mu}_{\text{reg}})$

### Poststratifiering under OSU u.å.:

Parametrar och punktskattning - se under **Stratifierat urval**

OBS! Populationsvikterna  $W_h$  måste vara kända.

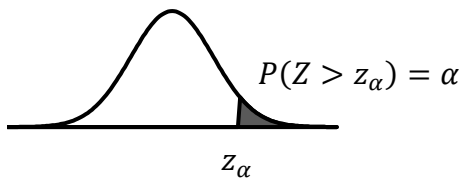
Variansskattning - *formler ingår inte i kursen*

## 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