Stockholms universitet

OBS! Läs noga igenom anvisningarna i tentamen, t.ex. hur du ska skriva svaren. Det är ditt ansvar som student att följa de anvisningar som ges.

NOTE! Read the examination instructions carefully, e.g. how to write the answers. It is your responsibility as a student to follow the given instructions.

Skriv din anonymiseringskod och dagens datum på allt material du lämnar in. (*Enter your anonymization code and todays date on all submitted materials*)												
Anonymiseringskod (*Anonymization code*)	3	1	1	-	0	0	0	5	-	H	A	T
Datum (*Date* YYYY-MM-DD)	2022-10-26							Plats nr. (*Seat No.*)		8		

Kurs/Kurskod (*Course/Course code*)	ST306G
Kursmoment (*Course component*)	

Fylls i av tentamensvärd (*To be filled in by invigilator*)

Direkt i skrivning: (kryss)		Svarsblankett: (kryss)		Lösa svarsblad: (antal)	11
Lämnat in blankt: (kryss)		Dator: (kryss)			

Inlämningstid: 18:27 Signatur tentamensvärd: Enad

Fylls i av lärare/examinator (*To be filled in by teacher/examinator*)

Betyg:	B	Poäng:	40

Signatur rättande lärare/examinator: JH

Stockholms universitet

Datum: (Date YYYY-MM-DD)	2022-10-26	Kurs/Kurskod: (Course/Course code)	ST306G	Sidnr.: (Page no.)	1
Anonymiseringskod (Anonymization code)	311-0005-HAT				

Uppg.nr.: (Task no.) 1

Lärarens kommentar: (Teacher's note)

1) b) R

2) ~~a)~~ D

3) c) R

4) e) R

5) a) R

8

---

6) I am gonna denote the incl.-probability of each sample with π_i.

a) $\pi_A = \pi_1 + \pi_2 + \pi_3 = 1/10 + 1/10 + 1/10$

$= 3/10 = 0.3$ R

1

b) ~~$\pi_i = n/N$~~

we have that $Pr(s_1) = Pr(s_2) = \dots = Pr(s_{10}) = 1/10$

$\hat{t}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}$, ~~where $\pi_i = n/N$~~

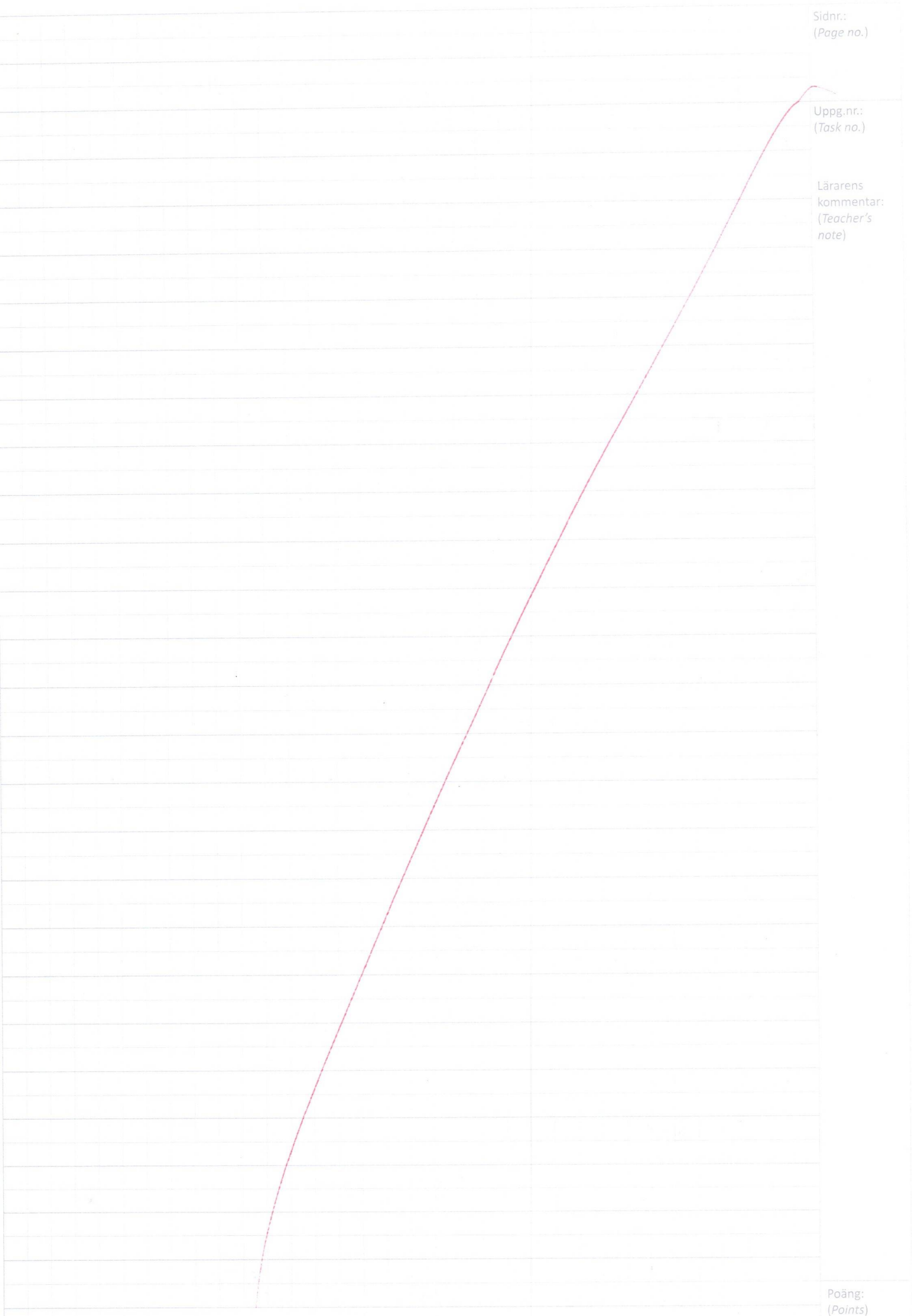
Poäng: (Points)

Tentamensservice/SU/typ_c

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

b) the inclusion probability for each unit:

$\pi_A = 3/10$, $\pi_B = 1/2$, $\pi_C = 1/2$, $\pi_D = 3/10$

$\pi_E = 3/10$, $\pi_F = 3/10$, $\pi_G = 4/10$, $\pi_H = 2/10$

$\pi_I = 2/10$

~~and as $\pi_i = n/N$, we~~

$\Rightarrow$

$\hat{t}_{S1} = \frac{10}{3}2 + \frac{2}{1}2 + \frac{2}{1}3 = 16.667$

$\hat{t}_{S2} = \frac{10}{3}2 + \frac{2}{1}2 + \frac{10}{3}4 = 24$

$\hat{t}_{S3} = \frac{10}{3}2 + \frac{2}{1}3 + \frac{10}{3}4 = 26$

$\hat{t}_{S4} = \frac{2}{1}2 + \frac{2}{1}3 + \frac{10}{3}4 = 23.333$

$\hat{t}_{S5} = \frac{2}{1}2 + \frac{2}{1}3 + \frac{10}{3}4 = 23.333$

$\hat{t}_{S6} = \frac{2}{1}2 + \frac{2}{1}3 + \frac{10}{3}4 = 23.333$

$\hat{t}_{S7} = \frac{10}{3}4 + \frac{10}{3}4 + \frac{10}{3}5 = 43.332$

$\hat{t}_{S8} = \frac{10}{3}4 + \frac{10}{3}5 + \frac{10}{2}5 = 55$

$\hat{t}_{S9} = \frac{10}{3}4 + \frac{10}{3}5 + \frac{10}{2}6 = 60$

$\hat{t}_{S10} = \frac{10}{3}5 + \frac{10}{2}5 + \frac{10}{2}6 = 71.667$

HT is unbiased so $E(\hat{t}_y) = 35$

No need to calculate anything

But in principle correct

$\Rightarrow$

$E(\hat{t}_y) = \left(\frac{1}{10}16.667\right) + \left(\frac{1}{10}24\right) + \left(\frac{1}{10}26\right) + \left(\frac{1}{10}23.333\right) +$

$+ \left(\frac{1}{10}23.333\right) + \left(\frac{1}{10}23.333\right) + \left(\frac{1}{10}43.332\right) + \left(\frac{1}{10}55\right) + \left(\frac{1}{10}60\right)$

$+ \left(\frac{1}{10}71.667\right) = \underline{36.6665}$ 35

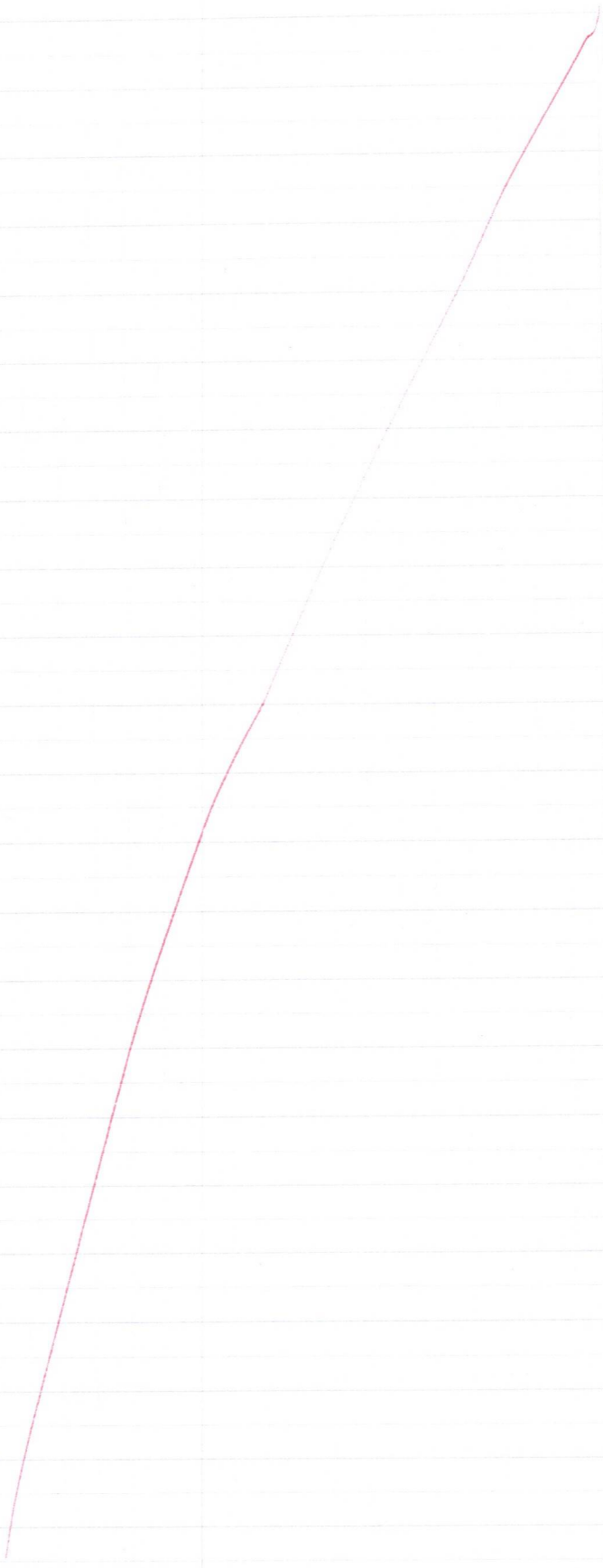
2

Poäng: (Points)

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

6. c) no it doesn't. R
(could be a special case of cluster sampling - but would say not).

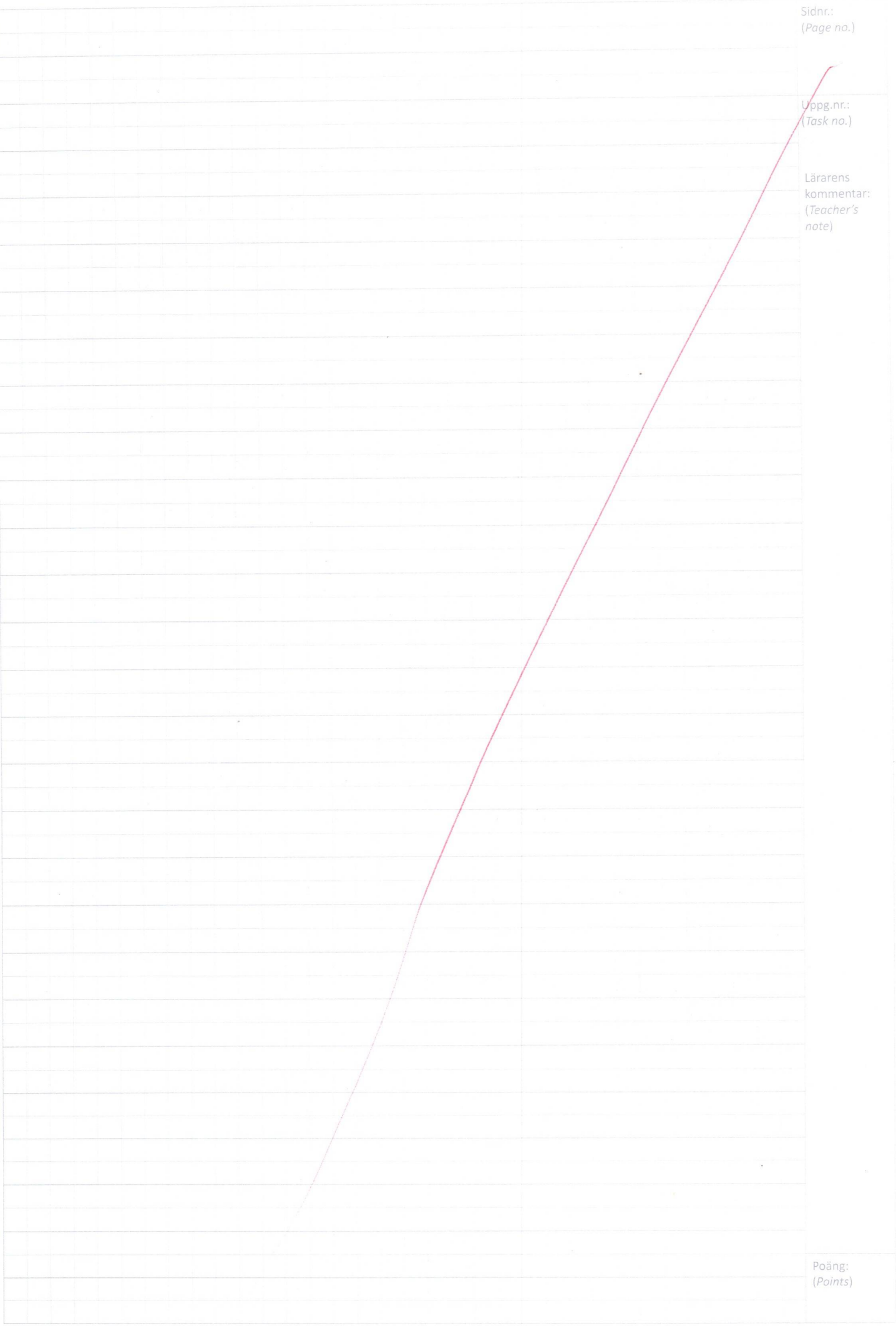
d) The population sum of inclusion probabilities is 1. No For example 0
$\sum_U \frac{n}{N} = \frac{n}{N} \sum_U 1$
$= \frac{n}{N} N = n$

e) As **x** is known to be strongly correlated with **y** - it would make sense to me to use stratified simple random sampling with one stratum being for $x \leq 3$, and and one stratum being for $x > 3$.
This would make sense as we have one unit with a value of **x** that is much larger than the rest of the units. R 4

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

a) domain - a subset of the target population which one wishes to estimate. Domains can overlap in a population. ok

~~stratum~~

stratum - also a subset of the target population, however, all strata combined makes up to the whole population and no strata can overlap within population. A sample from each stratum

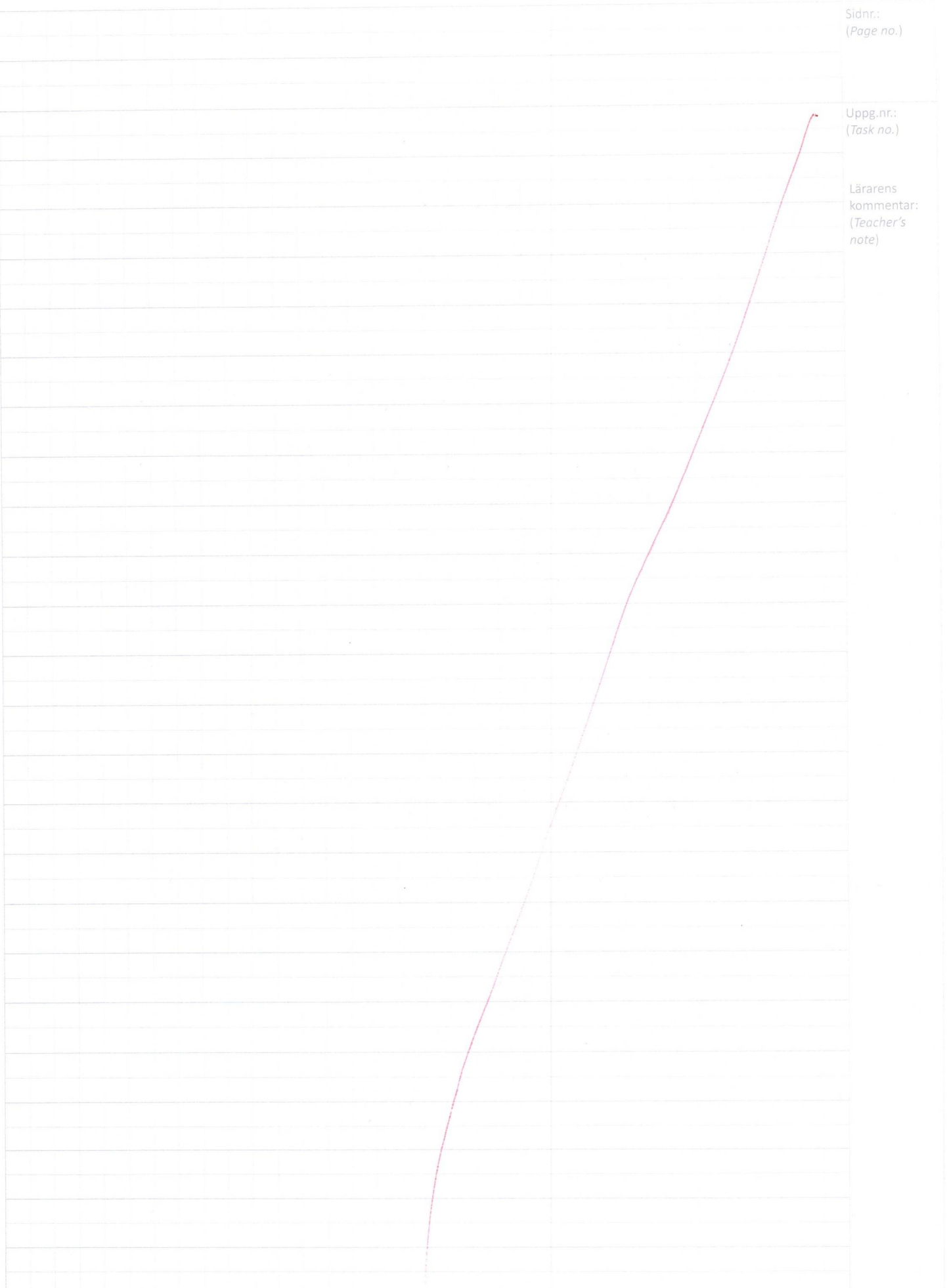
poststratum - is designed after a sample has been collected. For more precise variances estimations and/or to reduce non-response bias. The same holds for poststrata & strata when it comes to overlap and that all poststrata make up the whole population. ok

2

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

7 b) we have $N_d = 200$ (Renault made in 2021, owner)

$n_d = 76$

$\sum_{i \in s} u_i = 72$, where $u_i = y_i x_i$

where $x_i = \begin{cases} 1 - \text{if owner of a car Renault from 2021} \\ 0 - \text{otherwise} \end{cases}$

and $y_i = \begin{cases} 1 - \text{if pleased with car} \\ 0 - \text{otherwise} \end{cases}$

the proportion is given by

$$\bar{y}_d = \hat{P}_d = \frac{\bar{u}_s}{\bar{x}_s}$$

$$\bar{u}_s = \frac{72}{1000} = 0.072, \quad \bar{x}_s = \frac{76}{1000} = 0.076$$

$$\hat{P}_d = \frac{0.072}{0.076} \approx \underline{0.9474}$$

we also have that $s^2_{yd} = 0.10$

or we can compute an estimate by: $\hat{P}_d(1 - \hat{P}_d) = 0.0498$

extra

$\hat{t}_d = N_d \bar{y}_d \Rightarrow 189.48$

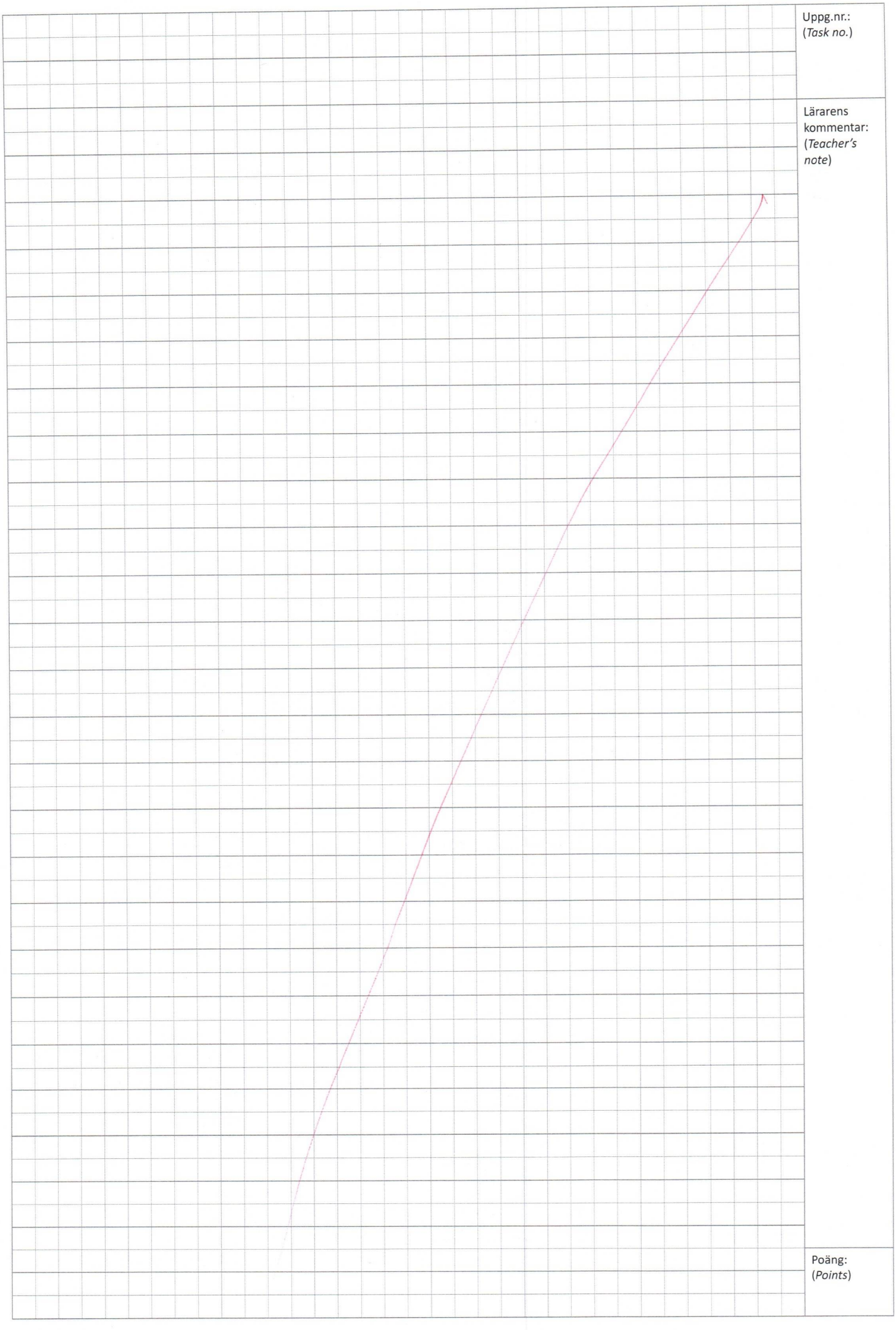
Variance formula:

$$\hat{V}(\bar{y}_d) = \left(1 - \frac{n_d}{N_d}\right) \frac{s^2_{yd}}{n_d}$$

$$\Rightarrow \hat{V}(\hat{P}_d) = \left(1 - \frac{76}{200}\right) \frac{0.0498}{76} = \underline{0.0004}$$

5

Poäng: (Points)



Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

c) One reason is that the will to respond might correlate with the response, meaning if you are pleased with your Renault you will respond, if you aren't you will not bother to respond. ok
→ a case of NMAR.

Second reason is that you are more willing to respond if you recently bought a Renault. ok
If you are a ~~new~~ recent owner you are probably more likely to respond to if you are satisfied with your purchase or not.
If you have been an owner for 5 or perhaps 10+ years, you are probably less eager to respond since the purchase was made long time ago and does not seem as relevant. ok, possible
→ This speaks for MAR

Both reasons speaks against MCAR.

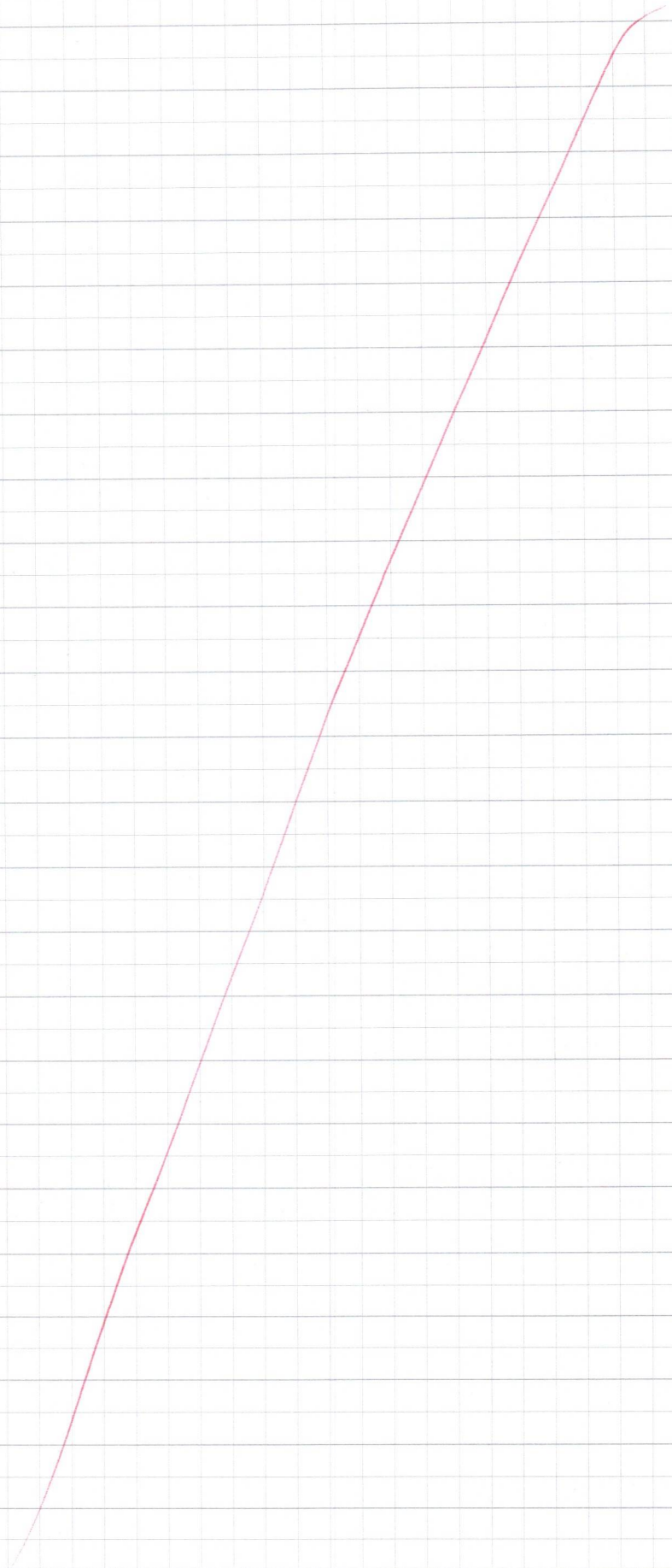
1

Poäng: (Points) 8

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

8 a) A likely reason for non-response is that an ill individual (likely in covid) does not want to be swabbed, nor meeting strangers. This would could a de negative bias - if a lot of the non-respondents are ill - meaning our variances and proportions would be under estimated.

2

b) Since sampling design was SRS, following formula will be used:

$$\hat{t}_y = \frac{N}{n}\sum_{i \in s} y_i, \Rightarrow \bar{y}_u = \hat{p}_u = \frac{\hat{t}_y}{N}$$

$$\hat{t}_y = \frac{20\,000}{269}\,25 = 1858.736$$

$$\Rightarrow \bar{y}_u = \hat{p}_u = \frac{\hat{t}_y}{N} = \frac{1858.736}{20000} \approx \underline{0.093}$$ R

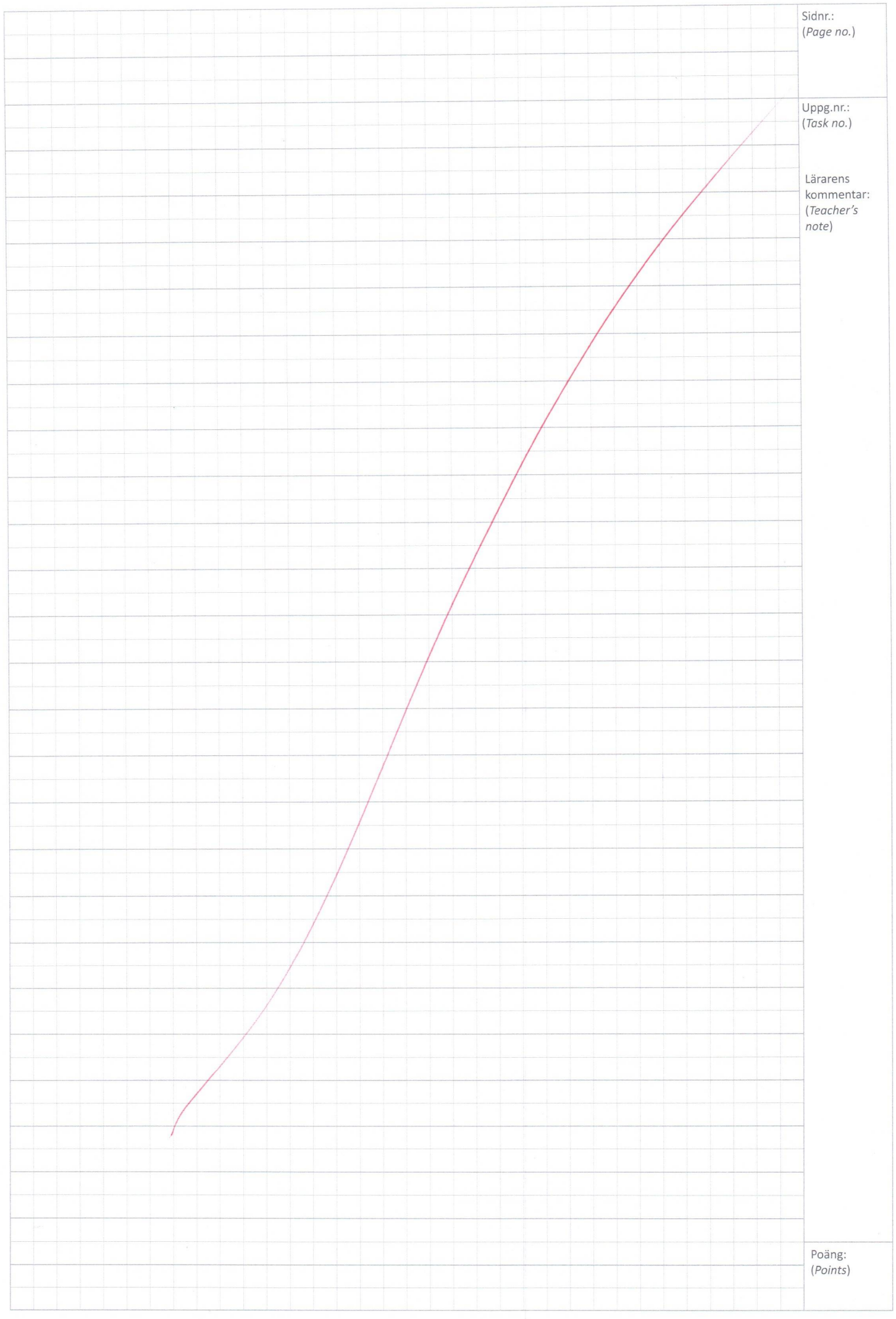
$$\hat{V}(\hat{p}_u) = \left(1 - \frac{n}{N}\right)\frac{\hat{p}_u(1-\hat{p}_u)}{n}$$

$$\Rightarrow \hat{V}(\hat{p}_u) = \left(1 - \frac{269}{20\,000}\right)\frac{0.0844}{269} \approx 0.00031$$

2

$$\hat{se}(\hat{p}_u) = \sqrt{\hat{V}(\hat{p}_u)} = \underline{0.0176}$$ R

Poäng: (Points)



Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

c) using the formulas for poststratification after SRS.

We first make use of Table 2 where we obtain $\sum_{g=1}^{G} \hat{t}_g$, the sum of totals in all strata. That is 2183.9 (or 2184).

The estimated proportion for the whole area is therefore:

$$\hat{p}_{post} = \bar{y}_{post} = \frac{1}{N} \sum_{g=1}^{G} \hat{t}_g$$

$$\Rightarrow \hat{p}_{post} = \frac{1}{20\,000} 2184 = 0.1092$$

for st.dev. we need the variance.

$$\hat{V}(\bar{y}_{post}) = \sum_{g=1}^{G} \frac{N_g^2}{N^2} \left(1 - \frac{n_g}{N_g}\right) \frac{s_g^2}{n_g}$$

$s_g^2 = \hat{p}_i(1-\hat{p}_i)$ values for $\left(1 - \frac{n_g}{N_g}\right)$ & $\frac{s_g^2}{n_g}$ can be found in the 3rd column & 5th column respectively. OK

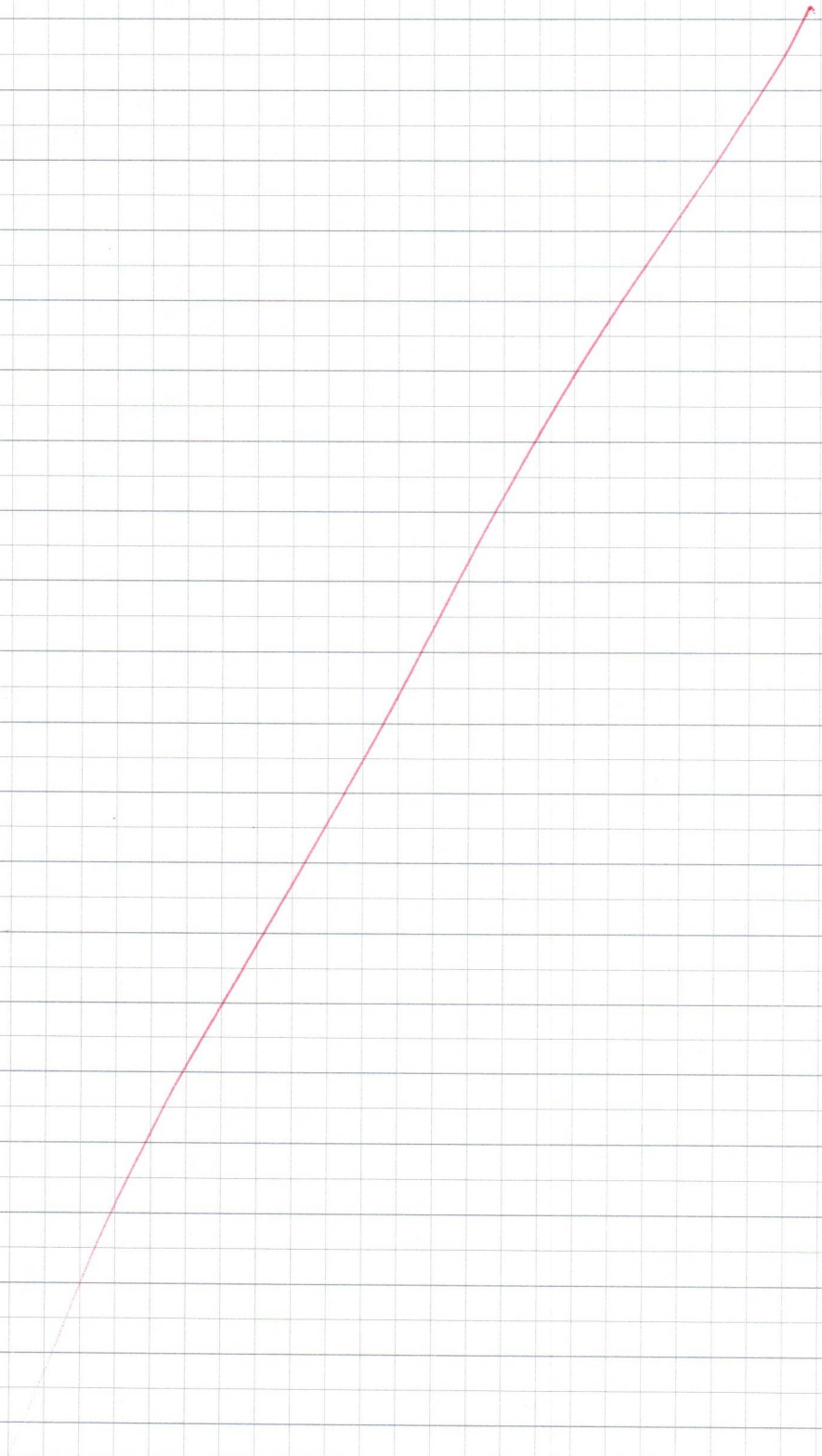
However, we could also use the sum of variances of all strata which is $\sum \hat{V}(\hat{t}_{post}) = 179349$ OK

Poäng: (Points)

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

8c). cont.

→ which we further divide by N^2 to obtain the variance for $\hat{p}_{post}$

$$\hat{V}(\hat{p}_{post}) = \frac{179349}{N^2} = 0.000448373$$

$$\hat{sd}(\hat{p}_{post}) = \sqrt{\hat{V}(\hat{p}_{post})} = 0.02117$$

The problem with the small numbers is that they are barely not taking into account in the estimated variance.

3

The action that could help is to redo a sample with stratified sampling with optimal allocation!

sum rows 3 and 4

d) Best sampling design would have been stratified SRS with optimal allocation, if possible.

Yes and what strata?

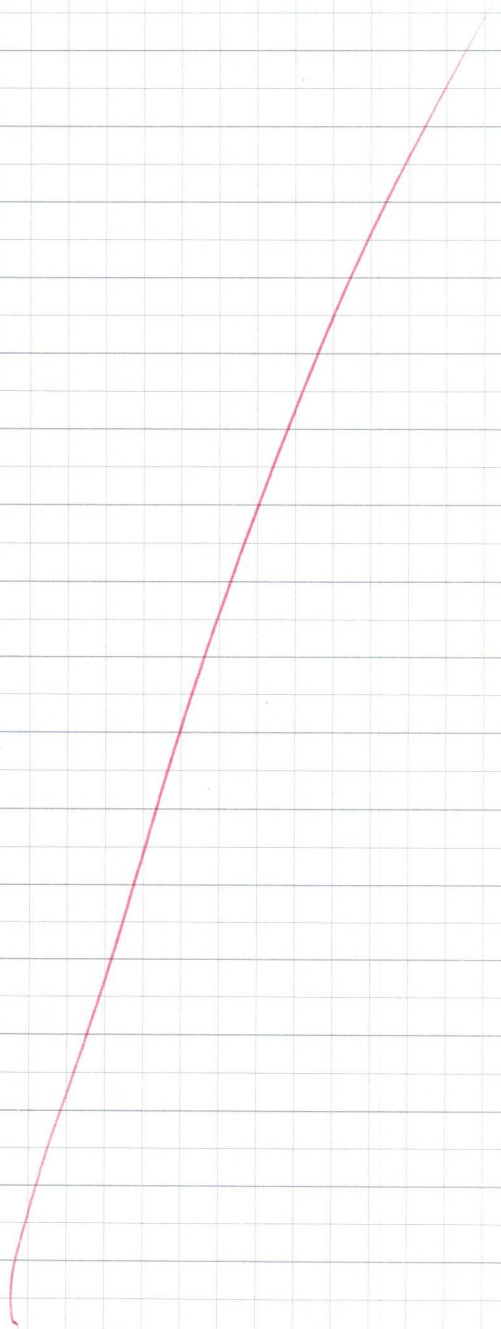
1

Poäng: (Points) 8

Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)



Poäng:
(*Points*)

9) a) we have $s_y^2 = 40\,000$ & $n = 1000$, we assume a 95% confidence-interval

$$n = \frac{1.96^2 s_y^2}{e^2} \Rightarrow e^2 = \frac{1.96^2 s_y^2}{n}$$

$$\Rightarrow \frac{1.96^2 \cdot 40\,000}{1000} = e^2 = 153.664$$

$$e = \sqrt{153.664} = 12.396$$

The C.I: $\bar{y}_0 = 20 \pm 12.396$ R 1

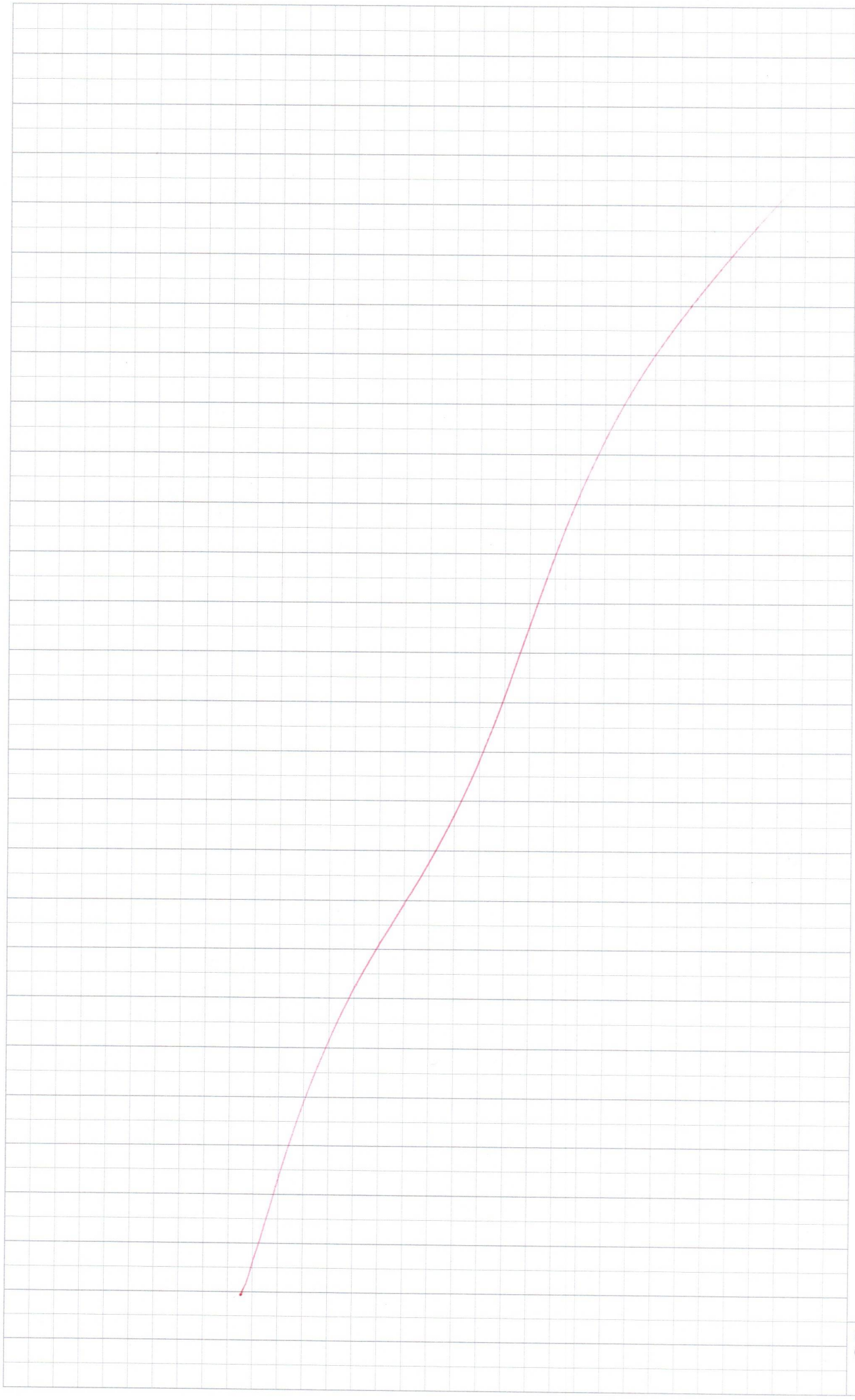
b) Since we know that $N_1 = N_2$ we could simplify the estimator to:

$$\hat{V}(\bar{y}_{str}) = 0.25\left[\left(1 - \frac{n_1}{N_h}\right)\frac{s_1^2}{n_1} + \left(1 - \frac{n_2}{N_h}\right)\frac{s_2^2}{n_2}\right] \quad 2$$

c) Since we have that $N_1 = N_2$, we use $s_1^2 = 5000$ & $s_2^2 = 20000$ to calculate proportions of n_1 & n_2

$$\Rightarrow n_h = n\frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} = \text{for } n_1 = n\frac{0.5(\sqrt{5000})}{0.5(\sqrt{5000}) + 0.5(\sqrt{20000})}$$

Poäng: (Points)



Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

9 c) cont.

$$\Rightarrow n_1 = n\frac{35.355}{35.355 + 70.7105} = n \cdot \underline{0.333}$$

meaning whatever size of $N_1 = N_2$ & n_h, strata 1 will have ≈ 0.33 proportion of n_h, and strata 2 will have ≈ 0.66 proportion of sample.

Meaning that $k = 1/2$

so we could re-express $\hat{V}(\bar{y}_{str}) \Rightarrow$

$$\hat{V}(\bar{y}_{str}) = 0.25\left[\left(1-\frac{n_2}{2N_h}\right)\frac{2s_1^2}{n_2} + \left(1-\frac{n_2}{N_h}\right)\frac{s_2^2}{n_2}\right]$$ R 2

d) $n = 2000$, $n_1 = 0.33 \cdot 2000 \approx 667$

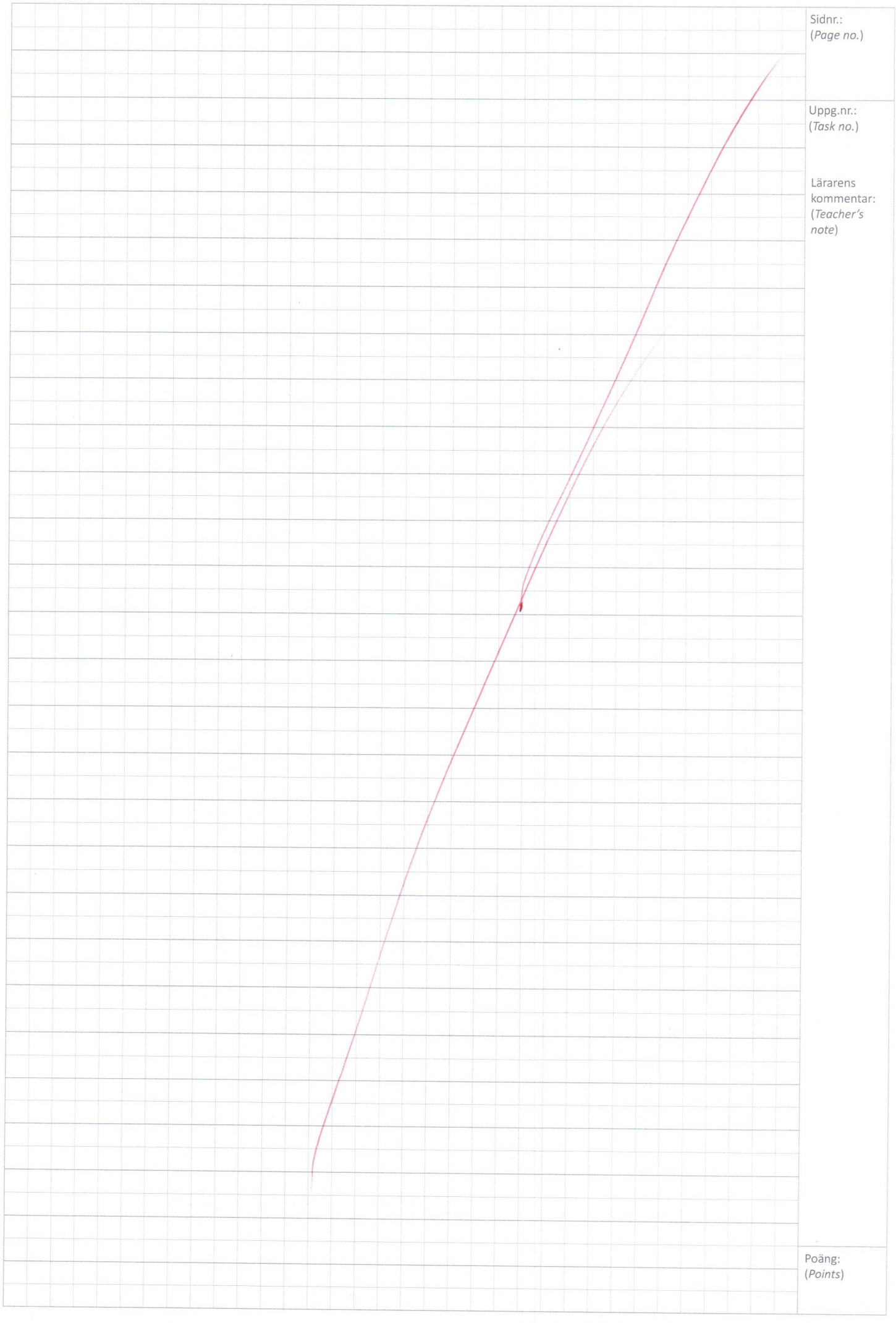
$n_2 = 0.66 \cdot 2000 \approx 1333$

we do not have a value for N_h, however, we can assume it being big enough so that $\left(1-\frac{n_h}{N_h}\right) \approx 1$ OK

$$\hat{V}(\bar{y}_{str}) = 0.25\left[\frac{5000(2)}{1333}\right] + 0.25\left[\frac{20\,000}{1333}\right]$$

$$= 1.875 + 3.75 = \underline{5.625}$$ R 2

Poäng: (Points) 7



Sidnr.:
(*Page no.*)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Regler i skrivsalen

- Följ tentamensvärds anvisningar.
- Väskor och ytterkläder ska placeras på anvisad plats.
- Placera ID-handling väl synlig på bordet framför dig.
- Ingen student får lämna skrivsalen under de första 30 minuterna.
- Endast en student i taget får besöka toaletten. Vid toalettbesök skriv ditt namn och klockslag på avsedd lista. Efter toalettbesöket ska du åter ange klockslag på listan.
- Elektronisk utrustning som mobiltelefon eller Smartwatch ska vara avstängd och placerad på anvisad plats.
- Under tentamen gäller tystnad – det är förbjudet att prata, eller på annat sätt kommunicera, med andra studenter under pågående tentamen.
- Innan tentamenshandlingarna lämnas in; skriv sidnummer, anonymiseringskod och datum på alla inlämnade papper.

Om något är oklart – fråga gärna tentamensvärden. Lycka till!

Rules in the examination hall

- Follow the invigilator's instructions.
- Bags and outerwear must be placed at the designated place.
- Place your ID document clearly visible on the table in front of you.
- No student may leave the examination hall for the first 30 minutes.
- Only one student at a time may visit the toilet. Before visiting the toilet, write your name and time on the intended list. After the toilet visit, enter the time on the list again.
- Electronic equipment such as a mobile phone or Smartwatch must be switched off and placed at the designated place.
- During the exam, silence applies – you are not allowed to talk, or otherwise communicate, with other students during the exam.
- Before submitting the examination documents; remember to write the page number, anonymization code, and date on all papers.

Please do not hesitate to ask the invigilator if anything is unclear. Good luck!

Tentamensservice
tentamensservice@su.se | www.su.se/tentamensservice

Stockholms universitet