STOCKHOLMS UNIVERSITET HT2016

Statistiska institutionen
Jonas Bjermo och Dan Hedlin

Sample surveys, ST306G

Examination 2016-10-26, 10.00 – 15.00

Approved aids:

1. Pocket calculator
2. Language dictionary

Separate pages with notes are not allowed.

The exam comprises 9 items, numbered 1 to 9. The maximum number of points is 50. 25 points will give you at least grade E. To obtain the maximum number of points full and clear motivations are required unless otherwise stated. You may write in English or Swedish. **There are some pages at the end of the exam with formulae that you may wish to use.**

In each of the five questions below one of the items a, b, c, d or e is incorrect. Which one? For each of the questions 1-5, answer with only one letter, a-e. Motivation is not required. Maximum 10 points.

1.

a) Six strata is usually enough if you want to stratify, for example, businesses in the ship building industry by size (employment) in order to estimate the total turnover in this industry.
b) Strata do not overlap.
c) You cannot use different sampling designs in different strata.
d) One common aim of stratification is to create strata that are internally more homogeneous than the whole population.
e) One common aim of stratification is to create strata that are similar to important domains.

2.

a) In design-based inference (which we have focused on in the course), a sample size of 30 units is nearly always enough for the central limit theorem to apply when constructing a 95% confidence interval.
b) In design-based inference the study variable is regarded as nonrandom.
c) When sampling from a finite population, the variance of an estimate usually contains a finite population correction.
d) When predicting what sample size you need to obtain the desired accuracy of an estimate, the size of the population is not important as long as the population is much larger than the sample.
e) One difference between a theoretical variance and a variance estimator is that the former may contain unknown parameters whereas in the latter unknown parameters are replaced with estimates.

3.

a) The ratio estimator is biased. However, the bias is usually small compared to the variance.
b) The regression estimator is biased. However, the bias is usually small compared to the variance.
c) Let x denote an auxiliary variable and y the study variable in a finite population and assume that y and x are strongly but negatively correlated. Suppose that the aim of a survey is to estimate the total of y. Then the variance of the regression estimator of the total is smaller than the variance of the estimator $\hat{t}_y = \sum_{i \in s} \frac{y_i}{\pi_i}$, where π_i is the inclusion probability of unit i.
d) Suppose that you take a simple random sample without replacement from a finite population that contains one extremely large outlier with the aim of estimating the population mean of a study variable and you use this estimator: $\bar{y} = \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i}$, where N is the population size. The bias of this estimator is greater than zero and may be very large if the sample happen to contain the outlier.
e) In design-based inference (which we have focused on in the course), the bias of an estimator of the finite population parameter θ is defined as $E(\hat{\theta}) - \theta$, where $E(\hat{\theta}) = \sum_s \theta_s P(s)$, θ_s is the estimate you obtain from sample s, and $P(s)$ is the probability of obtaining sample s, and the sum is taken over all possible samples.

4.

a) The term 'sampling design' refers to $P(s)$, the probability of obtaining sample s.
b) If you have a frame that lists all observation units in the population and the cost of collecting data from an observation unit is about the same for all population units, then simple random sampling without replacement is usually preferred to a cluster sampling design.
c) A cluster sampling design usually gives larger variance than simple random sampling without replacement.
d) Stratified simple random sampling without replacement does not usually give larger variance than simple random sampling without replacement.
e) Systematic sampling always gives a variance that is as large or larger than that of simple random sampling without replacement.

5.

a) Common nonresponse adjustment methods include imputation and poststratification.
b) A nonresponse rate of 60% does not necessarily produce a larger bias than a nonresponse rate of 30%.
c) In hot-deck imputation, values are taken from external sources, recomputed through a statistical model (e.g. a regression model) and imputed in the current data set to create a data matrix without missing values.
d) For simple random sampling, mean value imputation is the same as weighting-class estimation with only one class.

e) Suppose an elderly man has been included in the sample of a survey about happiness but due to dementia is unable to respond in a telephone interview. If the grown-up son of the elderly person gives the responses in place of his father, it can be considered nonresponse.

6. A survey is conducted to estimate the number of times people are victimized by different kinds of crimes, e.g. burglary, theft, assault, etc., during the previous 12 months. A stratified simple random sample of 800 persons was selected from a population of size 90000. The population was stratified by gender, according to data in the population register. The sample was allocated proportionally to stratum sizes. A total of 420 men and 380 women were included in the sample and asked about the number of times they have been victimized. The results on the question are given in the table below for women and men separately. Assume that everybody responded. Estimate the proportion of persons who have been victimized at least one time during the previous 12 months and give a 95 % confidence interval. You may wish to use the formulas attached to the exam.

Number of times a person has been victimized during the last 12 months	Number of men	Number of women
0	250	260
1	135	102
2	28	15
3 or more	7	3
Total	**420**	**380**

Maximum 10 points.

7. The universities in Uppsala and Stockholm want to estimate how many hours a week students spend on paid work. A survey with a total sample size of 1000 is planned. The sampling design will be stratified simple random sampling without replacement where Uppsala is one stratum and Stockholm is another. There are 28 000 students in Stockholm and 25 000 students in Uppsala. The population variance of the hours working in Stockholm is believed to be four times as that of Uppsala, that is, $S_S^2 = 4S_U^2$. Allocate the sample using

a) optimal allocation
b) proportional allocation.
c) After the first survey has been done, the universities want do the survey again. Now they want 95% confidence intervals for each of Stockholm and Uppsala for the average time spent on paid work with the total width 2 hours, that is, they should be of the form $average \pm 1$ hour. What total sample size (Stockholm + Uppsala) do they need? The plan is to use stratified simple random sampling (still Stockholm and Uppsala as the two strata) without replacement and optimal allocation. In the first survey it was estimated that $s_S^2 = 100$ and $s_U^2 = 36$.

Maximum 10 points.

8. The city of Stockholm (*Stockholms stad*) decided that they wanted to evaluate the knowledge in mathematics among 9 grade students (students in the last year of mandatory school). 20 schools with students in ninth grade were selected by simple random sampling and every student in the school was given a test about mathematics knowledge. The total number of schools in the population is 126 and the total number of students in the ninth grade is 8503. The summary statistics is given below. Also known is
$\sum_{i \in s}(t_i - \hat{\bar{y}}_{rat} M_i)^2 = 15994493.27$,
where t_i is the total number of points that all students in school i obtained as a group when they did the test. The other notation is left as part of the problem to understand the meaning of, if needed. Maximum 10 points.

School number	Number of students	Mean test result (maximum 100)	t_i
1	75	86.8	6510
2	13	96.8	1258.4
3	12	63.5	762
4	65	77.8	5057
5	52	88.1	4581.2
6	16	78	1248
7	153	73.5	11245.5
8	91	59.9	5450.9
9	55	55.7	3063.5
10	145	89.3	12948.5
11	65	58.5	3802.5
12	62	71.3	4420.6
13	116	79.4	9210.4
14	77	66.3	5105.1
15	18	93.5	1683
16	47	80.5	3783.5
17	83	94.2	7818.6
18	91	57.3	5214.3
19	19	62.7	1191.3
20	40	74.2	2968
Total	**1295**		**97322.3**

a. What is the sampling design called?
b. This is the actual sampling design used in this type of surveys in Sweden. Give one reason why this design is used instead of a simple random sample of students.
c. State the name of three types of nonsampling error, and for each give one likely reason why this sampling error may occur in this survey, or, if there is no likely reason why a specific type on nonsampling error should occur, explain why.
d. Estimate the mean test result in the population, using the estimator that gives the smallest variance, and give a 95 % confidence interval. You may assume that the set of all possible estimates follow a normal distribution.

9. The company Mooranalyses conducts an annual survey of number of private swimming pools using Google Maps. The have taken a simple random sample without replacement of 35 out of 290 municipalities in Sweden and they use the same sample every year. The observed number of pools is reported in one of the tables below. A is year 2015, B is year 2014 and C is year 2013. It is known that in 2013 the total number of private swimming pools were 7002. Mooranalyses fits an ordinary least squares model to the data from 2015 and 2013 containing the number of pools in each sampled municipality. It is part of the problem to interpret the print-outs below. Estimate the number of pools in Sweden in year 2015. Use two estimators, both of which should be unbiased or approximately unbiased. Estimate also the variance for each of them. Maximum 10 points.

Output

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
A	35	31.34	4.73	1097.0	12.0	47.0
B	35	28.91	6.33	1012.0	9.0	51.0
C	35	25.09	5.12	878.0	4.0	41.0

Correlation Matrix, n = 35

	A	B	C
A	1.00	0.28	0.73
B	0.28	1.00	0.12
C	0.73	0.12	1.00

The REG Procedure, Dependent Variable: A

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	96.455	96.455	37.649	<.0001
Error	33	84.545	2.562		
Corrected Total	34	181.000			

Root MSE	1.60	**R-Square**	0.5329
Dependent Mean	31.34	**Adj R-Sq**	0.5303

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.427	0.079	5.405	<.0001
x	1	0.871	0.102	8.539	<.0001

Formulae

Population

Population of size N: $U = \{1, \ldots, i, \ldots, N\}$

Sample, size n: $s = \{1, \ldots, i, \ldots, n\}$

Population total of study variable *y*: $t_y = \sum_{i \in U} y_i$

Population mean of study variable *y*: $\bar{y}_U = \frac{1}{N}\sum_{i \in U} y_i$

Population total of auxiliary variable *x*: $t_x = \sum_{i \in U} x_i$

Population variance: $S_y^2 = \frac{1}{N-1}\sum_{i\in U}(y_i - \bar{y}_U)^2$ (Lohr p. 32)

A **proportion** is a special case with $y_i = \begin{cases} 1 \text{ if unit } i \text{ has the relevant characteristic} \\ 0 \text{ otherwise} \end{cases}$ (compare Lohr p. 33).

For a proportion P the population variance $S^2 \approx P(1-P)$ (Lohr p. 38)

Formulas for SRS

Expansion estimator of t_y: $\hat{t}_y = \frac{N}{n}\sum_{i\in s} y_i$

Corresponding estimator of $\bar{y}_U$: $\frac{\hat{t}_y}{N} = \bar{y}_s$

$V(\hat{t}_y) = N^2\left(1 - \frac{n}{N}\right)\frac{S_y^2}{n}$ (Lohr (2.16))

For an estimator of $V(\hat{t}_y)$, replace S_y^2 with the following estimator of S_y^2:

$s_y^2 = \frac{1}{n-1}\sum_{i\in s}(y_i - \bar{y}_s)^2$ (Lohr (2.10) and (2.17))

Ratio estimator of t_y : $\hat{t}_{rat} = t_x\frac{\hat{t}_y}{\hat{t}_x} = t_x\hat{B}$ (Lohr (4.2))

$\hat{V}(\hat{t}_{rat}) = N^2\left(1 - \frac{n}{N}\right)\frac{s_e^2}{n}$, where $s_e^2 = \frac{1}{n-1}\sum_{i\in s}\left(y_i - \hat{B}x_i\right)^2 = \frac{1}{n-1}\left(s_y^2 + \hat{B}^2 s_x^2 - 2\hat{B}s_{xy}\right)$,

$s_{xy} = \frac{1}{n-1}\sum_{i\in s}(y_i - \bar{y}_s)(x_i - \bar{x}_s)$ (Lohr (4.8) and (4.11))

It is also ok (even rather better) to use $\hat{V}(\hat{t}_{rat}) = N^2\left(1 - \frac{n}{N}\right)\frac{s_e^2}{n}\left(\frac{\bar{x}_U}{\bar{x}_s}\right)^2$. (Lohr (4.11))

Regression estimator of t_y: $\hat{t}_{reg} = N\left(\bar{y}_s + \hat{B}_1(\bar{x}_U - \bar{x}_s)\right)$, where $\hat{B}_1 = \frac{\sum_{i\in s}(x_i - \bar{x}_s)(y_i - \bar{y}_s)}{\sum_{i\in s}(x_i - \bar{x}_s)^2}$ (Lohr (4.15))

$V\left(\hat{t}_{reg}\right) \approx N^2\left(1 - \frac{n}{N}\right)\frac{S_y^2}{n}(1 - R^2)$,

where $R = \frac{S_{xy}}{S_x S_y}$ is the finite population correlation coefficient. (Lohr (4.18))

A variance estimator is obtained by replacing the population quantities S_y^2 and R with sample quantities. (Lohr (4.20))

Alternative, equivalent, variance estimator: $\hat{V}\left(\hat{t}_{reg}\right) = N^2\left(1 - \frac{n}{N}\right)\frac{s_e^2}{n}$,

where $s_e^2 = \frac{1}{n-1}\sum_{i\in s}\left(y_i - \hat{B}_0 - \hat{B}_1 x_i\right)^2$, $\hat{B}_0 = \bar{y}_s - \hat{B}_1\bar{x}_s$ (Lohr p. 138-139)

Domain estimation in SRS

Let $u_i = y_i x_i$ with $x_i = \begin{cases} 1 \text{ if unit } i \text{ belongs to the domain} \\ 0 \text{ otherwise} \end{cases}$ (Lohr p. 134)

The part of the sample that falls in domain d is denoted by s_d and the number of units in s_d is denoted by n_d.

Estimation of the **mean** of study variable in domain d: $\bar{y}_d = \frac{\bar{u}_s}{\bar{x}_s}$

$\hat{V}(\bar{y}_d) = \left(1 - \frac{n}{N}\right)\frac{s_{yd}^2}{n_d}$, where $s_{yd}^2 = \frac{\sum_{i \in s_d}(y_i - \bar{y}_d)^2}{n_d - 1}$ (compare Lohr (4.13))

Estimation of the **total** of study variable in domain d, t_d, two cases:

1. If the population size of the domain, N_d, is known: $\hat{t}_d = N_d \bar{y}_d$ (Lohr p. 135)
2. N_d is unknown: $\hat{t}_d = N\bar{u}_s$, where $\bar{u}_s = \frac{1}{n}\sum_{i \in s} u_i$. $\hat{V}(\hat{t}_d) = N^2\left(1 - \frac{n}{N}\right)\frac{s_u^2}{n}$, where $s_u^2 = \frac{1}{n-1}\sum_{i \in s}(u_i - \bar{u}_s)^2$

Sample size estimation, SRS

We want this precision: $P(|\bar{y}_s - \bar{y}_U| \leq e) = 0.95$. Then, with the approximation $fpc = 1$, $n = \frac{1.96^2 s_y^2}{e^2}$. (compare Lohr (2.25))

Stratification and poststratification

The population is divided into nonoverlapping groups that will exhaust the population fully. I prefer subscript g as a generic notation of the number of one poststratum and subscript h for a generic notation of the number of one stratum. For example, the sample and total in stratum h is denoted by s_h and t_h, respectively. Lohr uses subscript h for both kinds of population subsets.

For **stratified simple random sampling** the population mean $\bar{y}_U$ is estimated as

$\bar{y}_{str} = \frac{1}{N}\sum_{h=1}^{H}\sum_{i \in s_h}\frac{N_h y_i}{n_h} = \frac{1}{N}\sum_{h=1}^{H}\hat{t}_h$ (Lohr (3.1) and (3.2))

and the variance as

$\hat{V}(\bar{y}_{str}) = \sum_{h=1}^{H}\frac{N_h^2}{N^2}\left(1 - \frac{n_h}{N_h}\right)\frac{s_h^2}{n_h}$ (Lohr (3.5))

With stratified simple random sampling with proportional allocation, that is, the sample size in each stratum h is $n_h = n\frac{N_h}{N}$, the variance of the estimate $\hat{V}(\bar{y}_{str}) = \frac{1}{Nn}\left(1 - \frac{n}{N}\right)\sum_{h=1}^{H} N_h s_h^2$ (Lohr p. 86)

Optimal allocation, equal costs: $n_h = n\frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h}$ (Lohr (3.14))

For **simple random sampling followed by poststratification**, if the sample sizes in poststrata are $n_g = n\frac{N_g}{N}$, the variance estimator is the same:

$\hat{V}(\bar{y}_{post}) = \frac{1}{Nn}\left(1 - \frac{n}{N}\right)\sum_{g=1}^{G} N_g s_g^2$ (Lohr (4.22))

For general poststratum sample sizes a variance estimator corresponding to the formula above marked as Lohr (3.5) can be used:

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

Poststratification estimator, SRS, general poststratum sample sizes:

$\bar{y}_{post} = \frac{1}{N}\sum_{g=1}^{G} \sum_{i \in s_g} \frac{N_g y_i}{n_g} = \frac{1}{N}\sum_{g=1}^{G} \hat{t}_g$

One-stage cluster sampling, unequal cluster sizes

N and n: number of clusters in the population and in the sample, respectively.

M_i and M_0: number of units in cluster *i* and in the population, respectively.

$t_i = \sum_{j=1}^{M_i} y_{ij}$ is the total of y_{ij} in cluster *i* (y_{ij} is the value of the study variable for unit *j* in cluster *i*).
$\hat{t}_i = t_i$ because in one-stage cluster sampling, all units in the clustered are sampled.

Unbiased estimator of t_y: $\hat{t}_{unb} = \frac{N}{n}\sum_{i \in s} t_i = \frac{N}{n}\sum_{i \in s}\sum_{j=1}^{M_i} y_{ij}$ (Lohr p. 169)

Corresponding estimator of $\bar{y}_U$: $\hat{\bar{y}} = \frac{\hat{t}_{unb}}{M_0}$ (M_0 must be known)

$\hat{V}(\hat{\bar{y}}) = \frac{N^2}{M_0^2}\left(1 - \frac{n}{N}\right)\frac{s_t^2}{n}$, where $s_t^2 = \frac{1}{n-1}\sum_{i \in s}\left(\hat{t}_i - \frac{\hat{t}_{unb}}{N}\right)^2$ (Lohr (5.13) and p. 170)

Ratio estimator of $\bar{y}_U$: $\hat{\bar{y}}_{rat} = \frac{\hat{t}_{unb}}{\hat{M}_0} = \frac{\sum_{i \in s}\hat{t}_i}{\sum_{i \in s} M_i}$

$\hat{V}(\hat{\bar{y}}_{rat}) = \left(1 - \frac{n}{N}\right)\frac{1}{n\bar{M}^2}\frac{\sum_{i \in s}(t_i - \hat{\bar{y}}_{rat}M_i)^2}{n-1}$, where $\bar{M} = \frac{1}{n}\sum_{i \in s} M_i$

Horvitz-Thompson estimator

General sampling design, inclusion probability π_i

Unbiased estimator of t_y: $\hat{t}_{HT} = \sum_{i \in s}\frac{y_i}{\pi_i}$ (Lohr (6.19))

Response rate

Response rate computed as $\frac{(6)}{(4)+(3A)}$, where (6) is number of sample units that responds, (4) is number of sample units that are established to be in scope (i.e. belong to target population) and (3A) is the number of unresolved sample units that are believed to be in scope.

Weighting class estimator

$\hat{t}_{WC} = N\sum_{c=1}^{C}\frac{n_c}{n}\bar{y}_{cR}$, where C is number of classes, n_c is sample size in class c, $\bar{y}_{cR} = \frac{\sum_{i \in s_{cR}} y_i}{n_{cR}}$ is the mean of the respondents in class c. (Lohr page 341)

Poststratified estimator to adjust for nonresponse

$\hat{t}_{post} = \sum_{g=1}^{G} N_g \bar{y}_{gR}$, , where G is number of poststrata, N_g is population size in poststratum g, $\bar{y}_{gR} = \frac{\sum_{i \in s_{gR}} y_i}{n_{gR}}$ is the mean of the respondents in poststratum g. (Lohr page 342)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 26/10-2016

Sal: Värtasalen

Tenta: Urvalsundersökningar

Kurs: Urvalsundersökningar

ANONYMKOD: URV-00K1 0019

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X	X	X	X	X	2
Lär.ant. 2	0	2	2	2	10	8	10	5	

POÄNG	BETYG	Lärarens sign.
41	B	[signature]

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: URV-0019 Blad nr: 1

1, c R 2

2, b V 0

3, d R 2

4, e R 2

5, c R 2

6, $\hat{p}_m$ = andel brottsutsatta män = (135+28+7)/420 = 0,405

$\hat{p}_k$ = andel brottsutsatta kvinnor = (102+15+3)/380 = 0,316

N_m = totalt antal män = 90000·420/800 = 47250

N_k = totalt antal kvinnor = 90000·380/800 = 42750

$\hat{t}_{str} = \hat{t}_m + \hat{t}_k = N_m \cdot \hat{p}_m + N_k \cdot \hat{p}_k$ =

47250·0,405 + 42750·0,316 = 32645,25

$\hat{p}_{str} = \hat{t}_{str}/N$ = 32645,25/90000 = <u>0,36</u> R 4

$$V(\hat{p}_{str}) = \frac{N_m^2}{N^2}\left(1-\frac{n_m}{N_m}\right)\frac{\hat{p}_m(1-\hat{p}_m)}{n_m-1} + \frac{N_k^2}{N^2}\left(1-\frac{n_k}{N_k}\right)\frac{\hat{p}_k(1-\hat{p}_k)}{n_k-1} =$$

↘

6 forts,
$$= \frac{47250^2}{90000^2}\left(1-\frac{420}{47250}\right)\frac{0{,}405(1-0{,}405)}{420-1} +$$

$$\frac{42750^2}{90000^2}\left(1-\frac{380}{42750}\right)\frac{0{,}316(1-0{,}316)}{380-1} =$$

$2{,}846 \cdot 10^{-4}$

$SE(\hat{p}_{str}) = \sqrt{2{,}846 \cdot 10^{-4}} = 0{,}017$ 4

$CI = \hat{p}_{str} \pm 1{,}96 \; SE(\hat{p}_{str}) = 0{,}363 \pm 1{,}96 \cdot 0{,}017 = [0{,}33, 0{,}40]$ R

2

7a,
$$n_u = n \cdot \frac{N_u \frac{1}{2} \hat{S}_s^2}{N_u \frac{1}{4} \hat{S}_s^2 + N_s \hat{S}_s^2} = n \cdot \frac{N_u \cdot \frac{1}{4}}{N_u \frac{1}{4} + N_s} =$$

$S_s^2 = 4 S_u^2$

$\Rightarrow S_s = 2 S_u$

$$\frac{1000 \cdot 25000 \cdot \frac{1}{4}}{25000 \cdot \frac{1}{4} + 28000} = \underline{\underline{182}}$$

(2)

$n_s = 1000 - 182 = \underline{\underline{818}}$

b, $n_u = n \cdot \frac{N_u}{N} = 1000 \cdot \frac{25000}{53000} = \underline{\underline{472}}$ R

2

$n_s = 1000 - 472 = \underline{\underline{528}}$

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: [illegible] Blad nr: 2

7, c, Antag oändlig population

$CI_u = \bar{y}_u \pm 1{,}96 \cdot SE(\bar{y}_u) \rightarrow$

$e = 1{,}96\ SE(\bar{y}_u) = 1{,}96 \sqrt{\frac{S_u^2}{n_u}} \rightarrow$

$n_u = \frac{1{,}96^2 \cdot S_u^2}{e^2} = \frac{1{,}96^2 \cdot 36}{1} = 138$

$n_s = \frac{1{,}96^2 \cdot S_s^2}{e^2} = \frac{1{,}96^2 \cdot 100}{1} = 384$ R 4

8a, One-stage cluster sampling R 1

b, Det är billigare R 2

c, 1, mätfel: testen kan ha rättats slarvigt B 3
2, bortfall: vissa elever var frånvarande
3, undertäckning: alla skolor var inte tillgängliga i register (mindre troligt)

d, $\hat{y}_{rat} = \frac{\hat{t}}{\hat{M}_0} = \frac{N\bar{t}}{N\bar{M}} = \frac{\sum t_i}{\sum M_i} = \frac{97322{,}3}{1295} = 75{,}15$ R

2

8d, forts. $\bar{M} = 1295/20 = 64,75$

$V(\bar{y}_{rat}) = \left(1-\frac{n}{N}\right)\frac{1}{n\bar{M}^2} \cdot \frac{\sum (t_i - \bar{y}_{rat} M_i)^2}{n-1} =$

$\left(1-\frac{20}{126}\right)\frac{1}{20 \cdot 64,75^2} \cdot \frac{1599449 3,27}{19} = 8,446$

R 2

$SE(\bar{y}_{rat}) = \sqrt{8,446} = 2,906$

$CI = \bar{y}_{rat} \pm 1,96 \cdot SE(\bar{y}_{rat}) = [69, 81]$

9, $\hat{t}_{exp} = N\bar{y} = 290 \cdot 31,34 = 9089$ R

$V(\hat{t}_{exp}) = \left(1-\frac{n}{N}\right)N^2\frac{s^2}{n} = \left(1-\frac{35}{290}\right) \cdot 290^2 \frac{4,73^2}{35} =$

47270,7 R

5

$\bar{x}_U = 7002/290 = 24,145$

$\hat{t}_{reg} = N(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_U) = 290(0,427 + 0,871 \cdot 24,145) =$

6222 ✓

$V(\hat{t}_{reg}) = N^2\left(1-\frac{n}{N}\right)\frac{s_e^2}{n} = 290^2\left(1-\frac{35}{290}\right)\frac{2,562^2}{35} =$ ✓

5413,1

0