Solutions final exam

I

1. a) - Strat. sample because: strata are homogeneous wrt antibiotics use which will give estimate w/ smaller variance
   - when stratum variances in popln are equal

b) Prop. allocation:

$n_h = n\frac{N_h}{N}$ $\qquad\qquad W_h = N_h/N$

$n_1 = 523.6 = 524$	0.5236
$n_2 = 327.539 \approx 327$	0.3275
$n_3 = 57.543 = 58$	0.0575
$n_4 = 91.3 = 91$	0.0913

Incl. prob. $\pi_i = \frac{n_h}{N_h} \approx \frac{n}{N} = .06738$

Exact: $\pi_1 = .06743$

$\pi_2 = .06727$

$\pi_3 = .06792$

$\pi_4 = .06716$

c) $Var(\hat{t}_{st}) = \sum_{h=1}^{H} N_h^2 \frac{\sigma_h^2}{n_h}(1-f_h) \overset{prop.}{=} N^2 \cdot \frac{1-f}{n} \sum_{h=1}^{4} W_h \sigma_h^2$

$f = \frac{n}{N}$

$= N^2 \frac{1-f}{n}[48.18] = N^2 \cdot 0.0449336 = 9,896,860.06$

$SE(\hat{t}_{st}) = 3145.93$

$Var(\hat{t}_{st}/N) = .0449 \qquad SE(\hat{t}_{st}/N) = \sqrt{.0449336} = .212$

d)

	$h=1$	$h=2$	$h=3$	$h=4$
$\mu_h =$	12.868	24.275	2.342	11.808

$$\sigma_t^2 = \sum_h W_h \sigma_h^2 + \sum_h W_h (\mu_h - \mu)^2 \qquad \mu = 15.902$$

$$= 48.18 + \{4.8198 + 22.9601 + 10.6008 + 1.5303\}$$

$$= 48.18 + 39.91 = \underline{88.09}$$

$$Var(N\bar{y}_s) = N^2 \frac{\sigma^2}{n}(1-f) = N^2 \cdot (.08215) \cancel{= 18{,}093{,}971.33}$$

Determine n such that

$$\frac{\sigma^2}{n}(1-f) = .04493 \Leftrightarrow \frac{\sigma^2}{n} - \frac{\sigma^2}{N} = .04493$$

$$\cancel{\Leftrightarrow n = \frac{\sigma^2(1-f)}{.04493}} = \underline{1828} \Leftrightarrow \frac{\sigma^2}{n} = \frac{\sigma^2}{N} + .04493 = .05087$$

$$\& \; \underline{n = 1732}$$

e) Optimal allocation.

$$n_h = n \cdot \frac{N_h \cdot \sigma_h}{\sum_{h=1}^{4} N_h \sigma_h} = \frac{1000}{95094} N_h \cdot \sigma_h$$

$\Rightarrow$ $n_1 = 408.60 = 409$

$n_2 = 511.18 = 511$

$n_3 = 8.98 = 9$

$n_4 = 71.25 = 71$

f) Cluster sample

I. yes, self-weighting

II. no.

g) $\hat{t}_{st} = \sum_{h=1}^{4} N_h \cdot \bar{y}_h = \underline{219.438.8}$

$\Rightarrow \hat{t}_{st}/N = \underline{14.79}$

Prop. allocation

$\hat{V}(\hat{t}_{st}) = N^2 \cdot \frac{1-f}{n} \sum_{h=1}^{H} W_h s_h^2$

$= N^2 \cdot \frac{1-f}{n}(27.6658) = N^2 \cdot (\overset{.0258}{\cancel{.0009326}})$

$SE(\hat{t}_{st}) = N \cdot \overset{.1606}{\cancel{0.030539}} = \cancel{453.2}\ 2383.46$

$SE(\hat{t}_{st}/N) = \sqrt{\overset{.0258}{\cancel{.0009326}}} = \cancel{.0305}\ \underline{.16}$

h) 95% CI:

$14.79 \pm 1.96 \times (0.16) = 14.79 \pm .31$

$= \underline{(14.48, 15.10)}$

usual intepretation

2. a) $\hat{T}_{exp} = n\bar{Y}_s + (N-n)\bar{Y}_s$

$= \sum_{i\in s} Y_i + \sum_{i\notin s} \bar{Y}_s \quad \Rightarrow \hat{Y}_i = \bar{Y}_s$

$\hat{T}_R = (\sum_{i\in s} x_i)\frac{\sum_{i\in s} Y_i}{\sum_{i\in s} x_i} + (\sum_{i\notin s} x_i)\cdot \underbrace{\frac{\sum_{i\in s} Y_i}{\sum_{i\in s} x_i}}_{\hat{R}} = \sum_{i\in s} Y_i + \sum_{i\notin s} \hat{R}\cdot x_i$

$\Rightarrow \hat{Y}_i = \hat{R}\cdot x_i = \hat{\beta}_{opt}\cdot x_i$

$\hat{T}_{reg} = \sum_{i\in s} Y_i + \sum_{i\notin s}(\hat{\beta}_1 + \hat{\beta}_2 x_i)$, $\hat{\beta}_1 = \bar{Y}_s - \hat{\beta}_2 \bar{x}_s$

$\Rightarrow \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i = \bar{Y}_s + \hat{\beta}_2(x_i - \bar{x}_s)$

Direkte: $\hat{T}_{reg} = n\bar{Y}_s + (N-n)\bar{Y}_s + \hat{\beta}_2(\sum_{i\notin s} x_i - (N-n)\bar{x}_s)$

$= \sum_{i\in s} Y_i + \sum_{i\notin s}[\underbrace{\bar{Y}_s + \hat{\beta}_2(x_i - \bar{x}_s)}_{\hat{Y}_i}]$

b) $\hat{T}_{exp} = N\bar{Y}_s$ is BLU predictor for the model $Y_i = \beta + \varepsilon_i$, $E(\varepsilon_i)=0$ & $V(\varepsilon_i)=\sigma^2$

$\hat{T}_R$ is BLU predictor for: $Y_i = \beta x_i + \varepsilon_i$, $E(\varepsilon_i)=0$ & $V(\varepsilon_i)=\sigma^2 x_i$

$\hat{T}_{reg}$ is BLU predictor for:

$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $E(\varepsilon_i)=0$ & $V(\varepsilon_i)=\sigma^2$

From the scatterplot: Either regr. or ratio-model

(Both are acceptable)

c) $\sum_{i\in s} y_i = 404.6$, $\sum^N x_i = 18.2\times 234 = \underset{t_x}{4258.8}$, $\sum_{i\in s} x_i = 194.3$, $\bar{x}_s = \cancel{16.175}$ 16.192

Merk: $< \bar{x}$

$\hat{T}_{exp} = N\cdot\bar{y}_s = 234\cdot\frac{404.6}{12} = \underline{7889.7}$

$\hat{T}_R = t_x\cdot\frac{\sum_s Y_i}{\sum_s x_i} = 4258.8\times\frac{404.6}{194.3} = \underline{8868.3}$

$$\hat{\beta}_2 = \frac{\sum_s x_i Y_i - n\bar{x}_s \cdot \bar{Y}_s}{\sum_s x_i^2 - n\bar{x}_s^2} = \frac{9387.16 - 6551.28}{4443.21 - 3146.17}$$

$$= 2835.88/1297.04 = 2.186$$

$$\hat{T}_{reg} = 7889.7 + 2.186 \times 234\,(18.2 - 16.192)$$

$$= 7889.7 + 1027.14 = \underline{8916.8}$$

d) i) Model: $Y_i = \beta + \varepsilon_i$

Impute: $Y_i^* = \hat{\beta}_r = \bar{Y}_r$ (= sample mean of the ~~[illegible]~~ units in the response sample

$\bar{Y}_r = \frac{1}{9} \sum_{i \in s_r} Y_i$

ii) Model: $Y_i = \beta x_i + \varepsilon_i$

$\hat{\beta}_r = \sum_{i \in s_r} Y_i \Big/ \sum_{i \in s_r} x_i$ $\qquad s_r = (2, \ldots, 8, 10, 11)$

Impute: $Y_i^* = \hat{\beta}_r \cdot x_i$

iii) Model: $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

Impute: $Y_i^* = \hat{\beta}_{1,r} + \hat{\beta}_{2,r} x_i$

where $\hat{\beta}_{1,r}$ and $\hat{\beta}_{2,r}$ are based on s_r.

~~This is enough to get full score, do not need to do any calculation!~~

Just in case: $\sum_{i \in s_r} Y_i = 281.7$ $\quad \sum_{s_r} x_i = 133.8$

Model i): $Y_i^* = 281.7/9 = 31.3$ for $i = 1, 9, 12$

Model ii): $Y_i^* = \hat{\beta}_r x_i = 2.105 x_i = \begin{cases} 6.7 & i=1 \\ 48.2 & i=9 \\ 72.4 & i=12 \end{cases}$

This is enough to get full score.