Homework 8 Nathan Bick

1. Exercises 12.A: 1

1. Use the data for the $N = 2{,}954$ white females and black females in example 12.2 to compute a 95% confidence interval for the population proportion with at least a bachelor's degree $P_{(B/WF\ SCHL \geq 21)}$.

complete $\hat{P}_{(B/WF\ SCHL \geq 21)str} = \frac{N_1}{N}\hat{p}_1 + \frac{N_2}{N}\hat{p}_2$

$= 0.469$ (verify)

then

$$\hat{V}(\hat{P}_{(B/WF\ SCHL \geq 21)str}) = \left(\frac{N_1}{N}\right)^2\left(\frac{N_1 - n_1}{N_1}\right)\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 - 1} + \left(\frac{N_2}{N}\right)^2\left(\frac{N_2 - n_2}{N_2}\right)\frac{\hat{p}_2(1-\hat{p}_2)}{n_2 - 1}$$

≈ 0.00185 (verify),

and finally (verify)

$$\left(\hat{P}_{(B/WF\ SCHL \geq 21)str} \pm 1.96\sqrt{\hat{V}(\hat{P}_{(B/WF\ SCHL \geq 21)str})}\right)$$

$= (0.38, 0.55)$.

As is reported in Example 12.2, we see

$\hat{p}_1 = \frac{37}{50} = 0.74$ and $\hat{p}_2 = \frac{12}{50} = 0.24.$

$N_1 = 1350$, $N_2 = 1604$.

so $$\hat{P}(B/W\ FSCHL \geq 21)_{str} = \frac{1350}{2954}(0.74) + \frac{1604}{2954}(0.24)$$

$$= 0.469, \text{ as stated.}$$

Next, we see $$\left(\frac{1350}{2954}\right)^2\left(\frac{1350-50}{1350}\right)\left(\frac{0.74(1-0.74)}{50-1}\right)$$

$$+\left(\frac{1604}{2954}\right)^2\left(\frac{1604-50}{1604}\right)\left(\frac{0.24(1-0.24)}{50-1}\right)$$

$$= 0.000787\ldots + 0.001063\ldots$$

$$= 0.00185$$

Finally, we see the C.I is

$$(0.469 - (1.96)(0.00185),\ 0.469 + (1.96)(0.00185))$$

$$= (0.38, 0.55).$$

2 Problems 12.A:1

1. Show that (12.1), (12.3), (12.4), and (12.5) are true.

$$(12.1)\quad \bar{y}_u = \frac{N_1}{N}\bar{y}_{1u} + \cdots + \frac{N_H}{N}\bar{y}_{Hu}$$

$$(12.2)\quad \bar{y}_{str} = \frac{N_1}{N}\bar{y}_1 + \cdots + \frac{N_H}{N}\bar{y}_H$$

$$(12.3)\quad E(\bar{y}_{str}) = \bar{y}_u$$

$$(12.4)\quad V(\bar{y}_{str}) = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)^2 V(\bar{y}_h) = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)^2\left(\frac{N_h - n_h}{N_h}\right)\frac{S_h^2}{n_h}$$

First we consider (12.1).

we know $\bar{y}_u = \dfrac{\sum_{i=1}^{N} y_i}{N}$, and $N_1 + \cdots + N_H = N$.

For each subpopulation, $\bar{y}_{hu} = \dfrac{\sum_{i=1}^{N_h} y_i}{N_h} \Leftrightarrow N_h\bar{y}_h = \sum_{i=1}^{N_h} y_i$.

we see $\displaystyle\sum_{i=1}^{N} y_i = \sum_{j=1}^{N_1} y_j + \cdots + \sum_{k=1}^{N_H} y_k = N_1\bar{y}_1 + \cdots + N_H\bar{y}_H$.

so $$\frac{\sum_{i=1}^{N} y_i}{N} = \frac{N_1}{N}\bar{y}_{1u} + \cdots + \frac{N_H}{N}\bar{y}_{Hu} = \bar{y}_u.$$

Next we consider $\bar{y}_{str}$ in (12.2).

The proof is almost exactly the same as above, but insted of pieces of the population we take weighted averge of means of pieces of the sample.

Next we consider (12.3). Prove $E(\bar{y}_{str}) = \bar{y}_U$.

$$E(\bar{y}_{str}) = E\left(\frac{N_1}{N}\bar{y}_1 + \cdots + \frac{N_H}{N}\bar{y}_H\right)$$ from previous section.

$$= E\left(\frac{N_1}{N}\bar{y}_1\right) + \cdots + E\left(\frac{N_H}{N}\bar{y}_H\right) = \frac{N_1}{N}E(\bar{y}_1) + \cdots + \frac{N_H}{N}E(\bar{y}_H).$$

since the sample mean is an unbiased estimator of the population mean, we know $E(\bar{y}_h) = \bar{y}_{hU}$.

so we have $\frac{N_1}{N}\bar{y}_{1U} + \cdots + \frac{N_H}{N}\bar{y}_{HU} = \bar{y}_U$, from previous section of homework.

Next we consider (12.4).

Prove $V(\bar{y}_{str}) = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)^2 V(\bar{y}_h) = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)^2\left(\frac{N_h - n_h}{N_h}\right)\frac{S_h^2}{n_h}$

$$V(\bar{y}_{str}) = V\left(\frac{N_1}{N}\bar{y}_1 + \cdots + \frac{N_H}{N}\bar{y}_H\right) = V\left(\frac{N_1}{N}\bar{y}_1\right) + \cdots + V\left(\frac{N_H}{N}\bar{y}_H\right)$$

due to independence of the strata.

$$= \left(\frac{N_1}{N}\right)^2 V(\bar{y}_1) + \cdots + \left(\frac{N_H}{N}\right)^2 V(\bar{y}_H).$$

For (12.5). we want $\hat{V}(\bar{y}_{str})$, an estimator for the above. we simply ~~do~~ replace V with $\hat{V}$ in all places above. This means we replace the population S_h^2 with estimator sampling s_h^2.