Comparison of large samples.

Let two large samples of size n_1, n_2 be drawn from two populations ~~of~~ giving proportions of attributes A's as P_1 & P_2 resp.

i) Hypothesis:

As regards the attribute A the two pop^n are siamilar. on combining two samples

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

where P is common proportion of attributes.

Let e_1 & e_2 are S.E in two samples then

$$e_1^2 = \frac{pq}{n_1}, \quad e_2^2 = \frac{pq}{n_2}$$

Let e be ~~combined~~ std. error of the combined samples then

$$e = e_1^2 + e_2^2 = \frac{pq}{n_1} + \frac{pq}{n_2} = pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

& $z = \frac{P_1 - P_2}{e}$. If z lies in critical region we are rejecting hypothesis. otherwise accept it at 5% & 1% LOS

Date: / /

(11) In the two population, the proportion of attribute A are not same then std. e of the difference $P_1 \sim P_2$ is

$$e^2 = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$$

~~If $z = \frac{P_1 - P_2}{e} < 3$ the difference could have arisen due to fluctuations of simple sampling.~~

find $z = \frac{P_1 - P_2}{e}$

If z lies in critical region at 5% & 1% LOS then reject hypothesis.

① In a city A 20% of a random sample of 900 school boys had a certain slight physical defect. In another city B, 18.5% of random sample of 1600 school boys had the same defect. Is the diff. betn proportion significant.

→ $n_1 = 900$, $n_2 = 1600$

$P_1 = \frac{20}{100} = \frac{1}{5}$, $P_2 = \frac{185}{100}$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{180 + 296}{2500} = 0.19$$

$q = 1 - 0.19 = 0.81$

$$e^2 = pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right) = 0.19 \times 0.81 \times \left(\frac{1}{900}+\frac{1}{1600}\right)$$ (22)

$$= 0.0017$$

$e = 0.04$ nearly.

Also $\quad p_1 - p_2 = \frac{1.5}{100} = 0.015$

$$z = \frac{p_1 - p_2}{e} = \frac{0.015}{0.04} = 0.37$$

As $z < 1$ the difference betn the proportion is not significant

(2) In two large popn there are 30% & 25% resp. of fair haired people. Is this difference likely to be hidden in the samples of 1200 & 900 resp. from two populations ?

$p_1 = 0.3, \quad p_2 = 0.25$ so that $p_1 - p_2 = 0.05$

$$e^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \Rightarrow \text{~~directly~~} \; e = 0.0195$$

$$z = \frac{p_1 - p_2}{e} = \frac{0.05}{0.0195} = 2.5 \text{ nearly}$$

diff. significant then it will be visible (not hidden)

~~Hence it is unlikely that the real diff. will be hidden~~. (not hidden at 5% 1% level)

(3) In a sample of 600 men from certain city, 450 are found smokers. In another sample of 900 men from another city 450 are smokers. Do the data indicate that the cities are significantly

different with respect to the habit of smoking among men

→ $n_1 = 600$ men, No. of smokers $= 450$, $P_1 = \frac{450}{600} = 0.75$

$n_2 = 900$ men, No. of smokers $= 450$, $P_2 = 0.5$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = 0.60$$

$q = 0.4$

$$e^2 = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 0.000667$$

$\Rightarrow e = 0.02582$

$$z = \frac{P_1 \sim P_2}{e} = \frac{0.75 - 0.5}{0.02582} = 9.68273$$

ie the diff is significant.

4) One type of aircraft is found to develop engine trouble in 5 flights out of total 100 & another type in 7 flights out of a total 200 flights. Is there a significant diff in two types of aircrafts so far as engine defects are concerned?

HW

not ~~di~~ significant

→ (Hint) $n_1 = 100$, No. of troubled flights $= 5$, $P_1 = 5/100$

$n_2 = 200$, $P_2 = 7/200$, $e = 0.0254$, $z = 0.59$

Diff is not significant