Difference of Proportions

Survey of voters in AZ (Arizona) and WV (West Virginia) by GSG in 2021 asked a variety of questions.

Question: Do you believe gov should invest more into clean energy solutions.

We want to know if the proportion strongly agreeing with the above differs between the states.

$H_0: p_{AZ} = p_{WV}$*

$p_{AZ} - p_{WV} = 0$

The proportions strongly agreeing are equal.

$H_a: p_{AZ} \neq p_{WV}$*

$p_{AZ} - p_{WV} \neq 0$

The proportions are not equal.

AZ

$n_{AZ} = 649$

strongly Agree = 331

$\hat{p}_{AZ} = 51\%$

WV

$n_{WV} = 600$

#SA = 216

$\hat{p}_{WV} = 36\%$

Could we....?

~~Test $\hat{p}_{AZ}$ against $p = 36\%$?~~

~~Test $\hat{p}_{WV}$ against $p = 51\%$?~~

not quite

The distribution of

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

generally

$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2,$

$SE = \sqrt{SE^2 + SE^2}$)

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{\quad}{n_1} \quad \frac{\quad}{n_2}}\right)$$

$$\hat{p}_1 \sim N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right), \quad \hat{p}_2 \sim N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right)$$

$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2,$

$SE = \sqrt{SE_1^2 + SE_2^2})$

Next week...

What is the standard error of $\hat{p}_{AZ} - \hat{p}_{WV}$

Exactly: $\sqrt{\dfrac{p_{AZ}(1-p_{AZ})}{n_{AZ}} + \dfrac{p_{WV}(1-p_{WV})}{n_{WV}}}$ But, we don't know p_{AZ} or p_{WV}

We could estimate with $\hat{p}_{AZ}$ and $\hat{p}_{WV}$

But.... According to H_0: $p_{AZ} = p_{WV} = p_?$

we could use this to calculate SE.

We can estimate $p_?$ with

$$\hat{p}_{pooled} = \frac{\#\text{success in group 1} + \#\text{success in group 2}}{\text{Sample size of } g_1 + \text{Sample size of } g_2}$$

$$= \frac{\hat{p}_{AZ} \cdot n_{AZ} + \hat{p}_{WV} \cdot n_{WV}}{n_{AZ} + n_{WV}} = \frac{331 + 216}{649 + 600} \approx .4380$$

$$SE_{\hat{p}_{AZ} - \hat{p}_{WV}} = \sqrt{\frac{\hat{p}_{pooled}(1-\hat{p}_p)}{n_{AZ}} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_{WV}}} \approx 2.81\% \quad .028098$$

Under H_0
$p_{AZ} = p_{WV}$

$$Z = \frac{(\hat{p}_{AZ} - \hat{p}_{WV}) - \cancel{(p_{AZ} - p_{WV})}^{\,0}}{SE_{\hat{p}_{AZ} - \hat{p}_{WV}}} \approx \frac{(.51 - .36)}{.0281}$$

$$Z = 5.34$$

$Z = 5.34$

If $p_{AZ} = p_{WV}$, the $\hat{p}_{AZ} - \hat{p}_{WV}$ values that we see should only happen in $P(Z > 5.34)$ samples.
4.68×10^{-8}

Our p-value is the prob of getting $|\hat{p}_{AZ} - \hat{p}_{WV}|$ as or greater than observed, which is equal to

$$P(Z > 5.34) + P(Z < -5.34)$$
$$= 9.35 \times 10^{-8}$$

We reject $H_0: p_{AZ} = p_{WV}$, and conclude that the proportion strongly agreeing with increased gov funding of clean energy is different for these two states. $(p_{AZ} \neq p_{WV})$

Neuanced Question: Is the proportion in AZ agreeing with increased funding higher than proportion in WV. → not necessarily strongly.

AZ

#agreeing = 526

WV

$\#a = 384$

p here represents proportion agreeing rather than just strongly agreeing.

$H_0: p_{AZ} = p_{WV}$

$p_{AZ} - p_{WV} = 0$

prop agreeing is equal in both states

$H_a: p_{AZ} > p_{WV}$

$p_{AZ} - p_{WV} > 0$

prop agreeing is higher in AZ than WV

prop agreeing is equal in both states

prop agreeing is higher in AZ than WV.

If H_0 is true, we can estimate $p = p_{AZ} = p_{WV}$ with

$$\hat{p}_{pooled} = \frac{526+384}{649+600} \approx 72.86\%$$

$$SE_{\hat{p}_{AZ}-\hat{p}_{WV}} = \sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_{AZ}} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_{WV}}} \approx 2.51\%$$

$$Z = \frac{\hat{p}_{AZ} - \hat{p}_{WV} - \cancel{(p_{AZ} - p_{WV})}^{0}}{SE_{\hat{p}_{AZ}-\hat{p}_{WV}}} = 6.77$$

If H_0 is true, we should see results as or more extreme than seen here $P(Z > 6.77) = 6.48 \times 10^{-12}$ or $6.48 \times 10^{-10}\%$ of the time.

We reject H_0 and conclude that $\underline{p_{AZ} > p_{WV}}$.

The proportion of AZ voters who with increased gov funding of clean energy is higher than in WV.