3Y03 - PROBABILITY AND STATISTICS FOR ENGINEERING

WS19 Lecture 28

Last Time

Hypothesis Tests on Population Proportion

$\underbrace{H_0 : p = p_0}$ $\quad H_1$: (I) $p \neq p_0$ (II) $p > p_0$ (III) $p < p_0$

If H_0 true, then $\hat{p} = \frac{X}{n} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$,

where $X \sim Bin(n, p_0)$, as long as np_0, $n(1-p_0) > 5$.

So $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1) \longrightarrow$ Z-test:

Reject H_0 if

(I) $|Z_0| > z_{\frac{\alpha}{2}}$ (II) $Z_0 > z_\alpha$ (III) $Z_0 < -z_\alpha$

Do this in case (II) of H_1: assume true value of prop. is $p' > p_0$

↳ Find β

If H_0 not true $Z_0 \not\sim N(0,1)$ so what is distr. of Z_0?

$$\beta = P(Z_0 < z_\alpha)$$

$$= P\left(\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_\alpha\right) = P\left(\hat{p} < p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

Or the question is what is distr. now of $\hat{p}$ if true proportion is p' (not p_0)?

standardizing

$$\beta = P\left(Z < \frac{p_0 + z_\alpha\sqrt{\frac{p_0(1-p_0)}{n}} - p'}{\sqrt{\frac{p'(1-p')}{n}}}\right)$$

$$\hat{p} = \frac{X}{n} \sim N\left(p', \frac{p'(1-p')}{n}\right)$$

$$= -z_\beta$$

uses : $P(Z < \sim) = -z_{\sim}$

Rearrange: $$n = \left[\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p_0 - p'}\right]^2$$

Case (III) : same formula

Case (I) : as above replacing $z_{\frac{\alpha}{2}}$ with z_α

__Example__ How many people should we sample to test

$H_0 : p = 90\% \rightarrow p_0 = 0.9$

$H_1 : p < 90\%$

at significance level $\alpha = 0.05$ when true proportion is 85% $\rightarrow p' = 0.85$

to get $\beta = 0.1$?

__Solution__ $z_\alpha = z_{0.05} = 1.64$, $z_\beta = z_{0.1} = 1.28$

So $$n = \left[\frac{1.64\sqrt{(0.9)(0.1)} + 1.28\sqrt{(0.85)(0.15)}}{0.9 - 0.85}\right]^2$$

$$= 360.279\ldots \rightsquigarrow \text{round up } \underline{\underline{n = 361}}.$$

10.2 Tests (& C.I.s) for difference between 2 means of 2 normal populations — variance UNKNOWN.

2 populations means μ_1 , μ_2

sample sizes n_1 , n_2

Test

$$H_0 : \mu_1 = \mu_2 \qquad \rightarrow \qquad H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \begin{Bmatrix} \mu_1 \neq \mu_2 \\ \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \end{Bmatrix} \qquad \rightarrow \qquad H_1 : \begin{Bmatrix} \mu_1 - \mu_2 \neq 0 \\ \mu_1 - \mu_2 > 0 \\ \mu_1 - \mu_2 < 0 \end{Bmatrix}.$$

$\bar{X}_1 - \bar{X}_2 \rightarrow$ has normal distr. $\sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}\right)$

$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$, $\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$

So if we knew the variances σ_1^2, σ_2^2

then use
$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \underbrace{(\mu_1 - \mu_2)}_{0 \text{ under } H_0}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \text{ under } H_0$$

BUT don't know variances.

2 cases:
1. Can assume $\sigma_1^2 = \sigma_2^2$ $(= \sigma^2)$
2. Cannot assume this, must assume $\sigma_1^2 \neq \sigma_2^2$.

1. We get an estimate for σ^2 from each sample: S_1^2 & S_2^2.

The <u>pooled estimator</u> for σ^2 is

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \boxed{\frac{(n_1 - 1)}{n_1 + n_2 - 2}} S_1^2 + \boxed{\frac{(n_2 - 1)}{n_1 + n_2 - 2}} S_2^2$$

$w \qquad\qquad 1-w$

S_p^2 is an unbiased estimator for σ^2:

$$E(S_p^2) = E(wS_1^2 + (1-w)S_2^2) = wE(S_1^2) + (1-w)E(S_2^2)$$
$$= w\sigma^2 + (1-w)\sigma^2$$
$$= \sigma^2.$$

So with this:

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \sim \text{ t-distrib. with } n_1+n_2-2 \text{ degrees of freedom}$$

$\sim$ under H_0 i.e. $\mu_1 - \mu_2 = 0$

$$= \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Now proceed as always with a t test but use t_{α, n_1+n_2-2} or $t_{\frac{\alpha}{2}, n_1+n_2-2}$ as required.

C.I. That $T_0 \sim t_{n_1+n_2-2}$ dist. is a special case of:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{ t-distr. with } n_1+n_2-2 \text{ degrees of freedom}$$

Therefore since $P\left(-t_{\frac{\alpha}{2}, n_1+n_2-2} < T < t_{\frac{\alpha}{2}, n_1+n_2-2}\right)$

$= 1-\alpha$,

So a $100(1-\alpha)\%$ C.I. for $\mu_1 - \mu_2$ is:

$$\bar{x}_1 - \bar{x}_2 \pm t_{\frac{\alpha}{2}, n_1+n_2-2} \, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

(2) When $\sigma_1^2 \neq \sigma_2^2$, we don't get to pool S_1^2, S_2^2.

We have to stick with

when H_0 is true

$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \text{ t-distribution with } \nu \text{ degrees of freedom}$$

where

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$$

← if this is an integer ✓

← if **not** an integer round **DOWN** to the nearest integer.

For C.I.s replace $t_{\frac{\alpha}{2}, n-1}$ from single mean case with $t_{\alpha/2, \nu}$ (or $t_{\alpha, n-1}$ with $t_{\alpha, \nu}$)

& use $\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ in place of $S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

e.g. $100(1-\alpha)\%$ C.I. for $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 \pm t_{\frac{\alpha}{2}, \nu} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$