§9.2: Two ~~Population~~ "Sample" t-Test

Testing difference between indep. population means.

Setup: X & Y are independent (but similar) r.v.

with unknown mean and variance $\begin{cases} \mu_X, & \sigma_X^2 \\ \mu_Y, & \sigma_Y^2 \end{cases}$

- Sample X n times for sample mean $\bar{x}$, sample variance s_x^2
- Sample Y m times for sample mean $\bar{y}$, sample variance s_y^2

If $n, m > 40$ then by "central limit thm"

$$\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \text{ and } \frac{\bar{Y}-\mu_Y}{\sigma_Y/\sqrt{m}} \approx \text{Normal}(0,1)$$

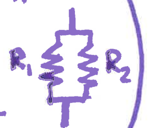
If not, then must use "t-distribution"

$\bar{X}$ has $(n-1)$ degrees of freedom
$\bar{Y}$ has $(m-1)$ degrees of freedom

What about $\bar{X}-\bar{Y}$?

(Idea: "Degrees of freedom" are like "resistance" they add by inverse

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$)



$$\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{s_X^2/n + s_Y^2/m}}$$ has ν degrees of freedom

Greek letter ν: "nu"

$$\frac{1}{\nu} = \left(\frac{s_x^2/n}{s_x^2/n + s_y^2/m}\right)\cdot\frac{1}{n-1} + \left(\frac{s_y^2/m}{s_x^2/n + s_y^2/m}\right)\cdot\frac{1}{m-1}$$

"Weighted average of $\frac{1}{n-1}$ & $\frac{1}{m-1}$"
"weighted" by variance of $\bar{X}$ & $\bar{Y}$

Hypothesis Test: $H_0: \mu_X - \mu_Y = \Delta$

Test Statistic:

$$\frac{(\bar{X}-\bar{Y})-\Delta}{\sqrt{s_x^2/n + s_y^2/m}} \sim t(\nu)$$

$$\nu = \frac{s_x^2/n + s_y^2/m}{\left(s_x^2/n\right)\cdot\frac{1}{n-1} + \left(s_y^2/m\right)\cdot\frac{1}{m-1}}$$

(Rounded down to nearest integer.)

Absolute value converts prob. to left tail for pt.

p-value:

$$pt\left(\frac{-\left|(\bar{x}-\bar{y})-\Delta\right|}{\sqrt{s_x^2/n + s_y^2/m}}, \nu\right)$$

2x this for Two-Tailed Test.

Note: ν is usually pretty close to the smaller of $(n-1)$ and $(m-1)$...

This is bad because

small # deg. of freedom $\Longrightarrow$ bigger p-value

Example: pt(-2, 4) ≈ .058

pt(-2, 8) ≈ .040

pt(-2, 12) ≈ .034

But there is one way to increase # deg. of freedom:

"Pooled Variance" t-Test

Same setup: X & Y independent r.v. with unknown means & variance.

- Sample X n times to get $\bar{x}$ and s_x^2
- Sample Y m times to get $\bar{y}$ and s_y^2

New: Assume that $Var[X] = Var[Y]$

$\sigma_X = \sigma_Y$

(Actually it is good enough if only approx. $\sigma_X \approx \sigma_Y$)

In this case we can combine s_x^2 and s_y^2 to get a better "pooled" estimate for variance of X and Y.

Theory: Pooled sample variance is

$$s^2 = \frac{\sum_i (X_i - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2}{n+m-2}$$

$n+m-2$ is so that it will be unbiased.

$$= \left(\frac{n-1}{n+m-2}\right) s_x^2 + \left(\frac{m-1}{n+m-2}\right) s_y^2$$

weighted average of s_x^2 and s_y^2 (weighted by # deg. of freedom)

Hypothesis Test: $H_0: \mu_X - \mu_Y = \Delta$

Test Statistic:

Sum of degrees of freedom!

Instead of $\frac{s_x}{\sqrt{n}} = \sqrt{\frac{s_x^2}{n}}$

$$\frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{s^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t(n+m-2)$$

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

"pooled sample variance"

p-value:

$$pt\left(-\left|\frac{(\bar{x} - \bar{y}) - \Delta}{\sqrt{s^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right|,\ n+m-2\right)$$

2× this for Two-Tailed Test.