How do we run these tests? Begin with an estimator for the parameter of interest. Here the parameter of interest is θ_1 - θ_2. What is a estimator for this quantity?

$$\hat{\theta}_1 - \hat{\theta}_2 = \bar{X}_1 - \bar{X}_2$$

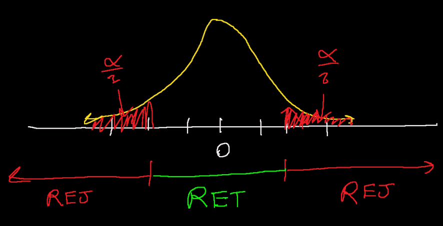
What is its sampling distribution exactly? From 241...

$$\hat{\theta}_1 - \hat{\theta}_2 \sim N\left(\theta_1 - \theta_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

What is its sampling distribution under H_0? E[θ_1 - θ_2] = 0 and the distribution is exactly...

Known

$$\hat{\theta}_1 - \hat{\theta}_2 \mid H_0 \sim N\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$



Does the estimate fall in RET? $\hat{\theta}_1 - \hat{\theta}_2 = \bar{x}_1 - \bar{x}_2 \in RET \Rightarrow$ Retain H_0

Let's test if male height and female height are unequal.

$$H_0: \theta_M - \theta_F = 0, \quad H_a: \theta_M - \theta_F \neq 0, \quad \alpha = 5\%$$

$\sigma_M^2 = 4''^2$, $\sigma_F^2 = 3.5''^2$. These are known values.

$x_1 = 62''$
$x_2 = 66''$
$x_3 = 63''$

$$n_M = 9, \ \bar{x}_M = 69.33'', \ n_F = 3, \ \bar{x}_F = 63.67'', \ \alpha = 5\% \Rightarrow RET = [-1.96, +1.96]$$

$$\hat{\theta}_M - \hat{\theta}_F \mid H_0 \sim N\left(0, \frac{4''^2}{9} + \frac{3.5''^2}{3}\right) = N(0, 2.42^2)$$

$$\bar{x}_M - \bar{x}_F = 69.33 - 63.67 = 5.66$$

$$z = \frac{\bar{x}_M - \bar{x}_F - 0}{\sqrt{\frac{\sigma_M^2}{n_M} + \frac{\sigma_F^2}{n_F}}} = \frac{5.66}{2.42} = 2.34 \notin RET \Rightarrow$$ we reject the null hypothesis!!

We conclude that male and female heights are different.

This was our first example of a 2-sample z-test.

Similar DGP: $X_{11}, \ldots, X_{1n_1} \overset{iid}{\sim} N(\theta_1, \sigma^2)$ indep. of
$X_{21}, \ldots, X_{2n_2} \overset{iid}{\sim} N(\theta_2, \sigma^2)$

This is the same situation as before except the variances are known to be equal and its value is known.

$$\hat{\theta}_1 - \hat{\theta}_2 \sim N\left(\theta_1 - \theta_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$\hat{\theta}_1 - \hat{\theta}_2 \mid H_0 \sim N\left(0, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) \Leftrightarrow \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

Consider the same DGP but σ² is unknown! We now have to estimate it. Recall from the one sample case:

$$\frac{\hat{\theta} - \theta}{\sigma\sqrt{\frac{1}{n}}} \sim N(0,1) \Rightarrow \frac{\hat{\theta} - \theta}{S\sqrt{\frac{1}{n}}} \sim T_{n-1}$$

So...

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1) \Rightarrow \frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{S_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1 + n_2 - 2}$$

To get an estimator of σ², we can take the weighted-by-degrees-of-freedom average of both sample standard deviation estimators:

$$S_{pooled}^2 = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2^2$$

pooled variance estimator

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

$$= \frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (X_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

$$E[S_{pooled}^2] = \frac{n_1 - 1}{n_1 + n_2 - 2} E[S_1^2] + \frac{n_2 - 1}{n_1 + n_2 - 2} E[S_2^2]$$

$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\sigma^2 + \frac{n_2 - 1}{n_1 + n_2 - 2}\sigma^2$$

$$= \sigma^2$$ unbiased!

Running a test with this T distribution is called a 2-sample t-test with equal variances

Consider the DGP: $X_{11}, \ldots, X_{1n_1} \overset{iid}{\sim} N(\theta_1, \sigma_1^2)$ indep. of
$X_{12}, \ldots, X_{1n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$

where both σ²'s are unknown and cannot assumped to be equal.

$$E[\hat{\theta}_1 - \hat{\theta}_2] = \theta_1 - \theta_2, \quad Var[\hat{\theta}_1 - \hat{\theta}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \overset{?}{\Rightarrow} \frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{?}{\sim} T_{n_1 + n_2 - 2}$$

this a reasonable conjecture!

It turns out this is wrong. There is no way to make it a T distrib.

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim$$ Behrens-Fisher distribution (1929/1935)

For 30 years, people thought its PDF was impossible to find in closed form. In 1966, it was proven it exists in closed form. And in 2018 it was found but it's very complicated.

However, we can use the "Welch-Satterthwaite Approximation" (1946/1947):

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\cdot}{\sim} T_{df}, \quad df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1 - 1)} + \frac{S_2^4}{n_2^2(n_2 - 1)}}$$

Using this distribution, the test is called "Welch's t-test" or the 2-sample t-test with unequal variances.

We will now return to estimation and explore it more generally. Consider the DGP: $X_1, \ldots, X_n \overset{iid}{\sim} f(\theta_1, \theta_2, \ldots, \theta_K)$ which has K parameters.

$$\hat{\theta} = w(X_1, \ldots, X_n), \quad \hat{\hat{\theta}} = w(x_1, \ldots, x_n)$$

statistical estimator, statistical estimate (statistic)

How we choose a function w to estimate an arbitrary parameter? If θ := E[X] then we used $\hat{\theta} = \bar{X} = \frac{1}{n}\sum X_i$.

How did we know to use this?

There's a principle that generated estimators called the "method of moments" (Karl Pearson, 1890's). Recall from probability theory that the kth moment of a rv is defined as:

$$\mu_K = E[X^k] \quad \text{e.g.} \quad \mu = \mu_1 = E[X^1] = E[X]$$

The mean is the "first moment" of a rv.

It seems reasonable that moments can be estimated from the data via:

$$\hat{\mu}_K = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad k = 1, \ \hat{\mu}_1 = \bar{X}$$ the average is a "method of moments estimator"

This is the logical extesion of the "average being a really good idea" for estimation.

The method of moments (MM) estimator of any parameter is defined as:

$$\hat{\theta}_j^{MM} = \beta_j(\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K) = w(X_1, \ldots, X_K)$$

Since there are K paramters and moments and parameters are 1:1 (a fact from probability), then you can set up a system of K equations and then invert them:

$$\begin{aligned} \mu_1 &= \alpha_1(\theta_1, \ldots, \theta_K) \\ &\vdots \\ \mu_K &= \alpha_K(\theta_1, \ldots, \theta_K) \end{aligned} \iff \begin{aligned} \theta_1 &= \beta_1(\mu_1, \ldots, \mu_K) \\ &\vdots \\ \theta_K &= \beta_K(\mu_1, \ldots, \mu_K) \end{aligned}$$

You set up these equations and solve for the one(s) you care about and then substite in the mu-hats and that becomes your estimator.

MM gives good estimators but not "great" estimators and sometimes its estimators produce illegal values in their support.

DGP: $X_1, \ldots, X_n \overset{iid}{\sim} f(\theta_1, \theta_2)$ mean variance

Let's derive MM estimators for mean and variance for the general case.

$$\theta_1 = \beta_1(\mu_1, \mu_2) = \mu_1$$

$$\theta_2 = \beta_2(\mu_1, \mu_2) = \mu_2 - \mu_1^2$$

$$\Downarrow$$

$$\hat{\theta}_1^{MM} = \beta_1(\hat{\mu}_1, \hat{\mu}_2) = \hat{\mu}_1 = \frac{1}{n}\sum X_i$$

$$\hat{\theta}_2^{MM} = \beta_2(\hat{\mu}_1, \hat{\mu}_2) = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2$$

$$= \frac{1}{n}\sum (X_i - \bar{x})^2 = \hat{\sigma}^2$$

Note: $\sum (X_i - \bar{x})^2 = \sum (X_i^2 - 2X_i\bar{x} + \bar{x}^2)$

$$= \sum X_i^2 - 2\bar{x}\underbrace{\sum X_i}_{n\bar{x}} + \sum \bar{x}^2$$

$$= \sum X_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum X_i^2 - n\bar{x}^2$$

$$= \sum X_i^2 - n\left(\frac{1}{n}\sum X_i\right)^2$$

$$\Rightarrow \frac{1}{n}\sum (X_i - \bar{x})^2 = \frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2$$

which is the biased estimator for the variance which we then correceted with Bessel's correction to get S^2