Math 455 Class 02 January 19 24

In summary regression output, we see F- and t-statistics.

These F- and t-distributions are sampling distributions connected to the normal distribution.

That is, if $Y_1, \dots, Y_n$ are iid $N(\mu, \sigma^2)$ and we cook up a statistic out of $Y_1, \dots, Y_n$ in exactly the right way, it can have the t- or F-distribution.

See Wackerly Thm 7.2, 7.3 and Def. 7.2, 7.3 and remarks following.

Recall Wackerly Thm. 6.4 of which a special case is: if $Z_1, \dots, Z_n$ are iid $N(0,1)$ what is the distribution of $Z_1^2 + \dots + Z_n^2$?

Ans: χ^2 with n dof.

Theorem 6.4 is more general:

Z_i can have dist $N(\mu_i, \sigma_i^2)$

and the statistic is

$$\left(\frac{Z_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{Z_2-\mu_2}{\sigma_2}\right)^2+\cdots+\left(\frac{Z_n-\mu_n}{\sigma_n}\right)^2 \sim \chi^2[n]$$

So we have seen how we can cook up a statistic from normal samples and get a χ^2-distribution.

The t- and F-distributions are defined (Def. 7.2 & 7.3) by using χ^2-distributions.

So further "cooking" will give us statistics with the t- and F-distributions and this is what we get in the summary regression output.

What does this have to do with regression?

Why do t- and F-statistics show up in the output?

A: This follows from the assumptions of linear regression:

- There is a linear relationship (approximate) between X and Y.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$.

β_0, β_1 are fixed but <u>unknown</u> parameters, which we are trying to estimate (math 448)

- Normality of errors: ε_i is normal.
- independence: ε_i are independent.
- Homoskedasticity: all ε_i have the same variance σ^2.

Now we can cook up ~~the~~ statistics

from X_i and Y_i that ~~the~~ ~~dep~~ depend

on the (normal) ε_i!

These statistics can be cooked up to have

the t- and F- distribution.

We will talk later about relaxing the

assumptions.

Q: Wait! The ε_i has distribution $N(0, \sigma^2)$

and we don't know what σ^2 is!

How can we cook up a statistic that doesn't

depend on σ?

A: You learned how to do this sort of thing

in Math 448.

See Wackerly, p. 359-360 as an example.

If $Y_1, \ldots, Y_n \sim N(0, \sigma^2)$ iid.

and $\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$

and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ (Wackerly p.357)

then $\bar{Y}$ and S are indep. (Thm. 7.3)

and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2[n-1]$ (Thm. 7.3)

and $Z = \sqrt{n}\,\frac{(\bar{Y})}{\sigma} \sim N(0,1)$

so $\sqrt{n}\,\frac{\bar{Y}}{S} \sim t$ with $n-1$ d.o.f.

i.e. the distribution of this statistic does <u>NOT</u> depend on σ.