M358K: November 10th, 2023.

χ^2: connections to normal samples.

Fact. Let $X_1, X_2, \dots, X_n$ be independent and Normal(mean$=\mu$, sd$=\sigma$).

Set, for all $i = 1..n$,

$$\boxed{Z_i := \frac{X_i - \mu}{\sigma}}$$

Note: The r.v. $Z_1, Z_2, \dots, Z_n$ are all independent and standard normal.

Define:

$$\boxed{Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi^2(df = n)}$$ ✓

Fact. Let $X_1, X_2, \dots, X_n$ be independent and Normal(mean$=\mu$, sd$=\sigma$)

μ unknown

Set $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$, i.e., the sample mean.

Then,

$$\boxed{\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(df = n-1)}$$

$$\Rightarrow \boxed{\frac{1}{\sigma^2}\boxed{\sum_{i=1}^{n}(X_i - \bar{X})^2} \sim \chi^2(df = n-1)} \quad ☆$$

Recall. The sample variance was defined as:

$$S^2 := \frac{1}{n-1}\boxed{\sum_{i=1}^{n}(X_i - \bar{X})^2} \longleftarrow$$

$$\Rightarrow \boxed{\frac{\boxed{(n-1)S^2}}{\sigma^2} \sim \chi^2(df = n-1)} \longleftarrow$$

Inference for Numerical Data.

So far: Normal population distribution
w/ an unknown mean μ
and a known std deviation σ.

The sample: $X_1, X_2, \dots, X_n$ independent and Normal(mean $=\mu$, sd $=\sigma$).

Set $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ to be the sample mean.

Its sampling distribution is

$$\bar{X} \sim \text{Normal}\left(\text{mean} = \mu, \text{sd} = \frac{\sigma}{\sqrt{n}}\right)$$

$$\Rightarrow \quad \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$\boxed{\text{Now:}}$ The std deviation will $\boxed{\text{not}}$ be known!

Idea: Use the sample std deviation s instead with

$$\boxed{s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

i.e., we want to use the following statistic

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

not standard normal

Random Variable

$$\boxed{\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t(df = n-1)} \quad \longleftarrow \quad \longleftarrow$$

t-distribution (aka Student's dist'n)

Def'n. Let $Z \sim N(0,1)$,

and $Y \sim \chi^2(df = n)$.

Assume they are independent r.v.s.

We define $\boxed{T = \dfrac{Z}{\sqrt{\frac{Y}{n}}}}$

We say that T has the t-distribution w/ n degrees of freedom.

Note: For a normal random sample:

We know:
$$Z := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

and
$$Y := \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(df = n-1)$$

independent

By def'n:
$$T = \frac{Z}{\sqrt{\frac{Y}{n-1}}} \sim \chi^2(df = n-1)$$

Consider:
$$T = \frac{\frac{\bar{X} - \mu}{\frac{\cancel{\sigma}}{\sqrt{n}}}}{\sqrt{\frac{\frac{\cancel{(n-1)}\, S^2}{\cancel{\sigma^2}}}{\cancel{n-1}}}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(df = n-1)$$