Some Normal-based Distribution Theory

Suppose \( X_1, \ldots, X_n \) i.i.d. r.v.'s with \( N(\mu, \sigma^2) \) distribution.

Any linear combination of the \( X_i \)'s also has a normal distribution.

In particular, \( \bar{X} \sim N(\mu, \sigma^2/n) \)

Let \( Z_i = \frac{X_i - \mu}{\sigma}, \quad Z_i \sim N(0,1) \)

\( Z_i^2 \sim \text{Chisquare with 1 df} \)

\( \text{Chisquare}(1) \) or \( K_i^2 \)

\( Z_1^2 + \cdots + Z_k^2 \sim \text{Chisquare } (k) \)

If \( s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \)
\( \frac{(n-1)s^2}{\sigma^2} \sim \text{Chi-square} \ (n-1) \)

If \( z \sim N(0,1) \), \( X \sim X^2 \)
\( z, X \) are independent
\( \frac{z}{\sqrt{X/m}} \sim t_m \) (t distribution with \( m \) d.f.)

In particular,
\[
\frac{\bar{X} - \mu}{\frac{s}{\sqrt{m}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{m}}} \sim t_{n-1}
\]

If \( X_1 \sim X^2 \), \( X_2 \sim X^2 \)
and \( X_1, X_2 \) independent
\[
\frac{X_1/m}{X_2/n} \sim F_{m, n} \]

F-distribution with \( m \) and \( n \) d.f.
The F distribution:
- Non-negative, right-skewed.

Note: the square of a r.v. with a tn distribution has a $F_{1,m}$ distribution.

F-test for equality of variance:

If we have 2 independent samples from 2 normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ of sizes $n_1$ and $n_2$ and calculate $s_1^2$ and $s_2^2$.

Suppose we want to test:

$H_0: \sigma_1^2 = \sigma_2^2$ versus $H_a: \neq$ ?
Test statistic:
\[
\left( \frac{(n_2-1)s_2^2}{s_1^2} \right) / n_2 - 1 \quad \text{under } H_0
\]
\[
= \frac{s_1^2}{s_2^2}
\]
\[
\sim F_{n_1-1, n_2-1}
\]

Problem with this test:
- Very sensitive to departures from normality (unlike the t-test)

- A significant test may mean \( s_1^2 \neq s_2^2 \) or it may mean that the distributions of the data weren't Normal.

(a better test: Levene's test)
**Data Collection**

3 methods:

1. Observational studies  
   - no design; no intervention  
   - H₀: θ₁₁.4.7 - graduate school admission  
   - coffee consumption and coronary disease  
   - Problem: confounding - can't separate effect of some variables from others - can't be generalized

2. Sample surveys  
   - data collected on a sample chosen from a population

3. Experiments  
   - high level of evidence to support conclusions
- Use sample statistics to make inferences about the population
  
  To avoid bias - randomly choose sample from the population

  Simplest method: Simple random sample - each subset of the population equally likely to be chosen

  - still observational, can still suffer from confounding

3 Experiment

- A value of a predictor variable is randomly assigned to the subjects or experimental units

- Can then make cause-effect conclusions