Statistical Concepts

Design of Experiments - Montgomery
Sections 2-1 through 2-3

Basic Statistical Concepts

- **Random Variable** - Y
  - Quantity (response) capable of taking on a set of values
  - Discrete or Continuous
  - Described by a probability distribution (density)
- **Numerical Summaries of Variable**
  - Center - Mean: μ, E()
  - Spread - Variance: σ^2, Var()

	Discrete	Continuous
μ :	$\sum y\text{Pr}(Y = y)$	$\int y\text{f}(y)$
σ^2 :	$\sum (y-\mu)^2\text{Pr}(Y = y)$	$\int (y-\mu)^2\text{f}(y)$

- **Elementary Results of Numerical Summaries**

  $\text{E}(aY \pm b) = a\text{E}(Y) \pm b$

  $\text{Var}(aY \pm b) = a^2\text{Var}(Y)$

  $\text{E}(Y_1 \pm Y_2) = \text{E}(Y_1) \pm \text{E}(Y_2)$

  $\text{Var}(Y_1 \pm Y_2) = \text{Var}(Y_1) + \text{Var}(Y_2) \pm 2\text{Cov}(Y_1, Y_2)$

  If Y_1 and Y_2 **independent** $\rightarrow$ $\text{Cov}(Y_1, Y_2) = 0$

  $\text{Var}(Y_1) = \text{E}(Y_1^2) - \text{E}(Y_1)^2$

  $\text{Cov}(Y_1, Y_2)$=$\text{E}(Y_1Y_2)$-$\text{E}(Y_1)\text{E}(Y_2)$.

Sampling/Reference Distribution

Statistical inference/testing: making decision in the presence of variability. Is result of experiment easily explained by chance variation or is it “unusual”?

- “Unusual”: Is it unlikely if only chance variation?
- Need dist of results assuming only chance variation
- Compare obs result with distribution of outcomes
- Example 1: Randomization test
  - Chance variation due to randomization
  - Generate all possible outcomes (each equally likely)
  - Compare observed result with dist of outcomes
- Example 2: t-test (comparing two means)
  - Calculate observed t test statistic
  - t dist summarizes outcomes under Null hypothesis
  - Compare observed result with distribution

Common Summaries

- Sample mean ($\overline{Y}$)

  Y_i independent with mean μ and variance σ^2

$$\text{E}(\frac{1}{n}\sum Y_i) = \frac{1}{n}\sum \text{E}(Y_i) = \frac{1}{n}n\mu = \mu$$
$$\text{Var}(\frac{1}{n}\sum Y_i) = \frac{1}{n^2}\sum \text{Var}(Y_i) = \frac{1}{n^2}n\sigma^2 = \sigma^2/n$$

What is distribution of $\overline{Y}$?

If Y_i Normal $\rightarrow \overline{Y}$ Normal
If Y_i Other $\rightarrow \overline{Y} \approx$ Normal

The Central Limit Theorem

If $Y_1, Y_2, \ldots, Y_n$ are independent r.v.'s with mean μ and variance σ^2.

$$\frac{\sum y - n\mu}{\sqrt{(n\sigma^2)}} \sim \text{N}(0, 1)$$

- Sample variance ($S^2 = \frac{1}{n-1}\sum(Y_i - \overline{Y})^2$)

$$\begin{aligned}
\text{E}(Y_i - \overline{Y}) &= \text{E}(Y_i) - \text{E}(\overline{Y}) = 0 \\
\text{Var}(Y_i - \overline{Y}) &= \text{Var}(Y_i) + \text{Var}(\overline{Y}) - 2\text{Cov}(Y_i, \overline{Y}) \\
&= \sigma^2 + \sigma^2/n - 2\sigma^2/n \\
&= \frac{n-1}{n}\sigma^2 \\
\text{E}((Y_i - \overline{Y})^2) &= \text{Var}(Y_i - \overline{Y}) + \text{E}(Y_i - \overline{Y})^2 \\
&= \frac{n-1}{n}\sigma^2 \\
\text{E}(S^2) &= \frac{1}{n-1} n \frac{n-1}{n}\sigma^2 \\
&= \sigma^2
\end{aligned}$$

What is distribution of S^2?

If Y_i Normal then

$$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$$

where $n-1$ is the degrees of freedom

Degrees of Freedom

Degrees of Freedom of a sum is equal to the number of elements in that sum that are independent (i.e., free to vary)

> For example, if you are told the sum of three elements equals five, you only need to know two of the three elements to know all of them

General Result:

If Y_i has variance σ^2 and SS=$\sum(Y_i - \overline{Y})^2$ with k degrees of freedom

$$\text{E}(\text{SS}/k) = \sigma^2$$

If Y_i is also Normally distributed

$$\text{SS}/\sigma^2 \sim \chi^2_k$$

Common Sampling Distributions

- **Normal Distribution**
  - Function of μ and σ^2
  - Standardize - $Z = (X - \mu)/\sigma$
  - Z is standard Normal
  - Only need probs associated with Z (Table 1)
- **Chi-square Distribution**
  - Function of degrees of freedom (k)
  - $Y = Z_1^2 + Z_2^2 + \ldots + Z_k^2 \sim \chi^2_k$ (Z's independent)
  - $\sum(Y_i - \overline{Y})^2/\sigma^2 \sim \chi^2_{n-1}$

- **t Distribution**
  - Function of degrees of freedom (k)
  - If Z and χ^2_k are independent, $t_k = Z/\sqrt{\chi^2/k}$
  - If Y Normal, $\overline{Y}$ and S^2 independent

$$\frac{(\overline{Y}-\mu)}{S/\sqrt{n}} \sim t_{n-1}$$

- **F distribution**
  - Function of degrees of freedom
  - Ratio of two independent χ^2 rvs
  - $F = \sqrt{\chi^2_1/k_1}/\sqrt{\chi^2_2/k_2}$
  - $S_1^2/S_2^2 \sim F_{n_1-1,n_2-1}$