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

- **Independence**

  Observations are statistically independent if the value of one of the observations does not influence the value of any other observations.

- **Elementary Results of Numerical Summaries**

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

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

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

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

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

  $\text{E}(Y_1 \times Y_2) = \text{E}(Y_1)\text{E}(Y_2)$, if Y_1, Y_2 independent

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

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

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

Populations/Samples

- A parameter is the true value of some numerical aspect of the population.
  - Examples: mean, median, variance, slope
- An estimator is a statistic that
  - Corresponds to a parameter
  - Is a random variable
  - An estimator $\widehat{\theta}$ of θ is unbiased if E($\widehat{\theta}$)=θ
- An estimate is a particular value of the estimator, computed from the sample data. It is considered fixed, **given the data**.

Common Estimators

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

  Y_i independent with mean μ and variance σ^2

$$
\begin{aligned}
\mathrm{E}(\frac{1}{n}\sum Y_i) &= \frac{1}{n}\sum \mathrm{E}(Y_i) = \frac{1}{n}n\mu = \mu \\
\mathrm{Var}(\frac{1}{n}\sum Y_i) &= \frac{1}{n^2}\sum \mathrm{Var}(Y_i) = \frac{1}{n^2}n\sigma^2 = \sigma^2/n
\end{aligned}
$$

  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.

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

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

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

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

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

If Y_i is also Normally distributed

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

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 Sampling Distributions

- **Normal Distribution**
  - Function of μ and σ : $X \sim N(\mu, \sigma)$
  - Z is standard Normal $\rightarrow \mu = 0, \sigma = 1$
  - Can standardize - $Z = (X - \mu)/\sigma$
  - Only need probs associated with Z (Table 1)
- **Chi-square Distribution**
  - Function of degrees of freedom (k)
  - $Z_1^2 + Z_2^2 + \ldots + Z_k^2 \sim \chi_k^2$ (Z's independent)
  - If $Y_i \sim N(\mu, \sigma)$, $\sum (Y_i - \overline{Y})^2/\sigma^2 \sim \chi_{n-1}^2$
  - For large k, by CLT $\chi^2(k) \approx N(k, 2k)$

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

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

- **F distribution**
  - Function of degrees of freedom
  - Ratio of two independent χ^2 rvs
  - $F = \left(\chi_u^2/u\right) / \left(\chi_v^2/v\right)$
  - $S_1^2/S_2^2 \sim F_{n_1-1, n_2-1}$

Noncentral Distributions

Will be used in power calculations!!!!

- **Noncentral t Distribution**
  - Recall $t_k = Z/\sqrt{\chi_k^2/k}$
  - Noncentral t of form

$$t_k(a) = \frac{N(a,1)}{\sqrt{\chi_k^2/k}}$$

  - Arises when $\mu \neq 0$ (*i.e.*, alternative hypothesis)
- **Noncentral Chi-square Distribution**
  - $X_i \sim N(a_i, 1)$, independent
  - $C = X_1^2 + X_2^2 + \ldots + X_k^2 \sim \chi_k^2(\phi)$, $\phi = \sum a_i^2$
- **Noncentral F Distribution**
  - Arises under alternative hypothesis of F test

$$F_{u,v}(\phi) = \left(\chi_u^2(\phi)/u\right) / \left(\chi_v^2/v\right)$$