Chapter 7: Sampling Distributions and the CLT

- In later chapters, a major focus will be on inference: making conclusions about a population based on sample data.
- Defn: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample if they are independent and identically distributed (iid) r.v.'s.
- In particular, we may wish to estimate a parameter using a statistic.
- Defn: A statistic is a function of the random variables in a random sample.
- Example: If the population mean μ is unknown, we may estimate it with the statistic:

- To aid inference, we should know the behavior of the statistic across repeated samples from the population.

Defn: The sampling distribution of a statistic is the distribution of possible values of the statistic across many repeated samples.

- The sampling distribution of a statistic will generally depend on the distribution of the original data $Y_1, Y_2, \ldots, Y_n$.

7.2 Sampling Distributions Related to the Normal Distn.

- In this section we assume the random variables in our random sample follow a normal distribution with mean μ and variance σ^2, i.e.,

$$Y_1, \dots, Y_n \overset{\text{indep}}{\sim} N(\mu, \sigma^2)$$

Theorem: If $Y_1, \dots, Y_n \overset{\text{indep}}{\sim} N(\mu, \sigma^2)$, then $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.

Proof: Write $\bar{Y} =$

$\Rightarrow$ Linear combination of indep. normal r.v.'s. We used method of mgf's to show this linear combination is:

- We write $\mu_{\bar{Y}} = \mu$, $\sigma^2_{\bar{Y}} = \frac{\sigma^2}{n}$. And the standard deviation of $\bar{Y}$ is $\sigma_{\bar{Y}} =$

Example 1(a): A factory's soft-drink machine is set to put 355 ml of soda, on average, in each can, with a standard deviation of 26 ml. What is the probability the mean amount per can in a 12-pack is less than 340 ml? Assume the amounts are normal.

Note:

Example 1 (b): How many observations must be in the sample for $\bar{Y}$ to be within 5 ml of μ with probability (at least) 0.95?

The Chi-Square (χ^2) distribution

- Recall the χ^2 distribution with ν degrees of freedom is simply the gamma distribution with $\alpha = \frac{\nu}{2}$ and $\beta = 2$.
- Thus the mgf of a χ^2_ν r.v. is:

$$M_Y(t) =$$

- We have shown that if $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$.

Theorem: If $Y_1, \ldots, Y_n$ are independent $\chi^2_{\nu_1}, \ldots, \chi^2_{\nu_n}$ r.v.'s, then $U = \sum_{i=1}^{n} Y_i$ has a χ^2 distribution with $\nu_1 + \cdots + \nu_n$ d.f.

Proof:

Theorem: If Y_1 and Y_2 are independent r.v.'s, and $Y_1 \sim \chi^2_{\nu_1}$ and $Y_1 + Y_2 \sim \chi^2_{\nu}$ (where $\nu > \nu_1$), then $Y_2 \sim \chi^2_{\nu - \nu_1}$.

Proof:

Sampling Distn. of the Sample Variance S^2 (when the Population is Normal)

- If $Y_1, \ldots, Y_n$ constitute a random sample from a population with variance σ^2, then the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

is an estimator of σ^2.

Theorem: If $Y_1, \dots, Y_n \overset{\text{indep}}{\sim} N(\mu, \sigma^2)$, then: $\bar{Y}$ and S^2 are independent; and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Proof: Details of the proof are given in a handout on the course web page. The basic steps are:

- Show that $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ can be written as a function of

- Use the multivariate transformation:

- It can be shown that the Jacobian of this transformation is n.
- By the multivariate transformation technique:

$f(u_1, \dots, u_n) =$

$\Rightarrow$ The joint pdf factors into a piece depending only on u_1 and a piece depending only on $(u_2, \dots, u_n)$.

$\Rightarrow$

Having shown independence, we now show that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$:

Example 2: At a factory, the standard deviation of a part's thickness is supposed to be 0.02 mm. A quality control inspector takes random samples of size 20 to make sure that the process maintains low variability. If the thickness is normally distributed, what is the probability the inspector would observe a sample standard deviation greater than 0.025 mm, if the process is in control?

The t-distribution

- We know that if $Y_1, \ldots, Y_n$ constitute a random sample from a $N(\mu, \sigma^2)$ population, then

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

- But in practice, the population standard deviation σ is unknown.
  - We typically replace σ with a sample estimate, such as s.
  - What is the sampling distribution of $\frac{\bar{Y} - \mu}{s/\sqrt{n}}$ when $Y_1, \ldots, Y_n \overset{\text{indep}}{\sim} N(\mu, \sigma^2)$?

Intuition: It seems replacing σ with s produces a statistic with _____ variability.