Memorize

- 68-95-99.7 rule for the normal distribution
- Central Limit Theorem
- confidence interval:
  estimate $\pm$ multiplier $\times$ SE(estimate)
- $$\text{test statistic} = \frac{\text{estimate - hypothesized value under null}}{SE(\text{estimate})}$$
- definitions of confidence interval and p-value

Example 1

The Indian False Vampire bat consumes frogs. Assume the consumption time has a normal distribution with mean 27 minutes and *known* standard deviation 8 minutes.For a single frog, what is the probability that consumption time is greater than 25 minutes?

Example 2

The eating habits of 12 bats were examined in the article “Foraging Behavior of the Indian False Vampire Bat” (*Biotropica*(1991):63-67). For these 12 bats the average time to consume a frog was $\overline{x}$ = 21.9 minutes. Assume that the standard deviation is known to be 8.8 minutes.

A. What is the probability that the mean consumption time is greater than 25 minutes?

B. Construct and interpret a 95% confidence interval for the mean suppertime of a vampire bat whose meal consists of a frog. Do this for two cases: (i) σ is known to be 8 minutes; (ii) σ is estimated from the data to be 8 minutes.

Sampling: From statistics to models

- simple random *sample* from a *population*
- sample quantity $\longrightarrow$ inference about population parameter
- population parameter: quantity chosen to specify a model (μ, σ)
- sample statistic: a summary found from the data, usually used to estimate a population parameter ($\overline{X}$, s_x)
- assumptions
- sampling distribution model
- inference

The normal distribution as a population model

- Why use the normal distribution?
  - Many variables follow an approximately normal distribution, such as scores on tests taken by many people, repeated careful measurements of the same quantity, characteristics of biological populations (heights, weights).
  - If a variable is non-normal, there is often a transformation to normality (3.5, *Sleuth*)
- Completely defined by the population parameters: μ (mean) and σ^2 (variance)
- For $Y \sim N(\mu, \sigma)$, $z = \frac{Y-\mu}{\sigma} \sim N(0, 1)$. z is called the “z-score”.
- Let $W \sim N(0, 1)$. Find z such that $P(-z \leq W \leq z) = 1 - \alpha$. This z is denoted $z_{1-\alpha/2}$ and is called a *z-value*, and α is called the confidence level. For example, let $\alpha = .05$. Here, $z_{0.025} = 1.96$.

A model for the sampling distribution of the mean

- Assumptions:
  1. *Random sampling condition:* The values must be sampled at random (or the concept of a sampling distribution makes no sense.)
  2. *Independence assumption:* The sampled values must be mutually independent. If the sample is drawn without replacement check that the sample size, n is no more than 10% of the population.
- **Central Limit Theorem:** As the sample size, n, increases, the average $\overline{X}$ of n independent values has a sampling distribution that tends toward a Normal model with mean equal to the population mean, μ, and standard deviation equal to $\sigma/\sqrt{n}$

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \tag{1}$$

- The distribution of sample means will be approximately normal *regardless of the distribution of values in the original population from which the samples were taken.*
- *Sample size issues:* Watch out for small samples from skewed populations.

Application: Confidence Limits for μ, σ known

- A 100(1-α)% confi dence interval for μ:$\bar{Y}\underline{+}z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$
- *Interpretation*: If we repeatedly drawing a sample of size n and form the interval each time, 95% of these *random* intervals will contain the true fi xed value μ.
- *An example*: Crop researchers plant 15 plots with a new variety of corn. The mean yield in bushels per acre is 130. Assume σ=10 bushels per acre. Find the 90% confi dence interval for the mean yield μ for this variety of corn. (Ans: [125.75,134.25])
- Sample size issues
  - In the above example, how many plots of corn are needed to estimate the average yield to within $\underline{+}$ 2 bushels per acre with probability 95%?
  - Ans: Solve $2 = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ for n. $n = 96$. *As the confidence interval on the mean becomes narrower, what happens to* n*? What is happening to* $SE(mean)$ *as* n *increases?*

The t-distribution

- In most applications, σ is unknown, and is estimated by s. This approximation is usually satisfactory for $n \geq 30$.
- *When we have to estimate* s*, the quantity* $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$ *is not normally distributed.*
- The quantity $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$ has a t distribution with $n-1$ degrees of freedom.

$$\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

  In chapter 3, we'll discuss the applicability of the t distribution and two-sample t-tests for different sample sizes, standard deviations, and departures from independence.

- For n large, the t distribution with n degrees of freedom approaches the standard normal distribution.

A t-model for the sampling distribution of the mean

- *Independence assumption*: Data values should be mutually independent
- *Randomization condition*: The data arise from a random sample or a randomized experiment.
- *10% condition*: When the sample is drawn without replacement, check that the sample is no more than 10% of the population.
- *Normal population assumption*: The data come from a population that follow a normal model. (Check unimodal and symmetric using histogram and normal probability plot.)
- A one-sample t-interval
  - The standard error of the mean is $SE(\overline{Y}) = \frac{s_y}{\sqrt{n}}$.
  - The interval is $\bar{Y} \pm t_{1-\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$