Basic statistical concepts

- Descriptive statistics (numeric/graphical)
- Population distribution vs. Sampling distribution
- Standard Deviation vs. Standard Error
- Estimation of population mean/proportion
- Confidence interval
- Hypothesis testing
- P-value
Confidence Interval for population mean

- An approximate 95% confidence interval for population mean \( \mu \) is:

\[
\bar{X} \pm 2 \times \text{SEM}
\]

or precisely

\[
\bar{X} \pm 1.96 \times \text{SEM}
\]

- \( \bar{X} \) is a random variable (vary from sample to sample), so confidence interval is random and it has 95% chance of covering \( \mu \) before a sample is selected.

- Once a sample is taken, we observe \( \bar{X} = \bar{x} \), then either \( \mu \) is within the calculated interval or it is not.

- The confidence interval gives the range of plausible values for \( \mu \).
Example

- 95% CI for $\mu$ (mean blood pressure in the population) is
  - $125 \pm 2 \times 1.4$
  - $125 \pm 2.8$

- Ways to write CI:
  - $122.2$ to $127.8$
  - $(122.2, 127.8)$
  - $(122.2 - 127.8)$

- The 95% error bound on $\bar{x}$ is 2.8.
- We are highly confident that the population mean falls in the range $122.2$ to $127.8$
Confidence Interval Interpretation

Technical interpretation

- The CI “works” (includes \( \mu \)) 95% of the time.
- If we were to take 100 random samples each of the same size, approximately 95 of the CIs would include the true value of \( \mu \).
Confidence Interval Interpretation

Each bar represents a 95% CI created from a random sample of size n.
In order to be able to use the formula

\[
\bar{x} \pm 1.96 \text{SEM}
\]

Assumptions:

- Random sample from population - important!
- Observations in the sample are independent.
- Sample size is large enough to support the Central Limit Theorem, how large depends on the population distribution.
Estimation of population proportion ($p$)

Examples:
- Proportion of patients who became infected
- Proportion of patients who are cured
- Proportion of individuals positive on a blood test
- Proportion of adverse drug reactions
- Proportion of premature infants who survive
Sampling Distribution of Sample Proportion

- Sampling distribution of sample proportion can be approximated by normal distribution when sample size is sufficiently large (central limit theorem).
- The standard error of a sample proportion \( \hat{p} \) is estimated by:
  \[
  SE(\hat{p}) = \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}}
  \]
- 95% Confidence Interval for a Proportion
  \[
  \hat{p} \pm 2 \times SE(\hat{p})
  \]
  The rule of thumb for good normal approximation is
  \[
  n \times \hat{p} \geq 5 \quad \text{and} \quad n \times (1-\hat{p}) \geq 5
  \]
Example

- In a study of 200 patients, 90 patients experienced adverse drug reactions
- The estimated proportion who experience an adverse drug reaction is
  \[ \hat{p} = \frac{90}{200} = 0.45 \]
- 95% confidence interval for the population proportion is
  \[ 0.45 \pm 2 \sqrt{\frac{0.45 \times 0.55}{200}} = (0.38, 0.52) \]
Hypothesis Testing

One-sample test

- Hypothesis specification
- Test statistics
- \( p \)-value
- Significance level
Hypothesis for blood pressure example

Suppose we want to know if the mean systolic blood pressure for the student population is different from the normal cutoff.

- Null hypothesis  \( H_0: \mu = \mu_0 \) (=120)
- Alternative hypothesis  \( H_A: \mu \neq 120 \)

- typically represents what you are trying to prove.

- We reject \( H_0 \) if the sample mean is far away from 120.
Hypothesis Testing Question

- Do our sample results allow us to reject $H_0$ in favor of $H_A$?
  - Sample mean $\bar{x}$ would have to be far from 120 to claim $H_A$ is true.
  - Is $\bar{x} = 125$ large enough to claim $H_A$ is true?
  - Maybe we have a large sample mean of 125 from a chance occurrence.
  - Maybe $H_0$ is true, and we just have an unusual sample.
  - We need some measure of how probable the result from our sample is, if the null hypothesis is true. $\Rightarrow$ p-value
Test Statistics

Test statistic is a score to measure how many standard errors the observed sample mean is away from null mean $\mu_0$.

If $H_0$ is true ($\mu = \mu_0$), consider

- **Z test statistic**
  \[ Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1) \]  
  (normal distribution)
  when (i) Population is normally distributed or sample size is large enough and (ii) Population variance $\sigma^2$ is known.

- **T test statistic**
  \[ T = \frac{\overline{X} - \mu_0}{s / \sqrt{n}} \sim t_{n-1} \]  
  (t-distribution)
  when (i) Population is normally distributed or sample size is large enough and (ii) Population variance $\sigma^2$ is unknown.
How are p-values calculated?

- In the SBP example, the observed value of T statistic is
  \[ t = \frac{125 - 120}{14 / \sqrt{100}} = 3.57 \]

- We observed a sample mean that was 3.57 standard errors away from what we would have expected the mean to be if we assume \( H_0 \) is true.

- Is a result of 3.57 standard errors above its mean unusual?
  - It depends on what kind of distribution we are dealing with.

- The p-value is the probability of getting a test statistic as (or more) extreme than what you observed (3.57) by chance if \( H_0 \) was true.

- The p-value comes from the sampling distribution of the test statistic.
Blood Pressure example

- T statistic follows a t-distribution with degrees of freedom = n-1 = 99
- p-value \( = P\{|T| \geq |t|\} = P\{|T| \geq 3.57\} = 0.0006 \) (red area)

If the mean SBP in the student population is the same as normal cutoff 120 mmHg, then the chance of seeing a sample mean as extreme or more extreme than 125 in a sample of 100 students is 0.0006.
Using the p-value to Make a Decision

- We need to decide if our sample result is unlikely enough to have occurred by chance if the null was true. Our measure of this “unlikeliness” is our p-value, $p = 0.0006$.

- We need to have a cutoff such that all p-values less than the cutoff result in a rejection of the null hypothesis.
  - The standard cutoff is 0.05, which is a somewhat arbitrary value.
  - The cutoff value is referred to as $\alpha$ or the significance level of the test.
Using the p-value to Make a Decision

- At the 0.05 level, the test results for the student SBP example is statistically significantly. There is sufficient evidence to conclude that the mean systolic blood pressure for the student population is different from the normal cutoff.

- The p-value alone imparts no information about the scientific importance or substantive content in a study.
More on the p-value

- Statistical significance is not the same as scientific significance.

- Suppose in the student SBP Example:
  - $n = 100,000$; $\bar{x} = 120.1$ mmHg; $s = 14$
  - $p$-value = 0.024

- A large $n$ can produce a small $p$-value, even though the magnitude of the difference is very small and may not be scientifically or substantively significant.
More on the p-value

- Not rejecting $H_0$ is not the same as accepting $H_0$
- Suppose in the student SBP example
  - $n = 5$; $\bar{x} = 135$; $s = 14$
  - p-value = 0.07
- We cannot reject $H_0$ at significance level $\alpha = 0.05$.
- But, are we really convinced mean SBP for student population is not different from normal cutoff, 120mmHg?
- Maybe we should have taken a bigger sample?
Connection Between Hypothesis Testing and Confidence Intervals

● The confidence interval gives a range of plausible values for the population parameter.

● If $\mu_0$ is not in the 95% CI, then we would reject the null hypothesis that $\mu = \mu_0$ at level $\alpha = 0.05$. (The p-value will be < 0.05.)

● In the student SBP example, the 95% confidence interval (122, 128) does not overlap 120, so we know that the result is statistically significant. Thus, the p-value is less than 0.05. But it doesn’t tell us that $p = 0.0006$. 
What if my data are clearly not normal?

- Is sample size large enough to apply the central limit theorem?
- Are there any obvious outliers?
- Nonparametric tests
  - Wilcoxon signed-rank test or signed test
    - Make few assumptions about the distribution of the data.
    - Test on the median instead of the mean.
Paired design

- Paired design
  - Self-pairing:
    - Measurements are taken at two distinct points in time from a single subject (e.g. Before vs. After)
  - Matched pairs (e.g., twins, eyes, subjects matched on important characteristics such as age and gender)

- Why pairing?
  - Control extraneous noise
  - Control confounding factors that affect the comparison
  - Make comparison more precise
### Example: Blood Pressure and Oral Contraceptive Use

<table>
<thead>
<tr>
<th>Participant</th>
<th>BP Before OC</th>
<th>BP After OC</th>
<th>After-Before</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>126</td>
<td>132</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>105</td>
<td>109</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>104</td>
<td>102</td>
<td>-2</td>
</tr>
<tr>
<td>4</td>
<td>115</td>
<td>117</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Paired samples**

1\(^{st}\) sample

2\(^{nd}\) sample
Scientific questions:

- What is the mean change in blood pressure after OC use in a population of women who use oral contraceptives?
  - Estimate the mean change by a confidence interval approach
- Is there any change in mean blood pressure after OC use in a population of women who use oral contraceptives?
  - Hypothesis testing
Inference on mean change

- Due to the design of the study, we can reduce the BP information on two samples (women’s BP prior to OC use and the same subject’s BP after OC use) into one piece of information: information on the differences in BP between the times points for the same subject.

- Perform the one sample inference on the difference for the relevant research question.
THE END

Want to learn more statistics or have more questions
http://ctsi.psu.edu/ctsi-programs/biostatisticsepidemiologyresearch-design/