UQ, STAT2201, 2017,
Lecture 6
Unit 6 – Statistical Inference Ideas.
Statistical Inference is the process of forming judgements about the parameters of a population typically on the basis of random sampling.
The random variables $X_1, X_2, \ldots, X_n$ are an (i.i.d.) random sample of size $n$ if

(a) the $X_i$’s are independent random variables and 
(b) every $X_i$ has the same probability distribution.

A statistic is any function of the observations in a random sample, and the probability distribution of a statistic is called the sampling distribution.
Any function of the observation, or any statistic, is also a random variable. We call the probability distribution of a statistic a **sampling distribution**. A **point estimate** of some population parameter $\theta$ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.
The most common statistic we consider is the sample mean, $\bar{X}$, with a given value denoted by $\bar{x}$. As an estimator, the sample mean is an estimator of the population mean, $\mu$. 
The Central Limit Theorem
Central Limit Theorem (for sample means):

If $X_1, X_2, \ldots, X_n$ is a random sample of size $n$ taken from a population with mean $\mu$ and finite variance $\sigma^2$ and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

as $n \to \infty$, is the standard normal distribution.

This implies that $\bar{X}$ is approximately normally distributed with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$. 
The **standard error** of $\bar{X}$ is given by $\sigma/\sqrt{n}$. In most practical situations $\sigma$ is not known but rather estimated in this case, the **estimated standard error**, (denoted in typical computer output as "SE"), is $s/\sqrt{n}$ where the sample standard deviation $s$ is the point estimator for the population standard deviation,

\[
s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}{n - 1}}.
\]
Central Limit Theorem (for sums):

Manipulate the central limit theorem (for sample means and use \(\sum_{i=1}^{n} X_i = n \bar{X}\). This yields,

\[
Z = \frac{\sum_{i=1}^{n} X_i - n \mu}{\sqrt{n\sigma^2}},
\]

which follows a standard normal distribution as \(n \to \infty\).

This implies that \(\sum_{i=1}^{n} X_i\) is approximately normally distributed with mean \(n \mu\) and variance \(n \sigma^2\).
Confidence Intervals
Knowing the sampling distribution (or the approximate sampling distribution) of a statistic is the key for the two main tools of statistical inference that we study:

(a) **Confidence intervals** – a method for yielding error bounds on point estimates.
(b) **Hypothesis testing** – a methodology for making conclusions about population parameters.
The formulas for most of the statistical procedures use **quantiles of the sampling distribution**. When the distribution is \( N(0, 1) \) (standard normal), the \( \alpha \)'s quantile is denoted \( z_{\alpha} \) and satisfies:

\[
\alpha = \int_{-\infty}^{z_{\alpha}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx.
\]

A common value to use for \( \alpha \) is 0.05 and in procedures the expressions \( z_{1-\alpha} \) or \( z_{1-\alpha/2} \) appear. Note that in this case \( z_{1-\alpha/2} = 1.96 \approx 2 \).
A **confidence interval** estimate for $\mu$ is an interval of the form $l \leq \mu \leq u$, where the end-points $l$ and $u$ are computed from the sample data. Because different samples will produce different values of $l$ and $u$, these end points are values of random variables $L$ and $U$, respectively. Suppose that

$$P(L \leq \mu \leq U) = 1 - \alpha.$$ 

The resulting **confidence interval** for $\mu$ is

$$l \leq \mu \leq u.$$ 

The end-points or bounds $l$ and $u$ are called the **lower- and upper-confidence limits** (bounds), respectively, and $1 - \alpha$ is called the **confidence level**.
If \( \bar{x} \) is the sample mean of a random sample of size \( n \) from a normal population with known variance \( \sigma^2 \), a 100(1 - \( \alpha \))% confidence interval on \( \mu \) is given by

\[
\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.
\]

Note that it is roughly of the form,

\[
\bar{x} - 2 \text{ SE} \leq \mu \leq \bar{x} + 2 \text{ SE}.
\]

Learn how to do back of the envelope calculations!
Confidence interval formulas give insight into the **required sample size**: If $\bar{x}$ is used as an estimate of $\mu$, we can be $100(1 - \alpha)\%$ confident that the error $|\bar{x} - \mu|$ will not exceed a specified amount $\Delta$ when the sample size is not smaller than

$$n = \left(\frac{z_{1-\alpha/2} \sigma}{\Delta}\right)^2.$$
Hypothesis Testing
A statistical hypothesis is a statement about the parameters of one or more populations.

The null hypothesis, denoted $H_0$ is the claim that is initially assumed to be true based on previous knowledge.

The alternative hypothesis, denoted $H_1$ is a claim that contradicts the null hypothesis.
For some arbitrary value $\mu_0$, a **two-sided alternative hypothesis** is expressed as:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

A **one-sided alternative hypothesis** is expressed as:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0$$

or

$$H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0.$$
The standard scientific research use of hypothesis is to “hope to reject” $H_0$ so as to have statistical evidence for the validity of $H_1$. 
An hypothesis test is based on a decision rule that is a function of the test statistic. For example: Reject $H_0$ if the test statistic is below a specified threshold, otherwise don’t reject.
Rejecting the null hypothesis $H_0$ when it is true is defined as a type I error. Failing to reject the null hypothesis $H_0$ when it is false is defined as a type II error.
<table>
<thead>
<tr>
<th></th>
<th>$H_0$ Is True</th>
<th>$H_0$ Is False</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fail to reject $H_0$:</strong></td>
<td>No error</td>
<td>Type II error</td>
</tr>
<tr>
<td><strong>Reject $H_0$:</strong></td>
<td>Type I error</td>
<td>No error</td>
</tr>
</tbody>
</table>

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true}).$$

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \mid H_0 \text{ is false}).$$
The **power** of a statistical test is the probability of rejecting the null hypothesis $H_0$ when the alternative hypothesis is true.

Desire: $\alpha$ is low and power $(1 - \beta)$ as high as can be.
Simple Hypothesis Tests
A typical example of a simple hypothesis test has

\[ H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1, \]

where \( \mu_0 \) and \( \mu_1 \) are some specified values for the population mean. This test isn’t typically practical but is useful for understanding the concepts at hand.

Assuming that \( \mu_0 < \mu_1 \) and setting a threshold, \( \tau \), reject \( H_0 \) if the \( \bar{x} > \tau \), otherwise don’t reject.
Explicit calculation of the relationships of $\tau$, $\alpha$, $\beta$, $n$, $\sigma$, $\mu_0$ and $\mu_1$ is possible in this case.
Practical Hypothesis Tests (focus of Units 7,8 of the course)
In most hypothesis tests used in practice (and in this course), a specified level of type I error, \( \alpha \) is predetermined (e.g. \( \alpha = 0.05 \)) and the type II error is not directly specified.

The probability of making a type II error \( \beta \) increases (power decreases) rapidly as the true value of \( \mu \) approaches the hypothesized value.

The probability of making a type II error also depends on the sample size \( n \) - increasing the sample size results in a decrease in the probability of a type II error.

The population (or natural) variability (e.g. described by \( \sigma \)) also affects the power.
The **P-value** is the smallest level of significance that would lead to rejection of the null hypothesis $H_0$ with the given data. That is, the P-value is based on the data. It is computed by considering the location of the test statistic under the sampling distribution based on $H_0$. 
It is customary to consider the test statistic (and the data) significant when the null hypothesis $H_0$ is rejected; therefore, we may think of the $P$-value as the smallest $\alpha$ at which the data are significant. In other words, the $P$-value is the observed significance level.
Clearly, the $P$-value provides a measure of the credibility of the null hypothesis. Computing the exact $P$-value for a statistical test is not always doable by hand.

It is typical to report the $P$-value in studies where $H_0$ was rejected (and new scientific claims were made). Typical (“convincing”) values can be of the order 0.001.
A General Procedure for Hypothesis Tests is

(1) **Parameter of interest:** From the problem context, identify the parameter of interest.

(2) **Null hypothesis,** $H_0$: State the null hypothesis, $H_0$.

(3) **Alternative hypothesis,** $H_1$: Specify an appropriate alternative hypothesis, $H_1$.

(4) **Test statistic:** Determine an appropriate test statistic.

(5) **Reject $H_0$ if:** State the rejection criteria for the null hypothesis.

(6) **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute the value.

(7) **Draw conclusions:** Decide whether or not $H_0$ should be rejected and report that in the problem context.