Review
Significance Test for a Population Mean $\mu$

Hypotheses
Null: $H_0 : \mu = \mu_0$
Alternative: $H_a : \mu > \mu_0$ (one-sided) or $H_a : \mu < \mu_0$ (one-sided) or $H_a : \mu \neq \mu_0$ (two-sided)

Test statistic

$$t = \frac{\bar{x} - \mu_0}{se_0} \quad \text{with } se_0 = \frac{s}{\sqrt{n}}$$

P-value (Use $t$ distribution with $df = n - 1$)

<table>
<thead>
<tr>
<th>Alternative hypothesis</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a : \mu &gt; \mu_0$</td>
<td>Right-tail probability</td>
</tr>
<tr>
<td>$H_a : \mu &lt; \mu_0$</td>
<td>Left-tail probability</td>
</tr>
<tr>
<td>$H_a : \mu \neq \mu_0$</td>
<td>Two-tail probability</td>
</tr>
</tbody>
</table>

Conclusion
More on P-values

Alternative hypothesis \( H_a : \mu > \mu_0 \) or \( H_a : \mu < \mu_0 \)

\( p \), where \( p \) is the probability corresponding to the absolute value of the t-score (the subscript for \( t \) in table A-3)

Alternative hypothesis \( H_a : \mu \neq \mu_0 \)

\( 2 \times p \), where \( p \) is the probability corresponding to the absolute value of the t-score if the t-score is negative
More on Conclusions
If the P-value is less than a preset value (threshold), e.g. 0.05, which corresponds to 95% confidence level, then we reject the null hypothesis. If the P-value is larger than a preset value (threshold), then we do not reject the null hypothesis.
Significance Test for a Population Proportion $p$

**Hypotheses**
- Null: $H_0 : p = p_0$
- Alternative: $H_a : p > p_0$ (one-sided) or $H_a : p < p_0$ (one-sided) or $H_a : p \neq p_0$ (two-sided)

**Test statistic**

$$z = \frac{\hat{p} - p_0}{se_0} \quad \text{with } se_0 = \sqrt{p_0(1-p_0)/n}$$

**P-value**

<table>
<thead>
<tr>
<th>Alternative hypothesis</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a : p &gt; p_0$</td>
<td>Right-tail probability</td>
</tr>
<tr>
<td>$H_a : p &lt; p_0$</td>
<td>Left-tail probability</td>
</tr>
<tr>
<td>$H_a : p \neq p_0$</td>
<td>Two-tail probability</td>
</tr>
</tbody>
</table>

**Conclusion**
More on P-values

Alternative hypothesis $H_a : p > p_0$

$1 - \text{prob}$, where \text{prob} is the probability corresponding to the z-score

Alternative hypothesis $H_a : p < p_0$

\text{prob}, where \text{prob} is the probability corresponding to the z-score

Alternative hypothesis $H_a : p \neq p_0$

$2 \times \text{prob}$, where \text{prob} is the probability corresponding to the z-score if z-score is negative;

$2 \times (1 - \text{prob})$, where \text{prob} is the probability corresponding to the z-score if z-score is positive
More on Conclusions
If the P-value is less than a preset value (threshold), e.g. 0.05, which corresponds to 95% confidence level, then we reject the null hypothesis. If the P-value is larger than a preset value (threshold), then we do not reject the null hypothesis.
A 95% confidence interval for the population mean $\mu$ is

$$\bar{x} \pm t_{.025,df} (se), \quad \text{with } se = s/\sqrt{n}$$

where $t_{.025,df}$ denotes the $t$-score of the $t$-distribution and $df = n - 1$ denotes the degrees of freedom of the corresponding $t$-distribution. (See Appendix A-3.)
A 95% confidence interval for a population proportion $p$ is

$$\hat{p} \pm 1.96(\text{se}), \quad \text{with } \text{se} = \sqrt{\hat{p}(1 - \hat{p})/n}$$

where $\hat{p}$ denotes the sample proportion based on $n$ observations.
### z-Scores for the Most Common Confidence Levels

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Error Probability</th>
<th>z-Score</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>0.10</td>
<td>1.645</td>
<td>$\hat{p} \pm 1.645$ (se)</td>
</tr>
<tr>
<td>0.95</td>
<td>0.05</td>
<td>1.96</td>
<td>$\hat{p} \pm 1.96$ (se)</td>
</tr>
<tr>
<td>0.99</td>
<td>0.01</td>
<td>2.58</td>
<td>$\hat{p} \pm 2.58$ (se)</td>
</tr>
</tbody>
</table>
• Mean and Standard Error of Sampling Distribution of Sample Mean $\bar{x}$

For a random sample of size $n$ from a population having mean $\mu$ and standard deviation $\sigma$, the sampling distribution of the sample mean $\bar{x}$ has

$$\text{mean} = \mu \quad \text{and} \quad \text{standard deviation} = \frac{\sigma}{\sqrt{n}}.$$ 

We call the standard deviation of a sampling distribution a standard error.
• Mean and Standard Deviation of the Sample Distribution of a Proportion

For a random sample of size \( n \) from a population with proportion \( p \) of outcomes in a particular category, the sampling distribution of the proportion of the sample in that category has

\[
\text{mean} = p \quad \text{and} \quad \text{standard deviation} = \sqrt{\frac{p(1-p)}{n}}.
\]

We call the standard deviation of a sampling distribution a **standard error**.
Review Session

Things need to know:

Observational Studies & Experimental Studies
Explanatory Variable & Response Variable
Treatment
Review Session

Things need to know:

Sampling

Simple Random Sampling
Cluster Random Sampling
Stratified Random Sampling
Margin of Error

\[
\text{approximate margin of error} = \frac{1}{\sqrt{n}} \times 100\%
\]
Review Session

Things need to know:

Bias: sampling bias, nonresponse bias & response bias
sampling bias – convenience sample & volunteer sample
Review Session

Things need to know:

**Sample Space**
The probability of each individual outcome is between 0 and 1. The total of all the individual probabilities equals 1.

**Event**
The probability of an event $A$, denoted by $P(A)$, is obtained by adding the probabilities of the individual outcomes in the event.

When all the possible outcomes are equally likely,

$$P(A) = \frac{\text{number of outcomes in event } A}{\text{number of outcomes in the sample space}}$$
Review Session

Things need to know:

**Complement** of an Event - - - $A^c$

$$P(A^c) = 1 - P(A)$$
Review Session

Things need to know:

**Intersection** of Two Event - - - $A \cap B$

$$P(A \cap B) = P(A | B) \times P(B)$$

If $A$ and $B$ are **independent**, then

$$P(A \cap B) = P(A) \times P(B).$$
Review Session

Things need to know:

Union of Two Events - - - \( A \cup B \)

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B).
\]

If \( A \) and \( B \) are disjoint, then

\[
P(A \cup B) = P(A) + P(B).
\]
Review Session

Things need to know:

Conditional Probability

\[ P(A \mid B) = \frac{P(A \cap B)}{P(B)} \]
**Independent Events**

Events $A$ and $B$ are **independent** if

$P(A \mid B) = P(A)$; or

$P(B \mid A) = P(B)$; or

$P(A \cap B) = P(A) \times P(B)$. 
Random Variable (discrete & continuous)
Probability Distribution
The probability distribution of a discrete random variable assigns a probability $P(x)$ to each possible value $x$.

The **mean** for a discrete random variable is

$$
\mu = \sum xP(x)
$$

The **standard deviation** for a discrete random variable is

$$
\sigma = \sqrt{\sum (x - \mu)^2 P(x)}
$$
The probability distribution of a continuous random variable assigns probabilities to each possible interval. Graphically, the probability for the random variable falling in certain interval is the area underneath the (density) curve for the corresponding interval.
Normal Distribution

z-Score Table (cumulative probabilities for standard normal distribution)

z-score

\[ z = \frac{x - \mu}{\sigma}, \]
Binomial Distribution
Probability Distribution of A Binomial Random Variable

\[ P(x) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \ldots, n \]

The mean is

\[ \mu = np \]

The standard deviation is

\[ \sigma = \sqrt{np(1-p)} \]
Review Session

Things need to know:

- Design, Description (Descriptive Statistics) & Inference (Inferential Statistics)
- Population & Sample (Sample size)
- Variable
  - Categorical
  - Quantitative — Discrete & Continuous
Review Session

Things need to know:

Frequency Table (Frequency, Relative Frequency)
Dot Plots
Stem-and-Leaf Plot
Histogram
Time Plots
Box Plots
Scatter Plots

For graphic summaries, one needs to know how to make graphs and how to get information from given graphs.
Review Session

Things need to know:

Mode (Unimodal, Bimodal)
Shape of the graph (Symmetric or Skewed)
Review Session

Things need to know:

Mean (know how to calculate)

\[ \bar{x} = \frac{\sum x}{n} \]

Median (know how to calculate)

Comparison between Mean and Median
Review Session

Things need to know:

Range \((= \text{max} - \text{min})\)

Deviations \((= x - \bar{x})\)

Variance (know how to calculate)

\[
s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}
\]

Standard Deviation (know how to calculate)

\[
s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}
\]
Review Session

Things need to know:

\( p \text{th Percentile} \)

Quartiles (Q1, Q2 & Q3) (know how to calculate)

Interquartile Range (IQR = Q3 − Q1)

1.5×IQR Criterion & Outliers

Five-Number Summary (Min, Q1, Median, Q3, Max)

\( z \text{-Score} \)

\[
z = \frac{x - \bar{x}}{s} = \frac{\text{observation} - \text{mean}}{\text{standard deviation}}
\]
Review Session

Things need to know:

- Explanatory Variable & Response Variable
- Association (Positive & Negative)
- Contingency Table (Conditional Proportions)
- Correlation (properties P108; formula for $r$ is not required)
Things need to know:

Regression Line

\[ \hat{y} = a + bx \]

\( a \) – y-intercept & \( b \) – slope

Make Predictions for a given Regression Line

Residual

Least Squares Method

\[ b = r \left( \frac{s_x}{s_y} \right) \quad a = \bar{y} - b(\bar{x}) \]