Review
Question

\[ \mu? \quad p? \]
Question

\( \mu \)?

C.I.?

Test?

1-S?

2-S?

1-S?

2-S?

1-S?

2-S?

1-S?

2-S?

1-S?

2-S?

1-S?

2-S?

\( \sigma \) known?

\( \sigma \) unknown?

P366

P476

P476

P505

P505

P526

P513

P532

P378

P447

P450
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Confidence intervals:

\[
(\bar{x} - z^* \sigma \sqrt{\frac{1}{n}}, \bar{x} + z^* \sigma \sqrt{\frac{1}{n}})
\]

$z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$. 

• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Confidence intervals:
  * form: estimate $\pm$ margin of error / interpretation
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Confidence intervals:
  * form: estimate $\pm$ margin of error / interpretation
  * $\left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}}\right)$
• Inference about $\mu$ with known $\sigma$ — z-procedures (confidence interval & test of significance)

• Confidence intervals:
  * form: estimate $\pm$ margin of error / interpretation
  * $(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}})$
  * $z^*$ is determined by the confidence level $C$ — the z-score corresponding to the upper tail $(1 - C)/2$
• Inference about \( \mu \) with known \( \sigma \) — z-procedures (confidence interval & test of significance)
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)
• Test of significance:

$H_0$ vs. $H_a$ / $H_0$:
$\mu = \mu_0$

Test statistics:
$z = \bar{x} - \mu_0 / \sigma / \sqrt{n}$

$P$-value:
$H_a$: $\mu > \mu_0$ — upper tail probability corresponding to $z$
$H_a$: $\mu < \mu_0$ — lower tail probability corresponding to $z$
$H_a$: $\mu \neq \mu_0$ — twice upper tail probability corresponding to $|z|$

Significance level $\alpha$ and conclusion
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  * test statistics: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a / H_0 : \mu = \mu_0$
  * test statistics: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
  * $P$-value:
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: \mu = \mu_0$
  * test statistics: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
  * $P$-value:
    * $H_a: \mu > \mu_0$ — upper tail probability corresponding to $z$
• Inference about $\mu$ with known $\sigma$ — z-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  * test statistics: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
  * $P$-value:
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $z$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $z$
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Test of significance:
  
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  
  * test statistics: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
  
  * $P$-value:
    
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $z$
    
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $z$
    
    * $H_a : \mu \neq \mu_0$ — twice upper tail probability corresponding to $|z|$
• Inference about $\mu$ with known $\sigma$ — $z$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  * test statistics: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
  * $P$-value:
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $z$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $z$
    * $H_a : \mu \neq \mu_0$ — twice upper tail probability corresponding to $|z|$
  * significance level $\alpha$ and conclusion
• Assumptions for z-procedures:

- the sample is an SRS
- the population has a normal distribution
- the population standard deviation \( \sigma \) is known

Margin of errors in confidence intervals are affected by \( C \), \( \sigma \) and \( n \) to get a level \( C \), C.I. with margin of \( m \), we need an SRS with sample size \( n = \left( \frac{z \ast \sigma}{m} \right)^2 \)

The significance of test will also be affected by sample size.
• Assumptions for z-procedures:
  * the sample is an SRS
Assumptions for $z$-procedures:

- the sample is an SRS
- the population has a normal distribution

Margin of errors in confidence intervals are affected by $C$, $\sigma$, and $n$ to get a level $C$. I with margin of $m$, we need an SRS with sample size $n = \left(\frac{z \sigma}{m}\right)^2$

The significance of test will also be affected by sample size
• Assumptions for z-procedures:
  ∗ the sample is an SRS
  ∗ the population has a normal distribution
  ∗ the population standard deviation $\sigma$ is known

• Margin of errors in confidence intervals are affected by $C$, $\sigma$ and $n$ to get a level $C$ C.I. with margin of $m$, we need an SRS with sample size $n = \left(\frac{z}{m}\right)^2$.

• The significance of test will also be affected by sample size.
• Assumptions for $z$-procedures:
  * the sample is an SRS
  * the population has a normal distribution
  * the population standard deviation $\sigma$ is known
• Margin of errors in confidence intervals are affected by $C$, $\sigma$, and $n$
  to get a level $C$ C.I. with margin of $m$, we need an SRS
  with sample size

$$n = \left( \frac{z^* \sigma}{m} \right)^2$$
• Assumptions for $z$-procedures:
  * the sample is an SRS
  * the population has a normal distribution
  * the population standard deviation $\sigma$ is known

• Margin of errors in confidence intervals are affected by $C$, $\sigma$ and $n$
  
  to get a level $C$ C.I. with margin of $m$, we need an SRS with sample size
  
  $$n = \left( \frac{z^* \sigma}{m} \right)^2$$

• The significance of test will also be affected by sample size
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Standard error: $\frac{s}{\sqrt{n}}$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Standard error: $\frac{s}{\sqrt{n}}$

• $t$-distribution; degrees of freedom ($n - 1$)
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Standard error: $\frac{s}{\sqrt{n}}$

• $t$-distribution; degrees of freedom ($n - 1$)

• Confidence intervals:
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Standard error: $\frac{s}{\sqrt{n}}$

• $t$-distribution; degrees of freedom $(n - 1)$

• Confidence intervals:
  
  $\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Standard error: $\frac{s}{\sqrt{n}}$

• $t$-distribution; degrees of freedom $(n - 1)$

• Confidence intervals:
  $\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}$

  $t^*$ is determined by the confidence level $C$ — the $t$-score corresponding to the upper tail $(1 - C)/2$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:

  $H_0$ vs $H_a$:
  
  $H_0$: $\mu = \mu_0$
  
  test statistics: $t = \bar{x} - \mu_0 \frac{s}{\sqrt{n}}$
  
  $P$-value:
  
  $H_a$: $\mu > \mu_0$ — upper tail probability corresponding to $t$
  
  $H_a$: $\mu < \mu_0$ — lower tail probability corresponding to $t$
  
  $H_a$: $\mu \neq \mu_0$ — twice upper tail probability corresponding to $|t|$
  
  significance level $\alpha$ and conclusion
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a / H_0 : \mu = \mu_0$
  * test statistics: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0 \ v.s \ H_a / H_0 : \mu = \mu_0$
  * test statistics: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
  * $P$-value:
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  * test statistics: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
  * $P$-value:
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  * test statistics: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
  * $P$-value:
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $t$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu = \mu_0$
  * test statistics: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
  * $P$-value:
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $t$
    * $H_a : \mu \neq \mu_0$ — twice upper tail probability corresponding to $|t|$
• Inference about $\mu$ with unknown $\sigma$ — $t$-procedures (confidence interval & test of significance)

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: \mu = \mu_0$
  * test statistics: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
  * $P$-value:
    * $H_a: \mu > \mu_0$ — upper tail probability corresponding to $t$
    * $H_a: \mu < \mu_0$ — lower tail probability corresponding to $t$
    * $H_a: \mu \neq \mu_0$ — twice upper tail probability corresponding to $|t|$
  * significance level $\alpha$ and conclusion
• Inference about two means — $\mu_1 - \mu_2$
• Inference about two means — $\mu_1 - \mu_2$

• Standard error for $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
• Inference about two means — $\mu_1 - \mu_2$

• Standard error for $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

• Confidence interval for $\mu_1 - \mu_2$:
• Inference about two means — $\mu_1 - \mu_2$

• Standard error for $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

• Confidence interval for $\mu_1 - \mu_2$:

$$\left( (\bar{x}_1 - \bar{x}_2) - t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$
• Inference about two means — $\mu_1 - \mu_2$
• Standard error for $\bar{x}_1 - \bar{x}_2$:

$$\sqrt{s_1^2 \frac{1}{n_1} + s_2^2 \frac{1}{n_2}}$$

• Confidence interval for $\mu_1 - \mu_2$:

$$\left( (\bar{x}_1 - \bar{x}_2) - t^* \sqrt{s_1^2 \frac{1}{n_1} + s_2^2 \frac{1}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t^* \sqrt{s_1^2 \frac{1}{n_1} + s_2^2 \frac{1}{n_2}} \right)$$

* $t^*$ is determined by the confidence level $C$ — the t-score corresponding to the upper tail $(1 - C)/2$
• Inference about two means — $\mu_1 - \mu_2$
• Standard error for $\bar{x}_1 - \bar{x}_2$:
  \[
  \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
  \]
• Confidence interval for $\mu_1 - \mu_2$:
  \[
  * \left( (\bar{x}_1 - \bar{x}_2) - t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)
  \]
  * $t^*$ is determined by the confidence level $C$ — the t-score corresponding to the upper tail $(1 - C)/2$
  * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
• Inference about two means — \( \mu_1 - \mu_2 \)
• Inference about two means — $\mu_1 - \mu_2$
• Test of significance:
  
  - Hypotheses:
    - $H_0$: $\mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
    - $H_a$: $\mu_1 > \mu_2$ — upper tail probability corresponding to $t^*$
    - $H_a$: $\mu_1 < \mu_2$ — lower tail probability corresponding to $t^*$
    - $H_a$: $\mu_1 \neq \mu_2$ — twice upper tail probability corresponding to $|t|$
  
  - Test statistics:
    - $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$
  
  - $P$-value:
    - $P$-value for $H_0$: $\mu_1 = \mu_2$
    - $P$-value for $H_a$: $\mu_1 > \mu_2$
    - $P$-value for $H_a$: $\mu_1 < \mu_2$
    - $P$-value for $H_a$: $\mu_1 \neq \mu_2$

  - Degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$

  - Significance level $\alpha$ and conclusion
• Inference about two means — $\mu_1 - \mu_2$
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  * test statistics: $t = \bar{x}_1 - \bar{x}_2 \sqrt{s_1^2/n_1 + s_2^2/n_2}$
  * $P$-value:
  * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
  * $H_a$: $\mu > \mu_0$ — upper tail probability corresponding to $t$;
  * $H_a$: $\mu < \mu_0$ — lower tail probability corresponding to $t$;
  * $H_a$: $\mu \neq \mu_0$ — twice upper tail probability corresponding to $|t|$;
  * significance level $\alpha$ and conclusion
• Inference about two means — $\mu_1 - \mu_2$
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
• Inference about two means — $\mu_1 - \mu_2$

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  * $P$-value:
• Inference about two means — $\mu_1 - \mu_2$

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0)$
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  * $P$-value:
    * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
• Inference about two means — $\mu_1 - \mu_2$

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  * $P$-value:
    * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
• Inference about two means — $\mu_1 - \mu_2$

• Test of significance:
  
  * hypotheses: $H_0$ v.s $H_a / H_0 : \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$

  * $P$-value:
    
    * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $t$
• Inference about two means — $\mu_1 - \mu_2$

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  * $P$-value:
    * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $t$
    * $H_a : \mu \neq \mu_0$ — twice upper tail probability corresponding to $|t|$
• Inference about two means — $\mu_1 - \mu_2$
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a / H_0 : \mu_1 = \mu_2$ ($\mu_1 - \mu_2 = 0$)
  * test statistics: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  * $P$-value:
    * degrees of freedom: smaller of $n_1 - 1$ and $n_2 - 1$
    * $H_a : \mu > \mu_0$ — upper tail probability corresponding to $t$
    * $H_a : \mu < \mu_0$ — lower tail probability corresponding to $t$
    * $H_a : \mu \neq \mu_0$ — twice upper tail probability corresponding to $|t|$
  * significance level $\alpha$ and conclusion
• Inference about population proportion $p$ — $z$-procedures
  (confidence interval & test of significance)
• Inference about population proportion $p$ — z-procedures (confidence interval & test of significance)

• Sampling distribution of the sample proportion $\hat{p}$ for an SRS:
• Inference about population proportion $p$ — $z$-procedures (confidence interval & test of significance)

• Sampling distribution of the sample proportion $\hat{p}$ for an SRS:
  * mean of $\hat{p}$ equals the population proportion $p$
• Inference about population proportion $p$ — $z$-procedures (confidence interval & test of significance)

• Sampling distribution of the sample proportion $\hat{p}$ for an SRS:
  * mean of $\hat{p}$ equals the population proportion $p$
  * standard deviation of $\hat{p}$ equals $\sqrt{\frac{p(1-p)}{n}}$
• Inference about population proportion $p$ — $z$-procedures (confidence interval & test of significance)

• Sampling distribution of the sample proportion $\hat{p}$ for an SRS:
  * mean of $\hat{p}$ equals the population proportion $p$
  * standard deviation of $\hat{p}$ equals $\sqrt{\frac{p(1-p)}{n}}$
  * If the sample size is large, $\hat{p}$ is approximately normal, i.e.
    $\hat{p} \overset{\text{approx}}{\sim} N(p, \sqrt{\frac{p(1-p)}{n}})$
• Inference about population proportion $p$ — $z$-procedures
  (confidence interval & test of significance)
• Sampling distribution of the sample proportion $\hat{p}$ for an SRS:
  * mean of $\hat{p}$ equals the population proportion $p$
  * standard deviation of $\hat{p}$ equals $\sqrt{\frac{p(1-p)}{n}}$
  * If the sample size is large, $\hat{p}$ is approximately normal, i.e.
    $\hat{p} \approx N(p, \sqrt{\frac{p(1-p)}{n}})$
• Standard error of $\hat{p}$: $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
• Inference about population proportion $p$ — $z$-procedures

* Large-sample confidence intervals:

$\left( \hat{p} - z^* \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z^* \sqrt{\hat{p}(1-\hat{p})/n} \right)$

* $z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1-C)/2$.

* Use it only when $n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$.

* Plus four confidence intervals:

$\left( \tilde{p} - z^* \sqrt{\tilde{p}(1-\tilde{p})/n} + 4, \tilde{p} + z^* \sqrt{\tilde{p}(1-\tilde{p})/n} + 4 \right)$

* $\tilde{p} = \text{number of successes in the sample} + 2/n + 4$.

* Use it when the confidence level is at least 90% and the sample size $n$ is at least 10.
• Inference about population proportion $p$ — $z$-procedures
• Large-sample confidence intervals:

$\left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$

$z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$.

Use it only when $n\hat{p} \geq 15$ and $n(1 - \hat{p}) \geq 15$.

$\left( \tilde{p} - z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}, \tilde{p} + z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} \right)$

$\tilde{p} = \frac{\text{number of successes in the sample} + 2}{n + 4}$

Use it when the confidence level is at least 90% and the sample size $n$ is at least 10.
• Inference about population proportion $p$ — z-procedures

• Large-sample confidence intervals:

$$
\hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \quad \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
$$

$z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$.

Use it only when $n\hat{p} \geq 15$ and $n(1 - \hat{p}) \geq 15$.

• Plus four confidence intervals:

$$
\tilde{p} - z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n} + 4}, \quad \tilde{p} + z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n} + 4}
$$

$\tilde{p} = \frac{\text{number of successes in the sample}}{n} + 2$.

Use it when the confidence level is at least 90% and the sample size $n$ is at least 10.
• Inference about population proportion $p$ — z-procedures

• Large-sample confidence intervals:
  \[
  \left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)
  \]
  * $z^*$ is determined by the confidence level $C$ — the z-score corresponding to the upper tail $(1 - C)/2$
• Inference about population proportion \( p \) — z-procedures
• Large-sample confidence intervals:
  \[
  \left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)
  \]
  * \( z^* \) is determined by the confidence level \( C \) — the z-score corresponding to the upper tail \( (1 - C)/2 \)
  * Use it only when \( n\hat{p} \geq 15 \) and \( n(1 - \hat{p}) \geq 15 \)
• Inference about population proportion \( p \) — z-procedures

• Large-sample confidence intervals:

\[
\left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)
\]

* \( z^* \) is determined by the confidence level \( C \) — the z-score corresponding to the upper tail \((1 - C)/2\)

* Use it only when \( n\hat{p} \geq 15 \) and \( n(1 - \hat{p}) \geq 15 \)

• Plus four confidence intervals:
• Inference about population proportion $p$ — $z$-procedures

• Large-sample confidence intervals:

$$\left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

* $z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$

* Use it only when $n\hat{p} \geq 15$ and $n(1 - \hat{p}) \geq 15$

• Plus four confidence intervals:

$$\left( \tilde{p} - z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}, \tilde{p} + z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} \right)$$

* $\tilde{p} = \text{number of successes in the sample} + 2$
• Inference about population proportion $p$ — $z$-procedures

• Large-sample confidence intervals:

$$\left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

* $z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$

* Use it only when $n\hat{p} \geq 15$ and $n(1 - \hat{p}) \geq 15$

• Plus four confidence intervals:

$$\left( \tilde{p} - z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}, \tilde{p} + z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} \right)$$

* $\tilde{p} = \frac{\text{number of successes in the sample} + 2}{n + 4}$
• Inference about population proportion $p$ — $z$-procedures

• Large-sample confidence intervals:
  \[
  \left( \hat{p} - z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)
  \]
  * $z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$
  * Use it only when $n\hat{p} \geq 15$ and $n(1 - \hat{p}) \geq 15$

• Plus four confidence intervals:
  \[
  \left( \tilde{p} - z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}, \tilde{p} + z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}} \right)
  \]
  * $\tilde{p} = \frac{\text{number of successes in the sample} + 2}{n + 4}$
  * Use it when the confidence level is at least 90% and the sample size $n$ is at least 10
• Inference about population proportion $p$ — $z$-procedures
• Inference about population proportion $p$ — $z$-procedures
• Test of significance:

$H_0$ vs $H_a$:

- $H_0$: $p = p_0$
- $H_a$: $p > p_0$ — upper tail probability corresponding to $z$
- $H_a$: $p < p_0$ — lower tail probability corresponding to $z$
- $H_a$: $p \neq p_0$ — twice upper tail probability corresponding to $|z|$

Significance level $\alpha$ and conclusion:

- Use this test when $np_0 \geq 10$ and $n(1 - p_0) \geq 10$
• Inference about population proportion $p$ — $z$-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: p = p_0$

$z =$ $\hat{p} - p_0 \sqrt{p_0(1 - p_0)} n$

$P$-value:

$\star H_a: p > p_0$ — upper tail probability corresponding to $z$

$\star H_a: p < p_0$ — lower tail probability corresponding to $z$

$\star H_a: p \neq p_0$ — twice upper tail probability corresponding to $|z|$

$\star$ significance level $\alpha$ and conclusion

$\star$ use this test when $np_0 \geq 10$ and $n(1 - p_0) \geq 10$
• Inference about population proportion $p$ — $z$-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ \( H_0 : p = p_0 \)
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

\( \star \) use this test when $np_0 \geq 10$ and $n(1-p_0) \geq 10$
• Inference about population proportion $p$ — $z$-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p = p_0$
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
  * $P$-value: $\star$
• Inference about population proportion $p$ — $z$-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p = p_0$
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
  * $P$-value:
    * $H_a : p > p_0$ — upper tail probability corresponding to $z$
• Inference about population proportion $p$ — $z$-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p = p_0$
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
  * $P$-value:
    * $H_a : p > p_0$ — upper tail probability corresponding to $z$
    * $H_a : p < p_0$ — lower tail probability corresponding to $z$
• Inference about population proportion $p$ — z-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p = p_0$
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
  * $P$-value:
    * $H_a : p > p_0$ — upper tail probability corresponding to $z$
    * $H_a : p < p_0$ — lower tail probability corresponding to $z$
    * $H_a : p \neq p_0$ — twice upper tail probability corresponding to $|z|$
• Inference about population proportion $p$ — $z$-procedures

• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p = p_0$
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
  * $P$-value:
    * $H_a : p > p_0$ — upper tail probability corresponding to $z$
    * $H_a : p < p_0$ — lower tail probability corresponding to $z$
    * $H_a : p \neq p_0$ — twice upper tail probability corresponding to $|z|$
  * significance level $\alpha$ and conclusion
• Inference about population proportion $p$ — z-procedures
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p = p_0$
  * test statistics: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
  * $P$-value:
    * $H_a : p > p_0$ — upper tail probability corresponding to $z$
    * $H_a : p < p_0$ — lower tail probability corresponding to $z$
    * $H_a : p \neq p_0$ — twice upper tail probability corresponding to $|z|$
  * significance level $\alpha$ and conclusion
  * use this test when $np_0 \geq 10$ and $n(1 - p_0) \geq 10$
• Inference about two proportions — $p_1 - p_2$
• Inference about two proportions — $p_1 - p_2$
• Sampling distribution of $\hat{p}_1 - \hat{p}_2$:
• Inference about two proportions — $p_1 - p_2$
• Sampling distribution of $\hat{p}_1 - \hat{p}_2$:
  * mean of $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$
• Inference about two proportions — $p_1 - p_2$

• Sampling distribution of $\hat{p}_1 - \hat{p}_2$:
  * mean of $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$
  * standard deviation of $\hat{p}_1 - \hat{p}_2$ is

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$
• Inference about two proportions — $p_1 - p_2$

• Sampling distribution of $\hat{p}_1 - \hat{p}_2$:
  * mean of $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$
  * standard deviation of $\hat{p}_1 - \hat{p}_2$ is
    \[\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\]
  * If the sample size is large, $\hat{p}_1 - \hat{p}_2$ is approximately normal
• Inference about two proportions — \( p_1 - p_2 \)

• Sampling distribution of \( \hat{p}_1 - \hat{p}_2 \):
  * mean of \( \hat{p}_1 - \hat{p}_2 \) is \( p_1 - p_2 \)
  * standard deviation of \( \hat{p}_1 - \hat{p}_2 \) is
    \[
    \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}
    \]
  * If the sample size is large, \( \hat{p}_1 - \hat{p}_2 \) is approximately normal

• Standard error of \( \hat{p} \):
  \[
  \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
  \]
• Inference about two proportions — $p_1 - p_2$
• Inference about two proportions — $p_1 - p_2$
• Large-sample confidence intervals:
• Inference about two proportions — $p_1 - p_2$

• Large-sample confidence intervals:

$$\left( (\hat{p}_1 - \hat{p}_2) - z^*SE, (\hat{p}_1 + \hat{p}_2) + z^*SE \right),$$

where $SE$ is the standard error of $\hat{p}_1 - \hat{p}_2$:

$$SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

$z^*$ is determined by the confidence level $C$ — the $z$-score corresponding to the upper tail $(1 - C)/2$.
• Inference about two proportions — $p_1 - p_2$
• Large-sample confidence intervals:
  
  * $\left( \hat{p}_1 - \hat{p}_2 - z^*SE, (\hat{p}_1 + \hat{p}_2) + z^*SE \right)$, where SE is the standard error of $\hat{p}_1 - \hat{p}_2$:

  $$SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

  * $z^*$ is determined by the confidence level $C$ — the z-score corresponding to the upper tail $(1 - C)/2$
• Inference about two proportions — $p_1 - p_2$

• Large-sample confidence intervals:
  \[
  \left( (\hat{p}_1 - \hat{p}_2) - z^*SE, (\hat{p}_1 + \hat{p}_2) + z^*SE \right), \text{ where } SE \text{ is the standard error of } \hat{p}_1 - \hat{p}_2:
  \]

  \[
  SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
  \]

  * $z^*$ is determined by the confidence level $C$ — the z-score corresponding to the upper tail $(1 - C)/2$
  * Use it only when $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$
• Inference about two proportions — $p_1 - p_2$

• Large-sample confidence intervals:

  $\left( (\hat{p}_1 - \hat{p}_2) - z^*SE, (\hat{p}_1 + \hat{p}_2) + z^*SE \right)$, where SE is the standard error of $\hat{p}_1 - \hat{p}_2$:

  \[
  SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}
  \]

  * $z^*$ is determined by the confidence level $C$ — the z-score corresponding to the upper tail $(1 - C)/2$

  * Use it only when $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$
• Inference about two proportions — $p_1 - p_2$
• Inference about two proportions — \( p_1 - p_2 \)
• Plus four confidence intervals:

\[
\left( \hat{p}_1 - \hat{p}_2 \right) - z^\star \text{SE}, \quad \left( \hat{p}_1 + \hat{p}_2 \right) + z^\star \text{SE},
\]
where SE is the standard error of \( \hat{p}_1 - \hat{p}_2 \):

\[
\text{SE} = \sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2} + \frac{1}{2}},
\]

\( \hat{p}_i \) = number of successes in the \( i \)th sample + 1

\( n_i \) = number of successes in the \( i \)th sample + 2,

\( i = 1, 2 \)

Use it when \( n_1 \geq 5 \) and \( n_2 \geq 5 \)
• Inference about two proportions — \( p_1 - p_2 \)
• Plus four confidence intervals:
  \[
  * \left( (\hat{p}_1 - \hat{p}_2) - z^*SE, (\hat{p}_1 + \hat{p}_2) + z^*SE \right),
  \]
  where SE is the standard error of \( \hat{p}_1 - \hat{p}_2 \):

  \[
  SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1 + 2} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2 + 2}}
  \]
• Inference about two proportions — \( p_1 - p_2 \)

• Plus four confidence intervals:
  \[
  (\hat{p}_1 - \hat{p}_2) - z^*SE, (\hat{p}_1 + \hat{p}_2) + z^*SE
  \]
  where SE is the standard error of \( \hat{p}_1 - \hat{p}_2 \):
  \[
  SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1 + 2} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2 + 2}}
  \]
  \[
  \hat{p}_i = \frac{\text{number of successes in the } i \text{ th sample} + 1}{n_i + 2}, \ i = 1, 2
  \]
Inference about two proportions — \( p_1 - p_2 \)

Plus four confidence intervals:

\[
(\hat{p}_1 - \hat{p}_2) - z^*SE, \ (\hat{p}_1 + \hat{p}_2) + z^*SE
\]

where SE is the standard error of \( \hat{p}_1 - \hat{p}_2 \):

\[
SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1 + 2} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2 + 2}}
\]

\[
\hat{p}_i = \frac{\text{number of successes in the } i \text{ th sample} + 1}{n_i + 2}, \ i = 1, 2
\]

Use it when \( n_1 \geq 5 \) and \( n_2 \geq 5 \)
Test of significance:

- Hypotheses:
  
  \[ H_0 \text{ v.s } H_a \]
  
  \[ p_1 = p_2 (p_1 - p_2 = 0) \]

- Pooled sample proportion \( \hat{p} \):
  
  \[ \hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}} \]

- Test statistic:
  
  \[ z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \]

- \( P \)-value:
  
  \( \star \)
  
  \[ H_a : p_1 - p_2 > 0 \] — upper tail probability corresponding to \( z \)
  
  \[ H_a : p_1 - p_2 < 0 \] — lower tail probability corresponding to \( z \)
  
  \[ H_a : p_1 - p_2 \neq 0 \] — twice upper tail probability corresponding to \( |z| \)

- Significance level \( \alpha \) and conclusion

- Use this test when counts of successes and failures are each 5 or more in both samples.
• Test of significance:
  * hypotheses: $H_0 \text{ v.s } H_a / H_0: p_1 = p_2 (p_1 - p_2 = 0)$
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0 : p_1 = p_2$ ($p_1 - p_2 = 0$)
  * pooled sample proportion $\hat{p}$:
    $$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$$
  * test statistic:
    $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
  * $P$-value:
    * $H_a$: $p_1 - p_2 > 0$ — upper tail probability corresponding to $z$
    * $H_a$: $p_1 - p_2 < 0$ — lower tail probability corresponding to $z$
    * $H_a$: $p_1 - p_2 \neq 0$ — twice upper tail probability corresponding to $|z|$
  * significance level $\alpha$ and conclusion
  * use this test when counts of successes and failures are each 5 or more in both samples
• Test of significance:
* hypotheses: $H_0$ v.s $H_a$ / $H_0: p_1 = p_2$ ($p_1 - p_2 = 0$)
* pooled sample proportion $\hat{p}$:
$$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$$
* test statistics:
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

* $P$-value:
  - $H_a$: $p_1 - p_2 > 0$ — upper tail probability corresponding to $z$
  - $H_a$: $p_1 - p_2 < 0$ — lower tail probability corresponding to $z$
  - $H_a$: $p_1 - p_2 \neq 0$ — twice upper tail probability corresponding to $|z|$

* significance level $\alpha$ and conclusion

* use this test when counts of successes and failures are each 5 or more in both samples
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: p_1 = p_2$ ($p_1 - p_2 = 0$)
  * pooled sample proportion $\hat{p}$:
    $$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$$
  * test statistics: 
    $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
  * $P$-value:
Test of significance:

* hypotheses: $H_0$ v.s $H_a$ / $H_0: p_1 = p_2$ ($p_1 - p_2 = 0$)

* pooled sample proportion $\hat{p}$:

$$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$$

* test statistics: $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

* $P$-value:

  * $H_a: p_1 - p_2 > 0$ — upper tail probability corresponding to $z$

  * $H_a: p_1 - p_2 \neq 0$ — twice upper tail probability corresponding to $|z|$ 

  * $\alpha$ and conclusion

* use this test when counts of successes and failures are each 5 or more in both samples
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: p_1 = p_2$ ($p_1 - p_2 = 0$)
  * pooled sample proportion $\hat{p}$:

  $\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$

  * test statistics: $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

  * $P$-value:
    * $H_a: p_1 - p_2 > 0$ — upper tail probability corresponding to $z$
    * $H_a: p_1 - p_2 < 0$ — lower tail probability corresponding to $z$

  * use this test when counts of successes and failures are each 5 or more in both samples
Test of significance:

* hypotheses: $H_0$ v.s $H_a / H_0: p_1 = p_2$ ($p_1 - p_2 = 0$)
* pooled sample proportion $\hat{p}$:

$$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$$

* test statistics: 

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

* $P$-value:

* $H_a: p_1 - p_2 > 0$ — upper tail probability corresponding to $z$
* $H_a: p_1 - p_2 < 0$ — lower tail probability corresponding to $z$
* $H_a: p_1 - p_2 \neq 0$ — twice upper tail probability corresponding to $|z|$
Test of significance:

* hypotheses: \( H_0 \) v.s \( H_a \) / \( H_0 : p_1 = p_2 \ (p_1 - p_2 = 0) \)

* pooled sample proportion \( \hat{p} \):

\[
\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}
\]

* test statistics: 

\[
z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}
\]

* \( P \)-value:

\* \( H_a : p_1 - p_2 > 0 \) — upper tail probability corresponding to \( z \)

\* \( H_a : p_1 - p_2 < 0 \) — lower tail probability corresponding to \( z \)

\* \( H_a : p_1 - p_2 \neq 0 \) — twice upper tail probability corresponding to \( |z| \)

* significance level \( \alpha \) and conclusion
• Test of significance:
  * hypotheses: $H_0$ v.s $H_a$ / $H_0: p_1 = p_2$ ($p_1 - p_2 = 0$)
  * pooled sample proportion $\hat{p}$:
    $$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}}$$
  * test statistics: $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
  * $P$-value:
    * $H_a: p_1 - p_2 > 0$ — upper tail probability corresponding to $z$
    * $H_a: p_1 - p_2 < 0$ — lower tail probability corresponding to $z$
    * $H_a: p_1 - p_2 \neq 0$ — twice upper tail probability corresponding to $|z|$
  * significance level $\alpha$ and conclusion
  * use this test when counts of successes and failures are each 5 or more in both samples
Variables - Categorical v.s. Quantitative
• Variables - Categorical v.s. Quantitative
• Graphs for distributional information: Pie chart, Bar graph, Histogram, Stemplot, Timeplot, Boxplot
• Variables - Categorical v.s. Quantitative
• Graphs for distributional information: Pie chart, Bar graph, Histogram, Stemplot, Timeplot, Boxplot
• Overall pattern of the graph: Symmetric/Skewed, Center, Spread, Outlier, Trend
• Measure of center: Mean/Median
• Measure of center: Mean/Median
• Measure of variability: Quartiles ($Q_1$, $Q_2$, $Q_3$), Range, IQR, 1.5×IQR rule, Outlier, Variance, Standard deviation
- Measure of center: Mean/Median
- Measure of variability: Quartiles \((Q_1, Q_2, Q_3)\), Range, IQR, \(1.5 \times \text{IQR}\) rule, Outlier, Variance, Standard deviation
- Five-number summary, Boxplot
• Density curve
• Density curve
• Normal distributions / Normal curves
• Density curve
• Normal distributions / Normal curves
• z-score, Standard normal distribution
• Density curve
• Normal distributions / Normal curves
• $z$-score, Standard normal distribution
• 68 – 95 – 99.7 rule, Probabilities for normal distribution
• Explanatory variable / Response variable
• Explanatory variable / Response variable
• Scatterplot: Direction (Positive / Negative), Form (Linear / Nonlinear), Strength, Outlier
• Explanatory variable / Response variable
• Scatterplot: Direction (Positive / Negative), Form (Linear / Nonlinear), Strength, Outlier
• Correlation
• Linear regression: \( \hat{y} = a + bx \); Slope \( b \), Intercept \( a \), Predication
• Linear regression: \( \hat{y} = a + bx; \) Slope \( b \), Intercept \( a \), Predication

• Correlation and regression, \( r^2 \), Residual
• Linear regression: \( \hat{y} = a + bx \); Slope \( b \), Intercept \( a \), Predication

• Correlation and regression, \( r^2 \), Residual

• Cautions for regression: Influential observations, Extrapolation, Lurking variables
• Sample / Population
• Sample / Population
• Random sampling design: Simple random sample (SRS), Stratified random sample, Multistage sample
• Sample / Population
• Random sampling design: Simple random sample (SRS), Stratified random sample, Multistage sample
• Bad samples: Voluntary response sample, Convenience sample
• Observational studies & Experimental studies (experiments)
• Observational studies & Experimental studies (experiments)
• Treatments / Factors
• Observational studies & Experimental studies (experiments)
• Treatments / Factors
• Design of experiments:
• Observational studies & Experimental studies (experiments)
• Treatments / Factors
• Design of experiments:
  control (comparison, placebo)
• Observational studies & Experimental studies (experiments)
• Treatments / Factors
• Design of experiments:
  control (comparison, placebo)
  randomization (table of random digits, double-blind)
• Observational studies & Experimental studies (experiments)
• Treatments / Factors
• Design of experiments:
  control (comparison, placebo)
  randomization (table of random digits, double-blind)
  matched pairs design / Block design
• Probability: Sample space \((S)\) & Events

• Rules for probability model:
  1. For any event \(A\), \(0 \leq P(A) \leq 1\)
  2. For sample space \(S\), \(P(S) = 1\)
  3. If two events \(A\) and \(B\) are disjoint, then \(P(A \text{ or } B) = P(A) + P(B)\)
  4. For any event \(A\), \(P(A \text{ does not occur}) = 1 - P(A)\)

• Discrete probability models / Continuous probability models
• Random variables / Distributions
• Probability: Sample space \((S)\) & Events
• Rules for probability model:
• Probability: Sample space $(S)$ & Events
• Rules for probability model:
  1. for any event $A$, $0 \leq P(A) \leq 1$
• Probability: Sample space \((S)\) & Events
• Rules for probability model:
  1. for any event \(A\), \(0 \leq P(A) \leq 1\)
  2. for sample space \(S\), \(P(S) = 1\)
- Probability: Sample space \((S)\) & Events
- Rules for probability model:
  1. for any event \(A\), \(0 \leq P(A) \leq 1\)
  2. for sample space \(S\), \(P(S) = 1\)
  3. if two events \(A\) and \(B\) are disjoint, then
     \[ P(A \text{ or } B) = P(A) + P(B) \]
• Probability: Sample space \((S)\) & Events
• Rules for probability model:
  1. for any event \(A\), \(0 \leq P(A) \leq 1\)
  2. for sample space \(S\), \(P(S) = 1\)
  3. if two events \(A\) and \(B\) are disjoint, then
     \[P(A \text{ or } B) = P(A) + P(B)\]
  4. for any event \(A\), \(P(A \text{ does not occur}) = 1 - P(A)\)
• Probability: Sample space ($S$) & Events
• Rules for probability model:
  1. for any event $A$, $0 \leq P(A) \leq 1$
  2. for sample space $S$, $P(S) = 1$
  3. if two events $A$ and $B$ are disjoint, then
     $P(A \text{ or } B) = P(A) + P(B)$
  4. for any event $A$, $P(A \text{ does not occur}) = 1 - P(A)$
• Discrete probability models / Continuous probability models
Probability: Sample space ($S$) & Events

Rules for probability model:
1. for any event $A$, $0 \leq P(A) \leq 1$
2. for sample space $S$, $P(S) = 1$
3. if two events $A$ and $B$ are disjoint, then $P(A \text{ or } B) = P(A) + P(B)$
4. for any event $A$, $P(A \text{ does not occur}) = 1 - P(A)$

Discrete probability models / Continuous probability models
Random variables / Distributions
• Population / Sample; Parameters / Statistics
  \( \mu / \bar{x}, \sigma / s, p / \hat{p} \)
• Population / Sample; Parameters / Statistics
  \[ \mu \ / \ \bar{x}, \ \sigma \ / \ s, \ \rho \ / \ \hat{p} \]
• Statistics are random variables
• Population / Sample; Parameters / Statistics
  \[ \mu / \bar{x}, \sigma / s, p / \hat{p} \]

• Statistics are random variables

• Sampling distribution of the sample mean \( \bar{x} \) for an SRS:
• Population / Sample; Parameters / Statistics
  \(\mu / \bar{x}, \sigma / s, p / \hat{p}\)
• Statistics are random variables
• Sampling distribution of the sample mean \(\bar{x}\) for an SRS:
  * mean of \(\bar{x}\) equals the population mean \(\mu\)
• Population / Sample; Parameters / Statistics
  \( \mu / \bar{x}, \sigma / s, p / \hat{p} \)

• Statistics are random variables

• Sampling distribution of the sample mean \( \bar{x} \) for an SRS:
  * mean of \( \bar{x} \) equals the population mean \( \mu \)
  * standard deviation of \( \bar{x} \) equals \( \frac{\sigma}{\sqrt{n}} \), where \( \sigma \) is the population standard deviation and \( n \) is the sample size
• Population / Sample; Parameters / Statistics
  $\mu / \bar{x}, \sigma / s, p / \hat{p}$

• Statistics are random variables

• Sampling distribution of the sample mean $\bar{x}$ for an SRS:
  * mean of $\bar{x}$ equals the population mean $\mu$
  * standard deviation of $\bar{x}$ equals $\frac{\sigma}{\sqrt{n}}$, where $\sigma$ is the population standard deviation and $n$ is the sample size
  * if the population has a normal distribution, then $\bar{x} \sim N(\mu, \sigma/\sqrt{n})$
• Population / Sample; Parameters / Statistics
  \( \mu / \bar{x}, \sigma / s, p / \hat{p} \)
• Statistics are random variables
• Sampling distribution of the sample mean \( \bar{x} \) for an SRS:
  * mean of \( \bar{x} \) equals the population mean \( \mu \)
  * standard deviation of \( \bar{x} \) equals \( \frac{\sigma}{\sqrt{n}} \), where \( \sigma \) is the population standard deviation and \( n \) is the sample size
  * if the population has a normal distribution, then \( \bar{x} \sim N(\mu, \sigma/\sqrt{n}) \)
  * **central limit theorem**: if the sample size is large (\( n \geq 30 \)), then \( \bar{x} \) is approximately normal, i.e. \( \bar{x} \approx \bar{x} \sim N(\mu, \sigma/\sqrt{n}) \)