Statistical Inference: Tests of Hypotheses

The population follows some distribution with unknown parameter(s). The main inference is to testing hypotheses about the parameter(s).

1. Proportion $p$: one group and difference between two groups.
2. mean $\mu$: one group, different between two groups, paired data (known and unknown variance)
3. variance $\sigma^2$: one group, ratio between two variances.

Duality between CIs and HT. Power and sample size.
Terms about hypotheses testing

1. Null and alternative hypotheses ($H_0$ vs $H_1$)
   
   $H_0 : \theta = \theta_0$ (no difference) against $H_1 : \theta > \theta_0$
   
   (Research H)

2. Test statistic

3. Rejection region (critical region) ($R$) and acceptance region

4. Significance level of the test $\alpha$

5. Type I and Type II error probabilities.

6. $p$-value

7. power
Test for proportion: Example (QC) p344

If the new method is better? Defective rate was 0.06, and we want to know does the new method decrease the defective rate?

Produce 200 circuits and check the number of defective products among these 200. Let $Y$ be the number of these 200 circuits that fail.
Test for proportion: Example (QC) p344

If the new method is better? Defective rate was 0.06, and we want to know does the new method decrease the defective rate?

Produce 200 circuits and check the number of defective products among these 200. Let $Y$ be the number of these 200 circuits that fail.

Decision rule: If $Y \leq 7$ or $Y/n \leq 0.035$, we could believe that the new procedure is an improvement. What is the significance level of the test? Will the engineers be happy about this rule is the new defective rate is 0.03.
The significance level of the test

\[ H_0 : p = 0.06 \text{ vs } H_1 : p < 0.06 \]
The significance level of the test

\[ H_0 : p = 0.06 \text{ vs } H_1 : p < 0.06 \]

Decision Rule: \( R : \{ y : y \leq 7 \text{ in 200 products} \} \)
The significance level of the test

\[ H_0 : p = 0.06 \text{ vs } H_1 : p < 0.06 \]

Decision Rule: \( R : \{ y : y \leq 7 \text{ in 200 products} \} \) If \( H_0 \) is true, then \( Y \sim Bin(200, 0.06) \)

The significance level of the test (Type I error rate)

\[
\alpha = P(T(x) \in R | H_0 : p = 0.06) \\
= P(y \leq 7 | H_0 : p = 0.06) \\
= \sum_{y=0}^{7} \binom{200}{y} (0.06)^y (0.94)^{200-y} = 0.0829
\]

\( Y \sim Bin(200, 0.06) \rightarrow Pos(12) \) so

\[
\alpha \approx \sum_{y=0}^{7} \frac{e^{-12}(12)^y}{y!} = 0.0895
\]
The probability of Type II error when $p = 0.03$

Decision Rule: $R : \{ y : y \leq 7 \text{ in 200 products} \}$ for testing

$H_0 : p = 0.06 \text{ vs } H_1 : p = 0.03$

$$\beta = P(\text{Retain a false } H_0 \text{ i.e. } p = 0.03) = P(Y > 7 | p = 0.03)$$

$$= \sum_{y=8}^{200} \binom{200}{y} (0.03)^y (0.97)^{200-y} = 0.254$$

Or use possion approximation $P(\text{Poi}(6) > 7) = 0.256$
Power when $p = 0.03$

power $= 1 - \beta = 0.744$.

The engineer and the statistician probably are not too pleased with this "decision rule" because the power of the test is quite low.
Test for proportion

one sided test and two-sided test.

\[ H_0 : p = p_0 \text{ vs } H_a : p > p_0 \]
\[ H_0 : p = p_0 \text{ vs } H_a : p < p_0 \]
\[ H_0 : p = p_0 \text{ vs } H_a : p \neq p_0 \]

Approximate Binominal distribution to a Normal distribution

\[ Z = \frac{(Y/n) - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim H_0 \mathcal{N}(0, 1) \]

if \( z_{\text{obs}} \geq z_\alpha \) Reject \( H_0 \) at \( \alpha \)

if \( z_{\text{obs}} \leq -z_\alpha \) Reject \( H_0 \) at \( \alpha \)

if \( |z_{\text{obs}}| \geq z_{\alpha/2} \) Reject \( H_0 \) at \( \alpha \)
Example: 7.1-1 [fair dice]

We want to know if dice favor a 6 more?

What is your null and alternative hypotheses?

Roll these dice $n = 8000$ times and observed $y = 1389$ successes (6s). What is your conclusion at a significance level of $\alpha = 0.05$?
Example: 7.1-1 [fair dice]

1. \( H_0 : p = 1/6 \) vs \( H_a : p > 1/6 \)
2. The significance level of the test \( \alpha = 0.05 \)
3. The Test statistic under \( H_0 \) is
   \[
   Z = \frac{Y/n - 1/6}{\sqrt{(1/6)(6/6)/n}} \sim H_0 \mathcal{N}(0, 1).
   \]
4. Decision rule: the rejection region is
   \[
   \text{If } z_{\text{obs}} \geq z_{0.05} = 1.645 \text{ reject } H_0
   \]
5. Data: \( n = 8000 \) and \( y = 1389 \) successes and we calculate the value of the test statistic \( z_{\text{obs}} = 1.670 \).
6. Conclusion: \( z_{\text{obs}} = 1.67 > 1.645 \), we reject the null hypothesis and conclude that the dice favor a 6 more than a fair dice would at \( \alpha = 0.05 \).
**P-value**

\[ H_0 : p = 1/6 \text{ vs } H_1 : p > 1/6 \]

\[ z_{\text{obs}} = 1.67 \]

The test statistic is more extreme than the observed value toward the direction of \( H_1 \).

\[ p\text{-value} = \Pr(Z \geq 1.67) = 1 - \text{pnorm}(1.67) = 0.0475. \]
\( p \)-value

\[
P(Z \geq z_{\text{obs}}), \ H_0: p = 1/6 \ vs \ H_1: p > 1/6
\]

\[
P(Z \leq z_{\text{obs}}), \ H_0: p = 1/6 \ vs \ H_1: p < 1/6
\]

\[
P(|Z| \geq |z_{\text{obs}}|), \ H_0: p = 1/6 \ vs \ H_1: p \neq 1/6
\]

Note:

change \( \alpha = 0.01 \)

one sided test and two-sided test.
Tests for difference between two proportions

\( H_0 : p_1 = p_2 \) vs \( H_a : p_1 > p_2 \)

\( H_0 : p_1 = p_2 \) vs \( H_a : p_1 < p_2 \)

\( H_0 : p_1 = p_2 \) vs \( H_a : p \neq p_0 \)

Let \( Y_i \) be the number of successes in \( n_i \) trials, \( i = 1, 2 \).

We have \( Y_1 \sim Bin(n_1, p_1) \) and \( Y_2 \sim Bin(n_2, p_2) \)

Estimate \( p_1 \) and \( p_2 \) by \( \hat{p}_1 = Y_1 / n_1 \), \( \hat{p}_2 = Y_2 / n_2 \). and we have

\[ \hat{p}_1 \sim \mathcal{N}(p_1, \frac{p_1(1 - p_1)}{n_1}) \]
\[ \hat{p}_2 \sim \mathcal{N}(p_2, \frac{p_2(1 - p_2)}{n_2}) \]
Tests for difference between two proportions

\[ \hat{p}_1 \sim \mathcal{N}(p_1, \frac{p_1(1 - p_1)}{n_1}), \quad \hat{p}_2 \sim \mathcal{N}(p_2, \frac{p_2(1 - p_2)}{n_2}) \]

The standard error of \( \hat{p}_1 \), \( \text{SE}(\hat{p}_1) \), is estimated by \( \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1}} \)

The standard error for \( \hat{p}_2 \), \( \text{SE}(\hat{p}_2) \), is estimated by \( \sqrt{\frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \)

Also we have \( \hat{p}_1 - \hat{p}_2 \sim \mathcal{N}(p_1 - p_2, \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}) \)

Test statistic

\[ Z = \frac{(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}) - (p_1 - p_2)}{\sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}} \]
Tests for difference between two proportions

under $H_0$, we have $p_1 = p_2 = p$, and we use the pooled estimate

$$\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2}$$

to estimate $p$

The Test statistic is

$$Z = \frac{\frac{Y_1}{n_1} - \frac{Y_2}{n_2} - 0}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}} \sim H_0 \mathcal{N}(0, 1)$$

where

$$\sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

is the estimated standard error for $\hat{p}_1 - \hat{p}_2$, $\text{SE}(\hat{p}_1 - \hat{p}_2)$. 
Tests for difference between two proportions

Alternatively, we could use the combined variance

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

to estimate the standard error for \( \hat{p}_1 - \hat{p}_2 \), and use the following test statistic

\[
Z = \frac{\left( \frac{Y_1}{n_1} - \frac{Y_2}{n_2} \right) - 0}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}
\]
Comparing two groups

Suppose the true proportions of students who like drinking coffee are $p_1$ for male students and $p_2$ for female students. Test the hypothesis $H_1 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$.

Data

$n_1 = 75, y_1 = 5, n_2 = 31, y_2 = 6.$

what is your conclusion at a significance level of $\alpha = 0.05$?
The duality between CIs and HT

Hypothesis testing: Is my parameter $\theta(\in \Omega)$ equal to $\theta_0$?
Confidence interval: the length for my parameter $\theta$?

For $H_0 : \rho = \rho_0$ vs $H_1 : \rho \neq \rho_0$ at $\alpha = 0.05$, if we reject $H_0$ at $\alpha$.

$$z_{\text{obs}} = \frac{|\hat{\rho} - \rho_0|}{\sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}}} \geq z_{\alpha/2}$$

This is equivalent to the statement that

$$\rho_0 \notin [\hat{\rho} \pm z_{\alpha/2} \sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}}]$$
The duality between CIs and HT

Example: $H_0 : p = 1/6$ against $H_1 : p \neq 1/6$.
Data: $n = 8000$ wiht $y = 1389$ successes.

We have $z_{\text{obs}} = 1.643 < z_{0.025} = 1.96$ don’t reject $H_0$

The 95% confidence intervals for $p$ given the data is

$$[\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p})/n}] = [0.1653, 0.1819]$$

The interval contains $p_0 = 1/6 = 0.1667$. 

The duality between CIs and HT

Fail to reject a null hypothesis $H_0 : \theta = \theta_0$ at the 5% significance level is $\theta_0$ equivalent to saying that the value $\theta_0$ lies in a 95% confidence interval for $\theta$.

Rejecting a null hypothesis $H_0 : \theta = \theta_0$ at the 5% significance level is equivalent to saying that the value $\theta_0$ is not included in a 95% confidence interval for $\theta$. 
The duality between CIs and HT

For \( H_0 : p = p_0 \) vs \( H_1 : p < p_0 \), if we reject \( H_0 \) at \( \alpha = 0.05 \), i.e.,

\[
    z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \geq 1.645
\]

This is equivalent to the statement that

\[
p_0 \notin [0, \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]
\]
example


text

\[ H_0 : p = 1/6 \text{ vs } H_1 : p < 1/6 \text{ data: } n = 8000 \text{ wiht } y = 1389 \text{ successes.} \]

we have \( z_{\text{obs}} = 1.643 < z_{0.05} = 1.645 \). Thus, we don’t reject \( H_0 \) at \( \alpha = 0.05 \).

The 95\% one-sided confidence intervals for \( p \) is

\[
[0, \frac{1389}{2000} + 1.645 \sqrt{\left(\frac{1389}{8000}\right) \left(1 - \frac{1389}{8000}\right)}] = [0, 0.185]
\]

We found that the interval contains \( p_0 = 1/6 = 0.1667 \).
The duality between CIs and HT

Suppose we have a test for testing the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_0 : \theta \neq \theta_0$, and that the test rejects iff $\{ T(X) \in R(\theta) \}$ at level $\alpha$, i.e.,

$$P(T(X) \in R(\theta)) = \alpha$$

Then, for a specified $\theta = \theta_0$, we have $P(T(X) \in R(\theta_0)) = \alpha$ i.e.,

$$P(T(X) \in A(\theta_0)) = 1 - \alpha$$

where $\{ T(X) \in A(\theta_0) \}$ is the acceptance region of the test. Then the set

$$\{ \theta : T(X) \in A(\theta) \}$$

is a $100(1 - \alpha)\%$ confidence interval for $\theta$. 
Test for a population mean

\[ H_0 : \mu = \mu_0 \ vs \ H_a : \mu > \mu_0 \]

\[ H_0 : \mu = \mu_0 \ vs \ H_a : \mu < \mu_0 \]

\[ H_0 : \mu = \mu_0 \ vs \ H_a : \mu \neq \mu_0 \]
Test for mean

The Test statistic is for a population mean for Normal distribution with known variance

\[ Z = \frac{X - \mu_0}{\sigma/\sqrt{n}}. \]

under \( H_0 \):

\[ Z \sim \mathcal{N}(0, 1) \]

\( H_0 : \mu = \mu_0 \) vs \( H_a : \mu > \mu_0 \), if \( z_{\text{obs}} \geq z_{\alpha} \) reject \( H_0 \) at \( \alpha \)

\( H_0 : \mu = \mu_0 \) vs \( H_a : \mu < \mu_0 \), if \( z_{\text{obs}} \leq -z_{\alpha} \) reject \( H_0 \) at \( \alpha \)

\( H_0 : \mu = \mu_0 \) vs \( H_a : \mu \neq \mu_0 \), if \( |z_{\text{obs}}| \geq z_{\alpha/2} \) reject \( H_0 \) at \( \alpha \)
Example: Test for a mean (known variance)

\[ X_1, \cdots, X_{16} \sim \mathcal{N}(\mu, 36) \]

1. \( H_0 : \mu = 50 \) vs \( H_1 : \mu \neq 50 \)
2. given \( \alpha = 0.05 \).
3. Test statistic \( Z = \frac{\bar{X} - 50}{\sqrt{36/16}} \sim H_0 \mathcal{N}(0, 1) \)
4. Rule: If \( z_{\text{obs}} > z_{\alpha/2} = 1.96 \) then we reject \( H_0 \) at \( \alpha \)
5. Data: \( \bar{x} = 52.75 \) \( z_{\text{obs}} = \frac{52.75 - 50}{6/\sqrt{16}} = 1.83 \)
6. Decision:
   a because \( z_{\text{obs}} = 1.83 < 1.96 \). We conclude that the data does not provide enough evidence to reject the null hypothesis \( \mu = 50 \) at \( \alpha = 0.05 \).
   b Or because \( p\text{-value} = \mathbb{P}(|Z| > 1.83) = 0.0668 > 0.05 \), don’t reject the null hypothesis.
The Test statistic is for a population mean with known variance

\[ Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}. \]

- \( H_0 : \mu = \mu_0 \) vs \( H_a : \mu > \mu_0 \)
  - \( p\)-value = \( P(Z \geq z_{obs}) \).

- \( H_0 : \mu = \mu_0 \) vs \( H_a : \mu < \mu_0 \)
  - \( p\)-value = \( P(Z \leq z_{obs}) \).

- \( H_0 : \mu = \mu_0 \) vs \( H_a : \mu \neq \mu_0 \)
  - \( p\)-value = \( P(|Z| \geq |z_{obs}|) = 2P(Z \geq |z_{obs}|) \).
Population is normal and the variance is unknown

\[ X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2) \]

The Test statistic is

\[ T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}. \]

Under \( H_0 \)

\[ T \sim t(n - 1) \]
Population is normal and the variance is unknown

1. $H_0 : \mu = \mu_0 \text{ vs } H_a : \mu > \mu_0$
2. $H_0 : \mu = \mu_0 \text{ vs } H_a : \mu < \mu_0$
3. $H_0 : \mu = \mu_0 \text{ vs } H_a : \mu \neq \mu_0$
7.2-3. Growth of a tumor in a mouse:
$n = 9, \bar{x} = 4.3, s = 1.2$

1. $H_0 : \mu = 4$ vs $H_a : \mu \neq 4$
2. Given a significant level of $\alpha = 0.10$ and $n = 9$
3. $T = \frac{\bar{x} - 4}{s/\sqrt{9}} \sim H_0 \ t(8)$
4. The critical region is: $|t_{\text{obs}}| \geq t_{0.05}(8) = 1.86$
5. Data: $n = 9, \bar{x} = 4.3, s = 1.2$ we have
   $$t_{\text{obs}} = \frac{|4.3 - 4|}{1.2/\sqrt{9}} = 0.75$$
6. Decision:
   a Because $|t_{\text{obs}}| < t_{0.05}(8) = 1.86$. Don’t reject the $H_0$ at the $\alpha = 10\%$ significant level.
   b $P(|T| \geq 0.75) = 2P(T \geq 0.75) = 0.475$. Because $p = 0.475 > 0.1$. Don’t reject the $H_0$ at the $\alpha = 10\%$ significant level.

**Remark:** Finding lack of significance does not
Example 7.2-4. If the oxygen-consuming power of their wastes is reduced. \( n = 25, \bar{x} = 308.8, s = 115.15 \)

1. \( H_0 : \mu = 500 \) vs \( H_a : \mu < 500 \)
2. Given a significant level of \( \alpha = 0.01 \)
3. \( t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \)
4. The critical region is : \( t_{\text{obs}} \leq -t_{0.01}(24) = -2.492 \)
5. Data: \( \bar{x} = 308.8, s = 115.15 \) we have \( t_{\text{obs}} = -8.3 \)
6. Decision: \( t_{\text{obs}} < -t_{0.01}(24) \). Reject the \( H_0 \) at the \( \alpha = 1\% \) significant level and accept that the improvement has been made.
Compare with the one-sided 99% confidence interval for \( \mu \).

\[
[0, 308.8 + 2.492(115.15/\sqrt{25})] = [0, 366.2]
\]

500 \( \not\in \) CI rejection \( H_0 \)
Paired $t$-test: one population, before and after

$X$ and $Y$ are measurements on the same person, they are dependent.

Blood pressure of an individual before and after exercise, they are dependent.
Paired $t$-test

7.2-5: Let $W$ be the difference in time “before” and “after” some program $n = 24$

1. $H_0 : \mu_W = 0$ vs $H_a : \mu_W > 0$

2. Given a significant level of $\alpha = 0.05$ and $n = 24$

3. Test statistic $T = \frac{\overline{W} - 0}{S_w / \sqrt{24}} \sim H_0 \ t(23)$

4. The critical region is : $T \geq t_{0.05}(23) = 1.714$

5. Data: $\overline{w} = 0.079$, $s = 0.255 \rightarrow T_{obs} = 1.518$

6. Decision: $t < t^*$. Don’t reject the $H_0$ at the $\alpha = 1\%$ significant level. There is no enough evidence to show that the program makes girls run faster.
Tests about equality of two means

equality of two normal distribution: variance known

\[ H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 \neq \mu_2 \]
equality of means of two normal distribution
variance unknown but assume common variance

Two-sample $t$-test

The estimate of common variance $S_p^2$
equality of means of two normal distribution

1. $H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 > \mu_2$

2. $H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 < \mu_2$

3. $H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 \neq \mu_2$
t-test

Growth of pea stem to two different levels of the hormone indoleacetic acid (IAA)

\[
x = 0.8 \ 1.8 \ 1.0 \ 0.1 \ 0.9 \ 1.7 \ 1.0 \ 1.4 \ 0.9 \ 1.2 \ 0.5
y = 1.0 \ 0.8 \ 1.6 \ 2.6 \ 1.3 \ 1.1 \ 2.4 \ 1.8 \ 2.5 \ 1.4
1.9 \ 2.0 \ 1.2
\]

\[
n = 11 \ , \ m = 13
\]

\[\bar{x} = 1.02727, \bar{y} = 1.66154 \text{ and } s_x^2 = 0.244, s_y^2 = 0.353 \text{ we have}
\]

\[s_p^2 = 0.3033 \text{ and}
\]

\[t = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2(1/n + 1/m)}} = -2.8112
\]

\[-t_{0.05(22)} = -1.7171, \text{ p-value } = P(T(22) \leq -2.8112) = 0.0051\]
One-sided t-test

\[ H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 < \mu_2 \]

t.test(x, y, var.equal = T, alternative = "less")

Two Sample t-test

data:  x and y

\[ t = -2.8112, \text{ df } = 22, \text{ p-value } = 0.005086 \]

alternative hypothesis: true difference in means is less than 0
two sided t-test

\[ H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 \neq \mu_2 \]
\[ t_{0.025}(22) = 2.0739 \]

p-value = 0.0102

t.test(x1, x2, var.equal = TRUE)

Two Sample t-test

data:  x and y
\( t = -2.8112, \text{ df } = 22, \text{ p-value } = 0.01017 \)
alternative hypothesis: true difference in means is not equal to 0
Welch’s t test

equality of two normal distribution: variance unknown and sample sizes are small

\begin{align*}
\text{Welch’s } t\text{-test} \\
H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 < \mu_2
\end{align*}

t.test(x,y)

Welch Two Sample t-test

data:  x and y
\begin{align*}
t & = -2.856, \text{ df } = 21.998, \text{ p-value } = 0.004593 \\
\text{alternative hypothesis: true difference in means is less than 0}
\end{align*}
Test about a population variance
Test about a population variance

Example

IQ score

\( H_0 : \sigma^2 = 100 \) vs \( H_a : \sigma^2 \neq 100 \)

under \( H_0 \) \( \frac{(n-1)S^2}{100} \sim \chi^2(n-1) \)
Test about the equality of variance
power and sample size

For a given $\alpha$, and $\beta$, we can determine the proper rejection region $c$ and $n$. 