1 Confidence intervals

measure	notes	confidence interval
mean	n is large or $X \sim N(\mu, \sigma^2)$ and σ known	$\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}},\ \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$
mean	$X \sim N(\mu, \sigma^2)$ and σ unknown	$\left[\bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}},\ \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$
proportion		$\left[\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$
proportion		$\left[\bar{X} - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}},\ \bar{X} + \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}\right]$
variance		$\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right]$

1.1 Code

Computing distributions

Given a confidence interval C or a significance level α ($C = 1 - \alpha$) and n samples. Let $a = 1 - \frac{\alpha}{2}$.
Then $z_{\frac{\alpha}{2}}$, $t_{\frac{\alpha}{2}}$, and $\chi^2_{\frac{\alpha}{2}}$ are computed by

```
qnorm(a)
qt(a, df = n-1)
qchisq(a, df = n-1)
```

Confidence interval of the mean when σ is unknown

Given some data, and confidence level $C = 1 - \alpha$.

```
data <- c(...)
t.test(data, conf.level = C)
```

Confidence interval of proportions

Given x successes out of n samples and a confidence interval C or a significance level α ($C = 1 - \alpha$). H is any number less than n not equal to x.

```
prop.test(x, n, H, conf.level = C)
```

2 Hypothesis testing

Given a measure of a sample with a certain set of hypotheses and a significance level α. Let w be the observed test statistic and T the appropriate statistical test with cumulative distribution function $F(x)$. If w is within the rejection region OR the p-value $< \alpha$, reject H_0.

test	hypotheses	rejection region	p-value
two-sided	$H_0 : a = b$ $H_1 : a \neq b$	$\lvert W \rvert > T_{\frac{\alpha}{2}}$	$2 \cdot F(-\lvert w \rvert)$
left-tail	$H_0 : a \geq b$ $H_1 : a < b$	$W < -T_\alpha$	$F(w)$
right-tail	$H_0 : a \leq b$ $H_1 : a > b$	$W > T_\alpha$	$1 - F(w)$

measure	notes	formulas	test statistic	method
one-sample mean	n is large or $X \sim N(\mu, \sigma^2)$ and σ known	S is σ or sample s.d.	$\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$	z-test
one-sample mean	$X \sim N(\mu, \sigma^2)$ and σ unknown	S is sample s.d.	$\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$	t-test, $n - 1$
one-sample proportion	$np_0 > 5$ and $np_0(1 - p_0) > 5$		$\dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$	z-test
two-sample means	independent, n is large		$\dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	z-test
two-sample means	independent, n is small	$s_p = \sqrt{\dfrac{S_1^2(n_1 - 1) + S_2^2(n_2 - 2)}{n_1 + n_2 - 1}}$	$\dfrac{\bar{X} - \bar{Y}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t-test, $n_1 + n_2 - 2$
paired samples		$D_i = X_i - Y_i$	$\dfrac{\bar{D}}{S_D/\sqrt{n}}$	t-test, $n - 1$
two-sample proportion	$H_0 : p_1 = p_2 = p_0$	$\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$	$\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	z-test

2.1 Code

Computing distributions

Given a probability density function, find the probability at a certain point a. The normal and t distributions (with b degrees of freedom) are as follows.

```
pnorm(a)
pt(a, df = b)
```

One-sample t-test

Given data with σ unknown, hypothetical mean μ_0 and confidence level C.

```
data <- c(...)
t.test(data, mu = m0, alternative = '...', conf.level = C)
```

Two-sample t-test

Given two sets of data and a confidence level C.

```
xdata <- c(...)
ydata <- c(...)
t.test(xdata, ydata, alternative = '...', paired = '...', var.equal = '...', conf.level = C)
```

One-sample proportion test

Given a proportion of x_0 successes out of n_0 trials and a hypothetical proportion p_0.

```
prop.test(x = x0, n = n0, p = p0, alternative = '...', correct = FALSE)
```

2.1.1 Two-sample proportion test

Given proportions x_1 and x_2 out of n_1 and n_2 trials and a confidence level C.

```
prop.test(x = c(x1,x2), n = c(n1,n2), alternative = '...', conf.level = C, correct = FALSE)
```

Options

```
alternative: corresponds to the alternative hypothesis
    'two.sided': two sided test
    'less': left-tailed test
    'greater': right-tailed test
```

```
paired: boolean whether or not a paired t-test should be performed
    TRUE: samples are dependent
    FALSE: samples are independent

var.equal: boolean whether or not the variances should be treated as equal
    TRUE: pooled variance is used
    FALSE: Welch approximation is used
```