Chapter 1

Confidence Intervals

A statistic is a **point estimator** used to estimate an unknown population parameter. A **point estimate** is a statistic's value, and is referred to as such because it is a single point. As such, it is almost never accurate.
A point estimate can be made made more reliable by either increasing the sample size or using a more accurate sampling procedure.
The **standard error** $\boldsymbol{s}$ (with the appropriate subscript denoting its statistic) of a statistic is the point estimate of the standard deviation of the sampling distribution.
A **confidence interval** provides an interval of plausible values for an unknown parameter based on sample data. It is equal to the point estimate plus or minus the **margin of error** (**ME**).

$$\text{confidence interval} = \text{point estimate} \pm \text{margin of error}$$

The probability that a confidence interval contains the true parameter value is the confidence interval's **confidence level** ($\boldsymbol{C}$). The margin of error describes the maximum deviation of the estimate from the parameter. It is the product of the critical value and the standard error of the statistic.

$$\text{ME} = \text{critical value} \times \text{standard error}$$

The **critical value** is equal to the the number of standard deviations from the mean within which the probability of a random variable falling is equal to C.

$$P(-\text{critical value} < \text{standardized test statistic} < \text{critical value}) = C$$

1.1 Confidence Intervals about Proportions

Confidence Intervals about Differences in Proportions

1.2 Confidence Intervals about Means

In order for Normality to be verified, the central limit theorem can be used, necessitating that n be at least 30, or a modified box plot can be created with the data and observed to be symmetrical without outliers. When the standard deviation of X (not of the sampling distribution of $\bar{x}$) is known, a confidence interval about $\bar{x}$ can be constructed using the templates for confidence intervals and margin of error, simply using the sampling distribution's standard deviation rather than the statistic's standard error.

$$\text{confidence interval} = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

This is a **one-sample z-interval for a population mean**. When σ is not known, the standard deviation can be replaced by the standard error to calculate the **standard error of $\bar{x}$**.

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

The margin of error is appropriately changed:

$$\text{ME} = z^* \frac{s_x}{\sqrt{n}}$$

This results in the intervals being too small, though, and the confidence level decreases from what would be expected given z^*. The critical value can also change, though, becoming t^* rather than z^*, making the intervals longer and making the confidence level representative. The reason that t is used is that a ***t* distribution** is used rather than a Normal one. (t^* can still be interpreted in the same way as z^*, though (the number of standard errors from the point estimate).)
The specific t distribution used is specified by **degrees of freedom (df)**, which is 1 less than the sample size.[1]

$$\text{df} = n - 1$$

The confidence interval constructed about $\bar{x}$ using t^* and s_x is a **one-sample *t* interval for a population mean**.

$$\text{ME} = t^* \frac{s_x}{\sqrt{n}}$$

Because t^* is dependent on df, which is 1 less than the sample size, and s_x is dependent on the data, which has not been produced, so the sample size cannot be solved for given a confidence level and the margin of error, so z^* and σ, a value of s_x from a previous study are instead used.

$$\text{ME} \geq z^* \frac{\sigma}{\sqrt{n}}$$
$$n \geq \left(\frac{\text{ME}}{z^*}\right)^2$$

Confidence Intervals about Differences in Means

A confidence interval about a difference in means is a **two-sample *t* interval for a mean difference**. In order for it be constructed about, Normality and independence must be justified and both samples must be independent.
The center of the confidence interval is the difference between the sample means, denoted by a subscript **diff**, while its standard error is simply the square root of the sum of the variances of the standard errors of the individual statistics.

$$\bar{x}_{\text{diff}} = \overline{x_1 - x_2} = \bar{x}_1 - \bar{x}_2 \qquad \text{confidence interval} = \bar{x}_{\text{diff}} \pm s_{\bar{x}_1 - \bar{x}_2} = (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

[1] The density curve of a t distribution with degrees of freedom ν is defined (using the gamma function Γ or the beta function B) as such:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu}\,\mathrm{B}\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

As df approaches infinity, the t distribution approaches a normal curve (the tails approaching 0 more quickly), so t^* approaches z^*. This is because a greater sample size means that s_x will be closer to σ.

If both standard deviations are known, they are used along with z^*.

$$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

The degrees of freedom of a difference of means is the equal to the estimated variance divided by the sum of the estimated variances of each statistic divided by 1 less than their sample sizes.

$$\text{df} = \frac{s^4_{\bar{x}_1 - \bar{x}_2}}{\frac{s^4_{\bar{x}_1}}{n_1 - 1} + \frac{s^4_{\bar{x}_2}}{n_2 - 1}} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

When the data is *paired*, the response being collected from the same set of individuals independently, the difference can be treated as a single mean, as each sample difference is known, and a **paired *t* interval** can be created. s_{diff} can therefore be used as well.[2]

$$s_{\text{diff}} = s_{1-2} = \sqrt{s_1^2 + s_2^2} \qquad \text{confidence interval} = \bar{x}_{\text{diff}} \pm t^* \frac{s_{\text{diff}}}{\sqrt{n}}$$

[2]When $n_1 = n_2$, it can be verified that the values of $s_{\bar{x}_1 - \bar{x}_2}$ derived by s_{diff} and by s_1 and s_2 individually are the same.

$$\begin{aligned} s_{\text{diff}} &= \sqrt{s_1^2 + s_2^2} \\ s_{\bar{x}_1 - \bar{x}_2} &= \frac{s_{\text{diff}}}{\sqrt{n}} & &= \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}} \\ &= \sqrt{\frac{s_1^2 + s_2^2}{n}} & &= \sqrt{\frac{s_1^2 + s_2^2}{n}} \end{aligned}$$

Chapter 2

Significance Tests

A **significance test** is a procedure that uses observed data to test between two claims, often made regarding parameters, about hypotheses.

> In order for a significance test to be performed, randomness, independence, and Normality must be verified.

The **null hypothesis** (**H_0**) claims that the parameter is equal to a **null value**, what it was previously assumed to be, denoted by a subscript 0 on the parameter. It is often a statement of no change or difference.

$$H_0 : \text{parameter} = \text{null value}$$

The claim that is attempting to be supported is the **alternative hypothesis** (**H_a**). It can either be **one-sided**, claiming that the parameter is greater or less than the null value, or **two-sided**, claiming simply that the parameter is not equal to the null value.

$$H_a : \text{parameter} \gtrless \text{null value} \vee \text{parameter} \neq \text{null value}$$

A test's ***P*-value** is the probability of *significant* evidence being found that supports H_a that is at least as strong as that observed assuming that H_0 is true.

$$P\text{-value} = P(\text{statistic supports } H_a \mid H_0)$$

The smaller the P-value, the lower the chances of receiving evidence of the alternative. A small P-value therefore supports the H_a.
If the P-value is less than the **significance level α**, H_0 can be rejected and it can be concluded that there is convincing evidence for H_a. If the P-value is greater than or equal to α, H_0 cannot be rejected, and it can be concluded that there is not convincing evidence for H_a.
The P-value is calculated using the **standardized test statistic**, which is the number of standard deviations from the null value of the parameter.
For a null hypothesis to be significant is for a significance test to provide a P-value less than the significance level.

$$\text{standardized test statistic} = \frac{\text{statistic} - \text{null parameter}}{\text{standard error of statistic from null parameter}} = \text{sts}$$

The P-value is equal to the probability of z satisfying the H_a assuming that H_0 is true. It can therefore be calculated as such using a *cumulative distribution function*, so long as Normality and independence are justified.

$$P\text{-value} = \begin{cases} P(\text{parameter} \gtrless \text{null parameter}) = P(\mathrm{S} \gtrless \mathrm{s}) & H_a : \text{parameter} \gtrless \text{null parameter} \\ P(|\text{parameter}| < |\text{null parameter}|) = P(\mathrm{S} < -|\mathrm{s}|) + P(\mathrm{S} > |\mathrm{s}|) & H_a : \text{parameter} \neq \text{null parameter} \end{cases}$$

Conclusions should only ever be made regarding the rejection of H_0 and the convincing support of H_a. H_0 should never be supported and H_a should never be rejected.

When answering a question regarding a significance test, the following process can be followed:

1. State State the hypotheses to be tested and the significance level and define any parameters.

2. Plan Identify the appropriate methods of inference and verify its conditions.

3. Do State the sample statistic(s), calculate the standardized test statistic(s), and calculate the P-value.

4. Conclude Make a conclusion regarding the hypotheses within the problem's context.

When performing significance tests, two types of errors may occur:

- A **Type I error** occurs when H_0 is rejected despite H_a being false; the data provided convincing evidence for H_a despite it being incorrect.
- A **Type II error** occurs when H_0 is not rejected despite H_a being true; the data did not provide convincing evidence for H_a despite it being correct.

	H_a is false	H_a is true
H_0 is rejected	Type I error	Correct conclusion
H_0 is not rejected	Correct conclusion	Type II error

The probability of a Type I error occurring is equal to α.
As α increases, the probability of a Type I error increases but that of a Type II error decreases.
A confidence interval about a statistic (using the *standard error*) can be used in tandem with a sample statistic to provide a set of plausible values for the true parameter, should the alternative hypothesis be convincingly supported.
A two-sided test of of H_0 at significance level α usually provides the same conclusion as a confidence level of the complement of α.

$$[P(|\text{STS}| < |\text{sts}|) < \alpha] \approx [\text{null parameter} \in (\text{statistic} \pm \text{ME})]$$

A test's **power** is the probability of convincing evidence being found that convincingly supports H_a given a value for the parameter being tested. This is also is equal to the probability of avoiding a Type II error.

$$\text{power} = 1 - P(\text{Type II Error})^C = P(\text{statistic convincingly supports } H_a \mid H_a \text{ is true})$$

Power can be increased in three ways:

1. Increasing the sample size
   - A large sample means that more data is collected and more information is given regarding the true population parameter. This also increases n, which decreases the standard error of the statistic, reducing the value of the standardized test statistic and therefore the P-value, making it more likely to fall below α.
2. Increasing the significance level
   - Increasing the significance level increases the probability of H_0 being rejected when H_a is true, as the maximum P-value for H_0 to be rejected increases.

3. Increasing the **effect size**, the minimum difference between the null parameter value and the alternative parameter value for the change to matter

   - Increasing the size of the difference that needs to be detected makes that difference more likely to be detected, as larger differences are easier to detect.

2.1 Significance Tests about Proportions

In order for a significance test of $H_0 : p = p_0$ to be performed, the following must be verified:

- Randomness (Sampling or Assignment)

  Independence (10% for Samples)

- Normality (Large Counts)

To perform **1-proportion *z* test**, a significance test about one proportion, z must be calculated.

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

Significance Tests about Differences in Proportions

A **2-proportion *z* test** can be performed to compare the proportions for two populations is based on the difference between sample proportions. They hypotheses for these tests typically take the following forms:

$$H_0 : p_1 - p_2 = p_0 \qquad H_a : p_1 - p_2 \gtrless p_0 \qquad H_a : p_1 - p_2 \neq p_0$$

Typically, p_0 is 0, so these hypotheses can be rewritten.

$$H_0 : p_1 = p_2 \qquad H_a : p_1 \gtrless p_2 \qquad H_a : p_1 \neq p_2$$

A significance test first assumes that the null hypothesis $H_0 : p_1 = p_2$ is true. This common value is referred to as p.
The **combined sample proportion** is denoted $\hat{p}_C$ and is equal to the total successes divided by the total sample size, making it effectively a weighted average. It is the sample proportion that assumes that the parameter values are equal.

$$\hat{p}_C = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

The *Large Counts* condition must be met with $\hat{p}_C$.

$$n_1\hat{p}_C, n_1(1 - \hat{p}_C), n_2\hat{p}_C, n_2(1 - \hat{p}_C) \geq 10$$

For a significance test to be run about a difference of proportions, the randomness, independence (*10%*) (for each proportion), and *Large Counts* conditions must be met.

The *standardized test statistic* is the z-score calculated using the difference in proportions and its standard error assuming the mean to be 0 (H_0 to be true).

$$z = \frac{\hat{p}_1 - \hat{p}_2 - \mu_{\hat{p}_1 - \hat{p}_2}}{s_{\hat{p}_1 - \hat{p}_2}} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{n_1} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_2}}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_C\left(1 - \hat{p}_C\right)\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}}$$

2.2 Significance Tests about Means

To perform a significance test for a population mean, a **1-sample *t* test**, randomness, independence (*10%*), and Normality (*CLT* or distribution) must be verified.

The *standardized test statistic* for a significance test about a mean is t.

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

The *t-distribution* used to calculate the P-value uses *degrees of freedom* 1 less than the sample size.

$$\text{df} = n - 1$$

Minute, practically unimportant changes in in μ_0 can drastically shrink the P-value when the sample size is always large enough. A very large sample size results in the null hypothesis almost always being rejected. ***P*-hacking** takes advantage of this fact.

Significance Tests about Differences in Means

A **2-sample *t* test** about a difference in means compares the means of two populations based on samples. These tests' hypotheses take the following forms:

$$H_0 : \mu_1 - \mu_2 = \mu_0 \qquad H_a : \mu_1 - \mu_2 \gtrless \mu_0 \qquad H_a : \mu_1 - \mu_2 \neq \mu_0$$

Typically, the null difference is 0, so these hypotheses can be rewritten in the following forms:

$$H_0 : \mu_1 = \mu_2 \qquad H_a : \mu_1 \gtrless \mu_2 \qquad H_a : \mu_1 \neq \mu_2$$

In order for a significance test to be performed about a difference in means, the randomness, independence (*10%*), and normality (*CLT* or distribution) must be verified for *each* sample.

The *standardized test statistic* for a difference in means is as follows:

$$t = \frac{\bar{x}_{\text{diff}} - \mu_0}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

When the data is *paired*, a **1-sample *t* test for a difference in means** (a **paired *t* test for a difference in means**) can be performed:

$$t = \frac{\bar{x}_{\text{diff}} - \mu_0}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_{\text{diff}} - 0}{s_{\text{diff}}/\sqrt{n}} = \frac{\bar{x}_{\text{diff}}}{s_{\text{diff}}/\sqrt{n}} \qquad \text{df} = n - 1$$

Chapter 3

Chi-Square Tests

3.1 Goodness of Fit

The **chi-square test for goodness of fit** tests the null hypothesis[1] (a theoretical distribution of *categorical* data) by comparing the observed and expected counts of each category.
The **chi-square test statistic** is the sum of the square of the difference between the observed and expected counts divided by the expected count over all categories.

$$\chi^2 = \sum \left[\frac{(\text{observed} - \text{expected})^2}{\text{expected}} \right]$$

In order for a chi-square test for goodness of fit to be run, the following conditions must be met:

- Random (sampling or assignment)

  Independent (10%)

- χ^2 Distribution

  The expected counts in each category must be at least 5.

For evidence to exist of a deviation from the null hypothesis is for any difference to exist between the observed and expected counts.
The **expected count** is the product of the null proportion and the sample size and *should not be rounded.*
The *degrees of freedom* is 1 less than the number of categories.
A **chi-square distribution** is denoted $\boldsymbol{\chi^2(k)}$ where $k = \text{df}$.

shape = skew right	$\min\{k\} = 0$	$\max\{k\} = \infty$
$\mu_{\chi^2(k)} = \chi^2(\text{df})$	$\max\{\chi^2(k)\} = \chi^2(\text{df} - 2)$	$\sigma_{\chi^2(k)} = \sqrt{2\,\text{df}}$

The P-value of a chi-square test is the probability of χ^2 falling above the standardized test statistic.

$$P\text{-value} = P(X^2 > \chi^2)$$

If the P-value is statistically significant, the largest contributions to the test statistic can be reported by comparing the observed and expected counts.

[1]The hypotheses for a distribution of categorical data can be described (for an arbitrary ordering scheme of categories using index i) as such:

$$H_0 : (p_{o,i} = p_{e,i})\forall(i \mid \exists p_i) \qquad H_a : (p_{o,i} \gtrless p_{e,i})\exists(i \mid \exists p_i) \qquad H_a : (p_{o,i} \neq p_{e,i})\exists(i \mid \exists p_i)$$

3.2 Homogeneity and Independence

The **chi-square test for homogeneity** compares the distributions of a single categorical variable for multiple populations/treatments.

The conditions for chi-square tests for homogeneity and independence are the same as those for one for goodness of fit.

The **chi-square test for independence** tests the association between two categorical variables in the same population of interest.

Chapter 4

Slopes