let μ_T be the mean appraised value of studio appartment in Toowong

μ_D ...

Example: Lets now perform the test of $H_0 : \mu_T = \mu_D$ against $H_1 : \mu_T \neq \mu_D$. Recall the sample data

	Toowong	Dutton Park
Sample Size	25	30
Sample Mean	\$ 226 716	\$ 206 634
Sample Standard Deviation	\$ 32 338	\$ 13 464

$X_i \sim N(\mu_T, \sigma^2)$

$Y_i \sim N(\mu_D, \sigma^2)$

To compute the test statistic we need the pooled variance estimator of σ^2.

$$\begin{aligned} s_p^2 &= \frac{(n_T - 1)s_T^2 + (n_D - 1)s_D^2}{n_T + n_D - 2} \\ &= \frac{24 \times 32338^2 + 29 \times 13464^2}{25 + 30 - 2} \\ &= 5.7274 \times 10^8 \end{aligned}$$

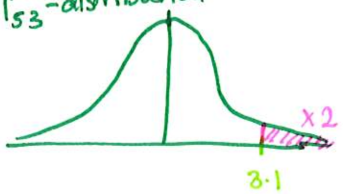
The test statistic is

under Ho

$$\begin{aligned} T(\mathbf{x}_T, \mathbf{x}_D) &= \frac{(\bar{x}_T - \bar{x}_D) - (\mu_T - \mu_D)}{s_p\sqrt{\frac{1}{n_T} + \frac{1}{n_D}}} \\ &= \frac{(226716 - 206634) - 0}{\sqrt{5.7274 \times 10^8}\sqrt{1/25 + 1/30}} \\ &= 3.0987 \end{aligned}$$

$$\frac{\text{estimate} - \text{hypothesised}}{\text{s.e.(estimate)}}$$

T_{53}-distribution



Under the null hypothesis, the test statistic has a t_{53}-distribution. The p-value is

$n_T + n_D - 2 = 25 + 30 - 2$

$$\begin{aligned} &2 \min\left[\mathbb{P}(T_{53} > 3.0987), \mathbb{P}(T_{53} < 3.0987)\right] \\ &= 2\mathbb{P}(T_{53} > 3.0987) \\ &= 2 \times (0.001, 0.005) \qquad \text{[from tables]} \\ &= (0.002, 0.01) \end{aligned}$$

$\mathbb{P}(T_{53} > 3.0987) = 0.0016$

p-value $= 2 \times 0.0016 = 0.0031$.

This is strong evidence against the null hypothesis in favour of the alternative hypothesis that the mean appraisal value for studio apartments is different for the two regions.

Paired t-test

There are situations where we have two samples $(X_1, \ldots, X_n)$ and $Y_1, \ldots, Y_n)$ and although $(X_1, \ldots, X_n)$ are independent and $(Y_1, \ldots, Y_n)$ are independent, X_i and Y_i are dependent for all i. To compare the means of the two populations in this setting, we first take difference $D_i = X_i - Y_i$ and then test the mean of D_i. This often arises when we have two measurements on a single subject before and after some treatment.

Assumptions

As in the previous chapter, we have worked under the assumption that our simple random samples came from a $\mathsf{Normal}(\mu, \sigma^2)$ distribution. Our inferences will still hold (approximately) when our samples are not from a normal distribution due to the central limit theorem. The general guidelines are the same as those for confidence intervals.

Test for a single proportion

Suppose you toss a coin 10 times and get 8 heads and 2 tails. Should you be suspicious that this is not a fair coin? Is this evidence that heads are the more likely outcome from the coin toss? We can address this problem using hypothesis testing. Assuming the outcome of each coin toss is independent of the other coin tosses, the number of heads has a $\mathsf{Binomial}(10, p)$ distribution where p is the probability of getting a head on a single coin toss. We want to test the hypothesis:

$$H_0 : p = 1/2 \qquad \text{against} \qquad H_1 : p > 1/2$$

We want to quantify how surprising it is to get 8 heads from 10 coin tosses which we will do using a p-value. Recall the p-value is the probability of getting data as extreme or more extreme than what we observed assuming the null hypothesis is true. Under the null hypothesis the number of heads has a $\boxed{\text{Binomial}(10, 1/2)}$ distribution. So the p-value is

$$p - \text{value} = \mathbb{P}(X \geqslant 8) \quad (\text{Under } H_0)$$
$$= \binom{10}{8}(0.5)^8(0.5)^2 + \binom{10}{9}(0.5)^9(0.5)^1 + \binom{10}{10}(0.5)^{10}\cdot(0.5)^0$$

where $\boxed{X \sim \text{Binomial}(10, 1/2)}$ The p-value is 0.0546875. This is weak ~~moderate~~ evidence against the null hypothesis, suggesting that the coin is biased towards heads.

If we had a two-sided alternative hypothesis, that is $\boxed{H_1 : p \neq 1/2}$, then the p-value would be computed by

$$p - \text{value} = 2 \min\left[\mathbb{P}(X \geqslant 8), \mathbb{P}(X \leqslant 8)\right]. \simeq 0.109$$
$$= 2 \times 0.05468\ldots$$

Question: What would we now conclude with the two-sided alternative?

inconclusive evidence against H_0.

When the number of Bernoulli trials is large, we can use the Central Limit Theorem to evaluate the p-value. Specifically, if $X \sim \mathsf{Binomial}(n, p)$, then

$$\frac{X - np}{\sqrt{np(1-p)}} \sim_{\text{approx}} \mathsf{Normal}(0, 1)$$

$\mathbb{E}X = np$

$\text{Var}\,X = np(1-p)$

Example. The National Health Interview Survey is conducted annually in the USA by the Center for Disease Control's National Center for Health Statistics. In the 1998-2002 NHIS dataset, there were 25,468 two-child families with children under 10 years old. Of these two-child families 5,844 had two girls. Is there evidence that the proportion of two-child families with two girls is different from 25%?

The null and alternative hypotheses are:

Let p be the probability that a two child family has two girls

$H_0 : p = 0.25$ against $H_1 : p \neq 0.25$

$\mathbb{E}X = 6367 \qquad Var(X) \simeq 4775$

The p-value is Under H_0, $X \sim$ Binomial(25 468, 0.25)

$$\text{p-value} = 2 \times \min\{\mathbb{P}(X \geq 5\,844), \mathbb{P}(X \leq 5844)\}$$

$$\mathbb{P}(X \leq 5844) = \mathbb{P}\left(\frac{X - 0.25 \times 25\,468}{\sqrt{25\,468 \times 0.25 \times 0.75}} \leq \boxed{\frac{5844 - 0.25 \times 25\,468}{\sqrt{25468 \times 0.25 \times 0.75}}}\right)$$

/ Test statistic.

$Z \sim N(0,1)$ $$\approx \mathbb{P}(Z \leq -7.5684) \simeq 2 \times 10^{-14}$$

$$\text{p-value} = 2 \times \mathbb{P}(Z \leq -7.5684) \simeq 4 \times 10^{-14}$$

Hence, we conclude ...

This is (very) strong evidence against the null hypothesis, suggesting the probability of two child family having two girls is different to 0.25.

The tail probabilities of the binomial distribution can be computed in MATLAB using the function `binocdf`.

Comparing two proportions

In the previous chapter we looked at a 2014 study aimed to assess the relationship between volume and type of alcohol consumed during pregnancy in relation to miscarriage. There we constructed a 95% confidence interval for the true difference in the rates of miscarriage between women 36+ and women < 36. We now want to formally test if there is any difference in the probability of miscarriage between women 36+ and women < 36. We introduce the following notation:

- let p_1 be the probability that a woman aged 36 years or above has a miscarriage;
- let p_2 be the probability that a woman aged under 36 years has a miscarriage.

We now state our null and alternative hypotheses.

$H_0: p_1 = p_2$ against $H_1: p_1 \neq p_2$

Of the 208 women who were aged 36 years or above (36+), 52 had a miscarriage, while for the remaining 853 women who were under 36 (< 36), 120 had a miscarriage. How can we decide if this represents evidence against the probability of miscarriage in the two groups being equal?

We know that if $X \sim \text{Binomial}(n, p)$, then

$$\widehat{P} = \frac{X}{n} \sim_{approx} \text{Normal}\left(p, \frac{p(1-p)}{n}\right).$$

So if $X_1 \sim \text{Binomial}(n_1, p_1)$ and $X_2 \sim \text{Binomial}(n_2, p_2)$ are two independent random variables, then

as in Chapter 6

$$\boxed{\widehat{P}_1 - \widehat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2} \sim_{approx} \text{Normal}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)}$$

Unfortunately, even under the null hypothesis that $p_1 = p_2$, the (approximate) distribution of $\widehat{P}_1 - \widehat{P}_2$ still depends on the unknown p_1 and p_2.

Let $\widehat{P} = (X_1 + X_2)/(n_1 + n_2)$. Then, assuming the null hypothesis that $p_1 = p_2$ holds,

pooled proportion

under H_0

$X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$

$p = p_1 = p_2$

$$\frac{\widehat{P}_1 - \widehat{P}_2 - 0}{\boxed{\sqrt{\widehat{P}(1-\widehat{P})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}} \sim_{approx} \text{Normal}(0, 1).$$

$\dfrac{\text{estimate} - \text{hypothesised}}{\text{s.e.(estimate)}}$

s.e. $(\hat{P}_1 - \hat{P}_2)$

The (approximate) distribution of this statistic does not depend on unknown parameters so we may use it as our test statistic. Returning to the study on miscarriage, the test statistic is given by

$\hat{P}_1 = \frac{52}{208} = 0.25$ $\quad \hat{P}_2 = \frac{120}{853} = 0.1407$ $\quad \hat{P} = \frac{52 + 102}{208 + 853} \simeq 0.1621$

Test statistic $\dfrac{(0.25 - 0.1407)}{\sqrt{0.1621 \times (1 - 0.1621)}\sqrt{1/208 + 1/853}} = \dfrac{0.1093}{0.0285} = 3.832$

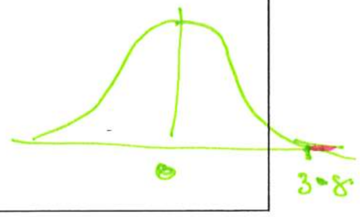
We are testing against the alternative hypothesis $H_1 : p_1 \neq p_2$. The p-value is

p-value $= 2 \min\{\mathbb{P}(Z \geq 3.832), \mathbb{P}(Z \leq 3.832)\}$ $\qquad Z \sim N(0,1)$

$= 2\,\mathbb{P}(Z \geq 3.832)$

$= 2 \times 6.355 \times 10^{-5}$

$= 1.273 \times 10^{-4}$



Hence, we conclude ...

This is strong evidence against the null hypothesis, suggesting there is difference in the probability of miscarriage between the two groups.