2. Basic Analyses for Continuous Measures

2.1 Randomization and Causal Inference

One of the advantages of randomization is that it justifies causal inference from statistical analysis rather than just association. Consider a randomized controlled trial in which patients have been randomized to either a new treatment (*T*) or a control treatment (*C*). For the i^{th} patient an outcome measure, Y_i , has been determined. A patient has two *potential outcomes*, say $Y_i(T)$ and $Y_i(C)$. The ideal way to estimate the effect of treatment would be to give both treatments to each patient, and calculate the benefit of treatment as the difference between the two potential outcomes. The treatment effect for the i^{th} patient would therefore be $\tau_i = Y_i(T) - Y_i(C)$. The expected treatment effect is therefore,

$$
\begin{aligned}
\tau = E[\tau_i] &= E[Y_i(T) - Y_i(C)] \\
&= E\left[Y_i(T) - Y_i(C)\middle|i \in T\right].\Pr[i \in T] + E\left[Y_i(T) - Y_i(C)\middle|i \in C\right].\Pr[i \in C] \\
&= \left(E\left[Y_i(T)\middle|i \in T\right].\Pr[i \in T] - \underline{E\left[Y_i(C)\middle|i \in T\right].\Pr[i \in T]}\right) \leftarrow \text{These cannot be observed as they are counter factual outcomes} \\
&\qquad + \left(\underline{E\left[Y_i(T)\middle|i \in C\right]}.\Pr[i \in C] - E\left[Y_i(C)\middle|i \in C\right].\Pr[i \in C]\right)
\end{aligned}
$$

In most trials a patient can only receive one treatment. If a patient receives treatment *T*, the outcome $Y_i(C)$ cannot be observed. $Y_i(C)$ is called a *counter-factual* outcome for patients that treatment. Similarly, $Y_i(T)$ is the *counter-factual* outcome for patients receiving treatment *C*. Randomization allows us to assert: $E\left[Y_i(C)\middle|i \in T\right] = E\left[Y_i(C)\middle|i \in C\right]$ and $E\left[Y_i(T)\middle|i \in C\right] = E\left[Y_i(T)\middle|i \in T\right]$.

Define $\mu_T = E\left[Y_i(T)\middle|i \in T\right]$ and $\mu_C = E\left[Y_i(C)\middle|i \in C\right]$.

Hence

$$\tau = (\mu_T.\Pr[i \in T] + \mu_T.\Pr[i \in C]) - (\mu_C.\Pr[i \in T] + \mu_C.\Pr[i \in C])$$
$$= \mu_T - \mu_C$$

The expected values, μ_T and μ_C, can be estimated by the sample means for each group, say $\bar{y}_T$ and $\bar{y}_C$. Hence, the expected treatment effect can be estimated by

$$\hat{\tau} = \bar{y}_T - \bar{y}_C \ .$$

If *Y is* a continuous normally distributed outcome measure, a statistical test of the null hypothesis $H_0 : \tau = 0$ can be carried out using a two independent samples t-test.

In observational studies there is no randomization. Other methods have to be used to allow one to assert that

$E\left[Y_i(T)\middle|i \in C\right] = E\left[Y_i(T)\middle|i \in T\right]$ and $E\left[Y_i(C)\middle|i \in T\right] = E\left[Y_i(C)\middle|i \in C\right]$

Design methods

Matching of cases with controls in case-control studies.

Stratification or matching exposed and unexposed subjects in cohort studies.

Data analysis

Using statistical modelling to adjust for confounding variables.

2.2 Glossary: Statistical Inference Terminology

Hypothesis test: A general term for the procedure of assessing whether "data" is consistent or otherwise with statements made about a "population".

Null Hypothesis: Represented by *H₀* meaning "no effect", "no difference" or "no association" .

Alternative hypothesis: Represented by H_1 that usually postulates non-zero "effect", "difference" or "association".

Significance test: A statistical procedure that when applied to a set of observations results in a *p*-value relative to a null hypothesis.
A common misinterpretation of significant test is that failure to reject the null hypothesis justifies acceptance of the null hypothesis.

***p*-value:** Probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.
A common misinterpretation of a p-value is to say it is the "probability of the null hypothesis".

Significance level (α) : The probability at which the null hypothesis (H_0) is rejected when the null hypothesis is actually true. Typical chosen to be 5%, 1%, or 0.1%. It is also referred to as the *test size*.

Critical value: This is the value of the test statistic corresponding to a given significance level.

Confidence interval: A range of values calculated from a sample of observations that are believed with a particular probability to contain the true population parameter value. A 95% confidence interval implies that if the process was repeated again and again 95% of intervals would contain the true value in the population.

2.3 The Two Samples t-test

If the outcome measure Y is normally distributed, a test statistic can be defined as $T=\dfrac{\bar{y}_T-\bar{y}_C}{\hat{SE}[\bar{y}_T-\bar{y}_C]}$ where

$\hat{SE}[\bar{y}_T-\bar{y}_C]=s.\lambda$, $\lambda=\sqrt{1/n_T+1/n_C}$, $s=\sqrt{\dfrac{(n_T-1)s_T^2+(n_C-1)s_C^2}{n_T+n_C-2}}$ with s_T and s_C being the sample standard deviations for the two treatment groups.

A two-sided test of the null hypothesis $H_0:\mu_T=\mu_C$ against the alternative hypothesis $H_1:\mu_T\neq\mu_C$ compares $|T|$ with a critical value, $t_{\alpha/2}(\nu)$, where α is the significance level and $\nu=n_T+n_C-2$ is the degrees of freedom. If $|T|>t_{\alpha/2}(\nu)$, the null hypothesis (H_0) is rejected.

$t_{\alpha/2}(\nu)$ is the percentage point of the central t-distribution with ν degrees of freedom such that upper tail probability $\Pr[t>t_\alpha(\nu)]=\alpha$.

Assumptions of the two-sample t-test

(i) Subjects are independent.

(ii) The variances of the two populations being compared are equal $(\sigma_T^2=\sigma_C^2=\sigma^2)$.

(iii) Data in each population are normally distributed.

One-sided and Two-sided Hypothesis Tests

A one-sided test restricts the alternative hypothesis to be either larger, that is $H_1:\mu_T>\mu_C$ or smaller $H_1:\mu_T<\mu_C$. For a two-sided test the alternative hypothesis is $H_1:\mu_T\neq\mu_C$. A two-sided test is in essence two one-sided tests each with significance level $\alpha/2$. Based on rejection of the null with a two-sided test one can conclude that $\mu_T<\mu_C$ or $\mu_T>\mu_C$.

It is recommended that two-sided tests be used unless there is a strong a-priori reason to believe rejection in one direction is of absolutely no interest. In medical studies this is rarely the case, so two-sided tests are recommended and generally used. The decision to use a one-sided test in preference to a two-sided test should be made prior to analysing the data to prevent statistical analysis bias.

Confidence Intervals for the Difference of Means

If the outcome measure Y is normally distributed satisfying the assumptions for the t-test , a (1-α) confidence interval for the treatment effect τ is given by $\bar{y}_T-\bar{y}_C\pm t_{\alpha/2}(\nu)\hat{SE}[\bar{y}_T-\bar{y}_C]$ where

$\hat{SE}[\bar{y}_T-\bar{y}_C]=s\sqrt{1/n_T+1/n_C}$ $s=\sqrt{\dfrac{(n_T-1)s_T^2+(n_C-1)s_C^2}{n_T+n_C-2}}$ and $\nu=n_T+n_C-2$.

Example 2.1: Ventilation Trial. *A trial of two ventilation methods during cardiac bypass surgery*. Seventeen patients undergoing cardiac bypass surgery were randomized to one of two ventilation schedules using 50% nitrous oxide 50% oxygen.

New For 24 hrs

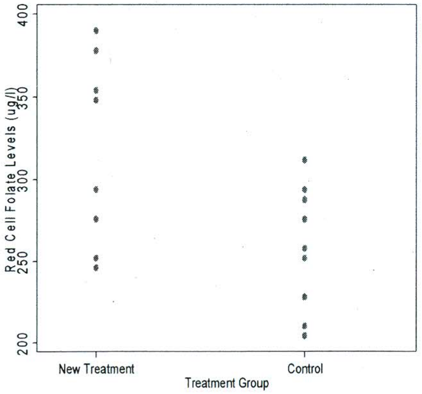
Control Only during operation

The outcome measure for the trial was red cell folate level at 24 hrs post-surgery.

Table 1.1 Red Cell Folate Level Data and Summary Statistics

Treatment Group	*New*	*Control*
Red Cell Folate Level (µg/l)	251	206
	275	210
	291	226
	293	249
	332	255
	347	273
	354	285
	360	295
		309
Mean	$\bar{y}_T = 312.9$	$\bar{y}_C = 256.4$
Standard deviation (s.d.)	$s_T = 40.7$	$s_C = 37.1$
Treatment group size	$n_T = 8$	$n_C = 9$

Figure 1.1 Dotplot of data



Ex 2.1 *Calculate the point estimate of the treatment effect of the New treatment compared to the Control treatment.*

Point estimate of the treatment effect is

$$\hat{\tau} = \bar{y}_T - \bar{y}_C = 312.9 - 256.4 = \underline{\underline{56.5}}$$

Ex 2.2 *Using a two-sample t-test, test whether there is a significant treatment effect using a 5% two-sided significance level.*

(i) Calculate the pooled standard deviation

$$s = \sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 2}} = \sqrt{\frac{7 \times 40.7^2 + 8 \times 37.1^2}{8 + 9 - 2}} = \underline{\underline{38.82}}$$

(ii) Calculate the standard error of the difference between means

$$\hat{SE}[\bar{y}_T - \bar{y}_C] = s\sqrt{1/n_T + 1/n_C} = 38.82 \times \sqrt{\frac{1}{8} + \frac{1}{9}} = \underline{\underline{18.86}}$$

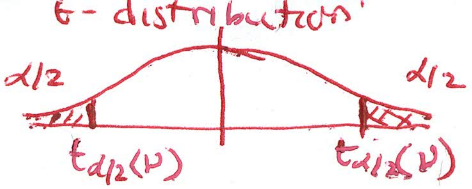
(iii) Calculate the test statistic

$$T = \frac{\bar{y}_T - \bar{y}_C}{\hat{SE}[\bar{y}_T - \bar{y}_C]} = \frac{56.5}{18.86} = 2.995$$

T is assumed to have a t-distribution with degrees of freedom $\nu = n_T + n_C - 2$. Hence $\nu = 9 + 8 - 2 = 15$, $\alpha = 0.05$

$t_{\alpha/2}(\nu) = t_{0.025}(15) = 2.1314$

Since $|T| > t_{\alpha/2}(\nu)$, we can reject the null hypothesis.



Using Statistical Tables

A copy of the School of Mathematics Statistical Tables is available on the module page. These give the cumulative distribution for the t-distribution $t_{v,q}$ where *q* is the cumulative probability for *q= 0.95,0.975, 0.99 and 0.995*. For a two-sided test of sizeα= 0.05, the critical values is the value of *t* that give a right tail probability equal to 0.025 (*α/2*), which corresponds to *q=1-α/2=0.975* in the table. The critical value for a two-sided 0.05 size test is therefore

$$t_{15\,,\,0.975}=t_{0.025}(15)=2.1314.$$

The null hypothesis of no treatment effect would therefore be rejected at a 5% level, because $|T|>2.1314$. The test statistic *T* is also larger than $t_{15,0.995}$=$t_{0.005}$(15)=2.9467. Hence, the null hypothesis would also be rejected with a two-sided 1% significance level. The p-value is therefore less than 0.01. Using statistical software on can calculate the p-value for $T=2.995$ to be 0.009.

Ex 2.3 *Calculate the 95% confidence interval of the treatment effect*

A (1-α) confidence interval for the treatment effect τ is given by

$\bar{y}_T-\bar{y}_C \pm t_{\alpha/2}(v)\hat{S}E[\bar{y}_T-\bar{y}_C]$ where $\hat{S}E[\bar{y}_T-\bar{y}_C]=s\sqrt{1/n_T+1/n_C}$

$\bar{y}_T-\bar{y}_C=$ 56.5 $t_{0.025}(v)=$ 2.1314 $\hat{S}E[\bar{y}_T-\bar{y}_C]=$ 18.86

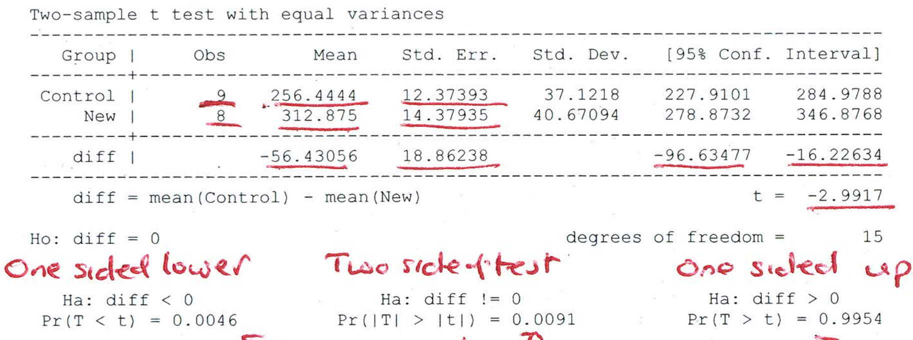
The 95% confidence interval is therefore

56.5 ± 2.1314 × 18.86

95% c.i. is 16.50 to 96.70

Figure 2.1 STATA Output for Two-Sample t-test and CI for

Ventilation Trial



```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 Control |       9    256.4444    12.37393     37.1218    227.9101    284.9788
     New |       8     312.875    14.37935    40.67094    278.8732    346.8768
---------+--------------------------------------------------------------------
    diff |           -56.43056    18.86238               -96.63477   -16.22634
------------------------------------------------------------------------------
    diff = mean(Control) - mean(New)                              t =  -2.9917
Ho: diff = 0                                     degrees of freedom =       15

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0046         Pr(|T| > |t|) = 0.0091          Pr(T > t) = 0.9954
```

One sided lower Two sided test One sided upper

p-values

Ex 2.4 *Briefly comment on the results of the ventilation trial.*

"For patients receiving bypass surgery there was evidence that ventilation for 24 hrs significantly increased post-operative red folate levels by 56.4 μgl (95% c.i. 16.2 to 96.6 μgl, p$\leq$ 0.0091 compared to ventilations restricted to the duration of the operation."

Note that there are small rounding errors in the hand calculation. Ex 2.4 uses the statistical output.

2.4 Assumptions of the two sample t-test and confidence interval

The two-sample t-test makes three assumptions:

I. **Subjects are independent.**

Independence relates to the design - are patients' outcomes independent or could patients be interacting in some way? In most but not all trials this is plausible.

II. **The variance of the two populations being compared are equal** $\left(\sigma_T^2 = \sigma_C^2 = \sigma^2\right)$.

It is sometimes suggested that one should carry out a test comparing variances $\left(H_0 : \sigma_T^2 = \sigma_C^2\right)$, such as Levene's test for equality of variances, to choose between using the t-test or tests such as the Satterthwaite or Welch test that do not assume $\sigma_T^2 = \sigma_C^2$. Unfortunately, this procedure has problems. First, the adverse effect of unequal variance on the results of a t-test is greatest when sample size is small, but in this circumstance the Levene's test will have low power to reject $H_0 : \sigma_T^2 = \sigma_C^2$. Secondly, this is a misuse of statistical test, as one cannot use a test to establish the null hypothesis $\left(H_0 : \sigma_T^2 = \sigma_C^2\right)$ only the alternative $\left(H_1 : \sigma_T^2 \neq \sigma_C^2\right)$.

III. **Data in each population are normally distributed.**

Sometimes tests of normality, such as the Kolmogorov-Smirnov test, are suggested to check the distributional assumptions. These have the same problem as the Levene's test as the assumptions of normality is most critical where sample size is small. A better alternative is to check the distributional assumption graphically. Alternatively one might consider external evidence from other studies using the same measure in similar subjects perhaps with a much larger sample size.

Where equality of variance is not plausible the Satterthwaite test or the Welch test can be used in place of a two-sample t-test.

Where data are non-normal data can be transformed to be closer to normally distributed, by taking the log, square-root, or reciprocal of the measure so that a t-test can be used. Note that inference now relates to the transformed values. For example if a log transformation is used inference now relates to the ratio of geometric means. Alternatively a non-parametric methods that make no distributional assumptions can be used such as the Fisher-Pitman permutation test or the Mann Whitney U-test can be used.

To simplify calculations equality of variance and normality can be assumed in all exercises and exam questions. It is important therefore only to be aware of the assumptions and the alternatives.