On Assumptions

An overview

<table>
<thead>
<tr>
<th>Detection</th>
<th>Normality</th>
<th>Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem-and-leaf plot</td>
<td>Normal scores plot</td>
<td>Study design</td>
</tr>
<tr>
<td>Correction</td>
<td>Transformation</td>
<td>More complex models</td>
</tr>
<tr>
<td>Robustness</td>
<td>Inference for $\mu$</td>
<td>Not robust</td>
</tr>
<tr>
<td></td>
<td>somewhat robust;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inference for $\sigma^2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>not robust</td>
<td></td>
</tr>
</tbody>
</table>

- Here robustness refers to the robustness of a procedure (e.g. t-test) against violation of a model assumption (e.g. non-normality).
- When a procedure is not robust, we may also say that the procedure is sensitive to the violation of a model assumption.
- Here we focus on the normality assumption. See Chapter 8 for other details.

Normal scores

- The main idea is to check if spacing among the obs is like a normal distribution.
- The following steps are used to compute the normal scores for a random sample of size $n$.
  1. Order the obs and think of them as representing quantiles.
  2. Divide the interval $[0,1]$ into $1/n$’s.
  3. Find the corresponding $z$-scores, which gives the desired spacing.
  4. Compare the observed spacing of data with the expected spacing for data that have normal distribution.

Normal scores example 1

Consider a random sample of size $n = 6$:
6.4, 5.1, 3.5, 5.8, 8.5, 6.9

<table>
<thead>
<tr>
<th>$y$</th>
<th>$z$</th>
<th>$P(Z \leq z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>-1.38</td>
<td>1/12 (0.0833)</td>
</tr>
<tr>
<td>5.1</td>
<td>-0.67</td>
<td>3/12 (0.25)</td>
</tr>
<tr>
<td>5.8</td>
<td>-0.21</td>
<td>5/12 (0.4167)</td>
</tr>
<tr>
<td>6.4</td>
<td>0.21</td>
<td>7/12 (0.5833)</td>
</tr>
<tr>
<td>6.9</td>
<td>0.67</td>
<td>9/12 (0.75)</td>
</tr>
<tr>
<td>8.5</td>
<td>1.38</td>
<td>11/12 (0.9167)</td>
</tr>
</tbody>
</table>

- Close to a straight line. This suggests that the data are close to normal.

Normal scores example 2

Consider another random sample of size $n = 6$:
3.7, 3.8, 4.0, 4.5, 5.9, 8.5

- Same $z$ and $P(Z \leq z)$ as in Example 1.
- Not a straight line. This suggests that the data are not normal.
- Caution: for normal data with a small sample size, it is very easy for data to appear non-normal.
On Assumptions

Key R commands

```r
> y = c(6.4, 5.1, 3.5, 5.8, 8.5, 6.9)
> stem(y)
The decimal point is at the |
2 | 5
4 | 18
6 | 49
8 | 5
> qnorm(y)
> qqline(y)
```

More on Hypothesis Testing

Rejection region example

- Here we take a look at hypothesis testing based on a "decision rule" and "rejection region".
- For example, with a random sample of size \( n = 16 \) from \( N(\mu, 36) \), test \( H_0 : \mu = 60 \) versus \( H_A : \mu \neq 60 \). Under \( H_0 \),
\[
\bar{Y} \sim N(60, (1.5)^2).
\]
Suppose \( \alpha = 0.10 \). Then we reject \( H_0 \) at the 10% level. What \( \bar{Y} \) would lead to rejection of \( H_0 \)?
- We can answer this question before having the data!

```r
> y = c(3.7, 3.8, 4.0, 4.5, 5.9, 8.5)
> stem(y)
The decimal point is at the |
3 | 78
4 | 05
5 | 9
6 |
7 |
8 | 5
> qnorm(y)
> qqline(y)
```

More on Hypothesis Testing

Rejection region

- Because \( P(Z \geq 1.645) = 0.05 \), \( H_0 \) will be rejected if the observed \( \bar{y} \) is more than 1.645 standard deviation \( (1.645 \times 1.5 = 2.47) \) away from 60.
- That is,
\[
\{ \bar{Y} \leq 57.53 \text{ or } \bar{Y} \geq 62.47 \}
\]
is called the critical region or rejection region.

Remarks

- The idea of rejection region can be applied to testing using
\[
Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}
\]
with \(|z| \geq z_{0.05} = 1.645\) in the example or testing using
\[
T = \frac{\bar{Y} - \mu}{S / \sqrt{n}}
\]
with \(|t| \geq t_{15,0.05} = 1.753\) if \( \sigma \) is unknown.
- These are equivalent ways of hypothesis testing. Recall that an observed \( \bar{y} \) close to the mean 60 relative to the standard error 1.5 would result in a large p-value and accepting \( H_0 \).
- Now, an observed \( \bar{y} \) not in the rejection region would result in a large p-value and accepting \( H_0 \). Similarly, an observed \( \bar{y} \) in the rejection region would result in a small p-value and rejecting \( H_0 \).
- Recall that
\[
P(\text{Type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})
= P(\bar{Y} \text{ is in the rejection region}) = \alpha
\]
More on Hypothesis Testing

Power example

- Continue with the rejection region example. We know that under $H_0$, $Y \sim N(60, (1.5)^2)$ and
  \[
  P(\text{reject } H_0|H_0) = P(\text{reject } H_0|\mu = 60) \\
  = P(Y \leq 57.53 \text{ or } Y \geq 62.47) \\
  = P(Y \leq 57.53) + P(Y \geq 62.47) \\
  = P(Z \leq \frac{57.53 - 60}{1.5}) + P(Z \geq \frac{62.47 - 60}{1.5}) \\
  = P(Z \leq -2.98) + P(Z \geq 0.31) \\
  = 0.0014 + 0.3783 = 0.3797
  \]

- Suppose that $\mu = 62$, what is the probability of rejecting $H_0$? If $\mu = 62$, then $Y \sim N(62, (1.5)^2)$ and
  \[
  P(\text{reject } H_0|\mu = 62) = P(Y \leq 57.53 \text{ or } Y \geq 62.47) \\
  = P(Y \leq 57.53) + P(Y \geq 62.47) \\
  = P(Z \leq \frac{57.53 - 62}{1.5}) + P(Z \geq \frac{62.47 - 62}{1.5}) \\
  = P(Z \leq -2.98) + P(Z \geq 0.31) \\
  = 0.0014 + 0.3783 = 0.3797
  \]

Power

- The power of a test is the probability of rejecting $H_0$ given some specific value of the parameter.
- The power at $H_0 : \mu = 60$ is $\alpha = 0.10$ by construction.
- The power at $\mu = 62$ is 0.3797.
- In hypothesis testing, we make two types of mistakes:
  - Type I error: reject $H_0$ when $H_0$ is true.
  - Type II error: accept $H_0$ when $H_0$ is false.
- $\alpha = P(\text{Type I error}) = P(\text{reject } H_0|H_0)$
- $\beta = P(\text{Type II error}) = P(\text{accept } H_0|H_A)$ at a given $H_A$.
- Thus power at a given $H_A$ is $P(\text{reject } H_0|H_A) = 1 - \beta$.
- Power is defined for all $\mu$ but $\beta$ is defined only for those $\mu$ that are within the alternative range.
- Thus far no data is needed for computing power. It suffices to have the rejection region, $\sigma$, and $n$ in the power example.

More on Hypothesis Testing

Power example continued

Now suppose $\mu = 64$, what is the power? Since $\mu = 64$, then $Y \sim N(64, (1.5)^2)$ and

\[
\begin{align*}
P(\text{reject } H_0|\mu = 64) &= P(Y \leq 57.53 \text{ or } Y \geq 62.47) \\
&= P(Y \leq 57.53) + P(Y \geq 62.47) \\
&= P(Z \leq \frac{57.53 - 64}{1.5}) + P(Z \geq \frac{62.47 - 64}{1.5}) \\
&= P(Z \leq -4.31) + P(Z \geq -1.02) \\
&= 0.000 + (1 - 0.1539) = 0.8461
\end{align*}
\]

This leads to a power curve (power vs $\mu$).
More on Hypothesis Testing

Power example continued

Remarks

• For a two-sided test, the power curve is symmetric.
• When the sample size $n$ increases, the power curve becomes steeper.
• Power curves are useful for exploring power at particular alternatives and planning for a study (e.g. sample size $n$ determination).

More on Hypothesis Testing

Pumpkin example

• A grower claims that the mean weight of pumpkins in a field is at least 20 lbs. To test this claim, let $Y = \text{the weight of a randomly selected pumpkin}$ and assume that $Y \sim N(\mu, 12)$. Also suppose a random sample of size $n = 7$ and the level of significance $\alpha = 0.01$.
• Here $H_0 : \mu \geq 20$ vs $H_A : \mu < 20$.
• Under $H_0 : \mu = 20$, $\bar{Y} \sim N(20, (1.309)^2)$.
• With $\alpha = 0.01$ and $z_\alpha = z_{0.01} = 2.326$, we know $P(Z \leq -2.326) = 0.01$.

Thus the rejection region is

\[ \{\bar{Y} \leq 20 - 2.326 \times 1.309\} \]

That is

\[ \{\bar{Y} \leq 16.95\} \]

What is the power of the test at $\mu = 18$?

\[
P(\text{reject } H_0 | \mu = 18) = P(\bar{Y} \leq 16.95)
= P(Z \leq \frac{16.95 - 18}{1.309})
= P(Z \leq -0.80) = 0.2119
\]

What is the power of the test at $\mu = 16$?

\[
P(\text{reject } H_0 | \mu = 16) = P(\bar{Y} \leq 16.95)
= P(Z \leq \frac{16.95 - 16}{1.309})
= P(Z \leq 0.73) = 0.7673
\]
More on Hypothesis Testing

Pumpkin example

Sample size determination

Continued with the pumpkin example. Suppose we want an \( \alpha = 0.01 \) test. That is, \( P(\text{Type I error}) = 0.01 \). Suppose we also want a power of 0.9 at \( \mu = 17 \). How large a sample should we take and what rejection region to use so that the two conditions above are met?

With two unknowns \( c \) and \( n \).

Since

\[
-2.326\sqrt{\frac{12}{n}} + 20 = c
\]

\[
1.282\sqrt{\frac{12}{n}} + 17 = c
\]

we have

\[
-2.326\sqrt{12/n} + 20 = 1.282\sqrt{12/n} + 17
\]

and

\[
\sqrt{n} = 3.608 \times \sqrt{12/3} = 4.166, \quad n = 17.37 \uparrow 18.
\]

Since \( n = 18 \),

\[
c = -2.326\sqrt{12/18} + 20 = 18.10.
\]
More on Hypothesis Testing

Remarks

• Suppose the alternative hypothesis is two-sided $H_A : \mu \neq 20$. The desired significance level is still $\alpha = 0.01$ and power is 0.9 at $\mu = 17$. Then use $-z_{0.005} = -2.567$ instead of $z_{0.01} = -2.326$ for achieving the significance level, but still use $z_{0.1} = 1.282$ for achieving the power.

• The two equations are now

$$\frac{\alpha}{2} = 0.005 = P(\bar{Y} \leq c | \mu = 20)$$

$$1 - \beta = 0.9 \approx P(\bar{Y} \leq c | \mu = 17)$$

More on Hypothesis Testing

Key R commands

```r
# power example
> n = 16
> sigma = 6
> sd = sigma/sqrt(n)
> sd
[1] 1.5
> alpha = 0.10
> mu = 60
> region = qnorm(c(alpha/2, 1-alpha/2), mu, sd)
> region
[1] 57.53272 62.46728
> mu = 62
> pnorm(region[1], mu, sd)
[1] 0.9944976
> pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.0055024
> mu = 64
> pnorm(region[1], mu, sd)
[1] 0.9994076
> pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.0005924
> mu = 65
> pnorm(region[1], mu, sd) + pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.4000000
> mu = seq(50, 70, by=0.5)
> power = pnorm(region[1], mu, sd) + pnorm(region[2], mu, sd, lower.tail=F)
> cbind(mu, power)
     mu power
[1,] 50.0 0.999997
[2,] 50.5 0.999996
[3,] 51.0 0.999994
[4,] 51.5 0.999971
[5,] 52.0 0.999972
```

More on Hypothesis Testing

Remarks

• For two-sided alternatives, the general formula for determining the sample size is:

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2}$$

where $\beta = 1 - \text{power}$ at $\mu = \mu_A$.

• For example $\alpha = 0.1, \text{power} = 0.9$. Thus $z_{\alpha/2} = z_{0.05} = -1.645, \beta = 1 - 0.9 = 0.1, z_\beta = z_{0.1} = 1.282$.

• Factors that determine the choice of $n$ are:
  - $H_0$ and $H_A$
  - $\alpha$
  - A particular alternative of interest
  - Power at a particular alternative.
  - $\sigma^2$
More on Hypothesis Testing

Key R commands

# pumpkin example
> n = 7
> sigma = sqrt(12)
> sd = sigma/sqrt(n)
> sd
[1] 1.309307
> alpha = 0.01
> mu = 20
> region = qnorm(alpha, mu, sd)
> region
[1] 16.95410
> mu = 18
> pnorm(region, mu, sd)
[1] 0.2121966
> mu = 16
> pnorm(region, mu, sd)
[1] 0.7669082
> mus = seq(10, 30, by=0.5)
> power = pnorm(region, mus, sd)
> cbind(mus, power)
     mus power
    [1,] 10.0 9.999999e-01
    [2,] 10.5 9.999999e-01
    [3,] 11.0 9.999973e-01
    [4,] 11.5 9.999845e-01
    [5,] 12.0 9.999228e-01
    [6,] 12.5 9.996654e-01
    [7,] 13.0 9.987361e-01
    [8,] 13.5 9.958315e-01
    [9,] 14.0 9.879718e-01
   [10,] 14.5 9.746606e-01
   [11,] 15.0 9.522114e-01
   [12,] 15.5 8.682365e-01
   [13,] 16.0 7.690082e-01

Comparing Two Samples

An overview

Two-sample comparisons are more common than one-sample studies. For example,
(a) Compare milk yield of cows on two different diets.
(b) Compare timber volumes of two species of trees in a forest.
(c) Compare heart rates of patients before and after a drug treatment.
(d) Compare blood flows in two arteries of dogs.

Comparing Two Samples

An overview

- There are two types of two-sample experiments: Paired two-sample versus independent (unpaired) two-sample.
- In an independent two-sample study, there are two treatments (A and B) but there is no direct relationship between an obs on trt A and an obs on trt B (e.g. (a) and (b)).
- In an paired two-sample study, there are two treatments (A and B) and each obs on trt A is naturally paired with an obs on trt B (e.g. (c) and (d)). Related or same experimental units are used for both treatments.
- Choice of paired versus independent two-sample experiment is an important design issue. Data analysis should follow the design.
- So when and why using paired experiment?
  - Paired two-sample studies are usually preferred, because of increased precision (i.e. reduced variability) in estimating treatment (trt) differences.
  - If there are more than 2 trts, use blocking instead of pairing.
Comparing Two Samples

An overview

- Example (c), consider
  (c) Heart rates of 10 patients before and after a drug treatment.
  (c*) Heart rates of 10 patients before the drug treatment and
  heart rates of another 10 patients after the drug treatment.

We would choose (c) over (c*) to remove the effect of variability among individuals and to increase the precision in estimating treatment differences.

- Example (a) can be done by pairing cows according to certain traits (e.g. body weight), but such pairing may not be effective in controlling variability.

- How about examples (b) and (d)?

Paired Two Samples

Blood pressure example

Example: A specific drug is thought to reduce blood pressure. An experiment is conducted on 15 middle-aged male hypertension patients to evaluate the effect of the drug. For each patient, blood pressure is measured, and then after 6 months of the drug treatment, blood pressure is measured again. The question of interest is whether there is any evidence that the drug has an effect on blood pressure.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Before ($Y_1$)</th>
<th>After ($Y_2$)</th>
<th>Difference ($D = Y_1 - Y_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90</td>
<td>88</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>92</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>92</td>
<td>82</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
<td>90</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
<td>78</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>86</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>92</td>
<td>88</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>98</td>
<td>72</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
<td>84</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>94</td>
<td>102</td>
<td>-8</td>
</tr>
<tr>
<td>11</td>
<td>94</td>
<td>94</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>102</td>
<td>70</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>94</td>
<td>94</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>88</td>
<td>92</td>
<td>-4</td>
</tr>
<tr>
<td>15</td>
<td>104</td>
<td>94</td>
<td>10</td>
</tr>
</tbody>
</table>

Let $\mu_1 = \text{the population mean blood pressure before the drug trt}$ and $\mu_2 = \text{the population mean blood pressure after the drug trt}$.

Let $\mu_D = \mu_1 - \mu_2$ denote the difference between the two blood pressure levels.

We want to test $\mu_1 = \mu_2$ or $\mu_1 \neq \mu_2$. Equivalent, test

\[
H_0 : \mu_D = 0 \quad \text{vs} \quad H_A : \mu_D \neq 0.
\]

Let $D = Y_1 - Y_2$ denote the blood pressure difference and assume a random sample $D_1, D_2, \ldots, D_{15}$ of sample size $n = 15$ from $N(\mu_D, \sigma_D^2)$.

We can use a one-sample t-test!
Paired Two Samples

Remarks

• Other possible hypotheses are $H_0 : \mu_D = 5$ vs $H_A : \mu_D \neq 5$, or $H_0 : \mu_D \geq 0$ vs $H_A : \mu_D < 0$.

• The assumptions are made about $D$: a random sample from normal distribution.

• As before, check normality by normal scores plot and design the study to ensure independence.

• However $Y_1$ and $Y_2$ are usually not independent due to pairing.

• Individual $Y_1$ and $Y_2$ need not be normal.

• A $(1 - \alpha)$ CI for $\mu_D$ is

$$\bar{d} - t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}} \leq \mu_D \leq \bar{d} + t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$

• In the blood pressure example, a 95% CI for $\mu_D$ is

$$8.80 - 2.145 \times \frac{10.98}{\sqrt{15}} \leq \mu_D \leq 8.80 + 2.145 \times \frac{10.98}{\sqrt{15}}$$

which is $[2.72, 14.88]$ or $8.80 \pm 6.08$

Paired Two Samples

Key R commands

```r
> # blood pressure example
> bpbefore = c(90, 100, 92, 96, 96, 92, 96, 96, 102, 94, 94, 102, 94, 98, 96, 92, 94)
> bpafter = c(88, 92, 90, 78, 86, 88, 72, 84, 102, 94, 70, 94, 92, 94)
> t.test(bpbefore-bpafter)
One Sample t-test
data: bpbefore - bpafter
t = 3.1054, df = 14, p-value = 0.00775 alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
2.722083 14.877917 sample estimates:
mean of x
8.8
```

• or equivalently

```r
> t.test(bpbefore, bpafter, paired=T)
Paired t-test
data: bpbefore and bpafter
t = 3.1054, df = 14, p-value = 0.00775 alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
2.722083 14.877917 sample estimates:
mean of the differences
8.8
```

• Another 90% CI

```r
> t.test(bpbefore, bpafter, paired=T, conf.level=0.90)
Paired t-test
data: bpbefore and bpafter
```

Independent Two Samples

Diet effect example

A researcher is interested in comparing the effects of two diets on the weight gains of juvenile rats. A total of 5 rats were assigned to diet type A and another 5 rats were assigned to diet type B. Recorded below are the weight gains (in gm) for a given period of time. The question of interest is whether there is evidence of a difference in the diet effect on weight gains.

Subject | Diet A ($Y_1$) | Diet B ($Y_2$)
--------|----------------|----------------
1       | 37.8           | 12.3           
2       | 27.5           | 14.3           
3       | 41.2           | 19.2           
4       | 26.5           | 4.0            
5       | 28.6           | 25.9           

• There is no pairing.

• One can permute the obs within each trt group.
Independent Two Samples

Diet effect example
- Let \( \mu_1 \) = the population mean weight gain of rats that are assigned diet A and \( \mu_2 \) = the population mean weight gain of rats that are assigned diet B.
- We want to test \( H_0: \mu_1 = \mu_2 \) vs \( H_A: \mu_1 \neq \mu_2 \)
- The main idea is to use \( \bar{Y}_1 - \bar{Y}_2 \)
- Let \( \mu_{\bar{Y}_1-\bar{Y}_2} = \mu_1 - \mu_2 \).
- For normal distribution, use the test statistic
  \[ T = \frac{\bar{Y}_1 - \bar{Y}_2 - \mu_{\bar{Y}_1-\bar{Y}_2}}{S_{\bar{Y}_1-\bar{Y}_2}} \]
- We could also use \( \bar{Y}_2 - \bar{Y}_1 \) for \( \mu_2 - \mu_1 \).

Notes
- Under (1)-(3), \( \bar{Y}_1 \sim N(\mu_1, \frac{\sigma^2}{n_1}), \quad \bar{Y}_2 \sim N(\mu_2, \frac{\sigma^2}{n_2}) \)
- The expectation of \( \bar{Y}_1 - \bar{Y}_2 \) is
  \[ \mu_{\bar{Y}_1-\bar{Y}_2} = E(\bar{Y}_1 - \bar{Y}_2) = E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2 \]
- Thus equivalently we want to test
  \[ H_0: \mu_{\bar{Y}_1-\bar{Y}_2} = 0 \quad \text{vs} \quad H_A: \mu_{\bar{Y}_1-\bar{Y}_2} \neq 0 \]

Remark:
- Essentially there are three assumptions: independence (both within a trt and between two trts), normality, and equal variance.
- There is no need to have equal sample size.
- There is an approximate test for \( \sigma_1^2 \neq \sigma_2^2 \).
Independent Two Samples

Diet effect example

• Recall that

\[ H_0: \mu_1 = \mu_2 \quad \text{vs} \quad H_A: \mu_1 \neq \mu_2 \]

• Under \( H_0 \), \( \bar{Y}_1 - \bar{Y}_2 = 0 \) and

\[ T = \frac{(\bar{Y}_1 - \bar{Y}_2) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1 + n_2 - 2} \]

• The observed \( \bar{y}_1 = 32.32, \bar{y}_2 = 15.14, s_1^2 = 44.96, s_2^2 = 66.28 \) with \( n_1 = n_2 = 5 \). Thus

\[ s_p^2 = \frac{4 \times 44.96 + 4 \times 66.28}{8} = 55.62, \]

\[ s_p = \sqrt{55.62} = 7.46, \text{ and the observed test statistic is} \]

\[ t = \frac{(32.32 - 15.14) - 0}{7.46 \sqrt{1/5 + 1/5}} = 3.64 \]

on df = \( n_1 + n_2 - 2 = 8 \).

• The p-value is \( 2 \times P(T_8 \geq 3.64) \), which is less than 0.01.

• Thus reject \( H_0 \) at the 5% level. There is strong evidence against \( H_0 \).

Confidence intervals

• A \((1 - \alpha)\) CI for \( \mu_1 - \mu_2 \) is

\[ \bar{Y}_1 - \bar{Y}_2 - t_{n_1 + n_2 - 2, \alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{Y}_1 - \bar{Y}_2 + t_{n_1 + n_2 - 2, \alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \]

• In the diet effect example, a 99% CI for \( \mu_1 - \mu_2 \) is

\[ 17.18 - 3.355 \times 7.46 \times \sqrt{\frac{2}{5}} \leq \mu_1 - \mu_2 \leq 17.18 + 3.355 \times 7.46 \times \sqrt{\frac{2}{5}} \]

which is \([1.35, 33.00]\) or \(17.18 \pm 15.83\).

• The result is consistent with the testing, as 0 is not in the 99% CI.

Key R commands

```r
> # diet effect example
> dietA = c(37.8, 27.5, 41.2, 26.5, 28.6)
> dietB = c(12.3, 14.3, 19.2, 4.0, 25.9)
> # independent two-sample t-test and 95% CI
> t.test(dietA, dietB, var.equal=T)
Two Sample t-test
data: dietA and dietB
t = 3.6423, df = 8, p-value = 0.006567
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
6.303089 28.056911
sample estimates:
mean of x mean of y
32.32 15.14

> # 99% CI
> t.test(dietA, dietB, var.equal=T, conf.level=0.99)
Two Sample t-test
data: dietA and dietB
t = 3.6423, df = 8, p-value = 0.006567
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
1.353381 33.006619
sample estimates:
mean of x mean of y
32.32 15.14
```

An example with unequal sample size

Suppose the purpose is to compare two trts using two independent samples. The summary statistics for each trt group is as follows.

\[
\begin{array}{l}
\text{trt 1} \\
\bar{y}_1 = 37.40 \\
\bar{y}_2 = 35.05 \\
s_1^2 = 9.86 \\
s_2^2 = 8.33 \\
n_1 = 7 \\
n_2 = 10
\end{array}
\]

1. Test \( H_0: \mu_1 = \mu_2 \) vs \( H_A: \mu_1 \neq \mu_2 \).

2. Construct a 90% CI for \( \mu_1 - \mu_2 \).

Note statistical significance vs a difference that is scientifically meaningful.
Independent Two Samples

Key R commands

```r
> # an example with unequal sample size
> ybar1 = 37.40
> ybar2 = 35.05
> var1 = 9.86
> var2 = 8.33
> n1 = 7
> n2 = 10
> sdpool = sqrt(((n1-1)*var1+(n2-1)*var2)/(n1+n2-2))
> sdpool
[1] 2.990318
> t.value = (ybar1-ybar2)/(sdpool*sqrt(1/n1+1/n2))
> t.value
[1] 1.594686
> 2*pt(t.value, n1+n2-2, lower.tail=F)
[1] 0.1316318
> alpha=0.1
> c(ybar1-ybar2-qt(alpha/2, n1+n2-2, lower.tail=F)*sdpool*sqrt(1/n1+1/n2),
+ ybar1-ybar2+qt(alpha/2, n1+n2-2, lower.tail=F)*sdpool*sqrt(1/n1+1/n2))
[1] -0.2333733 4.9333733
```