

On Assumptions

An overview

	Normality	Independence
Detection	Stem-and-leaf plot Normal scores plot	Study design
Correction	Transformation Nonparametric procedure	More complex models e.g. time series
Robustness	Inference for μ somewhat robust; Inference for σ^2 not robust	Not robust

- 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.

148

On Assumptions

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.

149

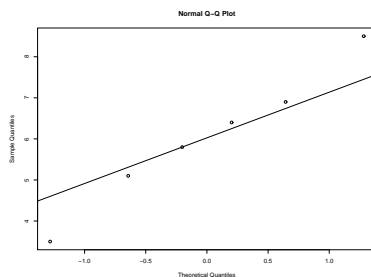
On Assumptions

Normal scores example 1

Consider a random sample of size $n = 6$:

6.4, 5.1, 3.5, 5.8, 8.5, 6.9

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



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

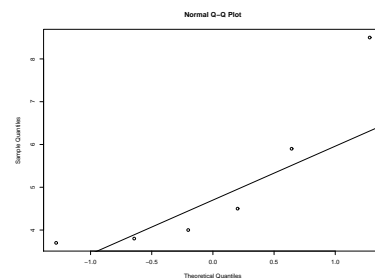
150

On Assumptions

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.

151

On Assumptions

Key R commands

```
> 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

> qqnorm(y)
> qqline(y)

> 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

> qqnorm(y)
> qqline(y)
```

152

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!

153

More on Hypothesis Testing

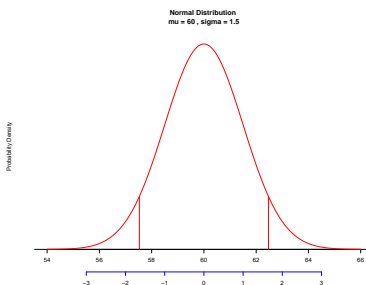
Rejection region

- 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*.



154

More on Hypothesis Testing

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 σ 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

$$\begin{aligned} P(\text{Type I error}) &= P(\text{reject } H_0 | H_0 \text{ is true}) \\ &= P(\bar{Y} \text{ is in the rejection region}) = \alpha \end{aligned}$$

155

More on Hypothesis Testing

Power example

- Continue with the rejection region example. We know that under H_0 , $\bar{Y} \sim N(60, (1.5)^2)$ and

$$\begin{aligned} & P(\text{reject } H_0 | H_0) \\ &= P(\text{reject } H_0 | \mu = 60) \\ &= P(\bar{Y} \leq 57.53 \text{ or } \bar{Y} \geq 62.47) \\ &= P(\bar{Y} \leq 57.53) + P(\bar{Y} \geq 62.47) \\ &= P\left(Z \leq \frac{57.53 - 60}{1.5}\right) + P\left(Z \geq \frac{62.47 - 60}{1.5}\right) \\ &= P(Z \leq -1.645) + P(Z \geq 1.645) \\ &= 0.10 \end{aligned}$$

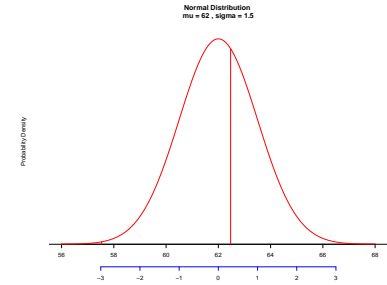
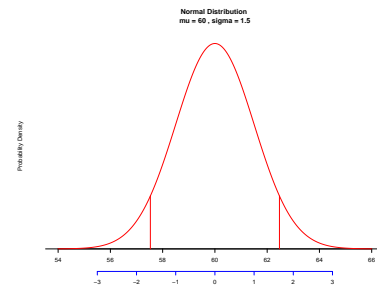
- Suppose that $\mu = 62$, what is the probability of rejecting H_0 ? If $\mu = 62$, then $\bar{Y} \sim N(62, (1.5)^2)$ and

$$\begin{aligned} & P(\text{reject } H_0 | \mu = 62) \\ &= P(\bar{Y} \leq 57.53 \text{ or } \bar{Y} \geq 62.47) \\ &= P(\bar{Y} \leq 57.53) + P(\bar{Y} \geq 62.47) \\ &= P\left(Z \leq \frac{57.53 - 62}{1.5}\right) + P\left(Z \geq \frac{62.47 - 62}{1.5}\right) \\ &= P(Z \leq -2.98) + P(Z \geq 0.31) \\ &= 0.0014 + 0.3783 = 0.3797 \end{aligned}$$

156

More on Hypothesis Testing

Power example



157

More on Hypothesis Testing

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 μ but β is defined only for those μ that are within the alternative range.
- Thus far no data is needed for computing power. It suffices to have the rejection region, σ , and n in the power example.

158

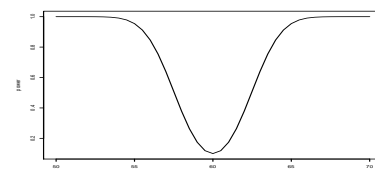
More on Hypothesis Testing

Power example continued

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

$$\begin{aligned} & P(\text{reject } H_0 | \mu = 64) \\ &= P(\bar{Y} \leq 57.53 \text{ or } \bar{Y} \geq 62.47) \\ &= P(\bar{Y} \leq 57.53) + P(\bar{Y} \geq 62.47) \\ &= P\left(Z \leq \frac{57.53 - 64}{1.5}\right) + P\left(Z \geq \frac{62.47 - 64}{1.5}\right) \\ &= P(Z \leq -4.31) + P(Z \geq -1.02) \\ &= 0.000 + (1 - 0.1539) = 0.8461 \end{aligned}$$

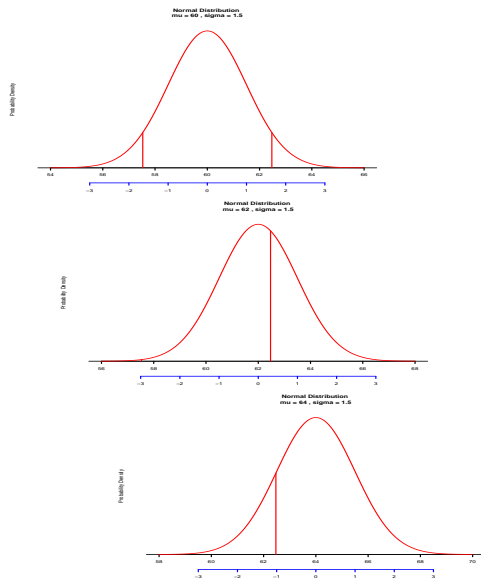
This leads to a power curve (power vs μ).



159

More on Hypothesis Testing

Power example continued



160

More on Hypothesis Testing

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).

161

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 = 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

$$\{\bar{Y} \leq 16.95\}.$$

162

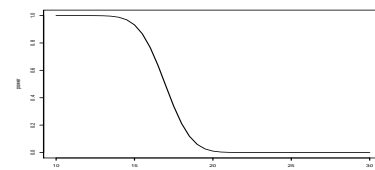
More on Hypothesis Testing

Pumpkin example

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

$$\begin{aligned} P(\text{reject } H_0 | \mu = 18) &= P(\bar{Y} \leq 16.95) \\ &= P\left(Z \leq \frac{16.95 - 18}{1.309}\right) \\ &= P(Z \leq -0.80) = 0.2119 \end{aligned}$$
- What is the power of the test at $\mu = 16$?

$$\begin{aligned} P(\text{reject } H_0 | \mu = 16) &= P(\bar{Y} \leq 16.95) \\ &= P\left(Z \leq \frac{16.95 - 16}{1.309}\right) \\ &= P(Z \leq 0.73) = 0.7673 \end{aligned}$$

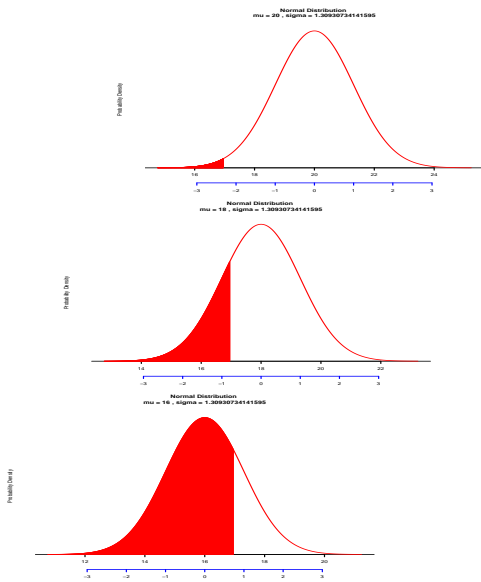


- For a one-sided test, the power curve is not symmetric.
- When the sample size n increases, the power curve becomes steeper.

163

More on Hypothesis Testing

Pumpkin example

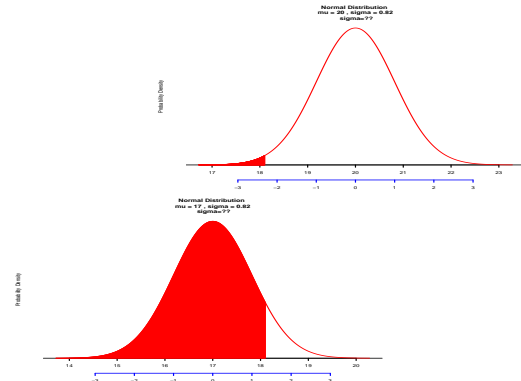


164

More on Hypothesis Testing

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?



165

More on Hypothesis Testing

Pumpkin example

- From Picture 1,

$$\begin{aligned} P(\text{reject } H_0 | \mu = 20) &= P(\bar{Y} \leq c) \\ &= P\left(Z \leq \frac{c - 20}{\sqrt{12/n}}\right) \\ &= 0.01 \end{aligned}$$

We know that $P(Z \leq -2.326) = 0.01$. Thus

$$-2.326 = \frac{c - 20}{\sqrt{12/n}}$$

- From Picture 2,

$$\begin{aligned} P(\text{reject } H_0 | \mu = 17) &= P(\bar{Y} \leq c) \\ &= P\left(Z \leq \frac{c - 17}{\sqrt{12/n}}\right) \\ &= 0.9 \end{aligned}$$

We know that $P(Z \leq 1.282) = 0.90$. Thus

$$1.282 = \frac{c - 17}{\sqrt{12/n}}$$

166

More on Hypothesis Testing

Pumpkin example

- There are two equations

$$\begin{aligned} -2.326 &= \frac{c - 20}{\sqrt{12/n}} \\ 1.282 &= \frac{c - 17}{\sqrt{12/n}} \end{aligned}$$

with two unknowns c and n .

- Since

$$\begin{aligned} -2.326\sqrt{12/n} + 20 &= c \\ 1.282\sqrt{12/n} + 17 &= c \end{aligned}$$

we have

$$\begin{aligned} -2.326\sqrt{12/n} + 20 &= 1.282\sqrt{12/n} + 17 \\ 3 &= 3.608 \times \sqrt{12/n} \end{aligned}$$

and $\sqrt{n} = 3.608 \times \sqrt{12}/3 = 4.166$, $n = 17.37 \uparrow 18$.

- Since $n = 18$,

$$c = -2.326\sqrt{12/18} + 20 = 18.10.$$

167

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

$$\begin{aligned}\alpha/2 = 0.005 &= P(\bar{Y} \leq c | \mu = 20) \\ 1 - \beta = 0.9 &\approx P(\bar{Y} \leq c | \mu = 17)\end{aligned}$$

168

More on Hypothesis Testing

Remarks

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

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

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

- For example $\alpha = 0.1$, 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
 - α
 - A particular alternative of interest
 - Power at a particular alternative.
 - σ^2

169

More on Hypothesis Testing

Key R commands

```
# 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] 0.001449795
> pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.3777026
> pnorm(region[1], mu, sd) + pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.3791524
>
> mu = 64
> pnorm(region[1], mu, sd)
[1] 8.106795e-06
> pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.8465653
> pnorm(region[1], mu, sd) + pnorm(region[2], mu, sd, lower.tail=F)
[1] 0.8465734
>
> mus = seq(50, 70, by=0.5)
> power = pnorm(region[1], mus, sd) + pnorm(region[2], mus, sd, lower.tail=F)
> cbind(mus, power)
      mus  power
[1,] 50.0 0.9999997
[2,] 50.5 0.9999986
[3,] 51.0 0.9999934
[4,] 51.5 0.9999711
[5,] 52.0 0.9998872
```

170

```
[6,] 52.5 0.9996034
[7,] 53.0 0.9987437
[8,] 53.5 0.9964111
[9,] 54.0 0.9907423
[10,] 54.5 0.9784022
[11,] 55.0 0.9543407
[12,] 55.5 0.9123162
[13,] 56.0 0.8465734
[14,] 56.5 0.7544593
[15,] 57.0 0.6388938
[16,] 57.5 0.5091654
[17,] 58.0 0.3791524
[18,] 58.5 0.2635973
[19,] 59.0 0.1743930
[20,] 59.5 0.1187948
[21,] 60.0 0.1000000
[22,] 60.5 0.1187948
[23,] 61.0 0.1743930
[24,] 61.5 0.2635973
[25,] 62.0 0.3791524
[26,] 62.5 0.5091654
[27,] 63.0 0.6388938
[28,] 63.5 0.7544593
[29,] 64.0 0.8465734
[30,] 64.5 0.9123162
[31,] 65.0 0.9543407
[32,] 65.5 0.9784022
[33,] 66.0 0.9907423
[34,] 66.5 0.9964111
[35,] 67.0 0.9987437
[36,] 67.5 0.9996034
[37,] 68.0 0.9998872
[38,] 68.5 0.9999711
[39,] 69.0 0.9999934
[40,] 69.5 0.9999986
[41,] 70.0 0.9999997
>
> plot(mus, power, type="l")
>
```

171

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 9.999996e-01
[3,] 11.0 9.99973e-01
[4,] 11.5 9.99845e-01
[5,] 12.0 9.99228e-01
[6,] 12.5 9.99654e-01
[7,] 13.0 9.987361e-01
[8,] 13.5 9.958315e-01
[9,] 14.0 9.879718e-01
[10,] 14.5 9.695587e-01
[11,] 15.0 9.322114e-01
[12,] 15.5 8.666263e-01
[13,] 16.0 7.669082e-01
```

172

```
[14,] 16.5 6.356372e-01
[15,] 17.0 4.860159e-01
[16,] 17.5 3.383607e-01
[17,] 18.0 2.121966e-01
[18,] 18.5 1.188602e-01
[19,] 19.0 5.907509e-02
[20,] 19.5 2.591961e-02
[21,] 20.0 1.000000e-02
[22,] 20.5 3.382165e-03
[23,] 21.0 1.000410e-03
[24,] 21.5 2.583061e-04
[25,] 22.0 5.813193e-05
[26,] 22.5 1.138928e-05
[27,] 23.0 1.940684e-06
[28,] 23.5 2.873705e-07
[29,] 24.0 3.695472e-08
[30,] 24.5 4.124747e-09
[31,] 25.0 3.994124e-10
[32,] 25.5 3.354056e-11
[33,] 26.0 2.441725e-12
[34,] 26.5 1.540544e-13
[35,] 27.0 8.421561e-15
[36,] 27.5 3.988017e-16
[37,] 28.0 1.635624e-17
[38,] 28.5 5.808966e-19
[39,] 29.0 1.786231e-20
[40,] 29.5 4.754909e-22
[41,] 30.0 1.095623e-23
>
> plot(mus, power, type="l")
>
```

173

Comparing Two Samples

An overview

Two-sample comparisons are more common than one-sample studies. For example,

- Compare milk yield of cows on two different diets.
- Compare timber volumes of two species of trees in a forest.
- Compare heart rates of patients before and after a drug treatment.
- Compare blood flows in two arteries of dogs.

174

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 a 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.

175

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 trt 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.

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

Paired Two Samples

Blood pressure example

- Let μ_1 = the population mean blood pressure before the drug trt and μ_2 = 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, \dots, 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

Blood pressure example

- Under $H_0 : \mu_D = 0$, the test statistic

$$T = \frac{\bar{D} - 0}{S_D/\sqrt{n}} \sim T_{n-1}.$$

- From the data, the observed $\bar{d} = 8.80$, $s_d = 10.98$. Thus the observed

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{8.80 - 0}{10.98/\sqrt{15}} = 3.10$$

on $df = n - 1 = 14$.

- The p-value is

$$2 \times P(T_{14} \geq 3.10)$$

which is between 0.002 and 0.01 from Table C.

- Reject H_0 at the 1% level. There is strong evidence against H_0 .

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 μ_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 μ_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 ± 6.08

180

```
t = 3.1054, df = 14, p-value = 0.00775
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 3.808783 13.791217
sample estimates:
mean of the differences
      8.8
```

182

Paired Two Samples

Key R commands

```
> # blood pressure example
> bpbefore = c(90, 100, 92, 96, 96, 96, 92, 98, 102, 94, 94, 102, 94, 88, 104)
> bpafter = c(88, 92, 82, 90, 78, 86, 88, 72, 84, 102, 94, 70, 94, 92, 94)
> # paired t-test and 95% CI
> 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
> 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
> t.test(bpbefore, bpafter, paired=T, conf.level=0.90)

      Paired t-test

data:  bpbefore and bpafter
```

181

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 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.

183

Independent Two Samples

Diet effect example

- Let μ_1 = the population mean weight gain of rats that are assigned diet A and μ_2 = the population mean weight gain of rats that are assigned diet B.
- We want to test

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

- 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 $\mu_2 - \mu_1$.

Independent Two Samples

Assumptions

- (1) Assume two independent samples Y_1 and Y_2 .
- (2) The first sample $Y_{11}, Y_{12}, \dots, Y_{1n_1}$ is a random sample of size n_1 from $N(\mu_1, \sigma_1^2)$. The second sample $Y_{21}, Y_{22}, \dots, Y_{2n_2}$ is a random sample of size n_2 from $N(\mu_2, \sigma_2^2)$.
- (3) The variances are the same $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Remarks

- 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

Test statistic

- Under (1)–(3),

$$\bar{Y}_1 \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right), \quad \bar{Y}_2 \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right)$$

- The expectation of $\bar{Y}_1 - \bar{Y}_2$ is

$$\begin{aligned} \mu_{\bar{Y}_1 - \bar{Y}_2} &= E(\bar{Y}_1 - \bar{Y}_2) \\ &= E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2 \end{aligned}$$

- Thus equivalently we want to test

$$H_0 : \mu_{\bar{Y}_1 - \bar{Y}_2} = 0 \quad \text{vs} \quad H_A : \mu_{\bar{Y}_1 - \bar{Y}_2} \neq 0$$

Independent Two Samples

Test statistic

- The variance of $\bar{Y}_1 - \bar{Y}_2$ is

$$\begin{aligned} \sigma_{\bar{Y}_1 - \bar{Y}_2}^2 &= \text{Var}(\bar{Y}_1 - \bar{Y}_2) \\ &= \text{Var}(\bar{Y}_1) + \text{Var}(\bar{Y}_2) \\ &= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \end{aligned}$$

- If S^2 is an estimator of σ^2 , then we can estimate $\sigma_{\bar{Y}_1 - \bar{Y}_2}^2$ by $S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$ and $\sigma_{\bar{Y}_1 - \bar{Y}_2}$ by $S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.
- What is S^2 ? We know that S_1^2 estimates σ^2 and so does S_2^2 . Thus a pooled estimate of σ^2 is

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- S_p^2 is a weighted average of the two sample variances, weighted by the d.f.'s.
- In the case of equal sample size $n_1 = n_2 = n$, $S_p^2 = \frac{S_1^2 + S_2^2}{2}$ and $S_{\bar{Y}_1 - \bar{Y}_2} = S_p \sqrt{\frac{2}{n}}$.

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 , $\mu_{\bar{Y}_1 - \bar{Y}_2} = 0$ 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

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

$s_p = \sqrt{55.62} = 7.46$, 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 .

188

Independent Two Samples

Confidence intervals

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

$$\begin{aligned} \bar{y}_1 - \bar{y}_2 - t_{n_1+n_2-2, \alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} &\leq \mu_1 - \mu_2 \\ &\leq \bar{y}_1 - \bar{y}_2 + t_{n_1+n_2-2, \alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \end{aligned}$$

- 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 ± 15.83 .

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

189

Independent Two Samples

Key R commands

```
> # 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
```

190

Independent Two Samples

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.

trt 1	trt 2
$\bar{y}_1 = 37.40$	$\bar{y}_2 = 35.05$
$s_1^2 = 9.86$	$s_2^2 = 8.33$
$n_1 = 7$	$n_2 = 10$

- Test $H_0 : \mu_1 = \mu_2$ vs $H_A : \mu_1 \neq \mu_2$.
- Construct a 90% CI for $\mu_1 - \mu_2$.

Note statistical significance vs a difference that is scientifically meaningful.

191

Independent Two Samples

Key R commands

```
> # 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
>
```