Hypotesis testing: Two samples (Chapter 8)

Medical statistics 2009

http://folk.ntnu.no/slyderse/medstat/medstatI_h09.html

1

Two sample test (def 8.1) vs one sample test :

- Two sample test: Compare the underlying parameters of two different groups, where the values in both groups are unknown.
- One sample test: Compare the underlying parameter in a group with a known value, for example 0 or a known population mean.

2

Example 8.2

- Is there a relationship bettween use of oral contraceptives (OC) and blood pressure (BP)?
- Several study design are possible.

3

Longitudinal study (follow-up study) - eqn 8.1

- Identify a group nonpregnant premenopausal women in childbearing age (16-49) who are not currently OC users and measure their BP (*baseline)*
- After 1 year: Identify a study group who have remained nonpregnant and have become OC users.
- Measure the BP in the study group.
- Compare baseline and 1 year values

4

Cross-sectional study - eqn 8.2

- Identify both a group of OC users and a group of non-OC users among nonpregnant premenopausal women in childbearing age (16-49), and measure their BP
- Compare the BP in the two groups

5

Matched pairs (Example 8.6)

- Does fertility differ between OC users and diaphragm (IUD) users?
- Group 1 consists of 20 OC users.
- For each woman in Group 1, identify an IUD user with same age (within 5 years), race, parity, socio-economic status.
- Registrer time to become pregnant after stopping contraception.

6

Matched versus independent samples: different methods

- Two samples are matched if every observation in the first sample is related to a specific observation in the second sample (for example longitudinal study or matched pairs)
- Two samples are independent if the observations in the first sample are not related to the observations in the second sample (for example cross-sectional study)

Matched pairs. Eample from Box, Hunter & Hunter: Statistics for Experimenters 2nd ed. (2005)

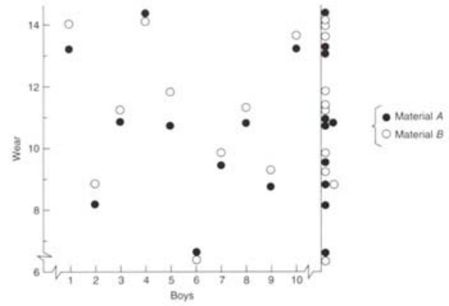


Figure 3.7. Data on two different materials A and B, used for making soles of boys' shoes.

Paired t-test or confidence interval:

- For each pair of observations, compute the difference $d = x_2 - x_1$
- Expected difference is $\Delta = E(D)$
- H_0: $\Delta = 0$ against H_1: $\Delta \neq 0$ (alt >0 or <0)
- Perform a one sample t-test or compute a confidence interval for Δ based on the differences $d_1, d_2, \ldots, d_n$

Repetition:

If $X_1, X_2, \ldots, X_n$ are independent $N(\mu, \sigma^2)$:

Then: $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$.

If σ is unknown, use $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \dfrac{1}{n-1}\left[\left(\sum_{i=1}^{n} X_i^2\right) - n\overline{X}^2\right]$

Then $T = \dfrac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$

Z or T is used to set up a hypotesis test or confidence interval for μ.

If n is "large", then T is approximately $N(\mu, \sigma^2)$

TABLE 8.1 Systolic blood-pressure levels (mm Hg) in 10 women while not using (baseline) and while using (follow-up) oral contraceptives

i	SBP level while not using OC's (x_{i1})	SBP level while using OC's (x_{i2})	d_i*
1	115	128	13
2	112	115	3
3	107	106	–1
4	119	128	9
5	115	122	7
6	138	145	7
7	126	132	6
8	105	109	4
9	104	102	–2
10	115	117	2

*$d_i = x_{i2} - x_{i1}$

Example 8.5 (Table 8.1)

n=10, $\bar{d} = 4.80$, $s^2 = 20.85 = 4.566^2$

Two sided test, t=3.32

Find 0.001 < p < 0.01 from Table 5 in Appendix

EXCEL: =TDIST(3,32;9;2) gives the value p=0.00894

95% confidence interval for Δ:

$$\bar{d} \mp t_{\underbrace{n-1}_{9},\underbrace{1-\alpha/2}_{0.975}} s/\sqrt{n}$$

$$4.8 \mp 2.262 \times 4.566/\sqrt{10}$$

$$= 4.80 \mp 3.27$$

That is, 1.53 to 8.07 (mmHg)

13

t-test and confidence interval for two independent samples

- n_1 observations, assumed independent $N(\mu_1, \sigma_1^2)$
- n_2 observations, assumed independent $N(\mu_2, \sigma_2^2)$
- H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 \neq \mu_2$
- Equivalent: H_0: $\mu_1 - \mu_2 = 0$ against H_1: $\mu_1 - \mu_2 \neq 0$
- Assume for the present equal variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$

14

Estimator for $\mu_1 - \mu_2$: $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

hence: $\dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

If $\sigma_1 = \sigma_2 = \sigma$ then $\dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$

15

But σ^2 is unknown and is estimated by "pooled estimate of the variance":

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2}(X_{i2} - \bar{X}_2)^2\right]$$
$$= \frac{n_1 - 1}{n_1 + n_2 - 2} S_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2^2$$

We use that

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

16

Example 8.9

Cardiovascular Disease, Hypertension
Suppose a sample of eight 35- to 39-year-old nonpregnant, premenopausal OC users who work in a company are identified who have mean systolic blood pressure of 132.86 mm Hg and sample standard deviation of 15.34 mm Hg. A sample of twenty-one 35- to 39 year-old nonpregnant, premenopausal non-OC users are similarly identified who have mean systolic blood pressure of 127.44 mm Hg and sample standard deviation of 18.23 mm Hg. What can be said about the underlying mean difference in blood pressure between the two groups?

17

Example 8.10 – equal variance

n_1=8, $\bar{x}_1$=132.86, s_1=15.34

n_2=21, $\bar{x}_2$=127.44, s_2=18.23

H_0: μ_1-μ_2=0

$$\bar{x}_1 - \bar{x}_2 = 5.42, \quad s^2 = \frac{7}{27}15.34^2 + \frac{20}{27}18.23^2 = 307.18 = 17.53^2$$

$$t = \frac{5.42 (-0)}{17.53\sqrt{\frac{1}{8} + \frac{1}{21}}} = 0.74$$

Degrees of freedom: 8+21-2=27, reject H_0 at 5% level if |0.74|>2.052

P-value f.ex. EXCEL TFORDELING(0,74;27;2)=0.47

18

$$\Pr(-t \le T \le t) = 1-\alpha, \quad \text{where} \quad t = t_{\underbrace{n_1+n_2-2}_{27},\underbrace{1-\alpha/2}_{0.975}} = 2.052$$

$$-t \le \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \le t$$

solve with respect to $\mu_1 - \mu_2$ (eqn 8.13):

$$-9.52 \le \mu_1 - \mu_2 \le 20.36$$

19

Two independent samples, unequal variance

- n_1 observations, assumed independent $N(\mu_1, \sigma_1^2)$
- n_2 observations, assumed independent $N(\mu_2, \sigma_2^2)$
- H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 \neq \mu_2$
- Unequal variance, $\sigma_1^2 \neq \sigma_2^2$

20

Two samples, $\sigma_1 \neq \sigma_2$: We use "Satterthwaite's method":

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t_{d'} \quad \text{approximately,}$$

Where the degrees of freedom d is computed from n_1, S_1, n_2, S_2.

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

21

Example 8.1 (extended) – Unequal variance

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = 0.81$$

$$d' = 15.04 \quad (d'' = 15)$$

$$p-value = 0.43$$

22

Two independent samples, test for unequal variance

- n_1 observations, assumed independent $N(\mu_1, \sigma_1^2)$
- n_2 observations, assumed independent $N(\mu_2, \sigma_2^2)$
- H_0: $\sigma_1^2 = \sigma_2^2$ against H_1: $\sigma_1^2 \neq \sigma_2^2$
- Equivalent: H_0: $\sigma_1^2/\sigma_2^2 = 1$ against H_1: $\sigma_1^2/\sigma_2^2 \neq 1$
- Reject H_0 if S_1^2/S_2^2 deviates "much" from 1
- Under H_0: $S_1^2/S_2^2 \sim F_{n1-1,\, n2-1}$ (Fisher distributed with n_1-1 and n_2-1 degrees of freedom)
- SPSS uses "Levene's test" instead of "Fisher's test"

23

Example 8.16

- $F = S_2^2/S_1^2 = 18.23^2/15.34^2 = 1.41$
- Reject H_0: $\sigma_2^2/\sigma_1^2 = 1$ at level α=0.05 if
  - $F > F_{20,7,0.975} = 4.47$ (FINV(0,025;20;7) in EXCEL)
  - or
  - $F < F_{20,7,0.0.025} = 0.33$ (FINV(0,975;20;7) in EXCEL)
- Alternatively: p-value = 2 * 0.335 = 0.67 (FDIST(1,41;20;7))
- Conclusion: Do not reject H_0

24

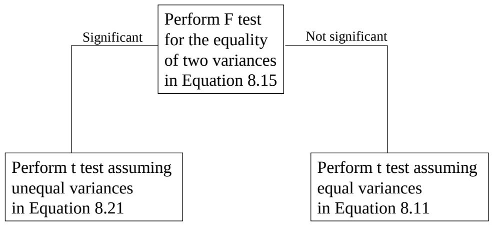
Equation 8.14

Lower p-percentile is an F-distribution with d_1 and d_2 degrees of freedom is the inverse of the upper p-percentile in an F-distribution with d_2 and d_1 degrees of freedom:

$$F_{d_1,d_2,p} = 1/F_{d_2,d_1,1-p}$$

(Useful if the table contains only upper percentiles)

Rosner, Figure 8.10
Strategy for testing the equality of means in two independent, normally distributed samples



BUT:

Navidi: “Statistics for Engineers and Scientists”, 2006, page 343-344:

“Don’t Assume the Population Variances are Equal Just Because the Sample Variances are Close”

” … the expression assuming equal variances requires that the population variances be equal, or nearly so. In situations where the *sample* variances are nearly equal, it is tempting to assume that the population variances are nearly equal as well. However, when the sample sizes are small, the sample variances are not necessarily good approximations to the population variances. Thus it is possible that the sample variances be close even when the population variances are fairly far apart. In general, population variances should be assumed equal only when there is knowledge about the processes that produced the data that justifies this assumption. “

“ … the expression not assuming equal variances produces good results in almost all cases, whether the population variances are equal or not. (Exceptions can occur when the sample sizes are very different.) Therefore, when in doubt, use the expression not assuming equal variances.”

Hence: t-test or confidence interval for difference between the means of two independent, normally distributed samples:

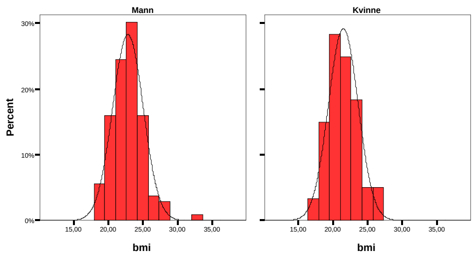
		The truth	
		Equal variance	unequal variance
Assume	Equal variance (eqn 8.11)	correct	Gives wrong answer
	unequal variance (eqn 8.21)	Approximately same answer as above	correct

Use t-test for unequal variance, or a non-parametric method, if in doubt!

If data are not normally distributed:

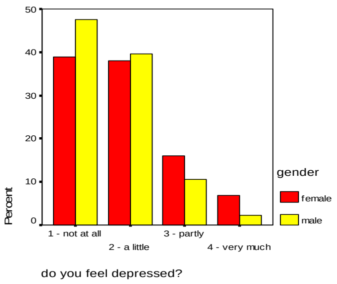
- t-tests give approximately correct results when there is limited variation in data
- t-tester are useless if extreme observations or outliers.
- Non-parametric methods can always be used, and are almost as powerful as the t-test (unless small sample sizes).
- F-test for comparing variances is not robust against departures from the normal distribution.

31



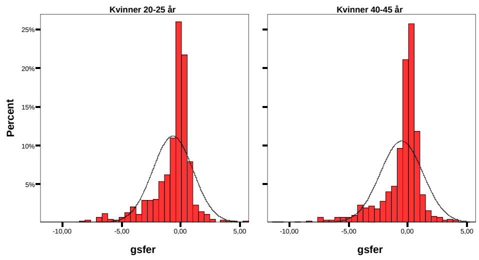
Approximately normally distributed - t-test is OK

32



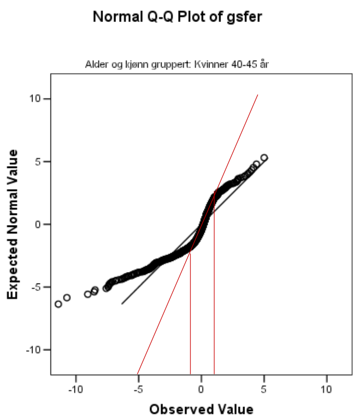
Limited variation in data.
T-test is OK – or use nonparametric methods

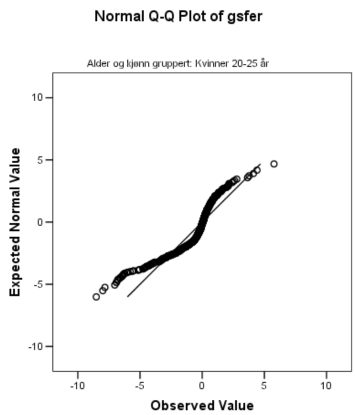
33

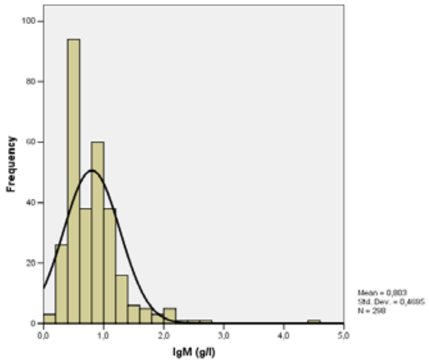


T-test is useless – use nonparametric methods

34







Concentration of serum IgM (g/l) in
298 healthy children, 6 months - 6 years old (Altman, 1991) 37