Two sample hypothesis testing

1. Suppose we want to study the relationship between use of oral contraceptives (OC) and level of blood pressure (BP) in women.

2. **Longitudinal Study:** i) Identify a group of non-pregnant, premenopausal women of child bearing age (16-49) who are not currently OC users, and measure their blood pressure (BP) which are called the *baseline blood pressure*. ii) Rescreen these women 1 year later to ascertain a subgroup who have remained non-pregnant throughout the year and have become OC users. This subgroup is the *study population*. iii) Measure the BP of the study population at the follow-up visit. Compare the baseline and follow-up BP of the women in the study population to determine the difference between the BP of women when they were using OC at the follow-up and when they are not using OC at baseline.

3. Also called follow-up study since the groups are followed over time

4. **Cross-sectional Study:** i) Identify a group of OC users and a group of non-OC users among non-pregnant, premenopausal women of childbearing age (16-49) and measure their BP. (ii) Compare the BP of the OC users and nonusers.

5. The participants are seen at only one point in time.

6. Two samples:

   (a) Paired: Each data point of the first sample is matched and is related to a unique data point of the second sample.

   (b) Independent: the data points in one sample are unrelated to the data points in the second sample.

Paired t-test

Denote the test statistics $\bar{d}/(S/\sqrt{n})$ by t, where S is the standard deviation of the differences d_i and n is the number of matched pairs, If $t > t_{n-1,1-\alpha/2}$ or $t < -t_{n-1,1-\alpha/2}$, then H_0 is rejected. If $-t_{n-1,1-\alpha/2} < t < t_{n-1,1-\alpha/2}$, then H_0 is accepted. A $100(1-\alpha)\%$ Confidence interval for the true difference (Δ) between the underlying means of two paired samples (two-sided) is $(\bar{d} - t_{n-1,1-\alpha/2}S/\sqrt{n}, \bar{d} + t_{n-1,1-\alpha/2}S/\sqrt{n})$.

Paired t Test

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

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

4

Two sample test for independent samples with equal variances

Suppose a sample of 8 35-39 old non-pregnant OC users are identified who have mean SBP of 132 mm Hg and sample sd of 15.34 mm Hg. A sample of twenty-one 35 to 39 year old non-pregnant, pre-menopausal non-OC users are similarly identified who have mean SBP 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?

$$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}\Big(\mu_1 - \mu_2, \sigma^2(1/n_1 + 1/n_2)\Big)$$

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim \mathrm{N}(0,1).$$

Suppose we want to test te hypothesis $H_0 : \mu_1 = \mu_2$ with a level α for two normally distributed populations, where $\sigma^2$is the same for each population , but unknown. The test statistic is

$$\frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{1/n_1 + 1/n_2}}$$

where

$$S = \sqrt{\{(n_1-1)S_1^2 + (n_2-1)S_2^2\}/\{n_1+n_2-2\}}$$

$t < -t_{n_1+n_2-2,1-\alpha/2}$, then H_0 is rejected. If $-t_{n_1+n_2-2,1-\alpha/2} < t < t_{n_1+n_2-2,1-\alpha/2}$, then H_0 is accepted. In the above example

$$\begin{aligned} S^2 &= \frac{7(15.34)^2 + 20(18.23)^2}{27} = 307.18 \\ t &= 0.74 \\ q(0.975, 27) &= 2.052 \end{aligned}$$

H_0 is accepted.

Testing equality of two variances

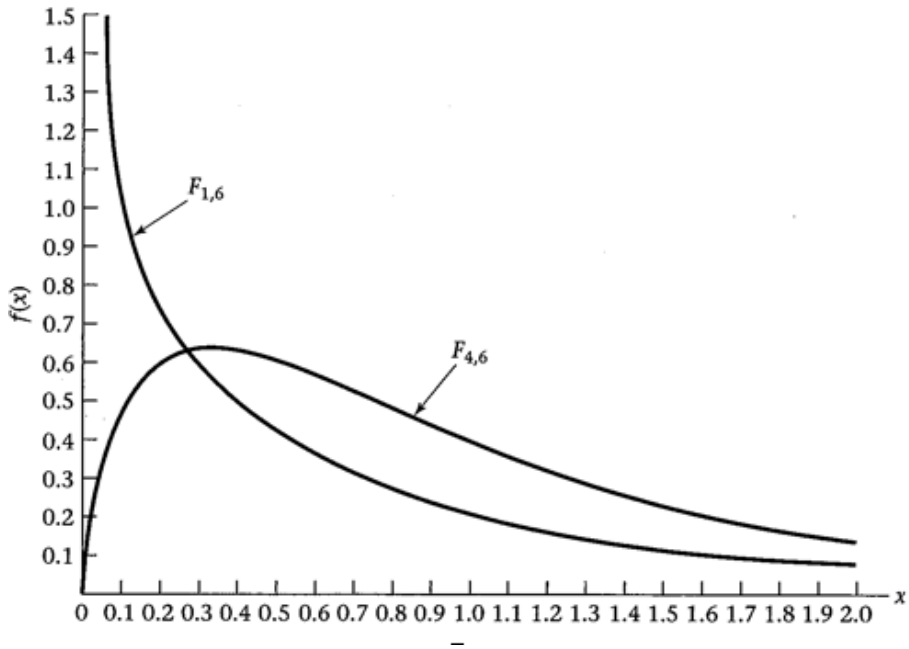
Familial aggregation of cholesterol levels. 100 children 2-14 years old, of men who have died from heart disease. Mean = 207.3 mg/dL and standard deviation = 35.62 mg/dL. Compare with the mean of general population, 175 mg/dL. Select a group of control children as the case children. 74 control children with mean = 193.4 mg/dL and standard deviation 17.3 mg/dL 35.62/17.32 = 4.23 Hypothesis: $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$. Best test is S_1^2/S_2^2. $S_1^2/S_2^2 \sim F_{n_1-1,n_2-1}$. The $100 \times p$th percentile of an F distribution with d_1 and d_2 degrees of freedom is denoted by $F_{d_1,d_2,p}$. $P(F \leq F_{d_1,d_2,p}) = p$. $F_{d_1,d_2,p} = 1/F_{d_2,d_1,1-p}$. If $F > F_{n_1-1,n_2-1,1-\alpha/2}$ or $F < F_{n_1-1,n_2-1,\alpha/2}$, then H_0 is rejected. If $F_{n_1-1,n_2-1,\alpha/2} \leq F_{n_1-1,n_2-1,1-\alpha/2}$, then H_0 is accepted.
Previous example: $F = S_1^2/S_2^2 = 35.6^2/17.3^2 = 4.23$. Under $H_0, F \sim F_{99,73}$. H_0 is rejected if $F > F_{99,73,0.975}$ or $F < F_{99,73,0.025}$. $qf(0.025, 99, 73) = 0.6547$ and $qf(0.975, 99, 73) = 1.549$.

Two-Sample Test for Independent Samples with Unequal Variances

$H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ $(\sigma_1 \neq \sigma_2)$
Fisher Behren's problem:

$$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}\Big(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)\Big)$$

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2)}} \sim \mathrm{N}(0,1).$$



Strategy for Testing for the Equality of Means in Two Independent, Normally Distributed Samples

Perform F test for the equality of two variances. If significant, then perform t test assuming unequal variances. If not, perform t test assuming equal variances

Sample size determination and power for two sample problem

Sample size needed for comparing the means of two normally distributed sample of equal size using a two-sided test with significance level α and power $1-\beta$

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(\mu_1 - \mu_2)^2}$$

For unequal size,

$$\begin{aligned} n_1 &= \frac{(\sigma_1^2 + \sigma_2^2/k)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(\mu_1 - \mu_2)^2} \\ n_2 &= \frac{(k\sigma_1^2 + \sigma_2^2)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(\mu_1 - \mu_2)^2} \end{aligned}$$

where $k = n_2/n_1$.

Two-Sample t Test for Independent Samples with Unequal Variances (Satterthwaite's Method)

(1) Compute the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(2) Compute the approximate degrees of freedom d', where

$$d' = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left(s_1^2/n_1\right)^2 \Big/ (n_1 - 1) + \left(s_2^2/n_2\right)^2 \Big/ (n_2 - 1)}$$

(3) Round d' down to the nearest integer d''.

If $t > t_{d'',1-\alpha/2}$ or $t < -t_{d'',1-\alpha/2}$

then reject H_0.

If $-t_{d'',1-\alpha/2} \le t \le t_{d'',1-\alpha/2}$

then accept H_0.

To test the hypothesis $H_0 : \mu_1 = \mu_2$ vs. $H_1 \mu_1 \neq \mu_2$ for the aulernative $|\mu_1 - \mu_2| = \Delta$, with significance level α, Power $= \Phi(-Z_{1-\alpha/2} + \Delta/(\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}))$.

1 Case Study

A group of children who lived near a lead smelter in El Paso, Texas, were identified and their blood levels of lead were measured. An exposed group of 46 children were identified who had blood-lead levels $\geq 40\mu g/ml$. A control group of 78 children were also identified who had blood-lead levels $< 40\mu g/ml$. Two outcome variables were studied. The number of finger-wrist taps in the dominant hand and the Wechsler full-scale IQ score.

Treatment of Outliers

Figure 8.11 Stem-and-leaf and box plots of finger–wrist tapping score by group, El Paso Lead Study

(a) Control

Stem	Leaf	#	Boxplot
8	4	1	0
7	69	2	|
7	224	3	|
6	55558	5	|
6	01122344	8	+-----+
5	566677778999	12	| |
5	000000011222334	15	*--+--*
4	566666888999	12	+-----+
4	02	2	|
3	8	1	|
3			
2	6	1	0
2	3	1	0
1			
1	3	1	0

----+----+----+----+

Multiply Stem.Leaf by 10**+1

(b) Exposed

Stem	Leaf	#	Boxplot
8	3	1	0
7			
7	0	1	|
6			|
6	2	1	|
5	567789	6	+-----+
5	0122244	7	| |
4	56889	5	*--+--*
4	0012244	7	+-----+
3	5788	4	|
3	4	1	|
2			
2			
1			
1	34	2	0

----+----+----+----+

Multiply Stem.Leaf by 10**+1

21

Detecting outliers

The Extreme studentized deviate (or ESD statistic) $= \max_{i=1,\ldots,n} |x_i - \bar{x}|/S$. Suppose we have a sample $x_1, \ldots, x_n \sim N(\mu, \sigma^2)$, but feel that there may be some outliers present. To test the hypothesis H_0 : no outliers present versus H_1 : that a single outlier is present, with a type I error of α.

1. We compute the Extreme Studentized Deviate test statistic (ESD). The sample value x_i, such that $ESD = |x_i - \bar{x}|/S$ is referred to as $x^{(n)}$.

2. We refer to Table 10 in the Appendix of BR to obtain the critical value $= ESD_{n,1-\alpha}$.

3. If $ESD > ESD_{n,1-\alpha}$, then we reject H_0 and declare that $x^{(n)}$ is an outlier. If $ESD < ESD_{n,1-\alpha}$, then we declare no outliers are present

Evaluate whether outliers are present for the finger-wrist tapping scores in the control group. The sample mean for control group is 54.4, sample standard deviation is 12.1, and n = 64. $ESD = |13-54.4|/12.1 = 3.44$ with 13 being the most extreme value. ($|84-54.4| < |13-54.4|$). From Table 10, $ESD_{70,.95} = 3.26$. $3.44 > ESD_{70,.95} = 3.26 > ESD_{64,.95}$. The finger-wrist tapping score of 13 is an outlier.