EE3211 Modelling Techniques

Lecture 4

Hypothesis Testing: Two-sample Inference

Overview

- Review the concepts of type I error and type II error
- Discuss how to perform hypothesis testing using two-sample inference:
  - Longitudinal Study Design
  - Cross-sectional Study

Recap

Hypothesis Testing

- Two hypotheses in hypothesis-testing framework:
  Null and **alternative hypothesis**

- Objective framework for making decisions
  - probabilities methods
  - not subjective impressions

- Uniform and consistent decision-making criterion

- **One-sample problem**: specify hypotheses about a single distribution

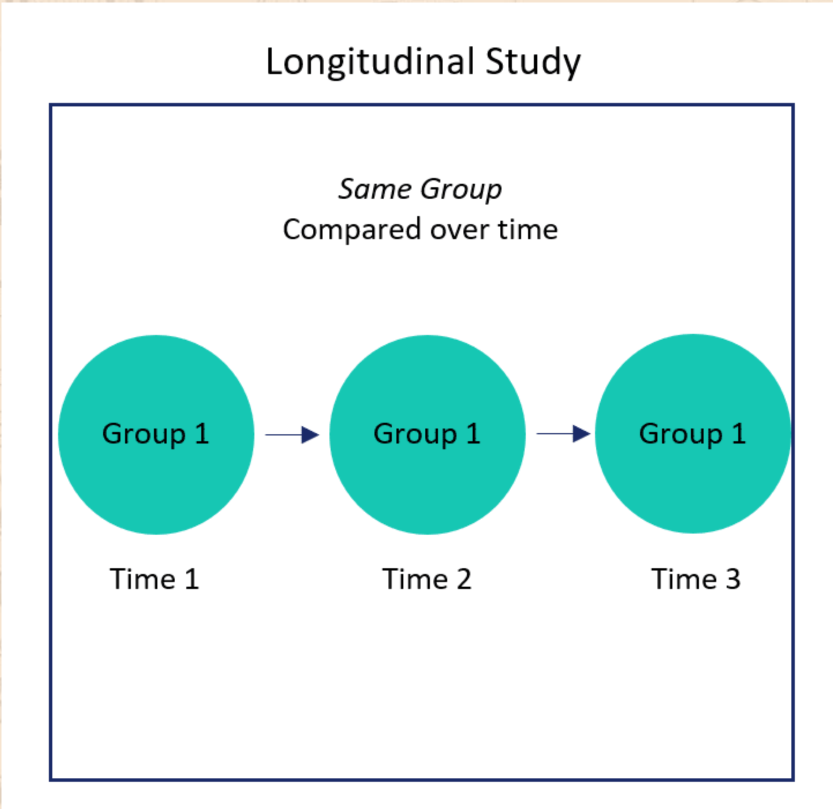
- **Two-sample problem**: compare two different distributions

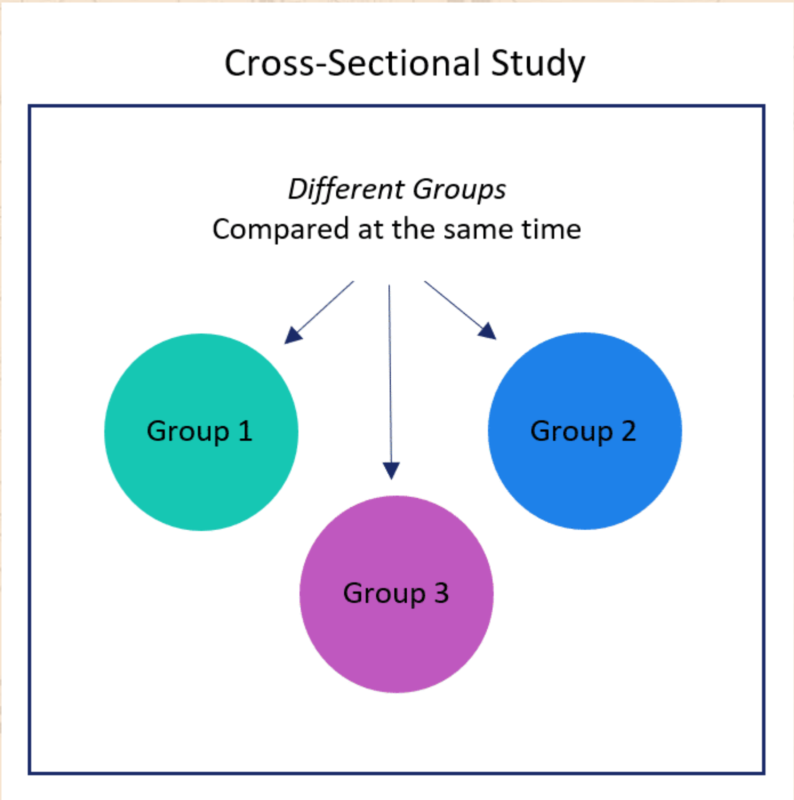
Two-sample hypothesis-testing

- Compare the parameters of two different populations
- *Neither of two sets of values is assumed known*

Example: Two-sample Hypothesis Testing

- Relationship between oral contraceptive (OC) use and blood pressure in women
  - two different experimental designs





Longitudinal Study Design:
OC use vs. Blood Pressure

Baseline blood pressure measurement

- Nonpregnant
- Premenopausal
- Childbearing age (16-49 years old)

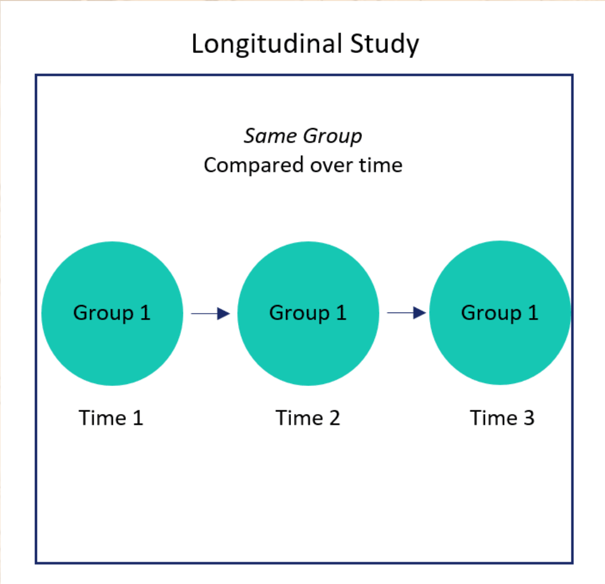
Study population formation

- Rescreen these women 1 year later
- Collect a subgroup who remained nonpregnant
- OC users

Blood pressure measurement at follow-up

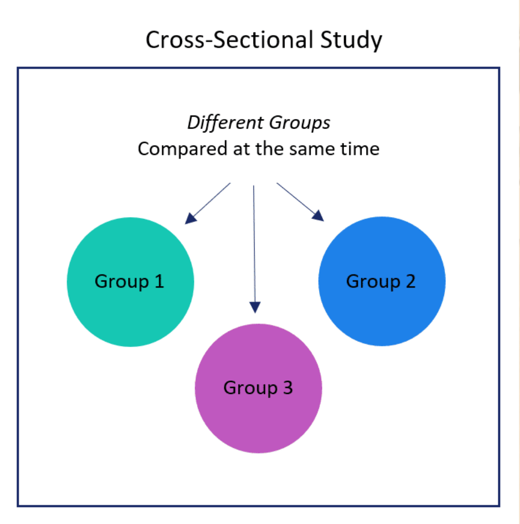
Compare base and follow-up blood pressure of the women

- Difference between blood pressure levels of women when they were using the pill at follow-up and when they were not using the pill at baseline



- **Longitudinal or follow-up** study: same group of people is followed *over time*
- *Paired-sample* design: each woman is used as her own control

Cross-Sectional Study



Blood pressure measurement

- Identify two groups of participants: OC users and non-OC users
- Nonpregnant, premenopaual women
- Bearing age (16-49 years)

Compare blood pressure levels between the two groups

- Difference between blood pressure levels of women who are OC users vs. non-OC users

- **Cross-sectional** study: participants are seen at only one point in time
- *Independent-sample* design: two completely different groups of women are being compared
- Less expensive than a follow-up (longitudinal) study

- **Paired sample**: when each data point in the first sample is matched to a unique data point in the second sample
  - two sets of measurements on the same people
  - different people selected using matching criteria e.g. age and gender

- **Independent samples**: when data points in one sample are unrelated to data points in the second sample

- Paired-study design: more definitive
  - Example: most influencing factors present at first screening will also be there at the second screening and will not influence the comparison of BP levels
  - A control group of non-OC users would completely rule out possible causes of BP change

- The second type of study (independent design): suggestive
  - other confounding factors may influence BP and cause an apparent difference to be found when none is actually present

Recap

H_o: hypothesis to be tested
H_1: hypothesis contradicts the null hypothesis

$$H_0: \mu = \mu_0 \quad \text{vs.} \quad H_1: \mu < \mu_0$$

- Only possible decisions : H_0 is true or H_1 is true
- Outcomes: refer to H_o
- Decision: H_0 is true→we accept H_0
  H_1 is true→H_0 is not true or, we reject H_0
- Four possible outcomes can occur:

Table 7.1 Four possible outcomes in hypothesis testing

Decision	Truth: H_0	Truth: H_1
Accept H_0	H_0 is true and H_0 is accepted	H_1 is true and H_0 is accepted
Reject H_0	H_0 is true and H_0 is rejected	H_1 is true and H_0 is rejected

Recap

Hypothesis Testing

Your Decision Based On a Random Sample	Given the Null Hypothesis Is: True	Given the Null Hypothesis Is: False
Reject	α Type I Error	Correct Decision
Do Not Reject	Correct Decision	β Type II Error

- **Probability of a type I error:**
  α : significance level of a test

- **Probability of a type II error:**
- **β: function of μ and other factors**

- **Power** of a test : 1 - β = 1 – probability of a type II error = Pr(rejecting $H_0|H_1$ true)

- **Objective of hypothesis testing**: use statistical tests that make α and β as small as possible

Paired t-test

- Compare two population means
- Two samples (observations in sample 1 can be paired with observations in sample 2)
- E.g.:
  - Pre- and post- test on the same subjects
  - Compare two treatments on the same subjects

Example: Paired *t* Test

Table 8.1 SBP levels (mm Hg) in 10 women while not using (baseline) and while using (follow-up) OCs

i	SBP level while not using OCs (x_{i1})	SBP level while using OCs (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

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

- Systolic blood pressure (SBP) of the *i*th woman:
  Baseline SBP ~$N(\mu_I, \sigma^2)$
  Follow-up SBP ~$N(\mu_i+\Delta, \sigma^2)$
- Δ: mean difference in SBP between follow-up and baseline
  - $\Delta = 0$: difference is 0
  - $\Delta > 0$: OC pills associated with increased mean SBP
  - $\Delta < 0$: OC pills associated with lowered mean SBP

H_0: $\Delta = 0$ vs. H_1: $\Delta \neq 0$

μ_1 is unknown

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

- Though BP levels are different for each woman, the difference in BP between baseline and follow-up have the same mean and variance over the entire population of women

  → One-sample t test based on the differences (d_i)

Mean difference $\bar{d} = (d_1 + d_2 + \ldots + d_n)/n$

$t = \bar{d}/(s_d/\sqrt{n})$

- s_d is the sample SD of the observed differences:

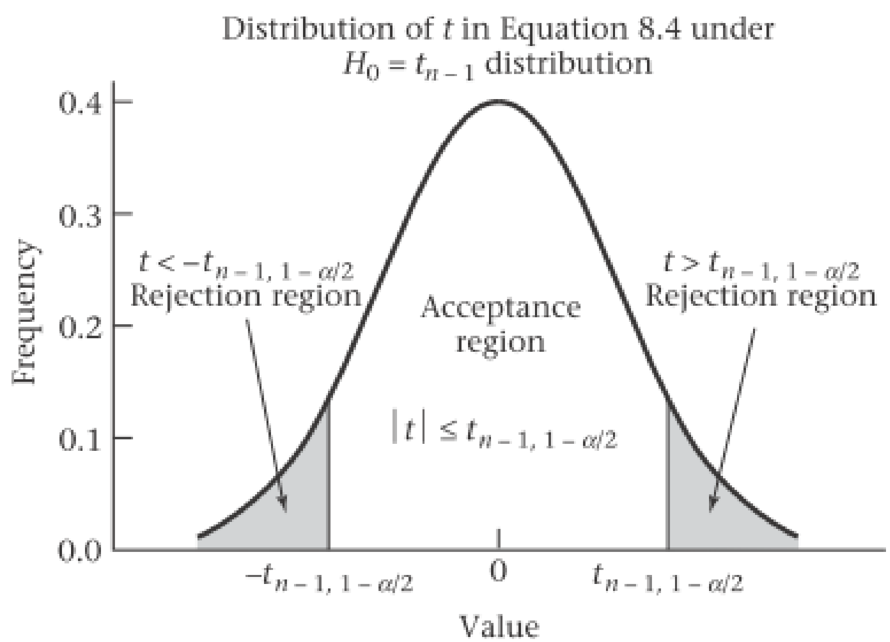
$$s_d = \sqrt{\left[\sum_{i=1}^{n} d_i^2 - \left(\sum_{i=1}^{n} d_i\right)^2 \Big/ n\right] \Big/ (n-1)}$$

or

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\text{sample variance}}$$

- n = number of matched pairs

Figure 8.1 Acceptance and rejection regions for the paired t test



- $t > t_{n-1,1-\alpha/2}$ or $t < -t_{n-1,1-\alpha/2}$ → reject H_0
- $-t_{n-1,1-\alpha/2} \le t \le t_{n-1,1-\alpha/2}$ → accept H_0

Computation of the p-Value for the Paired *t* Test

$t = \bar{d}/(s_d/\sqrt{n})$

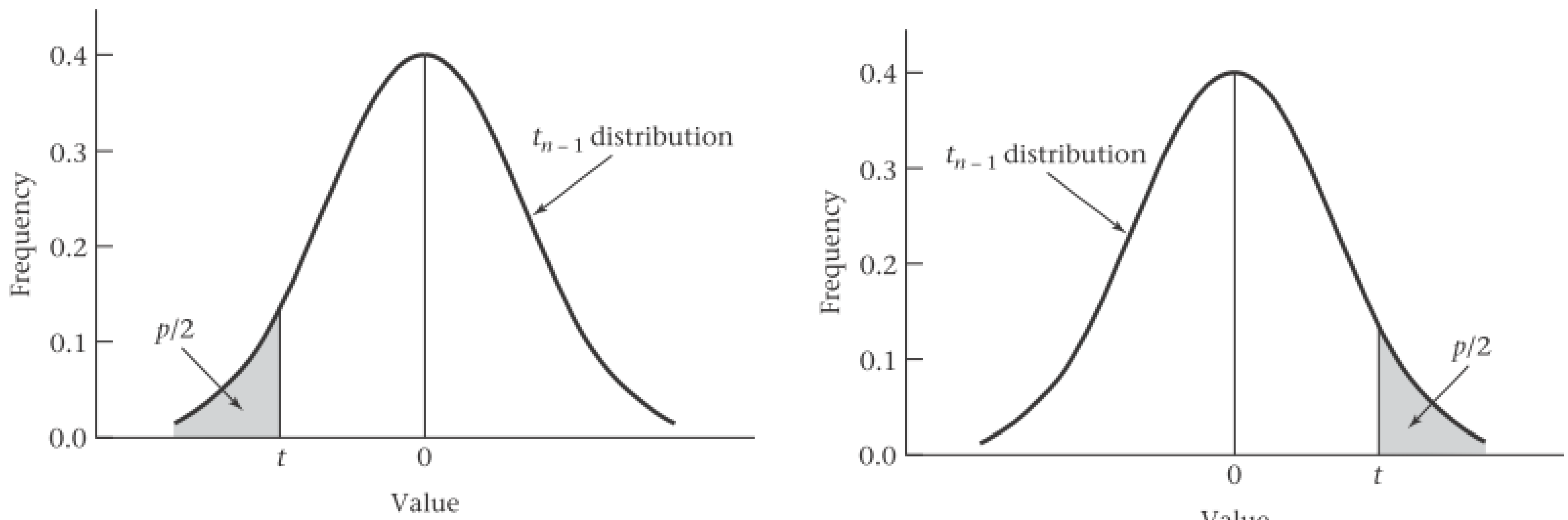
- $t < 0$

  $p = 2 \times$ [the area to the left of t under a t_{n-1} distribution]

- $t \geq 0$

  $p = 2 \times$ [the area to the right of t under a t_{n-1} distribution]

Figure 8.2 Computation of the *p*-value for the paired *t* test



If $t = \bar{d}/(s_d/\sqrt{n}) < 0$, then $p = 2 \times$ (area to the left of t under a t_{n-1} distribution).

Example on Paired T-test: Hypertension

Q: Assess the statistical significance of the OC–blood pressure data in the following table.

TABLE 8.1 SBP levels (mm Hg) in 10 women while not using (baseline) and while using (follow-up) OCs

i	SBP level while not using OCs (x_{i1})	SBP level while using OCs (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 on Paired T-test: Hypertension

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\text{sample variance}}$$

$$\bar{d} = \frac{13 + 3 + \cdots + 2}{10} = 4.80$$

$$s_d^2 = \frac{[(13 - 4.8)^2 + \cdots + (2 - 4.8)^2]}{9} = 20.844$$

$$s_d = \sqrt{20.844} = 4.566$$

$$t = 4.80/(4.566/\sqrt{10}) = 4.80/1.444 = 3.32$$

- Degree of freedom (df): 10-1 = 9 degrees of freedom (*df*)
- $t_{9,.975} = 2.262$
- Because $t = 3.32 > 2.262$ → Paired t Test that H_0 can be rejected using a two-sided significance test with α = .05
- approximate *p*-value:
  - $t_{9,.9995} = 4.781$ and $t_{9,.995} = 3.250$
  - because $3.25 < 3.32 < 4.781$ → $.0005 < p/2 < .005$ or $.001 < p < .01$
- R: t.test()

TABLE 5 Percentage points of the t distribution ($t_{d,u}$)[a]

Degrees of freedom, d	u = .75	.80	.85	.90	.95	.975	.99	.995	.9995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	$\boxed{2.262}$	2.821	$\boxed{3.250}$	$\boxed{4.781}$
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

TABLE 8.2 Using R to perform the paired *t* test based on the blood-pressure data in Table 8.1

```
> t.test(sbp.oc.no,sbp.oc.yes,paired=TRUE)

   Paired t-test

data: sbp.oc.no and sbp.oc.yes
t = -3.3247, df = 9, p-value = 0.008874
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-8.066013 -1.533987
sample estimates:
mean of the differences
   -4.8
```

Interval Estimation for the Comparison of Means from Two Paired Samples

Confidence Interval for the True Difference (Δ) Between the Underlying Means of Two Paired Samples (Two-Sided):

- A two-sided $100\% \times (1-\alpha)$ CI for the true mean difference (Δ) between two paired samples:

$$\left(\bar{d} - \boxed{t_{n-1,1-\alpha/2}\, s_d/\sqrt{n}},\ \bar{d} + \boxed{t_{n-1,1-\alpha/2}\, s_d/\sqrt{n}}\right)$$

*standard error (SE)

*margin of error

Example on Interval Estimation for Two Paired Samples: Hypertension

Q: compute a 95% CI for the true increase in mean SBP after starting OCs.

TABLE 8.1 SBP levels (mm Hg) in 10 women while not using (baseline) and while using (follow-up) OCs

i	SBP level while not using OCs (x_{i1})	SBP level while using OCs (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 on Interval Estimation for Two Paired Samples: Hypertension

Solution

We have $\bar{d}$ = 4.80mmHg, *sd* = 4.566 mm Hg, *n* = 10
A 95% CI for the true mean SBP change is given by:

$$\bar{d} \pm t_{n-1,.975} s_d/\sqrt{n} = 4.80 \pm t_{9,.975}(1.444) = 4.80 \pm 2.262(1.444) = 4.80 \pm 3.27 = (1.53, 8.07) \text{ mmHg}$$

- The true change in mean SBP is most likely between 1.5 and 8.1 mm Hg.
- R: t.test()

TABLE 8.2 Using R to perform the paired *t* test based on the blood-pressure data in Table 8.1

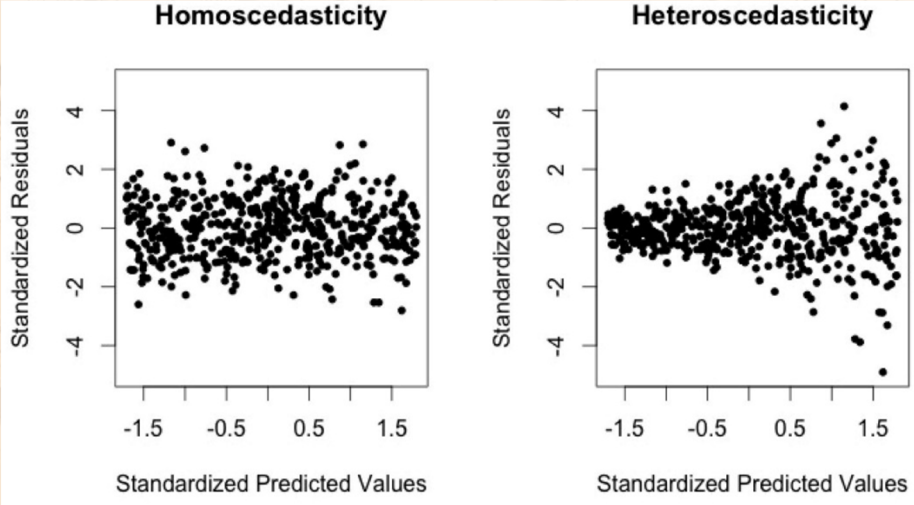
```
> t.test(sbp.oc.no,sbp.oc.yes,paired=TRUE)

   Paired t-test

data: sbp.oc.no and sbp.oc.yes
t = -3.3247, df = 9, p-value = 0.008874
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-8.066013 -1.533987
sample estimates:
mean of the differences
   -4.8
```

Equal vs. unequal variances

- Homoscedasticity: equal variances
- Heteroscedasticity: unequal variances

*Unequal variances → Type I error rate

*→ False positives (falsely reject H_0 when it is true)



Testing for the Equality of Two Variances

H_0: $\sigma_1^2 = \sigma_2^2$ vs. H_1: $\sigma_1^2 \neq \sigma_2^2$

Two independent random samples: $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$

- Best test: the ratio of the sample variances (s_1^2/s_2^2) rather than on difference $(s_1^2 – s_2^2)$
- If the variance ratio is either too large or too small
  → reject H_0
- Otherwise: → accept H_0

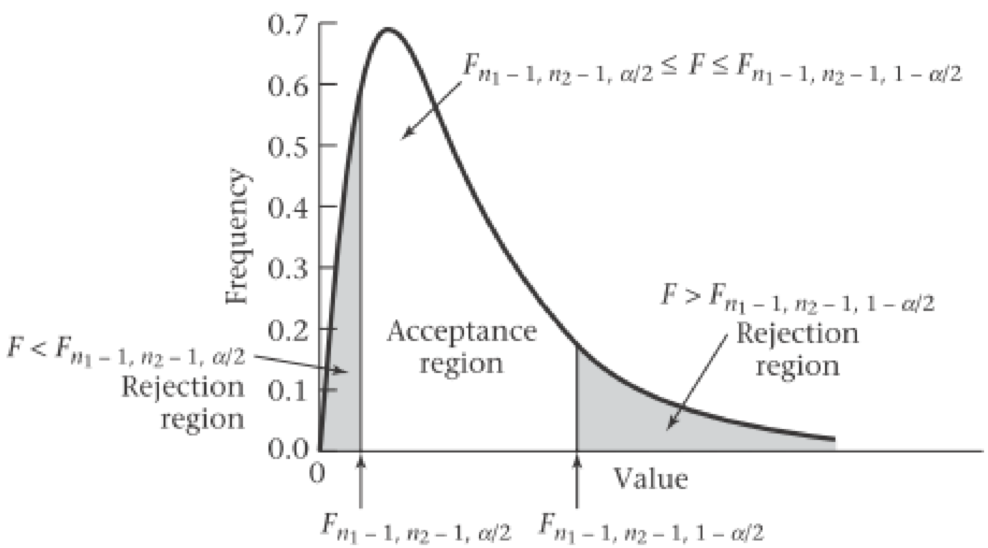
- Statisticians R. A. Fisher and G. Snedecor: distribution of the variance ratio (S_1^2/S_2^2)
- The variance ratio (S_1^2/S_2^2) follows an ***F* distribution** under the H_0 that $\sigma_1^2 = \sigma_2^2$
- A family of F distributions, rather than a unique *F* distribution
- Two parameters: *numerator and denominator degrees of freedom*.
  - n_1: first samples size
  - n_2: second samples size
  - variance ratio ~ *F* distribution with $n_1 - 1$ (numerator *df*) and $n_2 - 1$ (denominator *df*) → $F_{n1-1,n2-1}$ distribution
- *F* distribution: positively skewed
  - skewness : relative magnitudes of the two degrees of freedom
- Numerator *df* is 1 or 2 → distribution has a mode at 0
  - otherwise: mode > 0

F Test for the Equality of Two Variances

H_0: $\sigma_1^2 = \sigma_2^2$ vs. H_1: $\sigma_1^2 \neq \sigma_2^2$
with significance level α
$F = s_1^2/s_2^2$

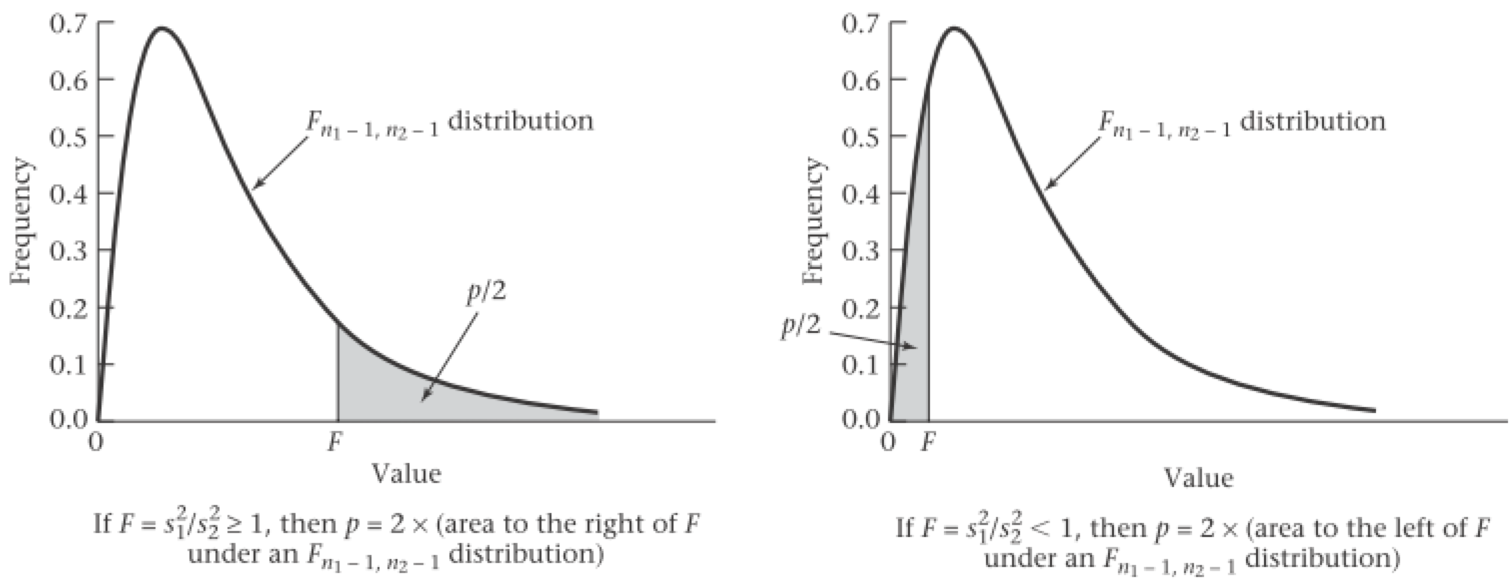
- $F > F_{n1-1,n2-1,1-\alpha/2}$ or $F < F_{n1-1,n2-1,\alpha/2}$ →reject H_0
- $F_{n1-1,n2-1,\alpha/2} \leq F \leq F_{n1-1,n2-1,1-\alpha/2}$ →accept H_0

Acceptance and rejection regions for the *F* test for the equality of two variances



$F_{n_1-1,\,n_2-1}$ distribution = distribution of $F = S_1^2/S_2^2$ under H_0

Figure 8.7 Computation of the *p*-value for the *F* test for the equality of two variances



- $F \geq 1 \rightarrow \quad p = 2 \times Pr(F_{n1-1,n2-1} > F)$
- $F < 1 \rightarrow \quad p = 2 \times Pr(F_{n1-1,n2-1} < F)$

Example on F Test for the Equality of Two Variances: Cardiovascular Disease and Pediatrics

- Familial aggregation of cholesterol levels: suppose cholesterol levels are assessed in 100 children, 2 to 14 years of age, of men who have died from heart disease and it is found that the mean cholesterol level in the group ($\bar{x}_1$) is 207.3 mg/dL.
- Suppose the sample standard deviation in this group (s_1) is 35.6 mg/dL.
- Previously, the cholesterol levels in this group of children were compared with 175 mg/dL, which was assumed to be the underlying mean level in children in this age group based on previous large studies.
- The case and control children come from the same census tract (a geographic region defined for the purpose of taking a census) but are not individually matched → two independent samples but not two paired samples
- Suppose the researchers found that among 74 control children, the mean cholesterol level ($\bar{x}_2$) is 193.4 mg/dL with a sample standard deviation (s_2) of 17.3 mg/dL. We would like to compare the means of these two groups using the two-sample t test for independent samples, but we are hesitant to assume equal variances because the sample variance of the case group is about four times as large as that of the control group: $35.6^2 / 17.3^2 = 4.23$.

Q: Test for the equality of the two variances.

Example on F Test for the Equality of Two Variances: Cardiovascular Disease and Pediatrics

Solution:

$$F = \frac{s_1^2}{s_2^2} = \frac{35.6^2}{17.3^2} = 4.23$$

The two samples have 100 and 74 people, respectively under $H0$, $F \sim F_{99,73}$

→ H_0 is rejected if

$$F > F_{99,73,.975} \text{ or } F < F_{99,73,.025}$$

cannot obtain relevant values from F table

- obtain the percentiles using R:
- find the value $c1 = F_{99,73,.025}$ and $c2 = F_{99,73,.975}$ such that:

$$Pr\,(F_{99,73} \le c_1) = .025 \text{ and } Pr\,(F_{99,73} \ge c_2) = .975$$

- use the qf function of R :

c_1 = qf(0.025, 99, 73)

c_2 = qf(0.975, 99, 73)

Example on F Test for the Equality of Two Variances: Cardiovascular Disease and Pediatrics

```
> qf(0.025, 99, 73)
[1] 0.65476
> qf(0.975, 99, 73)
[1] 1.549079.
```

Thus, $c_1 = 0.655$, $c_2 = 1.549$. Because $F = 4.23 > c_2$ it follows that $p < 0.05$. Alternatively, we could compute an exact *p*-value. This is given by:

$p = 2 \times Pr(F_{99,73} > 4.23) = 2 \times [1 - pf(4.23, 99, 73)]$. The result is shown as follows:

```
> p.value < -2 * (1 - pf(4.23, 99, 73))
> p.value
[1] 8.839514e-10
```

Two-Sample *t* Test for Independent Samples with Equal Variances

Difference between the two sample means: $\bar{x}_1 - \bar{x}_2$

- far from 0 → reject H_0
- Close to 0 → accept H_0

- The two samples are independent:

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

$$\bar{X}_1 - \bar{X}_2 \sim N\left[\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

Under H_0, $\mu_1 = \mu_2$:

$$\bar{X}_1 - \bar{X}_2 \sim N\left[0, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]$$

- σ^2 were known:
  $\overline{X1} - \overline{X2}$ could be divided by $\sigma\sqrt{(1/n_1 + 1/n_2)}$:

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

- σ^2 is unknown (generally):
  need to be estimated from the data using sample variances s_1^2 and s_2^2
  The pooled estimate of the variance from two independent samples:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$

significance level of α for two normally distributed populations

σ^2 is assumed to be the same for each population

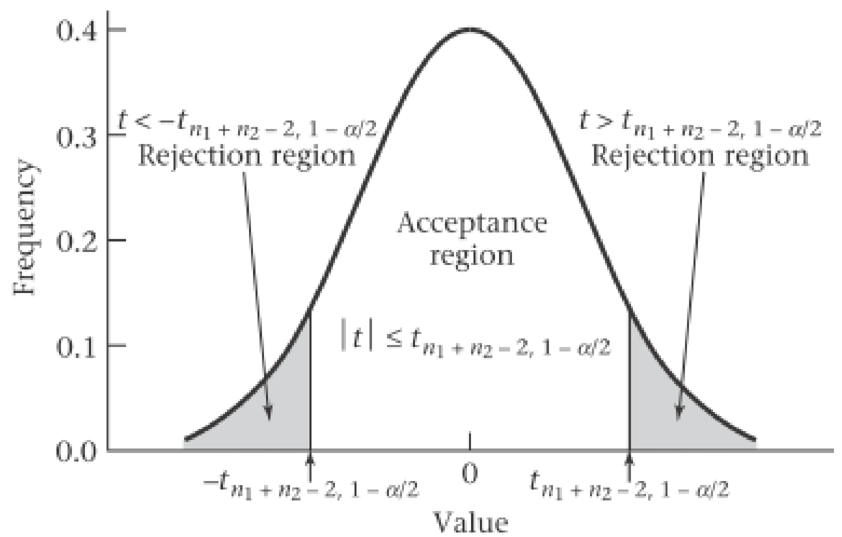
Test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s = \sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right]/(n_1 + n_2 - 2)}$

- $t > t_{n1+n2-2,1-\alpha/2}$ or $t < -t_{n1+n2-2,1-\alpha/2}$ → reject H_0
- $-t_{n1+n2-2,1-\alpha/2} \le t \le t_{n1+n2-2,1-\alpha/2}$ → accept H_0

Figure 8.3 Acceptance and rejection regions for the two-sample *t* test for independent samples with equal variances



Distribution of t in Equation 8.11 under $H_0 = t_{n_1+n_2-2}$ distribution

Computation of the *p*-value for the Two-Sample *t* Test for Independent Samples with Equal Variances

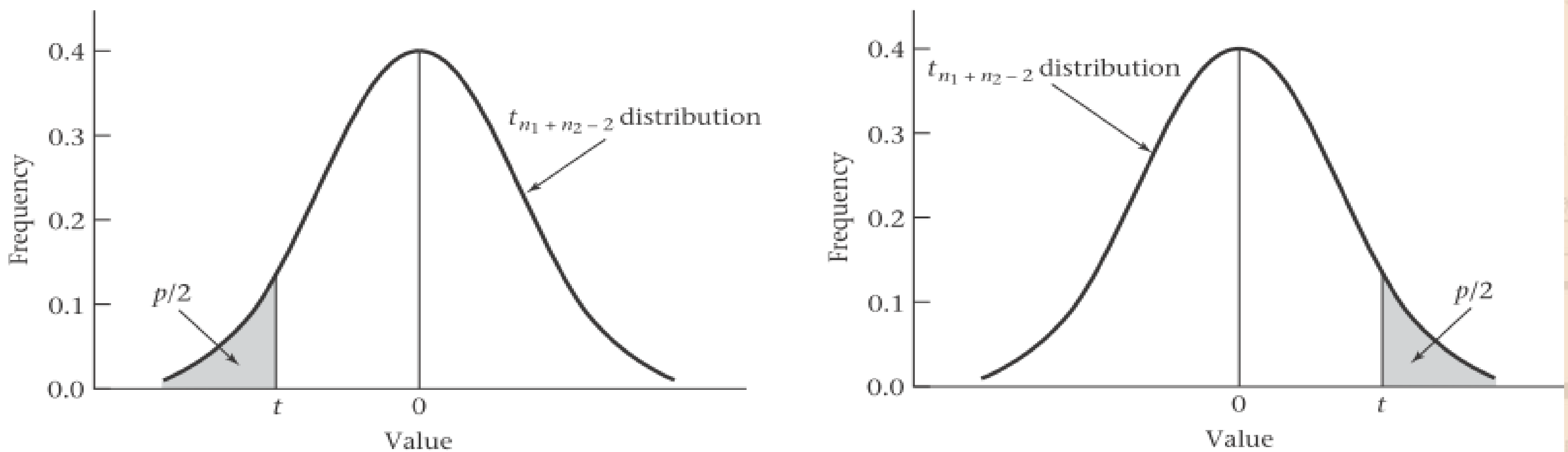
Test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

where $s = \sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right] / (n_1 + n_2 - 2)}$

- $t \le 0$: $p = 2 \times$ (area to the left of t under a $t_{n1+n2-2}$ distribution)
- $t > 0$: $p = 2 \times$ (area to the right of t under a $t_{n1+n2-2}$ distribution)

Computation of the *p*-value for the two-sample *t* test for independent samples with equal variances



If $t = (\bar{x}_1 - \bar{x}_2) \Big/ \left(s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \le 0$, then $p = 2 \times$ (area to the left of t under a $t_{n_1+n_2-2}$ distribution).

If $t = (\bar{x}_1 - \bar{x}_2) \Big/ \left(s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) > 0$, then $p = 2 \times$ (area to the right of t under a $t_{n_1+n_2-2}$ distribution).

Example on Two-sample *t* test for Independent Samples with Equal Variances: Hypertension

Suppose a sample of eight 35- to 39-year-old non-pregnant, premenopausal OC users who work in a company and have a mean systolic blood pressure (SBP) of 132.86 mm Hg and sample standard deviation of 15.34 mm Hg are identified. A sample of 21 nonpregnant, premenopausal, non-OC users in the same age group are similarly identified who have mean SBP of 127.44 mm Hg and sample standard deviation of 18.23 mm Hg.

Q: Assess the statistical significance of the data.

Example on Two-sample *t* test for Independent Samples with Equal Variances: Hypertension

Solution

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s = \sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right]/(n_1 + n_2 - 2)}$

Let's first estimate the common variance:

$$s^2 = \frac{7(15.34)^2 + 20(18.23)^2}{27} = \frac{8293.9}{27} = 307.18, \qquad s = 17.527$$

The following test statistic is then computed:

$$t = \frac{132.86 - 127.44}{17.527\sqrt{1/8 + 1/21}} = \frac{5.42}{17.527 \times 0.415} = \frac{5.42}{7.282} = 0.74$$

- <u>Critical-value method</u>:
  - Under H_0: t comes from a t_{27} distribution, $t_{27,.975} = 2.052$
  - Because $-2.052 \le 0.74 \le 2.052 \rightarrow H_0$ is accepted using a two-sided test at the 5% level
- Conclusion: the mean blood pressures of the OC users and non-OC users do not significantly differ from each other
- <u>*p*-value approximation</u>: $t_{27,.75} = 0.684$, $t_{27,.80} = 0.855$
- Because $0.684 < 0.74 < 0.855 \rightarrow .2 < p/2 < .25$ or $.4 < p < .5$.
- Exact *p*-value from statistical software: $p = 2 \times P(t_{27} > 0.74) = .46$

$1 - .8 = .2$

$1 - .75 = .25$

R command:
>pt(-0.74, 27)

TABLE 5 Percentage points of the t distribution ($t_{d,u}$)[a]

Degrees of freedom, d	u								
	.75	.80	.85	.90	.95	.975	.99	.995	.9995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Interval Estimation for the Comparison of Means from Two Independent Samples (Equal Variance Case)

Two-sided 100% × (1-α) CI for true mean difference μ_1 - μ_2 (two independent samples):

$$\left(\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2,1-\alpha/2}\, s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2,1-\alpha/2}\, s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

- Wide interval: need much larger sample to accurately assess the true mean difference

Example on Interval Estimation for the Comparison of Means from Two Independent Samples

(Equal Variance Case):

Hypertension

Q: Compute a 95% CI for the true mean difference in systolic blood pressure (SBP) between 35- to 39-year-old OC users and non-OC users.

Solution:

95% CI for the underlying mean difference in SBP between the population of 35- to 39-year-old OC users and non-OC users is:

$$[5.42 - t_{27,.975}(7.282), 5.42 + t_{27,.975}(7.282)] = [5.42 - 2.052(7.282), 5.42 + 2.052(7.282)]$$

$$\left(\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2,1-\alpha/2}\, s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2,1-\alpha/2}\, s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

$$= (-9.52, 20.36)$$

- Wide interval
  → much larger sample is needed to accurately assess the true mean difference

Two-Sample *t* Test for Independent Samples with Unequal Variances

Behrens-Fisher problem

Two normally distributed samples

- first sample: random sample of size n_1, $N(\mu_1,\sigma_1^2)$
- second sample: $N(\mu_2,\sigma_2^2)$
- $\sigma_1^2 \neq \sigma_2^2$
- H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$

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

Under H_0, $\mu_1 - \mu_2 = 0$: $$\bar{X}_1 - \bar{X}_2 \sim N\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

If σ_1^2 and σ_2^2 were known, test statistic: $$z = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

σ_1^2 and σ_2^2 are unknown and estimated by s_1^2 and s_2^2:

$$t = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

- Difficult to derive the exact distribution of t under H_0
- To determine t: **Satterthwaite's approximation method** or the two-sample t test for independent samples with unequal variances

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

  - Compute the approximate degrees of freedom d' :

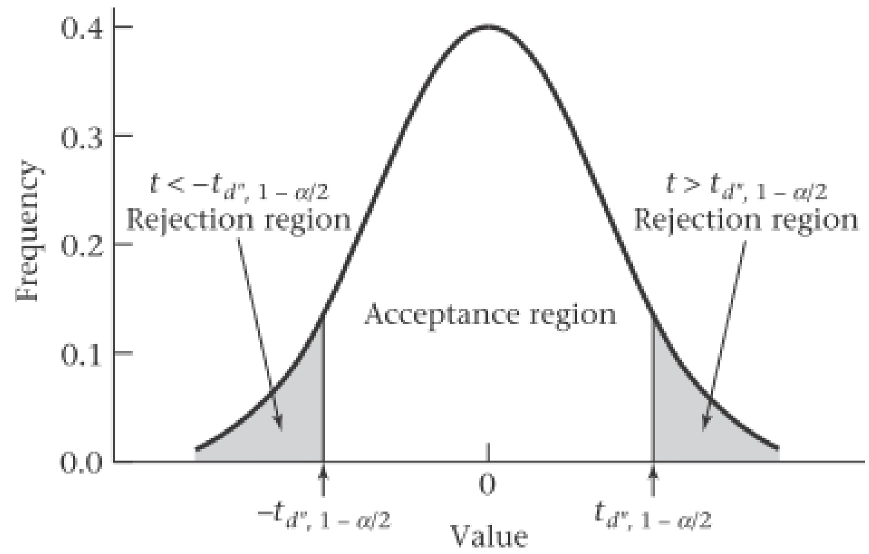
$$d' = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left(s_1^2/n_1\right)^2/(n_1 - 1) + \left(s_2^2/n_2\right)^2/(n_2 - 1)}$$

  - Round d' down to the nearest integer d'' :

$t > t_{d'',1-\alpha/2}$ or $t < -t_{d'',1-\alpha/2}$ → reject H_0

$-t_{d'',1-\alpha/2} \le t \le t_{d'',1-\alpha/2}$ → accept H_0

Figure 8.8 Acceptance and rejection regions for the two-sample *t* test for independent samples with unequal variances



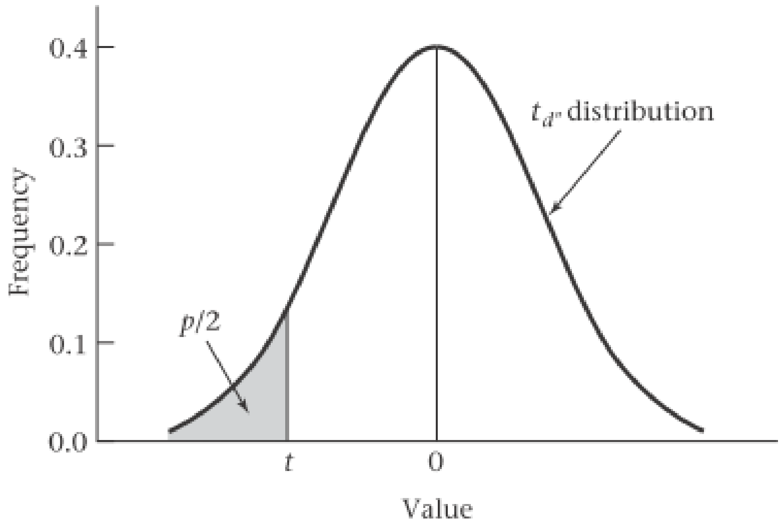
$t_{d''}$ distribution = approximate distribution of t in Equation 8.21 under H_0

- Test statistic:

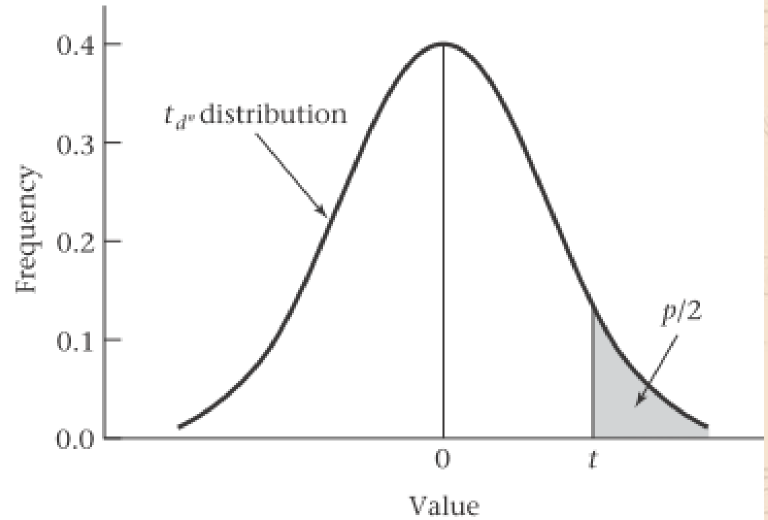
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $t \le 0$ → $p = 2 \times$ (area to the left of t under a $t_{d''}$ distribution)
- $t > 0$ → $p = 2 \times$ (area to the right of t under a $t_{d''}$ distribution)

Computation of the *p*-value for the two-sample *t* test for independent samples with unequal variances



If $t = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2} \le 0$, then $p = 2 \times$ (area to the left of t under a $t_{d''}$ distribution)



If $t = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2} > 0$, then $p = 2 \times$ (area to the right of t under a $t_{d''}$ distribution)

Two-sided 100% × (1-α) CI for μ_1 - μ_2 ($\sigma_1^2 \neq \sigma_2^2$)

$$\left(\bar{x}_1 - \bar{x}_2 - t_{d'',1-\alpha/2}\sqrt{s_1^2/n_1 + s_2^2/n_2},\ \bar{x}_1 - \bar{x}_2 + t_{d'',1-\alpha/2}\sqrt{s_1^2/n_1 + s_2^2/n_2}\right)$$

Example on Two-Sample *t* Test for Independent Samples with Unequal Variances: Cardiovascular Disease, Pediatrics

Q: Consider the cholesterol data. **Test for the equality of the mean cholesterol levels of the children whose fathers have died from heart disease vs. the children whose fathers do not have a history of heart disease.**

Solution:

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

- Tested for equality of the two variances: significantly different
- Need to use two-sample *t* test for unequal variances
- Test statistic:

$$t = \frac{207.3 - 193.4}{\sqrt{\frac{35.6^2}{100} + \frac{17.3^2}{74}}} = \frac{13.9}{4.089} = 3.40$$

- Degrees of freedom are now computed:

$$d' = \frac{(s_1^2/n_1 + s_1^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)} = \frac{(35.6^2/100 + 17.3^2/74)^2}{(35.6^2/100)^2/99 + (17.3^2/74)^2/(73)} = \frac{16.718^2}{1.8465} = 151.4$$

- Critical value method: $t = 3.40 > \boldsymbol{t}_{\mathbf{120,.975}} = 1.980 > \boldsymbol{t}_{\mathbf{151,0.975}}$

TABLE 5 Percentage points of the t distribution ($t_{d,u}$)[a]

Degrees of freedom, d	u = .75	.80	.85	.90	.95	.975	.99	.995	.9995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Example on Two-Sample *t* Test for Independent Samples with Unequal Variances: Cardiovascular Disease, Pediatrics

- use the qt command of R to evaluate $t_{151,.975}$ directly as follows:

```
> qt(0.975, 151)
[1] 1.975799
```

- t = 3.40 > 1.976→ reject H0 using a two-sided test at the 5% level
- exact *p*-value: use pt command of R :

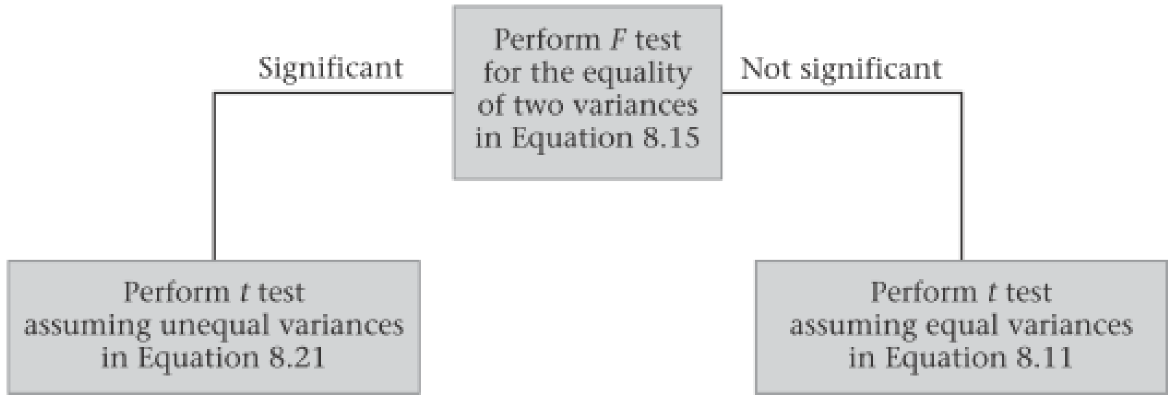
```
> p.value < -2 * (1 - pt(3.40, 151))
> p.value
[1] 0.0008622208
```

- two-sided *p*-value = 0.0009
- Conclusion: mean cholesterol levels in children whose fathers have died from heart disease are significantly higher than mean cholesterol levels in children of fathers without heart disease

  →cause of this difference? genetic factors, environmental factors such as diet, or both?

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

$F = s_1^2/s_2^2$



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

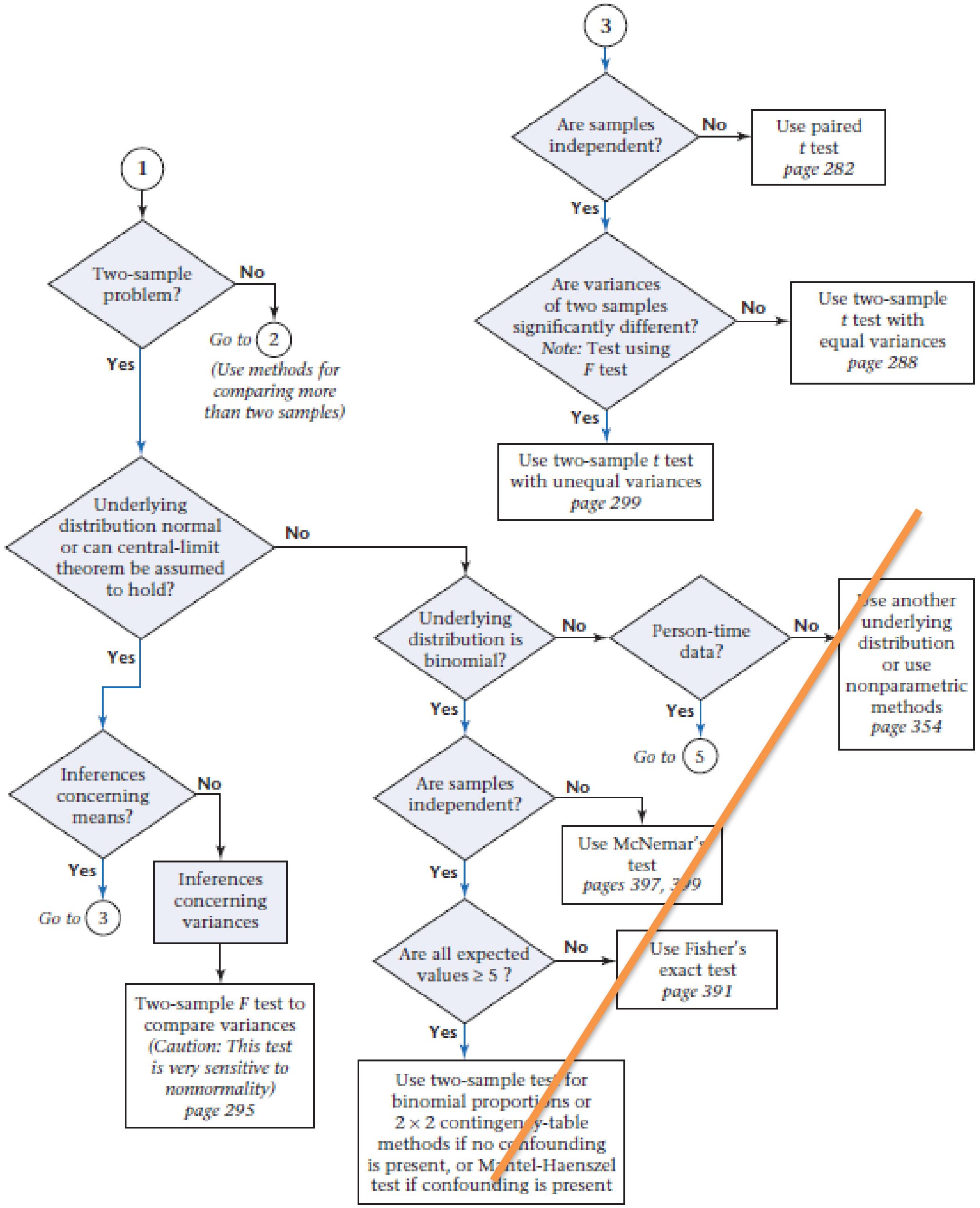
$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s = \sqrt{\left[(n_1-1)s_1^2 + (n_2-1)s_2^2\right]/(n_1+n_2-2)}$

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

- Two procedures for comparing two means from independent and normally distributed
- First step: test for the equality of the two variances ***F test***
  - If this test is not significant →use the *t* test with equal variances
  - If this test is significant → use the *t* test with unequal variances

FIGURE 8.13 Flowchart summarizing two-sample statistical inference—normal-theory methods



Summary

- Methods of hypothesis testing for comparing the means and variances of two normally distributed samples

- Paired t test and F test for two-sample problem:
  - Two samples are paired → paired t test is appropriate
  - samples are independent →F test for the equality of two variances is used to decide whether the variances are significantly different
  - If the variances are not significantly different →two-sample t test with equal variances
  - If the variances are significantly different →two sample t test with unequal variances