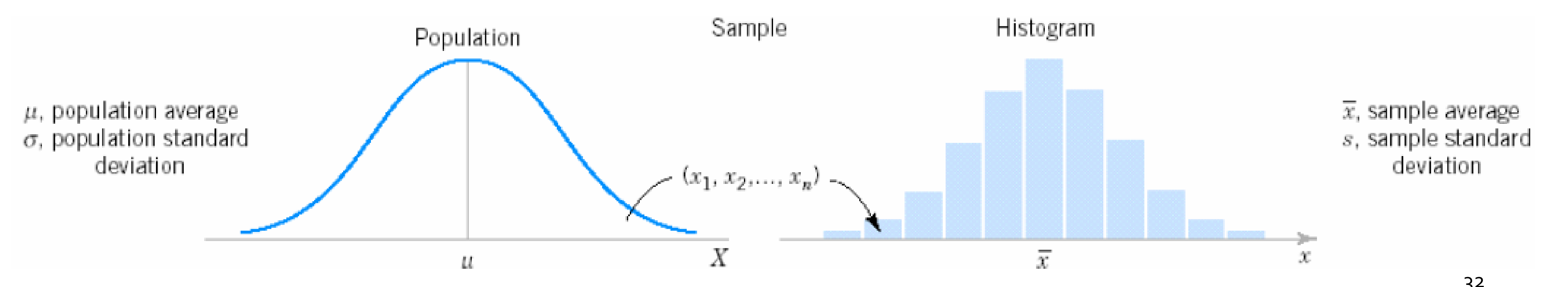
Decision making

Decision making

- The field of statistical inference consists of those methods used to make decisions or draw conclusions about a population.
  - Parameter estimation
  - Hypothesis testing



Estimation of μ and σ

- Is $\overline{X}$ an unbiased estimator of population average μ?
- Is s^2 the unbiased estimation of σ^2 (population variance)?

Point estimation

- A point estimate of some population parameter θ is a single numerical value θ of a statistic Θ.

Unknown Parameter θ	Statistic $\hat{\Theta}$	Point Estimate $\hat{\theta}$
μ	$\bar{X} = \frac{\Sigma X_i}{n}$	$\bar{x}$
σ^2	$S^2 = \frac{\Sigma (X_i - \bar{X})^2}{n - 1}$	s^2
p	$\hat{P} = \frac{X}{n}$	$\hat{p}$
$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2 = \frac{\Sigma X_{1i}}{n_1} - \frac{\Sigma X_{2i}}{n_2}$	$\bar{x}_1 - \bar{x}_2$
$p_1 - p_2$	$\hat{P}_1 - \hat{P}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$	$\hat{p}_1 - \hat{p}_2$

Hypothesis testing

- Statistical hypothesis
  - A statement about the parameters of one or more populations
- Hypothesis testing
  - To accept or reject a statement (hypothesis) about some parameters

Example

- Assume monkey temperature is a random variable that can be described by a probability distribution.
- Suppose that our interest focuses on the mean body temperature.
- Specifically, we are interested in deciding whether or not the mean body temperature is 35°C.
  - H_0: μ = 35°C (null hypothesis)

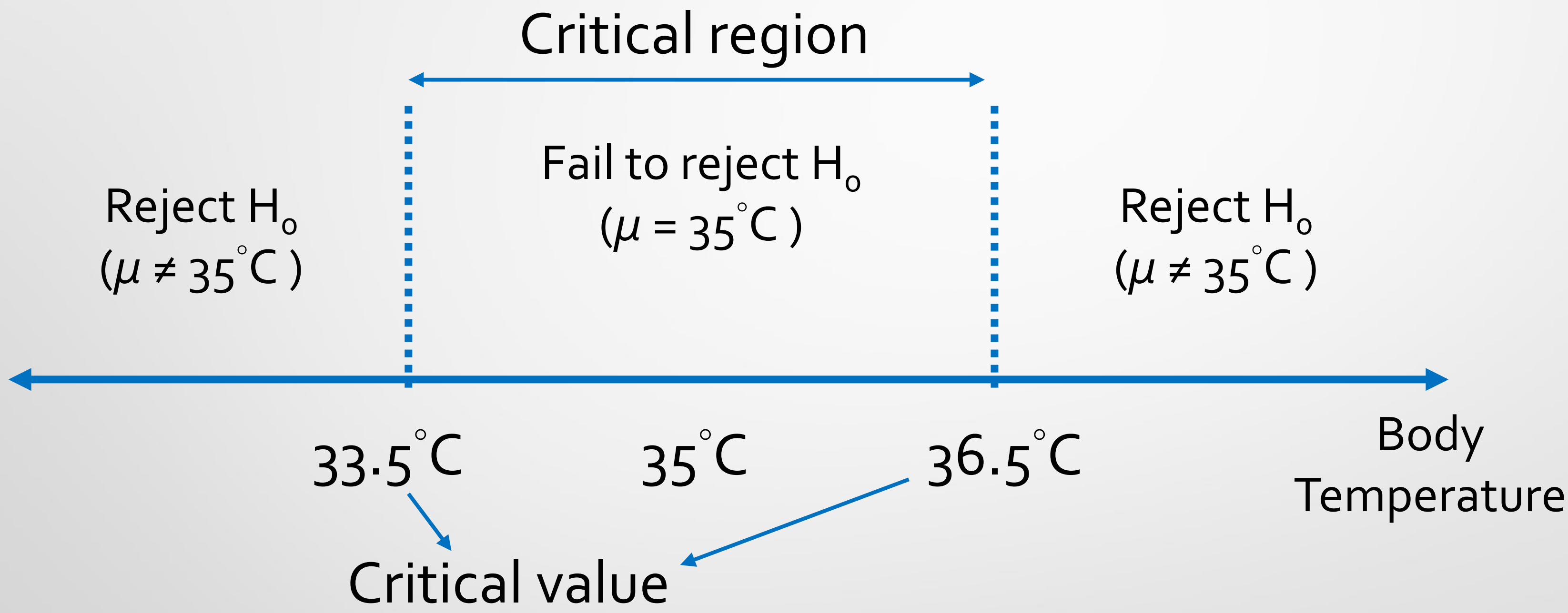
Statistical hypothesis

- Null hypothesis
  - H_0: $\mu = 35^{\circ}C$
- Two-sided alternative hypothesis
  - H_1: $\mu \neq 35^{\circ}C$
- One-sided Alternative Hypotheses
  - H_1: $\mu > 35^{\circ}C$ or $\mu < 35^{\circ}C$
- <u>Hypotheses are always statements about the population, not the sample.</u>

Test of a hypothesis

- Hypothesis-testing procedures rely on using the information in a random sample from the population of interest.
- If this information is consistent with the hypothesis, then we will conclude that the hypothesis is true; else, it is false.

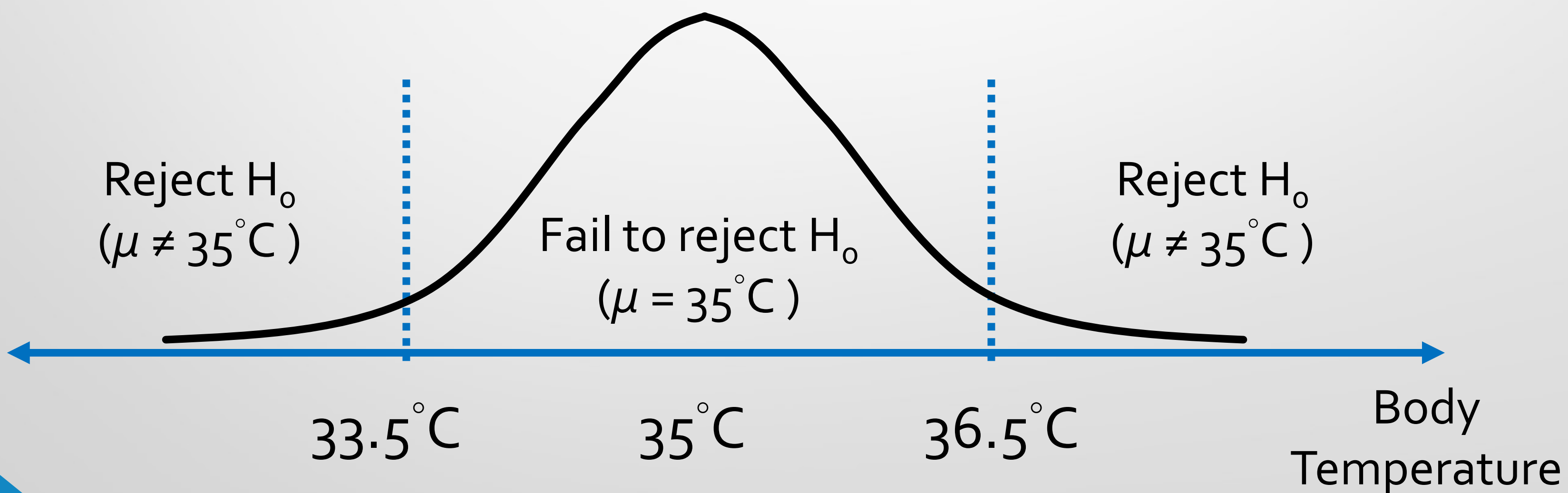
Decision criteria



A critical region has to be defined to reject null hypothesis.

Significance test

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error



Confusion matrix

Actual condition

Prediction

	Actually Positive	Actually Negative
Predictive Positive	**True Positive (TP)**	**False Positive (FP)**
Predictive Negative	**False Negative (FN)**	**True Negative (TN)**

Sensitivity (recall, true positive rate) = $\frac{TP}{TP+FN}$

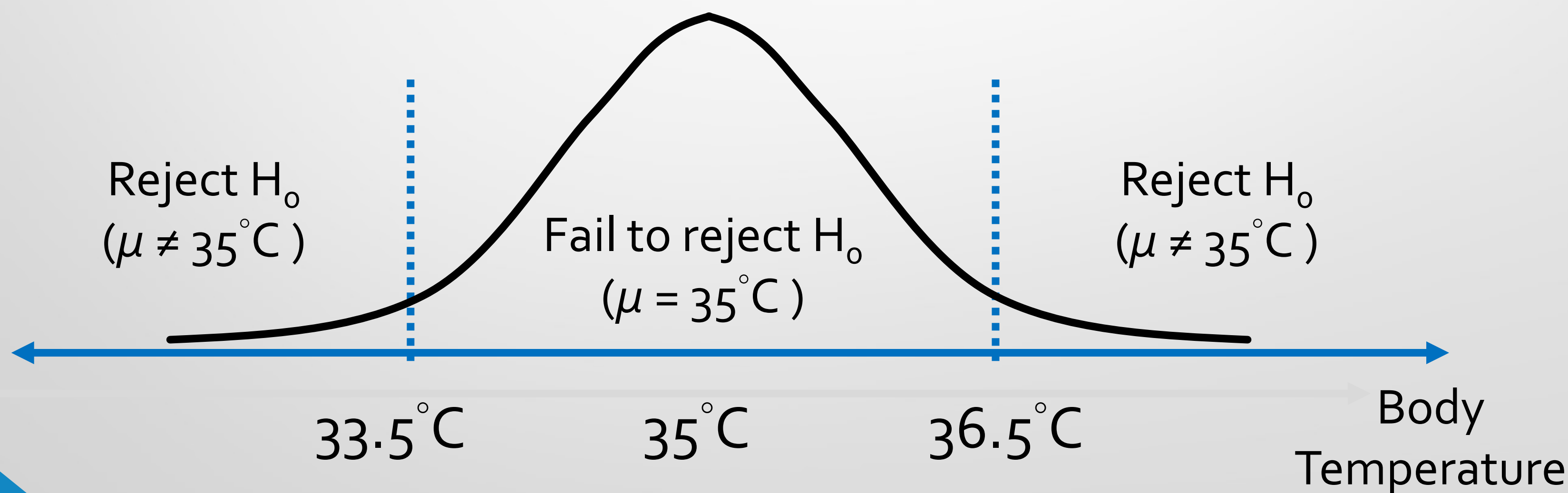
Precision (positive predictive rate) = $\frac{TP}{TP+FP}$

Harmonic mean = F1 score

Specificity (selectivity, true negative rate) = $\frac{TN}{FP+TN}$

Types of error

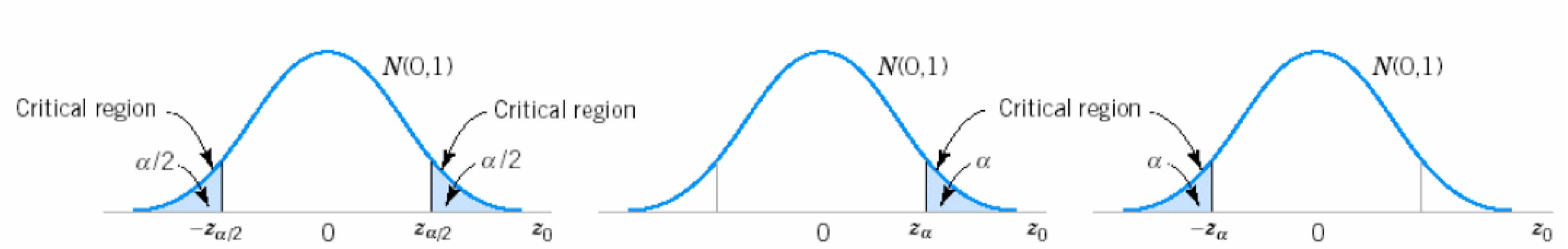
- Type-I error (α)
  - Significance level, α-error
  - $\alpha = P$(rejecting H_o when H_o is true)



Type-I error

Two-sided

One-sided



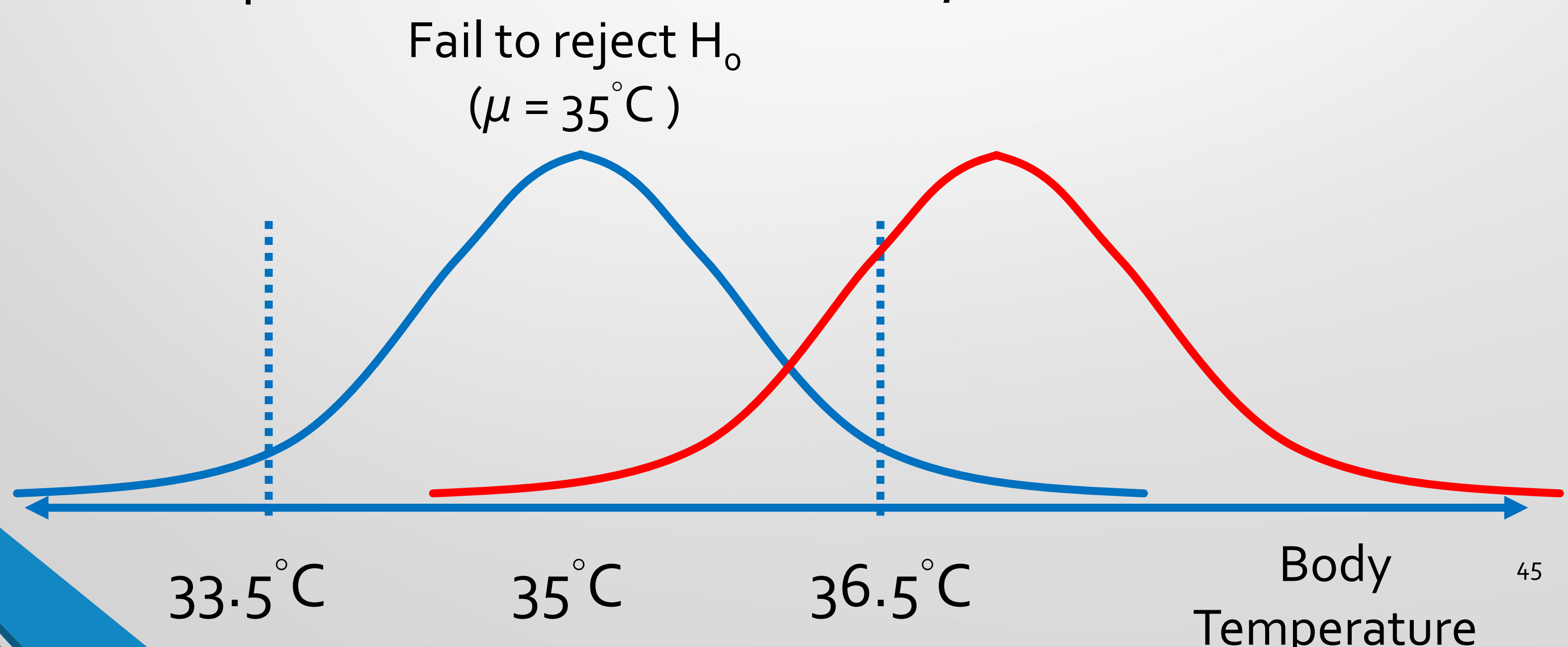
- *P*-value: The smallest level of significance that would lead to rejection of H_o

Type-I error

- Example: The body temperature is measured on 16 healthy monkeys and the critical region is set between 33.5~36.5°C. What is the probability of type-I error when the true mean temperature is 35°C and σ = 3°C?

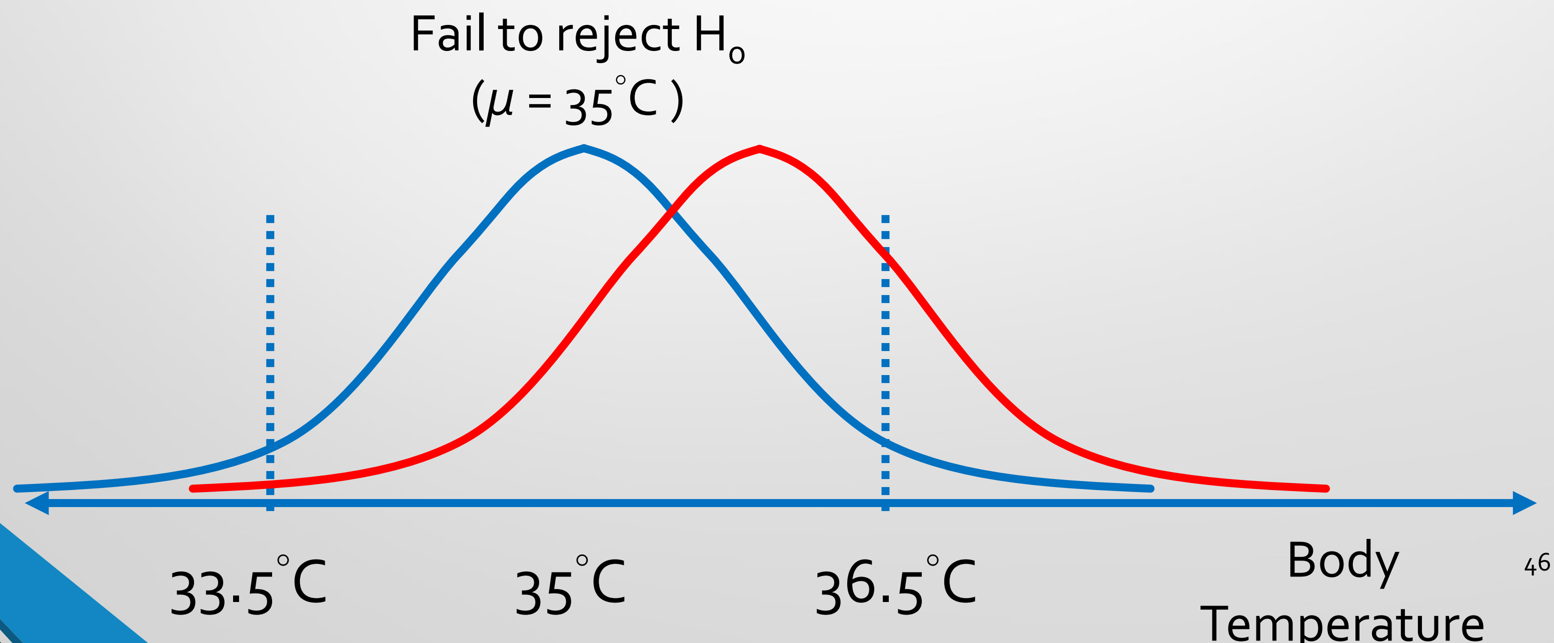
Types of error

- Type-II error (β)
  - β = P(fail to reject H_o when H_o is false)
  - The power of a statistic test: 1- β



Types of error

- β increases rapidly as sample mean approaches the hypothesized value



How to reduce error?

- To reduce type-I error?
  - Push the critical values further toward the tails
  - Increase the sample size
- To reduce type-II error?
  - Push the critical region away from the sample mean
  - Increase the sample size

- Type I and type II errors are related. A decrease in α always results in an increase in β (given constant n)

General procedure for hypothesis testing

1. Identify the parameter of interest
2. State the null hypothesis H_0
3. Specify an appropriate alternative hypothesis H_1
4. Choose a significance level α
5. State the corresponding rejection region
6. Decide whether or not H_0 should be rejected

- Example of propellant burning rate:
  - Sample size = 25, mean = 51.3 cm/sec
  - Population σ = 2 cm/sec, specification = 50 cm/sec
  - Type-I error probability (α) = 0.05

- Solution procedure (reject or not?)
  - Parameter of interest = the mean burning rate μ
  - H_0 :
  - H_1 :
  - Test statistics
  - Reject criterion
  - Conclusion?

- In the previous example, we reject H_0 at 0.05 significance level.
- Still no information about how far off the sample mean is.
- What is the smallest α (significance level) to result in a rejection of H_0?

- *P*-value
  - The smallest level of significance that would lead to rejection of H_0

Confidence interval

- If X is the sample mean of a random sample of size n from a population with known variance σ^2, a 100×(1- α) % confidence interval on μ is given by

$$\bar{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the upper $100 \times \alpha/2$ percentage point of the standard normal distribution.

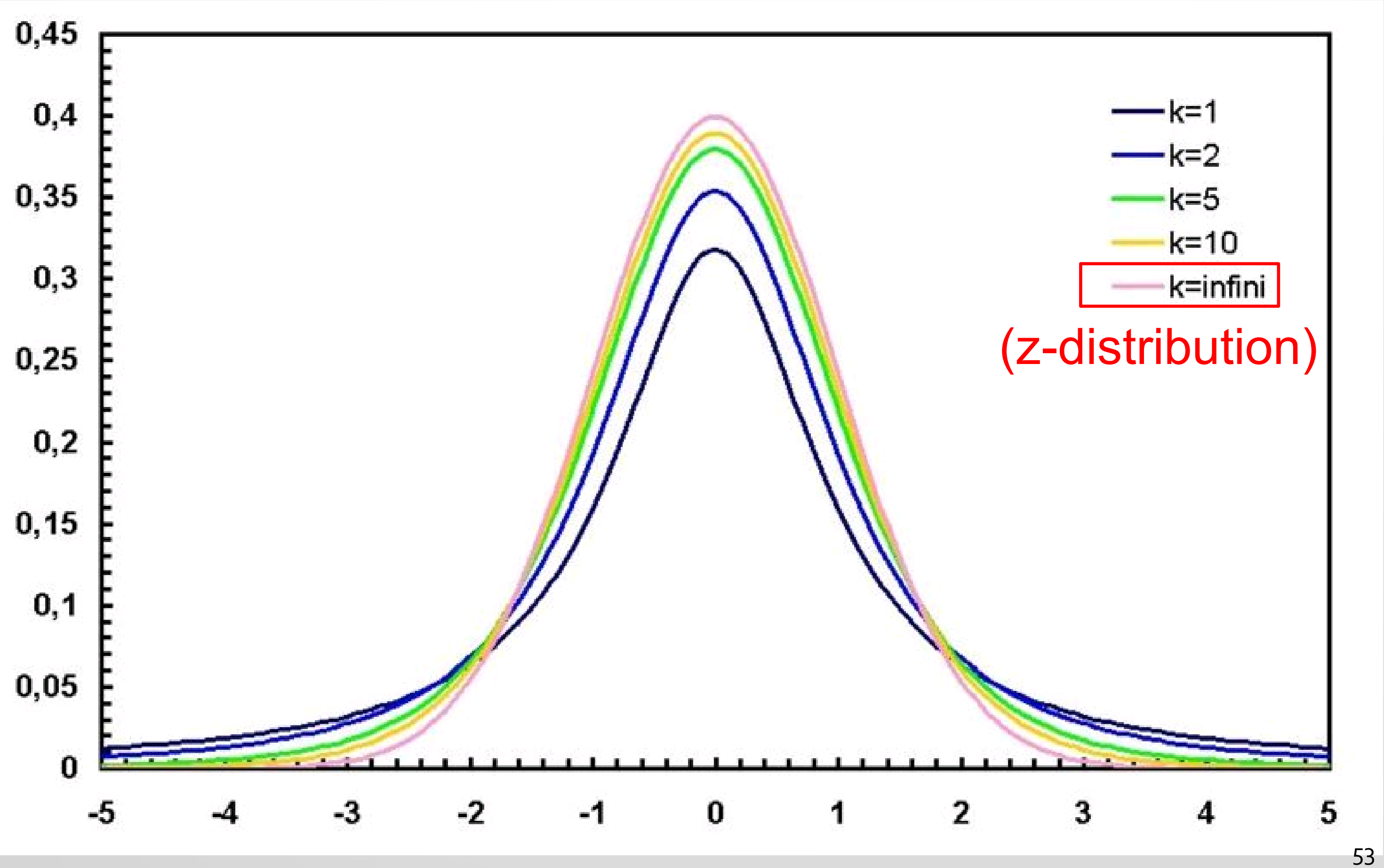
What if variance is unknown?

- When the population mean μ and variance σ^2 is known → z-test
- When the population variance σ^2 is unknown, and $n > 30$ → z-test
- When the population mean μ and variance σ^2 is unknown, and $n < 30$,

  → t-test (assuming a normal distribution)

  – $t = \dfrac{\overline{X} - \mu}{S / \sqrt{n}}$, degree of freedom: $k = n - 1$

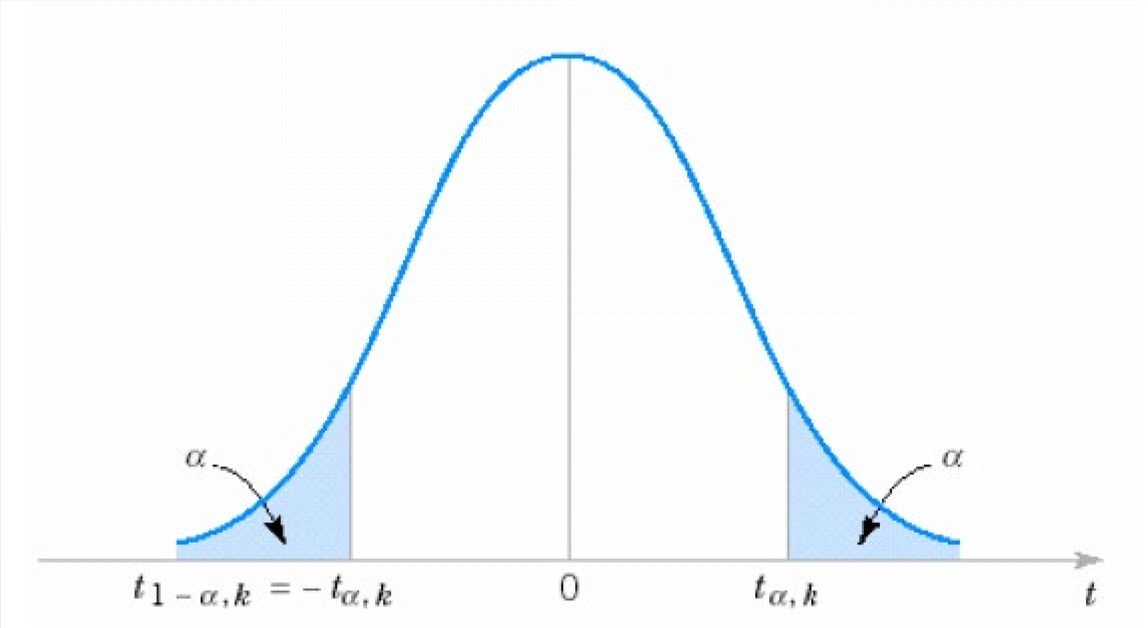
PDF of t distributions



T-test (one sample)

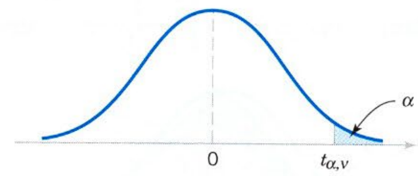
- H_o ($\mu = \mu_o$) is rejected only when

  $t_o > t_{\alpha/2,n-1}$ or $t_o < -t_{\alpha/2,n-1}$ (two-sided hypothesis)



T-distribution

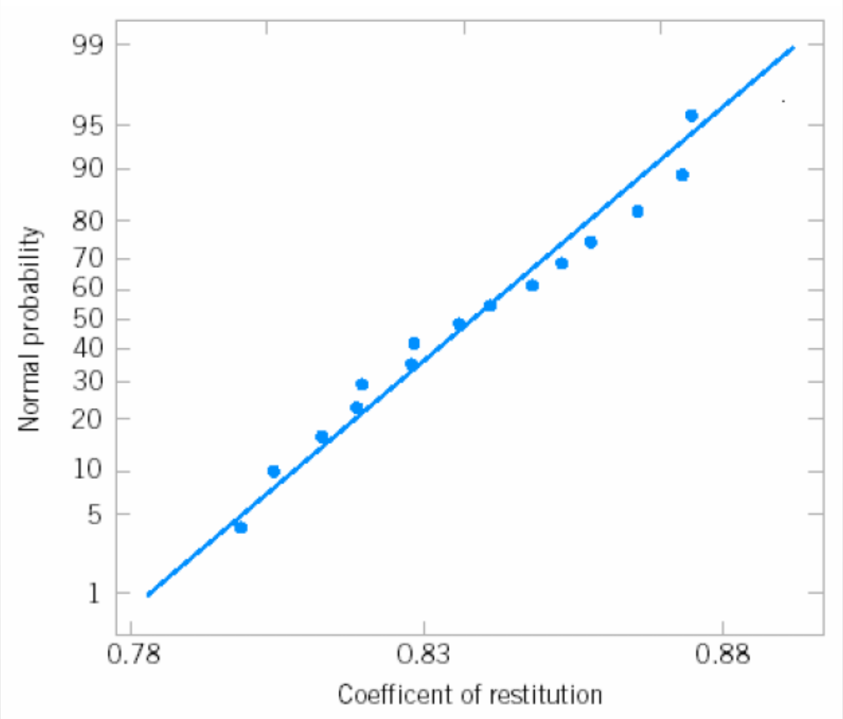
v: degrees of freedom



v \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

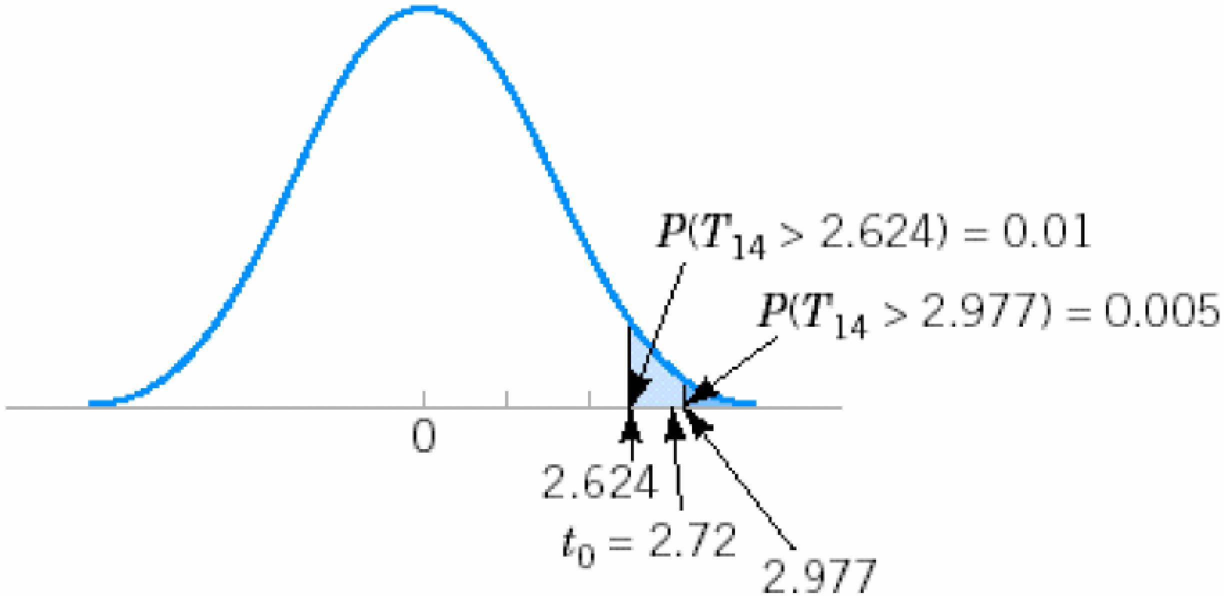
- Example of golf club performance:
  - coefficient of restitution = outgoing speed / incoming speed
  - 15 balls are measured, mean = 0.8375, s = 0.02456
  - Does the mean exceed 0.82?

- Solution procedure
  - Check normality of sample
  - One-sided test: H_0 and H_1 ?
  - $\alpha = 0.05$
  - Test statistics
  - Reject criterion
  - Conclusion?



P-value of a t-test

Critical value	**1.761**	**2.145**	**2.624**	**2.977**
tail area	**0.05**	**0.025**	**0.01**	**0.005**



Confidence interval

- If $\overline{X}$ and S is the sample mean and standard deviation of a random sample of size n from a population with unknown variance σ^2, a 100×(1- α) % confidence interval on μ is given by

$$\overline{X} - \frac{t_{\alpha/2,n-1}s}{\sqrt{n}} \le \mu \le \overline{X} + \frac{t_{\alpha/2,n-1}s}{\sqrt{n}}$$

where $t_{\alpha/2,n-1}$ is the upper $100 \times \alpha/2$ percentage point of the t distribution with n-1 DoF.

Hypothesis testing on the variance of a normal population

- Let $X_1, X_2..., X_n$ be a random sample from normal distribution with unknown mean μ and unknown variance σ^2. The quantity

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

has a chi-square distribution with n-1 degrees of freedom.

Chi-square distribution

$$f(x,k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, x > 0$$

$$\Gamma(m) = \int_0^{\infty} x^{m-1} e^{-x} dx$$

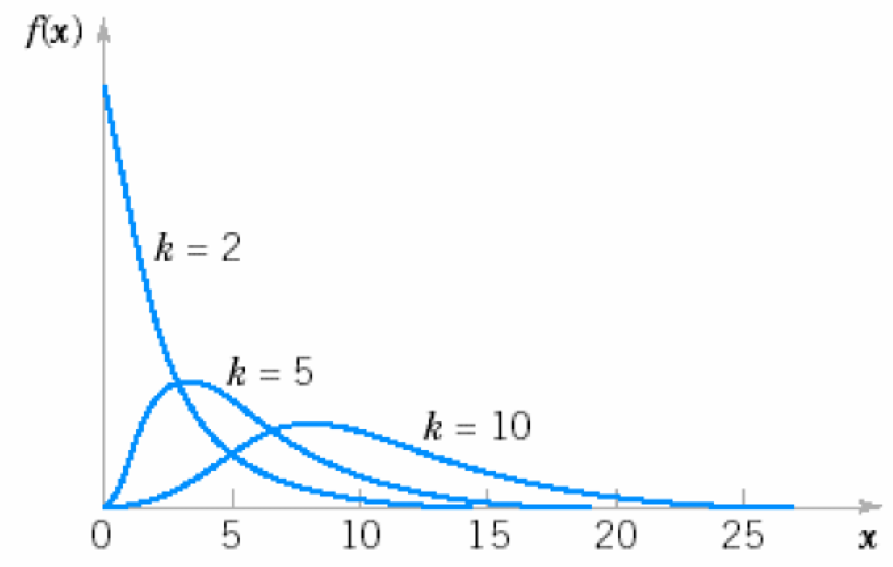
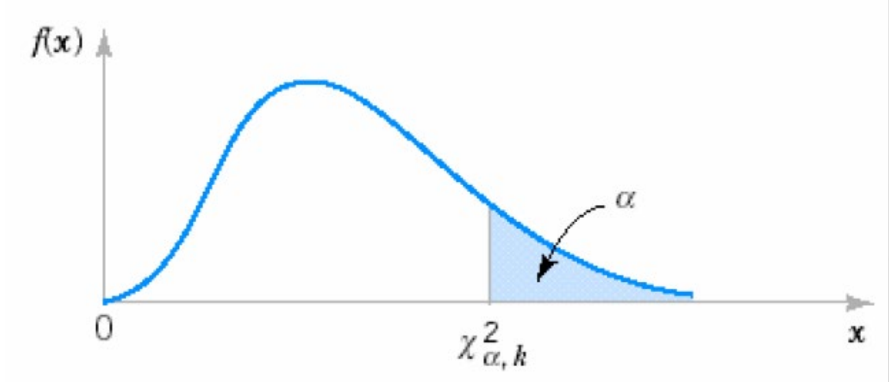


Figure 4-19 Probability density functions of several χ^2 distributions.



Hypothesis testing on the variance

$P(X^2 > \chi^2_{\alpha,k})$

Testing Hypotheses on the Variance of a Normal Distribution

Null hypothesis: $H_0: \sigma^2 = \sigma_0^2$

Test statistic: $$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Alternative Hypotheses	Rejection Criterion
$H_1: \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi_{\alpha/2,n-1}^2$ or $\chi_0^2 < \chi_{1-\alpha/2,n-1}^2$
$H_1: \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi_{\alpha,n-1}^2$
$H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi_{1-\alpha,n-1}^2$

The locations of the critical region are shown in Fig. 4-21.

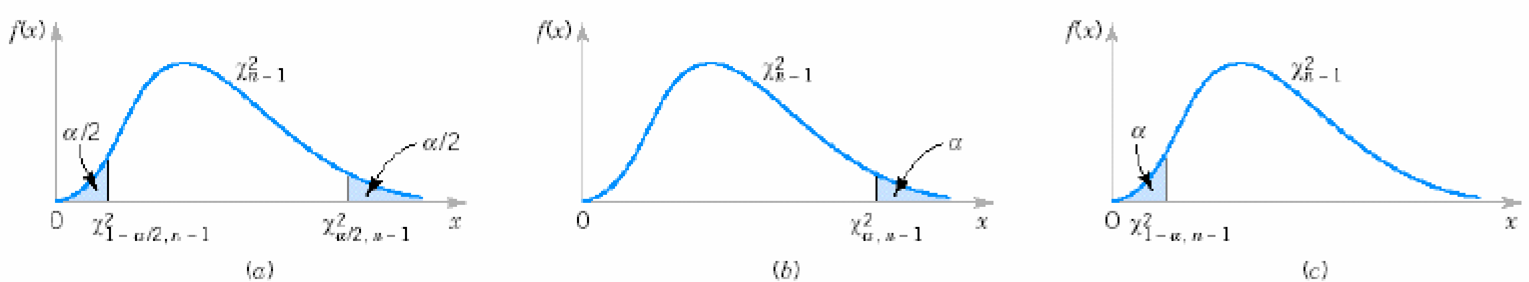


Figure 4-21 Distribution of the test statistic for $H_0: \sigma^2 = \sigma_0^2$ with critical region values for (a) $H_1: \sigma^2 \neq \sigma_0^2$, (b) $H_0: \sigma^2 > \sigma_0^2$, and (c) $H_0: \sigma^2 < \sigma_0^2$.

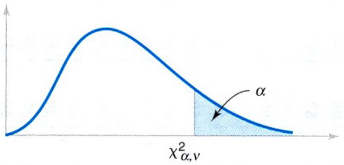
- Example: automatic filling machine fills bottles with detergent
  - 20 samples are measured, $s^2 = 0.0153$
  - Variance > 0.01 will cause unacceptable proportion of under- and over-filled bottles ← From t-test
  - Do the data suggest a problem in filling ($\alpha = 0.05$)?

- Solution procedure
  - H_0 and H_1 ?
  - $\alpha = 0.05$
  - Test statistics
  - Reject criterion
  - Conclusion?

Chi-square distribution table

v: degrees of freedom

Table III Percentage Points $\chi^2_{\alpha,v}$ of the Chi-Square Distribution



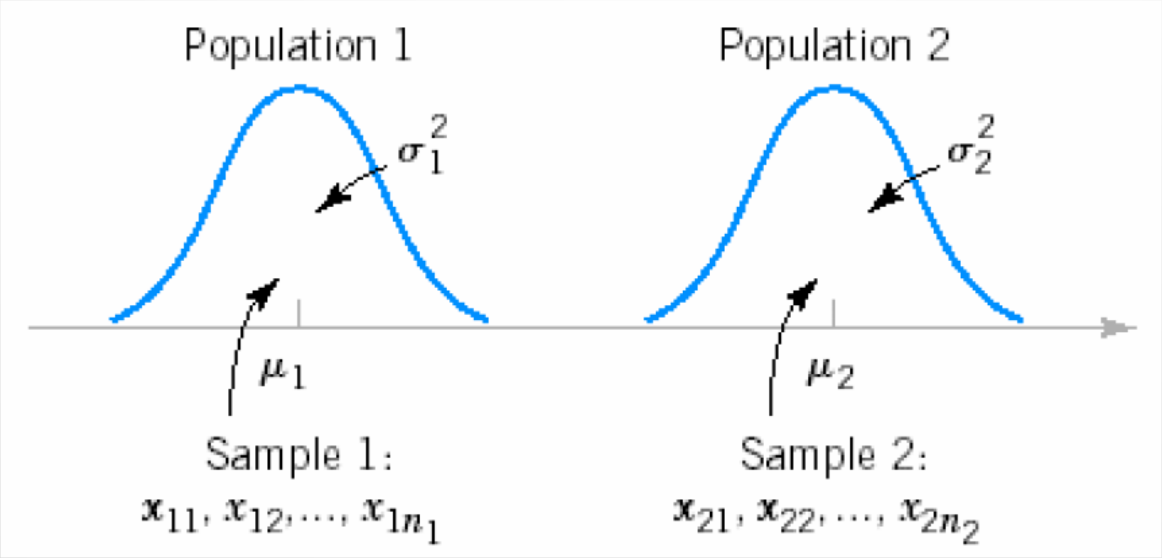
v \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

v = degrees of freedom.

Decision making for two samples

Two-sample tests

- Purpose: to extend the hypothesis testing for the population parameter to the case of two independent populations



Inference on the means of two populations

- When the 2 populations' variances σ_1^2 and σ_2^2 are known → z-test
- When the 2 populations' variances σ_1^2 and σ_2^2 are unknown, and $n > 30$ → z-test
- When the 2 populations' variances σ_1^2 and σ_2^2 are unknown, and $n < 30$ → t-test (assuming a normal distribution)

Case 1: Hypothesis testing on the difference of means, variance known

- Assumptions:
  - The two populations, X_1 and X_2, are independent.
  - Both are normal.

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) = \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}$$

- Treat X_1 - X_2 as a parameter,

$$z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

Standard normal distribution

Two-sample z-test

Testing Hypotheses on the Difference in Means, Variances Known

Null hypothesis: $H_0\colon \mu_1 - \mu_2 = \Delta_0$

Test statistic: $$Z_0 = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Alternative Hypotheses	**Rejection Criterion**
$H_1\colon \mu_1 - \mu_2 \neq \Delta_0$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1\colon \mu_1 - \mu_2 > \Delta_0$	$z_0 > z_\alpha$
$H_1\colon \mu_1 - \mu_2 < \Delta_0$	$z_0 < -z_\alpha$

- Example:
  - A product developer is working to shorten a primer paint's drying time
  - formulation 1 = standard chemistry; formulation 2 = new chemistry
  - σ of drying time = 8 min, independent of formulation
  - 10 specimens of each formulation are randomly tested
  - $\alpha = 0.05$
- Solution procedure
  - One-sided test: H_0 and H_1 ?
  - Test statistics
  - Reject criterion
  - Significance level?

Case 2: Hypothesis testing on the difference of means, variance unknown and equal

- Assumptions:
  - The two populations, X_1 and X_2, are independent.
  - Both are normal.
  - $\sigma_1 = \sigma_2 = \sigma$. The pooled estimator of σ^2 is defined by

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

- Treat X_1 - X_2 as a parameter,

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

t distribution

(DoF = $n_1 + n_2$ - 2)

Case 3: Hypothesis testing on the difference of means, variance unknown and not equal

- Assumptions:
  - The two populations, X_1 and X_2, are independent.
  - Both are normal.
  - $\sigma_1 \neq \sigma_2$
- No exact t-statistic available. Use the following approximation:

t distribution

$$t^* = \frac{\bar{X}_1 - \bar{X}_2 - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

(DoF: $v = \dfrac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$)

Case 4: Hypothesis testing on the difference of means, pair t-test

- Assumptions:
  - The observations, X_1 and X_2, on the two populations are collected in pair.

$$\mu_D = E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$$

S_D: standard deviation of difference

The Paired t-Test

Null hypothesis: $H_0\colon \mu_D = \Delta_0$

Test statistic: $T_0 = \dfrac{\overline{D} - \Delta_0}{S_D/\sqrt{n}}$

Alternative Hypothesis	Rejection Region
$H_1\colon \mu_D \neq \Delta_0$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1\colon \mu_D > \Delta_0$	$t_0 > t_{\alpha,n-1}$
$H_1\colon \mu_D < \Delta_0$	$t_0 < -t_{\alpha,n-1}$

What if we have more than two samples?

- We have learned about testing differences between 2 levels of a factor of interest
- However, there are usually more than 2 levels for a factor
  - Ex: effect of different medications
- How to distinguish?
  - Analysis of variance (ANOVA)

如果生命只剩下最後一小時，我願意花在統計課上，

因為統計課總是讓我覺得度日如年…

生醫影像處理：生醫統計學