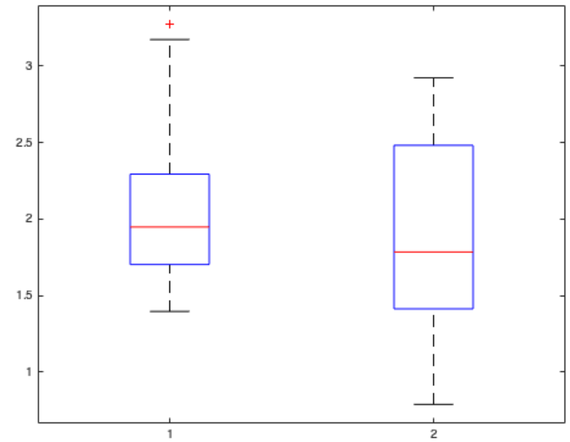
Slides 10: Comparing two populations

- Comparing two independent samples
- Two independent samples: $H_0 : \mu_1 = \mu_2$
- Two independent binary samples: $H_0 : p_1 = p_2$
- Paired samples
- Sample correlation coefficient
- Paired samples: $H_0 : \mu_1 = \mu_2$
- Paired binary samples: $H_0 : p_1 = p_2$
- McNemar's test
- Simpson's paradox



Suppose we wish to compare two population distributions with means and standard deviations (μ_1, σ_1) and (μ_2, σ_2) based on two random samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$ from these two populations.

The difference $\mu_1 - \mu_2$ is estimated by $\bar{x} - \bar{y}$, where

$$\bar{x} = \tfrac{x_1 + \ldots + x_n}{n}, \quad s_{\bar{x}} = \tfrac{s_1}{\sqrt{n}}, \quad s_1^2 = \tfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

$$\bar{y} = \tfrac{y_1 + \ldots + y_m}{m}, \quad s_{\bar{y}} = \tfrac{s_2}{\sqrt{m}}, \quad s_2^2 = \tfrac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{y})^2.$$

If $(X_1, \ldots, X_n)$ is independent from $(Y_1, \ldots, Y_m)$, then

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \tfrac{\sigma_1^2}{n} + \tfrac{\sigma_2^2}{m}$$

and we may compute the standard error of $\bar{x} - \bar{y}$ as

$$s_{\bar{x}-\bar{y}} = \sqrt{\tfrac{s_1^2}{n} + \tfrac{s_2^2}{m}}$$

Question. Is $\bar{x} - \bar{y}$ an unbiased estimate of $\mu_1 - \mu_2$?

Large sample test for two means

If n and m are large, we can use a normal approximation

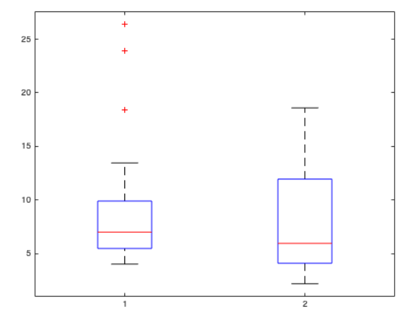
$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_{\bar{X}-\bar{Y}}} \approx \mathrm{N}(0,1).$$

Under the hypothesis of no difference $H_0 : \mu_1 = \mu_2$ the distribution of the test statistic $z = \frac{\bar{x}-\bar{y}}{s_{\bar{x}-\bar{y}}}$ is approximated by the standard normal.

$$\boxed{\text{Approximate confidence interval } I_{\mu_1-\mu_2} \approx \bar{x} - \bar{y} \pm z_{\alpha/2} \cdot s_{\bar{x}-\bar{y}}}$$

Example: iron retention

Percentage of Fe^{2+} and Fe^{3+} retained by mice data at concentration 1.2 millimolar. From the summary of the data:



Fe^{2+}: $n = 18$, $\bar{x} = 9.63$, $s_1 = 6.69$, $s_{\bar{x}} = 1.58$

Fe^{3+}: $m = 18$, $\bar{y} = 8.20$, $s_2 = 5.45$, $s_{\bar{y}} = 1.28$

we obtain

$$\bar{x} - \bar{y} = 1.43, \quad s_{\bar{x}-\bar{y}} = \sqrt{s_{\bar{x}}^2 + s_{\bar{y}}^2} = 2.03$$

According to the large sample test we cannot reject H_0: $\mu_1 = \mu_2$.

For the binomial model $X \sim \mathrm{Bin}(n, p_1)$, $Y \sim \mathrm{Bin}(m, p_2)$, two independently generated values (x, y) give sample proportions

$$\hat{p}_1 = \frac{x}{n}, \qquad \hat{p}_2 = \frac{y}{m},$$

which are unbiased estimates of p_1, p_2 and have standard errors

$$s_{\hat{p}_1} = \sqrt{\tfrac{\hat{p}_1(1-\hat{p}_1)}{n-1}}, \quad s_{\hat{p}_2} = \sqrt{\tfrac{\hat{p}_2(1-\hat{p}_2)}{m-1}}.$$

If the samples sizes m and n are large, then an approximate 95 % confidence interval for the difference $p_1 - p_2$ is given by

$$I_{p_1-p_2} \approx \hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\tfrac{\hat{p}_1(1-\hat{p}_1)}{n-1} + \tfrac{\hat{p}_2(1-\hat{p}_2)}{m-1}}.$$

With help of this formula we can test the null hypothesis of equality

$$H_0 : p_1 = p_2.$$

Question. Would you reject $H_0 : p_1 = p_2$ given a 95 % confidence interval $I_{p_1-p_2} \approx (-0.3, -0.1)$, at what significance level?

Consider two consecutive monthly poll results $\hat{p}_1$ and $\hat{p}_2$ with $n \approx m \approx 5000$ interviews. A change in support to a major political party from $\hat{p}_1$ to $\hat{p}_2$ (with both numbers being close to 40%) is deemed significant at 5% level, if

$$|\hat{p}_1 - \hat{p}_2| > 1.96 \cdot \sqrt{2 \cdot \tfrac{0.4 \cdot 0.6}{5000}} \approx 1.9\%.$$

This should be compared to comparing one poll result with the previous election result $p_0 = 0.4$. Here we apply the one-sample hypothesis for testing $H_0 : p = 0.4$ vs $H_0 : p \neq 0.4$. In view of

$$I_p \approx \hat{p} \pm 1.96 \cdot \sqrt{\tfrac{\hat{p}(1-\hat{p})}{n-1}},$$

with $\hat{p} \approx 0.4$, we conclude that the difference from the election result is significant if

$$|\hat{p} - 0.4| > 1.96 \cdot \sqrt{\tfrac{0.4 \cdot 0.6}{5000}} \approx 1.3\%.$$

Question. Where the difference between two margines of error 1.9% vs 1.3% come from?

Examples of paired observations

different drugs for two patients matched by age, sex,
a fruit weighed before and after shipment,
two types of tires tested on the same car.

Two paired samples can be viewed as one 2D random sample

$$(x_1, y_1), \ldots, (x_n, y_n).$$

Two estimate $\mu_1 - \mu_2$, turn to a 1D sample of differences

$$(d_1, \ldots, d_n), \quad d_i = x_i - y_i.$$

Its sample mean is $\bar{d} = \bar{x} - \bar{y}$. It is an unbiased estimate of $\mu_1 - \mu_2$ whose standard error is computed based on

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \tfrac{1}{n}(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho).$$

taking into account the correlation coefficient $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_1\sigma_2}$.

Question. Why pairing should ensure $\rho > 0$?

To study the effect of cigarette smoking on platelet aggregation, Levine (1973) drew blood samples from $n = 11$ individuals before and after they smoked a cigarette and counted the platelets.

Before y_i	After x_i	$d_i = x_i - y_i$
25	27	2
25	29	4
27	37	10
28	43	15
30	46	16
44	56	12
52	61	9
53	57	4
53	80	27
60	59	−1
67	82	15

Sample correlation coefficient

$$r = \frac{(x_1-\bar{x})(y_1-\bar{y})+\ldots+(x_n-\bar{x})(y_n-\bar{y})}{(n-1)s_1 s_2} = 0.90$$

We test H_0: $\mu_1 = \mu_2$ by applying the large sample test for the mean to

H_0: $\mu = 0$ against H_1: $\mu \neq 0$

where $\mu = \mu_1 - \mu_2$. The test statistic

$$z_{\text{obs}} = \frac{\bar{d}}{s_{\bar{d}}} = \frac{10.27}{2.40} = 4.28$$

gives a very small two-sided p-value, $2 \cdot (1 - \Phi(4.28) = 0.00002$, showing that smoking has a significant health effect.

Question. Where the value $r = 0.9$ was used?

Suppose we have two dependent random variables

$$X \sim \mathrm{Bin}(1, p_1), \quad Y \sim \mathrm{Bin}(1, p_2).$$

Vector (X, Y) takes one of the four possible values $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ with probabilities $\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}$. Observe that

$$p_1 = \pi_{10} + \pi_{11}, \quad p_2 = \pi_{01} + \pi_{11}, \quad p_1 - p_2 = \pi_{10} - \pi_{01}$$

With n independent paired observations, we count $(w_{00}, w_{01}, w_{10}, w_{11})$ the numbers of different outcomes.

An unbiased point estimate of $p_1 - p_2$ is given by

$$\hat{p}_1 - \hat{p}_2 = \hat{\pi}_{10} - \hat{\pi}_{01}, \quad \hat{\pi}_{10} = \frac{w_{10}}{n}, \quad \hat{\pi}_{01} = \frac{w_{01}}{n}.$$

Example. In terms of opinion polls, paired sampling corresponds to asking the same n individuals in January and then in February about their opinion towards a certain political party. The important counts are w_{01} and w_{10} of how many people have changed their preferences.

Paired binary samples: confidence intervall for $p_1 - p_2$

Using the multinomial $\mathrm{Mn}(n, \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$ distribution, we find

$$\mathrm{Var}(W_{10} - W_{01}) = n(\pi_{10} + \pi_{01} - (\pi_{10} - \pi_{01})^2).$$

This yields a formula for the standard error

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01} - (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n-1}}.$$

Again, referring to the CLT we arrive at a 95% confidence interval

$$I_{p_1 - p_2} \approx \hat{\pi}_{10} - \hat{\pi}_{01} \pm 1.96\sqrt{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01} - (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n-1}}$$

Example. Suppose the same $n = 2000$ individuals were asked first in January and then in February about their opinion towards a certain political party. The counts were $w_{01} = 100$ and $w_{10} = 60$. In this case, $\hat{\pi}_{01} = 0.05$ and $\pi_{10} = 0.03$, so that

$$I_{p_2 - p_1} \approx \hat{\pi}_{01} - \hat{\pi}_{10} \pm 1.96\sqrt{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01} - (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n-1}} = 0.03 \pm 0.012$$

Significant difference at 5% level.

The hypothesis of no difference $H_0 : p_1 = p_2$ is equivalent to $H_0 : \pi_{10} = \pi_{01}$. According to

$$I_{p_1 - p_2} \approx \hat{\pi}_{10} - \hat{\pi}_{01} \pm 1.96 \sqrt{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01} - (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n-1}}$$

the rejection region for against $H_0 : \pi_{10} \neq \pi_{01}$ has the form

$$\mathcal{R} = \left\{ \frac{|\hat{\pi}_{10} - \hat{\pi}_{01}|}{\sqrt{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01} - (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n-1}}} > 1.96 \right\}$$

Now notice that the squared left hand side equals

$$\frac{(\hat{\pi}_{10} - \hat{\pi}_{01})^2}{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01} - (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n-1}} \approx \frac{1}{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01}}{n(\hat{\pi}_{10} - \hat{\pi}_{01})^2} - \tfrac{1}{n}} \approx \frac{1}{\tfrac{\hat{\pi}_{10} + \hat{\pi}_{01}}{n(\hat{\pi}_{10} - \hat{\pi}_{01})^2}} = \frac{(w_{10} - w_{01})^2}{w_{10} + w_{01}}.$$

This leads to the McNemar test statistic

$$\mathrm{X}^2 = \frac{(w_{10} - w_{01})^2}{w_{10} + w_{01}}$$

whose null distribution is approximately χ^2_1-distribution.

Double-blind, randomised controlled experiments are used to balance out such external factors as

placebo effect,
time factor,
background variables like temperature,
location factor.

Example. Portocaval shunt is an operation used to lower blood pressure in the liver. People believed in its high efficiency until the controlled experiments were performed.

Enthusiasm level	Marked	Moderate	None
No controls	24	7	1
Nonrandomized controls	10	3	2
Randomized controls	0	1	3

Hospital A has higher overall death rate than hospital B.

However, if we split the data in two parts, patients in good (+) and bad (−) conditions, for both parts hospital A performs better.

Hospital:	A	B	A+	B+	A−	B−
Died	63	16	6	8	57	8
Survived	2037	784	594	592	1443	192
Total	2100	800	600	600	1500	200
Death Rate	.030	.020	.010	.013	.038	.040

Here, the external factor, patient condition, is an example of a confounding factor:

$$\text{Hospital performance} \leftarrow \text{Patient condition} \rightarrow \text{Death rate}$$

Always remember that correlation does not imply causation.