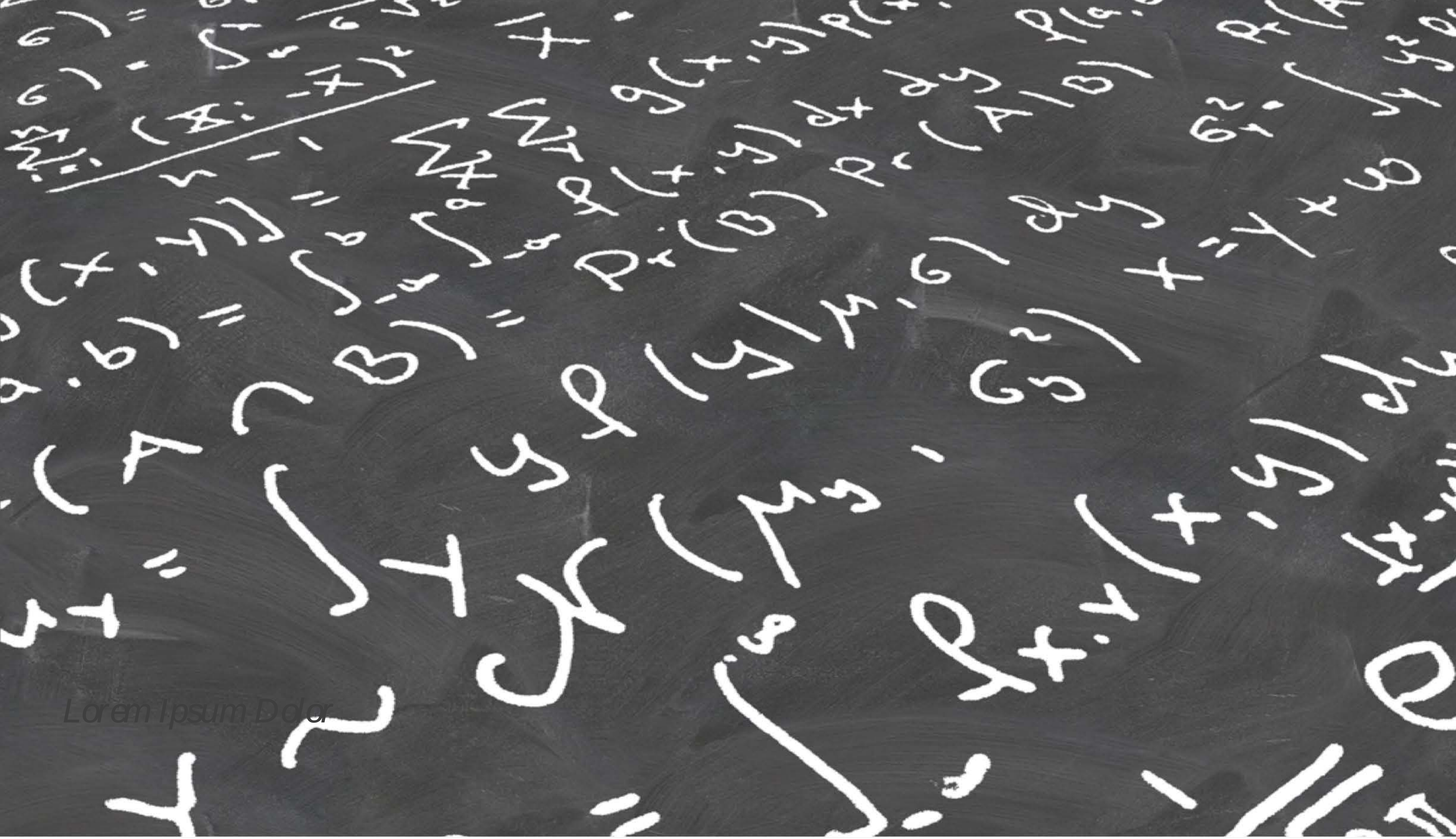


12. Comparison of the Mean of Two Sample Sets

Gunvor Elisabeth Kirkelund
Lars Mandrup
Slides and material provided in parts by
Henrik Pedersen

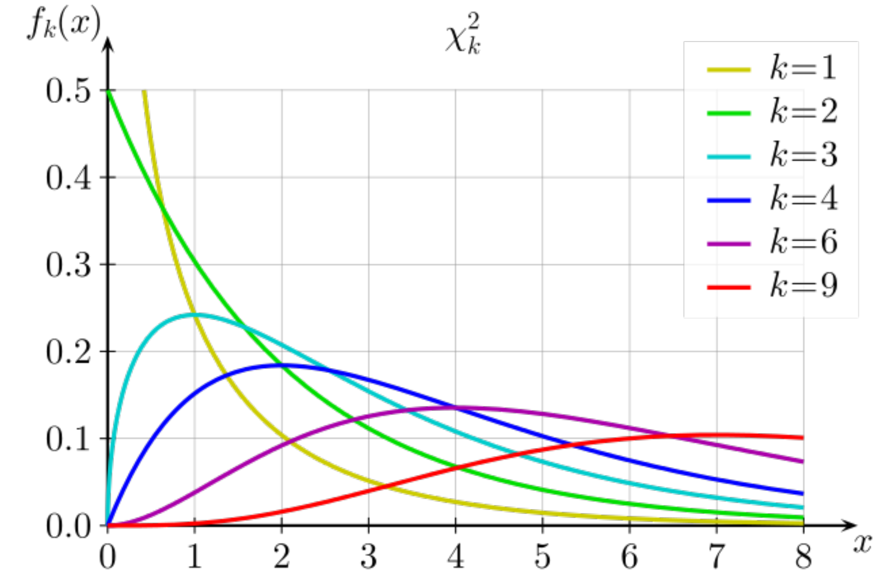
Todays Content

- Repetition from last time
- Comparison the mean of two populations
  - With known variance
  - With unknown variance
- Paired and unpaired data.
- Scientific experimental test

Chi-Square Distribution

- If we have a set of i.i.d. data $X_1, X_2, \dots, X_n$ distributed according to:

$$X_i \sim \mathcal{N}(0,1)$$



- Then we have that: $Q = \sum_{i=1}^{n} X_i^2$

  is χ^2_k distributed with k degrees of freedom.

- **χ^2 -test for independence:** $X^2 = \sum_{i=1}^{n} \frac{(observed - expected)^2}{Expected}$

  How well does the observed data fits the expected values

- **χ^2 -test for variance: Test statistics** $T = (N-1) \cdot \frac{s^2}{\sigma_0^2}$

  Hypothesis test of the variance σ_0^2

The Binomial Distribution

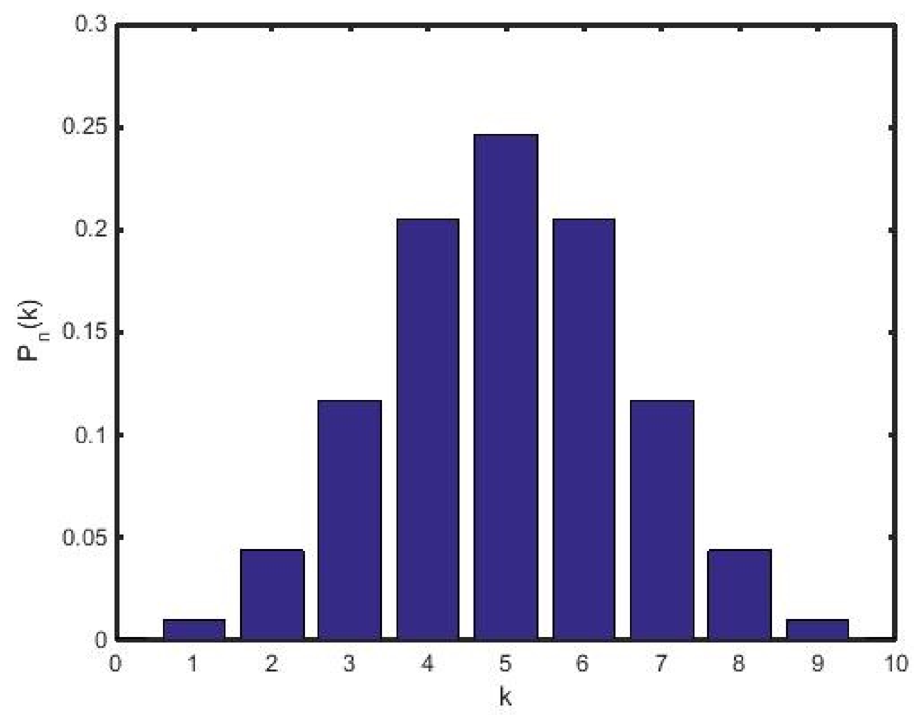
- We have n repeated trials.
- Each trial has two possible outcomes   Bernoulli Event
  - **Success** — probability p
  - **Failure** — probability 1-p
- We write the mass function as:

  X =Number of successes in n trials

$$Pr(X = k) = f(k|n,p)$$

$$f(k|n,p) = \frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$



Test catalog for the Binomial Distribution

- **Statistical model:**
- $X \sim binomial(n, p)$
- Parameter estimate: $\hat{p} = x/n$
- Where the observation is *x* = 'number of successes out of *n* trials'

- **Hypothesis test (two-tailed):**
- $H_0\colon p = p_0$
- $H_1\colon p \neq p_0$
- Test size: $z = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \sim N(0,1)$ (if $np > 5$ and $n(1-p) > 5$)
- Approximate p-value: $2 \cdot |1 - \Phi(|z|)|$

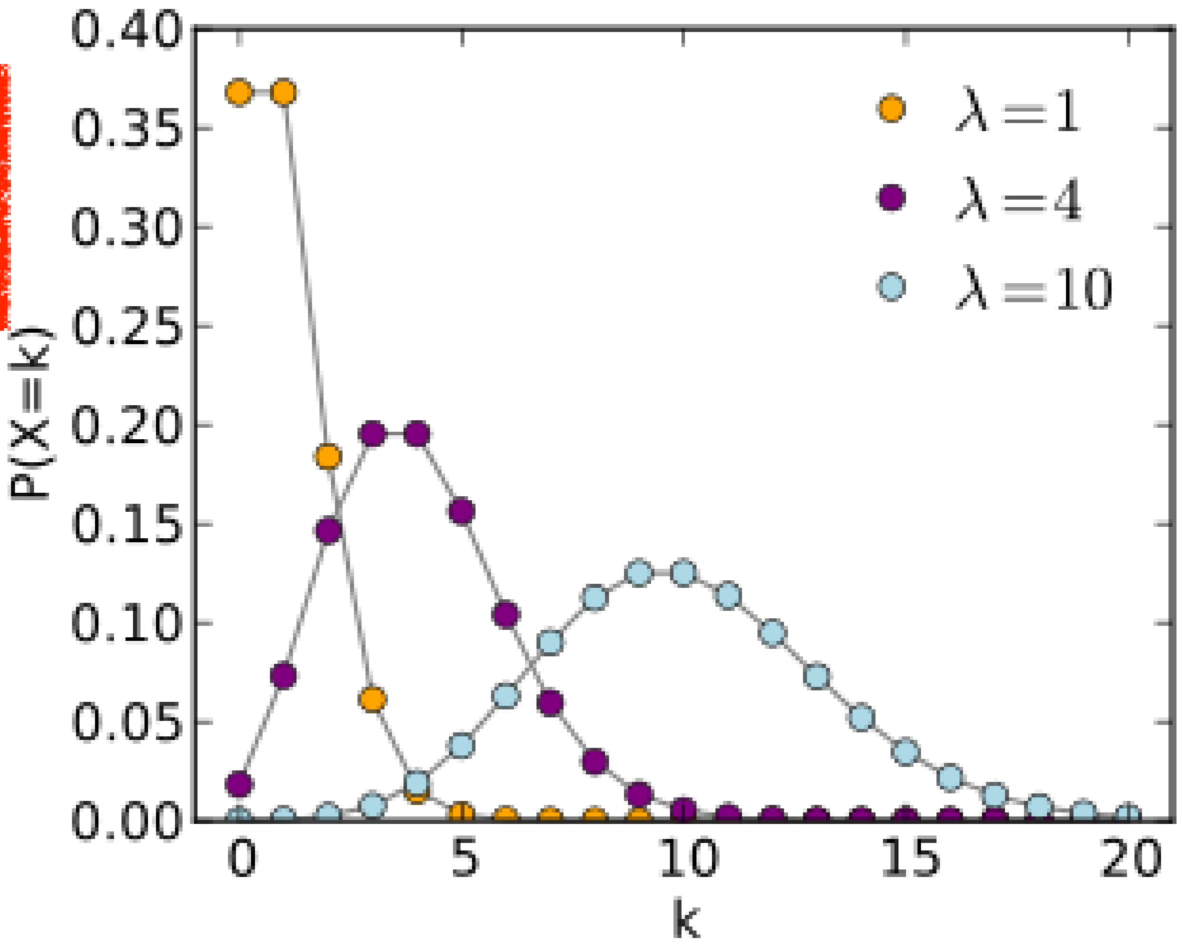
- **95% confidence interval:**
- $p_- = \frac{1}{n+1.96^2}\left[x + \frac{1.96^2}{2} - 1.96\sqrt{\frac{x(n-x)}{n} + \frac{1.96^2}{4}}\right]$
- $p_+ = \frac{1}{n+1.96^2}\left[x + \frac{1.96^2}{2} + 1.96\sqrt{\frac{x(n-x)}{n} + \frac{1.96^2}{4}}\right]$

The Poisson Distribution

- The Poisson distribution is a discrete probability distribution.
- The probability of a given number of events k occurring in a fixed interval of time t.
- If these events occur with a known average rate λ.
- And events are independently of the time since the last event.

- $$\boxed{Pr(X = k) = \frac{(t \cdot \lambda)^k}{k!} e^{-t \cdot \lambda} = \frac{\gamma^k}{k!} e^{-\gamma}}$$

- where $\gamma = t \cdot \lambda$ are the expected number of events in time interval t.



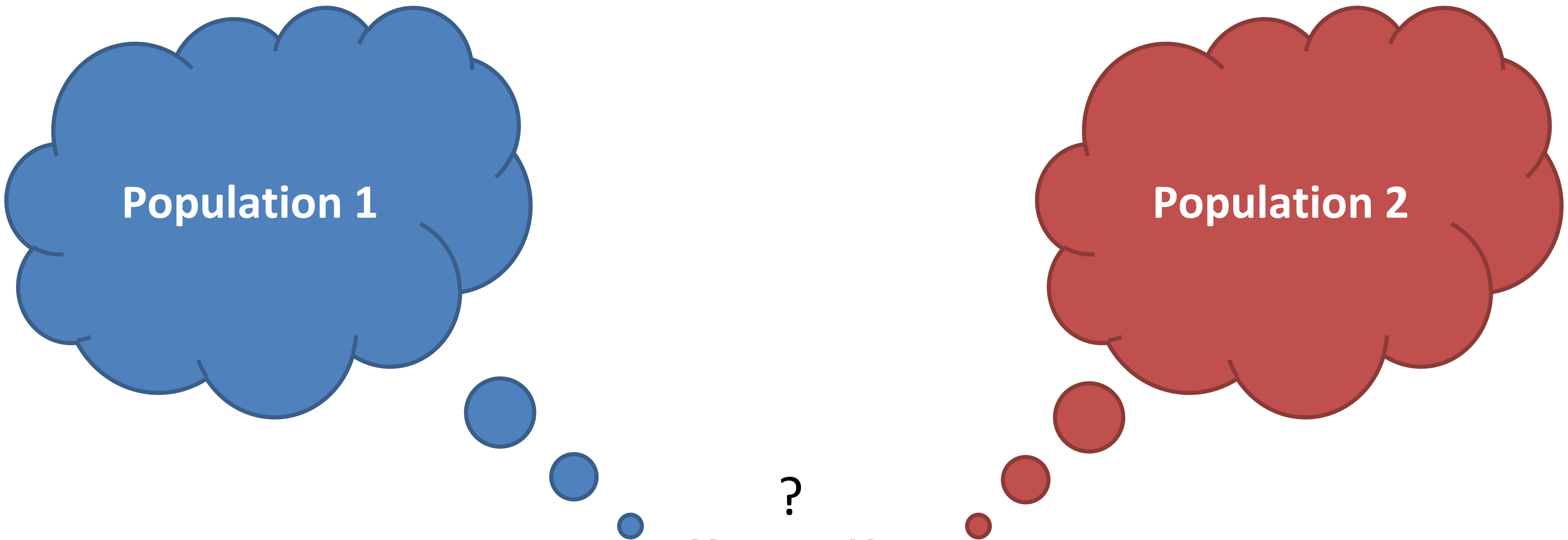
Test Catalog for the Poisson Distribution

- **Statistical model:**
- $X \sim poisson(\lambda \cdot t)$
- Parameter estimate: $\hat{\lambda} = x/t$
- Where the observation is x = 'number of arrivals/events observed over a period of time t'

- **Hypothesis test (two-tailed):**
- $H_0 : \lambda = \lambda_0$
- $H_1 : \lambda \neq \lambda_0$
- Test size: $z = \frac{x-\lambda \cdot t}{\sqrt{\lambda \cdot t}} \sim N(0,1)$ (if $\gamma = t \cdot \lambda > 5$)
- Approximate p-value: $2 \cdot |1 - \Phi(|z|)|$

- **95% confidence interval:**
- $\lambda_- = \frac{1}{t}\left[x + \frac{1.96^2}{2} - 1.96\sqrt{x + \frac{1.96^2}{4}}\right]$
- $\lambda_+ = \frac{1}{t}\left[x + \frac{1.96^2}{2} + 1.96\sqrt{x + \frac{1.96^2}{4}}\right]$

Comparing two population means



$$\mu_1 \stackrel{?}{=} \mu_2$$

- Fx. The height of people from Funen (μ_1) and Jutland (μ_2)

Statistical Model

- Suppose we have two populations with samples $X_{11}, X_{12}, \dots, X_{1n_1}$ drawn from a normally distributed population

$$X_{1i} \sim N\left(\mu_1, \sigma_1^2\right), i = 1,2, \dots, n_1$$

- and samples $X_{21}, X_{22}, \dots, X_{2n_2}$ drawn from a second normally distributed population

$$X_{2i} \sim N\left(\mu_2, \sigma_2^2\right) \;\; i = 1,2, \dots, n_2$$

The exact models not important

- Statistical question of interest:

$$\mu_1 = \mu_2?$$

- Note that in general, the two samples are of different size (i.e., $n_1 \neq n_2$).

Parameter Estimate

- This is the same as asking whether $\mu_1 - \mu_2 = 0$?
- From the central limit theorem, we know that the estimates of the two population means are

$$\hat{\mu}_1 = \bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} \sim N(\mu_1, \sigma_1^2/n_1)$$

$$\hat{\mu}_2 = \bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i} \sim N(\mu_2, \sigma_2^2/n_2)$$

- The estimate of the difference between the two population means is

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

- (the sum of two Gaussian PDFs is another Gaussian).

$E(aX + bY) = a(E(X) + bE(Y);$ $\quad Var(aX + bY) = a^2Var(X) + b^2Var(Y)$

Test Size

Known and identical variances!

- Here we will assume that the variances of the two populations are equal

$$\sigma^2 = \sigma_1^2 = \sigma_2^2$$

Assumption

- Under the null hypothesis H_0: $\boldsymbol{\delta = \mu_1 - \mu_2 = 0}$, **we must have**

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N\left(0, \sigma^2/n_1 + \sigma^2/n_2\right)$$

- Standardizing, we get

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - E[\bar{x}_1 - \bar{x}_2]}{\sqrt{Var(\bar{x}_1 - \bar{x}_2)}} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$$

Hypothesis Test for Comparing Two Population Means

with known and identical variances

- Suppose we observe sample means $\bar{x}_1 = 3$ and $\bar{x}_2 = 4$ from two normally distributed populations with standard deviation 1.
- Furthermore, let us assume that $n_1 = 10$ and $n_2 = 20$.
- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = 0$$

Null Hypothesis

- Z-score

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} = \frac{3 - 4}{1 \cdot \sqrt{1/10 + 1/20}} = -2.582$$

- P-value

$$2 \cdot \left(1 - \Phi(|z|)\right) = 2 \cdot \left(1 - \Phi(2.582)\right) = 2 \cdot (1 - 0.9951) = 0.0098$$

- Since p<0.05, we reject the null hypothesis and conclude that the two populations do not have identical mean values.

for the difference between two population means

95% Confidence Interval

Known and identical variances!

- True difference

$$\delta = \mu_1 - \mu_2$$

Statistical parameter in question

- Estimator

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2\right)$$

- To find the 95% confidence interval of δ, we need to find limits δ_- and δ_+ such that

$$\Pr(\delta_- \leq \delta \leq \delta_+) = 0.95$$

- Standardizing, we get

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

for the difference between two population means

95% Confidence Interval

Known and identical variances!

- Same trick as in the previous chapters

$$0.95 = \Pr(-1.96 \leq Z \leq 1.96) = \Pr\left(-1.96 \leq \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sigma\sqrt{1/n_1 + 1/n_2}} \leq 1.96\right)$$

- Isolating, we get

$$0.95 = \Pr\Big((\bar{x}_1 - \bar{x}_2) - 1.96 \cdot \sigma\sqrt{1/n_1 + 1/n_2} \leq \delta \\ \leq (\bar{x}_1 - \bar{x}_2) + 1.96 \cdot \sigma\sqrt{1/n_1 + 1/n_2}\Big)$$

- Hence, the endpoints of the 95% confidence interval are

$$\delta_- = (\bar{x}_1 - \bar{x}_2) - 1.96 \cdot \sigma\sqrt{1/n_1 + 1/n_2}$$
$$\delta_+ = (\bar{x}_1 - \bar{x}_2) + 1.96 \cdot \sigma\sqrt{1/n_1 + 1/n_2}$$

for the difference between two population means

95% Confidence Interval

Known and identical variances!

- We observe sample means $\bar{x}_1 = 3$ and $\bar{x}_2 = 4$ from two normally distributed populations with standard deviation 1, $n_1 = 10$, and $n_2 = 20$.

- The endpoints of the 95% confidence interval for the true difference between the population means (δ), are

Lower bound:

$$\delta_- = (\bar{x}_1 - \bar{x}_2) - 1.96 \cdot \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (3 - 4) - 1.96 \cdot 1 \cdot \sqrt{\frac{1}{10} + \frac{1}{20}} = -1.7591$$

Upper bound:

$$\delta_+ = (\bar{x}_1 - \bar{x}_2) + 1.96 \cdot \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (3 - 4) + 1.96 \cdot 1 \cdot \sqrt{\frac{1}{10} + \frac{1}{20}} = -0.2409$$

- Since $\delta = 0$ is not included in the 95% confidence interval, we reject the null hypothesis stating that the population means are equal, i.e., $H_0: \mu_1 - \mu_2 = 0$

Test catalog for Comparing Two Means (known variance)

Statistical model:

- $X_{1i} \sim N\left(\mu_1, \sigma_1^2\right), i = 1,2,\ldots,n_1$ and $X_{2i} \sim N\left(\mu_2, \sigma_2^2\right)$ $i = 1,2,\ldots,n_2$
- Parameter estimate: $\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2\right)$
- Where the observation is $\bar{x}_1 - \bar{x}_2$ = 'the difference between two sample means'.

Hypothesis test (two-tailed):

- $H_0: \mu_1 = \mu_2$
- $H_1: \mu_1 \neq \mu_2$
- Test size: $z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$
- Approximate p-value: $2 \cdot |1 - \Phi(|z|)|$

95% confidence interval:

- $\delta_- = (\bar{x}_1 - \bar{x}_2) - 1.96 \cdot \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- $\delta_+ = (\bar{x}_1 - \bar{x}_2) + 1.96 \cdot \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

also called pooled variance

Sample Variance

Unknown variances!

- In the case, where the true variance is unknown we have to estimate it.
- The formula for the pooled variance estimator is

Pooled variance

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

weighted mean of s_1 and s_2

- where

Sample variance for 1: $s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2$

Sample variance for 2: $s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2$

Test Size

Unknown variances!

- The effect of using the empirical variance instead of the true variance is that we have to use the t-score instead of the z-score

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{s\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$$

- where *s* is the empirical standard deviation. Pooled variance
- Notice that the number of degrees of freedom for the t distribution is $n_1 + n_2 - 2$.

Hypothesis Test and Confidence Interval for Comparison of two Means with Unknown and Identical Variances

- Like before, suppose we observe sample means $\bar{x}_1 = 3$ and $\bar{x}_2 = 4$ from two normally distributed populations with empirical variances

$$s_1^2 = 1.4 \text{ and } s_2^2 = 1.1$$

- where $n_1 = 10$ and $n_2 = 20$. Assuming equal variances, the pooled estimate of the variance is

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right)$$
$$= \frac{1}{10 + 20 - 2}\left((10 - 1) \cdot 1.4 + (20 - 1) \cdot 1.1\right) = 1.1964$$

- and the empirical standard deviation is $\sqrt{1.1964} = 1.0938$.

Hypothesis Test and Confidence Interval for Comparison of two Means with Unknown and Identical Variances

- Null hypothesis:

$$H_0: \delta = \mu_1 - \mu_2 = 0$$

- t-score

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s\sqrt{1/n_1 + 1/n_2}} = \frac{3 - 4}{1.0938 \cdot \sqrt{1/10 + 1/20}}$$
$$= -2.3606 \sim t(n_1 + n_2 - 2)$$

degrees of freedom

- P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right) = 2 \cdot \left(1 - t_{cdf}(2.3606, 30 - 2)\right)$$
$$= 2 \cdot (1 - 0.9873) = 0.0254$$

- Since $p<0.05$, we reject the null hypothesis and conclude that the two populations do not have identical mean values.

Hypothesis Test and Confidence Interval for Comparison of two Means with Unknown and Identical Variances

- The endpoints of the 95% confidence interval for the true difference between the population means (δ), are

Lower bound:

$$\delta_- = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (3-4) - 2.0484 \cdot 1.0938 \cdot \sqrt{\frac{1}{10} + \frac{1}{20}}$$
$$= -1.8678$$

Upper bound:

$$\delta_+ = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (3-4) + 2.0484 \cdot 1.0938 \cdot \sqrt{\frac{1}{10} + \frac{1}{20}}$$
$$= -0.1322$$

- where t0 = tinv(0.975,n1+n2-2) = tinv(0.975,30-2) = 2.0484.
- Since $\delta = 0$ is not included in the 95% confidence interval, we reject the null hypothesis stating that the population means are equal.

Test Catalog for Comparing Two Means (unknown variance)

Statistical model:

- $X_{1i} \sim N\left(\mu_1, \sigma_1^2\right), i = 1,2,\ldots,n_1$ and $X_{2i} \sim N\left(\mu_2, \sigma_2^2\right)\ i = 1,2,\ldots,n_2$
- Parameter estimate:

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

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

- Where the observation is $\bar{x}_1 - \bar{x}_2$ = 'the difference between two sample means'.

Hypothesis test (two-tailed):

- $H_0 : \mu_1 = \mu_2$
- $H_1 : \mu_1 \neq \mu_2$
- Test size: $t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{1/n_1+1/n_2}} = \sim t(n_1 + n_2 - 2)$
- Approximate p-value: $2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right)$

95% confidence interval:

- $\delta_- = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- $\delta_+ = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- where t0 = tinv(1-0.05/2,n1+n2-2)

OBS:
t-test (compared with Z-test)

- *Less knowledge*
- *Larger uncertaincy*
- *Confidence interval larger*
- *More difficult to reject H_o*

Paired vs. Unpaired Tests

- The best way to look at the effect of, say, a medical treatment is to measure some physiological parameter *in the same patient* before and after treatment.
- Consider the alternative; one group of patients gets the treatment, the other group doesn't.
- Comparing the mean of the physiological parameter between the two groups could be both due to the treatment *and* differences between the patient groups.
- If a difference was detected, there would be no way of telling whether that difference was due to the treatment or differences between the patient groups.

Paired Difference Test

Each point in one data-set correspond to a point in the other

- Paired difference tests are often used to compare "before" and "after" scores in experiments to determine whether significant change has occurred.
- The data are paired.
  - If $X_{11}, X_{12}, \ldots, X_{1n}$ and $X_{21}, X_{22}, \ldots, X_{2n}$ are the two samples, then X_{1i} corresponds to X_{2i}.

Stalk	1	2	3	4	5	6	7	8	9	10
Before height	35.5	31.7	31.2	36.3	22.8	28.0	24.6	26.1	34.5	27.7
After height	45.3	36.0	38.6	44.7	31.4	33.5	28.8	35.8	42.9	35.0

Corn stalk height before and after using a fertilizer.

sample size n1=n2

Paired Test

Statistical Model

- For paired samples, we look at the difference $d_i = X_{1i} - X_{2i}$ and make the assumption that

$$d_i \sim N(\delta, \sigma^2), i = 1,2, \dots, n$$

- where δ is the true (unknown) difference between X_1 and X_2.
- Estimator

Maximum Likelihood: $$\hat{\delta} = \bar{d} = \frac{1}{n}\sum_{i=1}^{n}(X_{1i} - X_{2i})$$

- Since the individual terms in the sum are normally distributed, it follows from the central limit theorem that

$$\bar{d} \sim N(\delta, \sigma^2/n)$$

Paired Test

Test Size

- In general, we cannot assume that we know the true variance (σ^2), so we will have to estimate it.
- We will use the unbiased estimate of the variance,

Sample variance:

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{1i} - X_{2i} - \overline{d})^2$$

- Under the null hypothesis, $H_0\colon \delta = \delta_0$, the standardized test size is

$$t = \frac{\overline{d} - \delta_0}{s_d/\sqrt{n}} \sim t(n-1)$$

Distributed according to the students t-distribution, because the variance is unknown

Does Fertiliser have an Effect on Corn Growth?

- A farmer decides to try out a new fertiliser on a test plot containing 10 stalks of corn.
- Before applying the fertiliser, he measures the height of each stalk.
- Two weeks later, he measures the stalks again, being careful to match each stalk's new height to its previous one.
- The stalks would have grown an average of 6 inches during that time even without the fertiliser.
- Did the fertiliser change the expected mean?

Stalk	1	2	3	4	5	6	7	8	9	10
Before height	35.5	31.7	31.2	36.3	22.8	28.0	24.6	26.1	34.5	27.7
After height	45.3	36.0	38.6	44.7	31.4	33.5	28.8	35.8	42.9	35.0
d_i	9.8	4.3	7.4	8.4	8.6	5.5	4.2	9.7	8.4	7.3

Does Fertiliser have an Effect on Corn Growth?

- The null hypothesis is

$$H_0\!: \delta = 6$$

Mean difference in height

- If the fertilizer is thought to increase corn growth, the correct alternative hypothesis is $H_1\!: \delta > 6$, which would result in a one-tailed p-value.
- However, for simplicity let us just choose the two-tailed test, $H_1\!: \delta \neq 6$.

It would be possible to include a second alternative hypothesis, saying $\delta < 6$

Does Fertiliser have an Effect on Corn Growth?

- The average difference in stalk height before and after is

ML estimator $$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}\left(X_{1i} - X_{2i}\right) = 7.36.$$

- The estimated variance is

Unbiased ML estimator $$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2 = 4.216$$

- and the estimated standard deviation becomes $s_d = \sqrt{4.216} = 2.05$.

Does Fertiliser have an Effect on Corn Growth?

- Test size

$$t = \frac{\bar{d} - \delta}{s_d/\sqrt{n}} = \frac{7.36 - 6}{2.05/\sqrt{10}} = 2.09 \sim t(10-1)$$

- P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n-1)\right) = 2 \cdot \left(1 - t_{cdf}(2.09, 10-1)\right) = 2 \cdot (1 - 0.9669)$$
$$= 0.0662$$

The p-value is very small, so it is not likely, but we cannot reject the null-hypothesis.

- Since p>0.05 the null hypothesis cannot be rejected. Hence, the test does not provide evidence that the fertilizer caused the corn to grow more or less than if it had not been fertilized.
- Note: A one-tailed test would have resulted in the opposite conclusion (p<0.05).

Because the 5% is placed in one tail instead of dividing it between two.

Paired Test

95% Confidence Interval

- Recall that the estimator of δ is

ML estimator
$$\hat{\delta} = \overline{d} = \frac{1}{n}\sum_{i=1}^{n}\left(X_{1i} - X_{2i}\right) \sim N(\delta, \sigma^2/n)$$

- After standardizing it is straight forward to show that the endpoints of the 95% confidence interval are

$$\delta_- = \overline{d} - t_0 \cdot s_d/\sqrt{n} \quad = 5.89$$

$$\delta_+ = \overline{d} + t_0 \cdot s_d/\sqrt{n} \quad = 8.83$$

- where t_0 = *tinv*(0.975,*n*-1).

Which Sowing Machine is the Better One?

- In an agricultural research study from 1934, two sowing machines were compared in terms of the yield after harvesting.
- A total of twenty fields of equal size were sowed; ten fields with machine 1 and ten fields with machine 2.
- The fields were paired such that the two fields in a pair were neighbours.
- One field in a pair was sowed with machine 1 and the other field was sowed with machine 2.
- By pairing and sowing the fields in this way, potential field-effects could be removed.
- Hence, any differences in yield between two paired fields could be attributed to a difference between the machines (not a difference between the fields).

Which Sowing Machine is the Better One? (Paired test)

Field	*Machine 1*	*Machine 2*	*Difference* $d_i = x_{1i} - x_{2i}$	$d_i - \bar{x})^2$
1	8.0	5.6	2.4	2,46
2	8.4	7.4	1.0	0,03
3	8.0	7.3	0.7	0,02
4	6.4	6.4	0.0	0,69
5	8.6	7.5	1.1	0,07
6	7.7	6.1	1.6	0,59
7	7.7	6.6	1.1	0,07
8	5.6	6.0	−0.4	1,51
9	5.6	5.5	0.1	0,53
10	6.2	5.5	0.7	0,02
$\bar{x}$	7.22	6.39	0.83	
s^2	1.33	0.62	0.67	

Paired test, the fields are assumed to be the same.

Which Sowing Machine is the Better One? (Paired test)

Hypothesis testing

- The null hypothesis in this experiment states that there is no difference between the two machines:

$$H_0: \delta = 0$$

- and the alternative hypothesis states that there is a difference

$$H_1: \delta \neq 0$$

- If the data suggest that there is indeed a difference between the two machines, the p-value should be smaller than, say, 0.05.

Which Sowing Machine is the Better One? (Paired test)

- Test size

Variance unknown, thus t distribution

$$t = \frac{\bar{d} - \delta}{s_d/\sqrt{n}} = \frac{0.83 - 0}{0.8166/\sqrt{10}} \sim t(10-1)$$

t = 3.214

- P-value

Remember p-value is a probability.

$$2 \cdot \left(1 - t_{cdf}(|t|, n-1)\right) = 0.0106$$

- Since $p<0.05$, we reject the null hypothesis that there is no difference between the yield of the two machines.

- The average difference $\bar{d}$ is positive, and we conclude that the data suggest that machine 1 outperforms machine 2.

A positive difference can only be concluded on, after the hypothesis testing.

Which Sowing Machine is the Better One? (Paired test)

- The endpoints of the 95% confidence interval δ for are

$$\delta_- = \overline{d} - t_0 \cdot \frac{s_d}{\sqrt{n}} = 0.2459$$

$$\delta_+ = \overline{d} + t_0 \cdot \frac{s_d}{\sqrt{n}} = 1.4141$$

- Since $\delta = 0$ is not included in the 95% confidence interval, we reject the null hypothesis.

Paired vs. Unpaired Test

Consider what would happen if we performed an *unpaired* comparison between the two sowing machines.

Which Sowing Machine is the Better One? (Unpaired test)

- The sample means are $\bar{x}_1 = 7.22$ and $\bar{x}_2 = 6.39$, and the empirical variances are $\bar{s}_1^2 = 1.33$ and $\bar{s}_2^2 = 0.62$.

- With $n_1 = n_2 = 10$, the pooled variance is:

Pooled sample variance

$$s^2 = \frac{1}{10 + 10 - 2}\big((10 - 1) \cdot 1.33 + (10 - 1) \cdot 0.62\big) = 0.97$$

- And the empirical standard deviation is: $s = \sqrt{0.97} = 0.99$

Which Sowing Machine is the Better One? (Unpaired test)

- Null hypothesis: $H_0: \mu_1 - \mu_2 = 0$
- t-score

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{1/n_1 + 1/n_2}} = \frac{7.22 - 6.39}{0.99 \cdot \sqrt{1/10 + 1/10}} = 1.88 \sim t(n_1 + n_2 - 2)$$

unknown variance.

- P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right) = 2 \cdot \left(1 - t_{cdf}(1.88\,, 20 - 2)\right)$$
$$= 2 \cdot (1 - 0.9619) = 0.076$$

p-value is still small.

- Since p>0.05, we fail to reject the null hypothesis and conclude that the two machines are <u>not</u> different!

Which Sowing Machine is the Better One? (Unpaired test)

- We see that the conclusions drawn with the paired test and the unpaired test are contradictory.
  - The paired test suggests a difference between the two sowing machines.
  - The unpaired test does not.
- Explanation:
  - In the unpaired test, the difference between the two population means is due to a combination of machine effects and field effects.
  - By pairing the data, such that we look at the difference between the two machines in similar fields, the field-effects are removed, and we can detect a difference between the machines.

Variance for estimate decreases

Paired vs. Unpaired test

- The best way to look at the effect of a medical treatment is to measure some physiological parameter *in the same patient* before and after treatment.
- Consider the alternative; one group of patients gets the treatment, another group does not.
- Comparing the mean of the physiological parameter between the two groups could be both due to the treatment *and* differences between the patient groups.
- If a difference was detected, there would be no way of telling whether that difference was due to the treatment or differences between the patient groups.

Paired vs. Unpaired test

- **Unpaired:**
  - No one-to-one correspondance between X_1 and X_2 data
  - Sample size n_1 and n_2 could be different
  - Many different factors could influence the result
  - Larger uncertaincy
  - More difficult to reject the H_0 hypothesis

- **Paired:**
  - A one-to-one correspondance between X_1 and X_2 data
  - Sample size n_1 and n_2 equal
  - Elimination of factors not related to the test
  - Reducing uncertaincy
  - Easier to reject the H_0 hypothesis

Test Catalog for Paired Data

Statistical model:

- $d_i = X_{1i} - X_{2i}$, where $d_i \sim N(\delta, \sigma^2), i = 1,2, \dots, n$
- Parameter estimate:

$$\hat{\delta} = \overline{d} = \frac{1}{n}\sum_{i=1}^{n} X_{1i} - X_{2i}$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2$$

- Where the observation is $\overline{d}$ = 'the average of the differences between paired samples'.

Hypothesis test (two-tailed):

- $H_0: \delta = \delta_0$
- $H_1: \delta \neq \delta_0$
- Test size: $t = \frac{\overline{d} - \delta_0}{s_d/\sqrt{\mathrm{n}}} = \sim t(\mathrm{n}-1)$
- Approximate p-value: $2 \cdot \left(1 - t_{cdf}(|t|, \mathrm{n}-1)\right)$

95% confidence interval:

- $\delta_{-} = \overline{d} - t_0 \cdot \frac{s_d}{\sqrt{n}}$
- $\delta_{+} = \overline{d} + t_0 \cdot \frac{s_d}{\sqrt{n}}$
- where t0 = tinv(1-0.05/2,n-1)

Words and Concepts to Know

Pooled variance

Paired test

Unpaired test

Comparing two population means