02409 Multivariate Statistics

Presentation A, September 1 2025

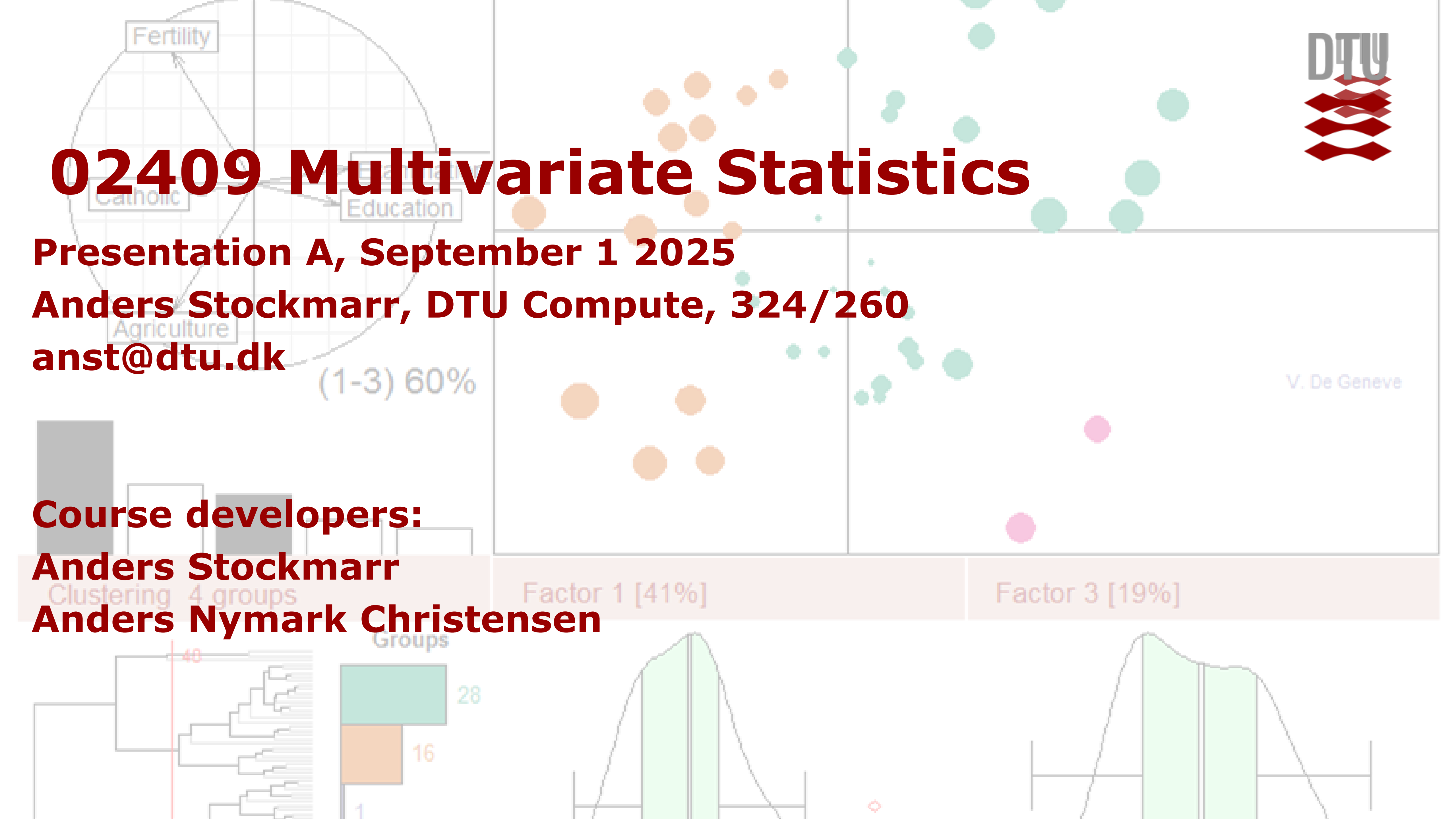
Anders Stockmarr, DTU Compute, 324/260

anst@dtu.dk

Course developers:

Anders Stockmarr

Anders Nymark Christensen



The TA team

Letizia Bottani

Mathies Brink Sørensen

Todays lecture

- Practicalities
- The Challenger case
- Summary of concepts from one, two and k-dimensional random variables
- Introducing the multivariate Gaussian distribution

Course Content

- Multivariate Random Variables
- The General Linear Model
- Regression Analysis
- Multivariate Linear Model
- Discriminant Analysis
- Principal Component and Factor Analysis, Canonical Correlation Analysis

- Statistics is about modelling and interpreting real data.
  - To that end, we use R
- It is dependent on mathematics and (scientific) computing.
  - Pen and paper exercises (Maple)
  - R
- Python, SPSS, SAS, Matlab etc. are welcome but not supported
  - Interpret outputs at the exam
  - Use statistical software at the exam

Exam

- 4 hours multiple choice
- All aids + internet

- Questions
  - ”Pen & paper”
  - Interpretation of outputs
  - Data processing
    - You need to use statistical software for the exam!
    - Recommend R

How do I get **12**?

- Read the material
  - During the course – not in the last week

- Do the weekly exercises
  - During the course – not in the last week

- Do the questions in the lectures
  - Often problems picked from old exams

- Old exams
  - Go through as many as you can

How do I get **at least 02**?

- Read the material
  - During the course – not in the last week

- Do the weekly exercises
  - During the course – not in the last week

- Do the questions in the lectures
  - Often problems picked from old exams

- Old exams
  - Go through as many as you can

Practicalities

- Book: Conradsen et al: **Multivariate Statistics for the Technical Sciences**.
  - Can be bought at Polyteknisk Boghandel.
  - A pdf is on DTU Learn.

- Alternative readings: ”Applied Multivariate Statistical Analysis”, Johnson & Wichern

- R – will be used for the exam

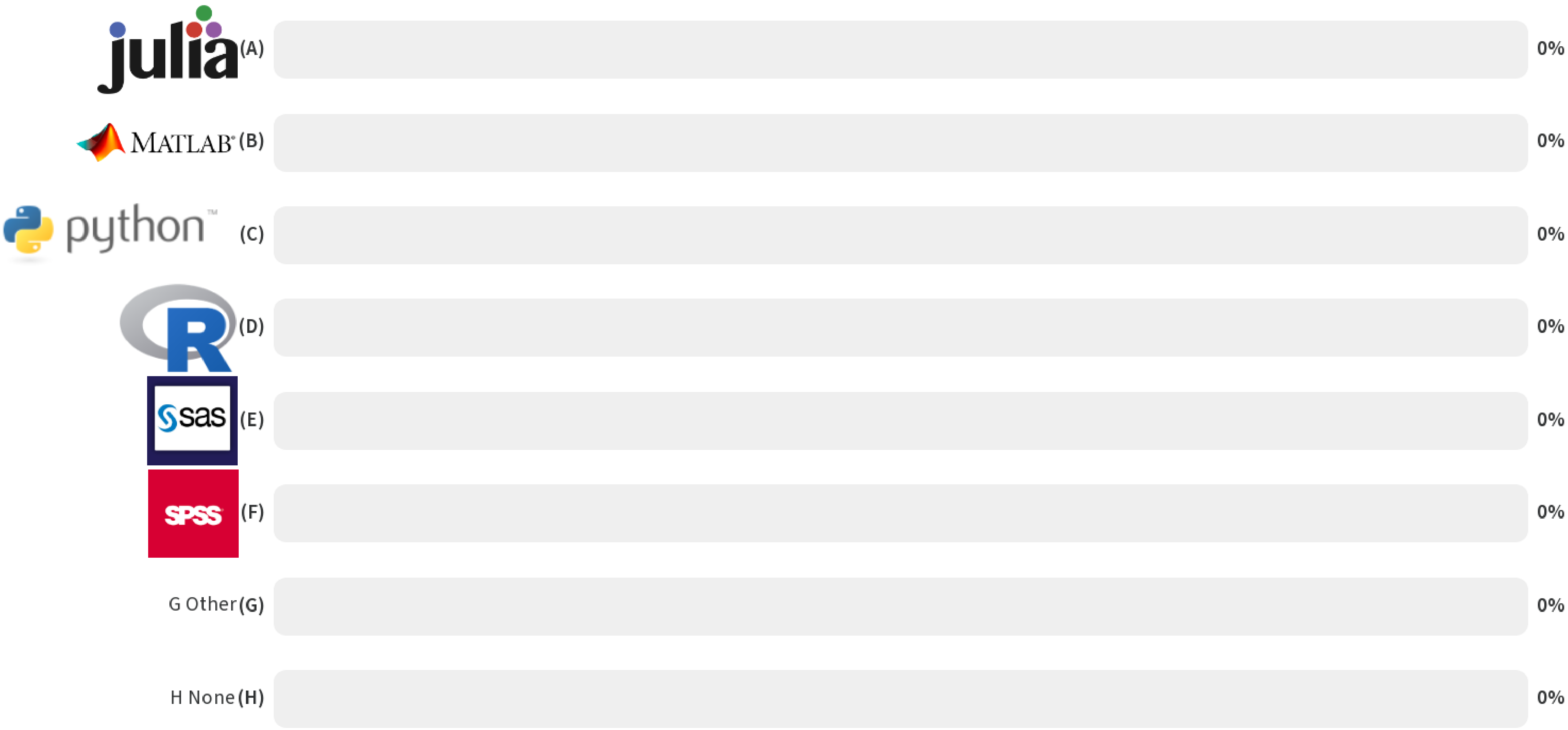
- Exercises in Building 308, rooms 001, 009 and 017. Please use your own laptop!

- Other ressources:
  - Intro Stat: https://02402.compute.dtu.dk
  - Mat 1, linear algebra: https://math1a.compute.dtu.dk/intro.html
  - **3 Brown 1 Blue, linear algebra:** https://youtu.be/fNk_zzaMoSs

Polls:

- Go to pollev.com and register;
- Join presenter andersstockmarr442.
- You can also scan the QR code in the presentation and send A-B-C….

I have experience with statistical analysis in



Multivariate Statistics

Generally we shall distinguish between three different types of data approaches:

$$\begin{array}{c|ccc} Y_1 & X_{11} & \cdots & X_{k1} \\ \vdots & \vdots & & \vdots \\ Y_n & X_{1n} & \cdots & X_{xn} \end{array}$$

Multiple regression

$$\begin{array}{ccc|ccc} Y_{11} & \cdots & Y_{d1} & X_{11} & \cdots & X_{k1} \\ \vdots & & \vdots & \vdots & & \vdots \\ Y_{1n} & \cdots & Y_{dn} & X_{1n} & \cdots & X_{xn} \end{array}$$

Multivariate regression

$$\begin{array}{ccc} X_{11} & \cdots & X_{k1} \\ \vdots & & \vdots \\ X_{1n} & \cdots & X_{xn} \end{array}$$

Structural data analysis

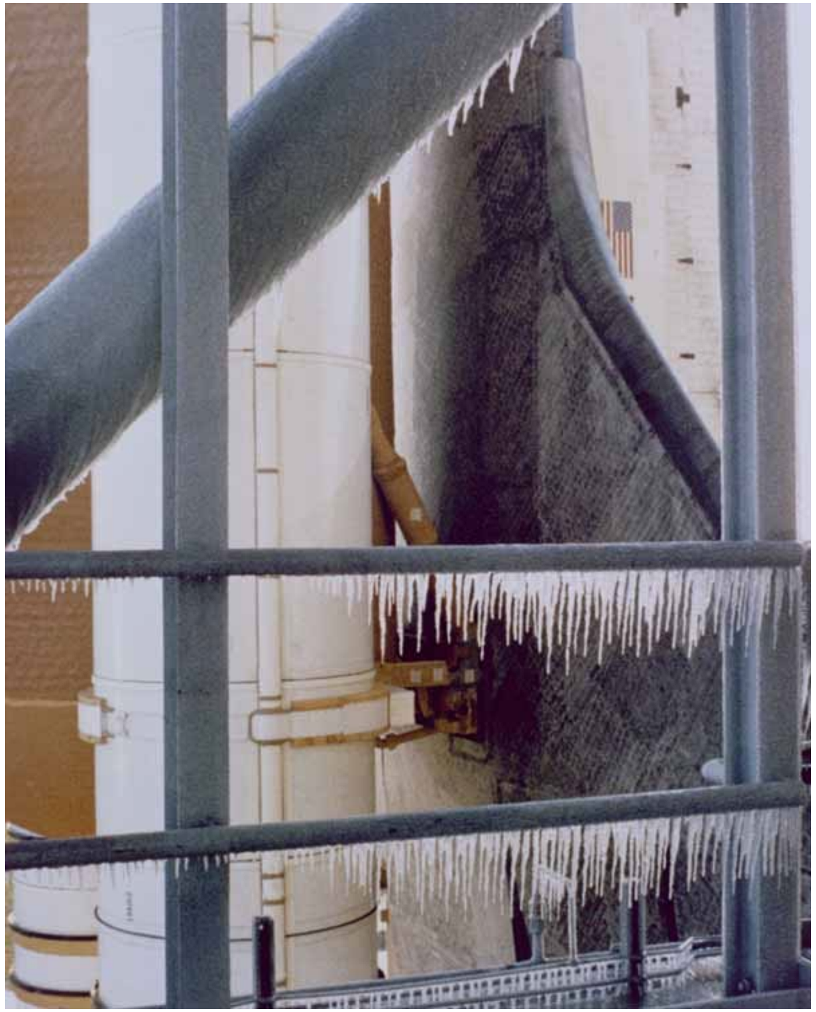
Introductory example:

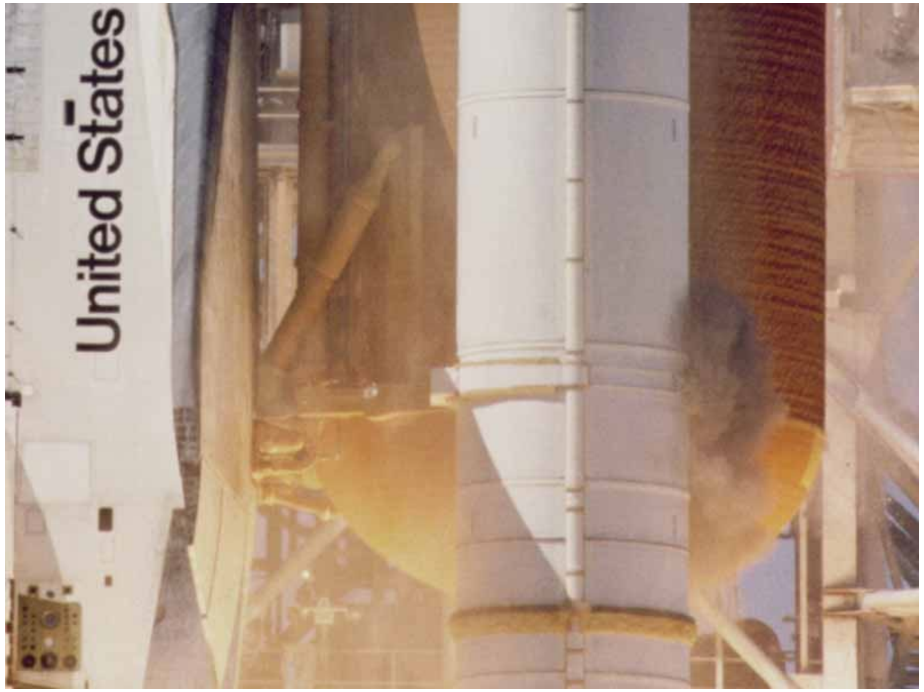
$$\begin{array}{c|c} Y_1 & X_1 \\ \vdots & \vdots \\ Y_n & X_n \end{array}$$

The Challenger Disaster

- Source, images and facts: Wikipedia.
- Data: Dalal, Fowlkess and Hoadley, JASA 1989.

Cape Kennedy January 28, 1986





- Temperature at launch time: 28 to 29 ° F (-2.2 to -1.7 ° C).

- Redline of Thiokol Engineers: 40 ° F (4 ° C).

Cape Kennedy January 28, 1986

SLIDE THERMAL DISTRESS DATA

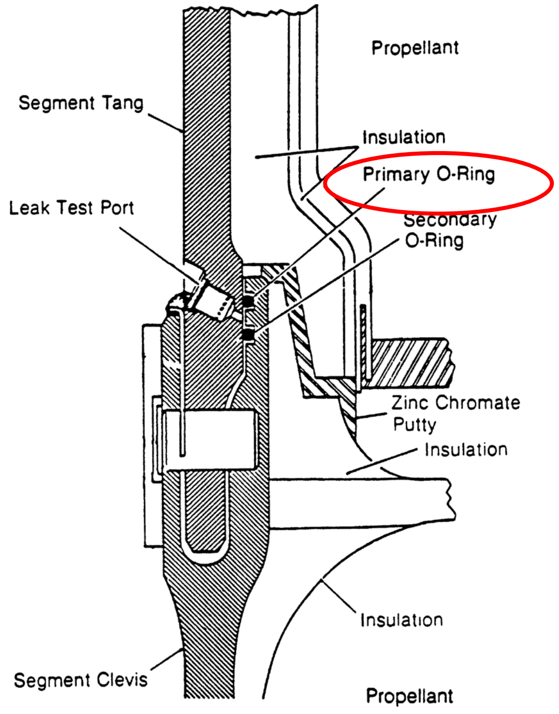
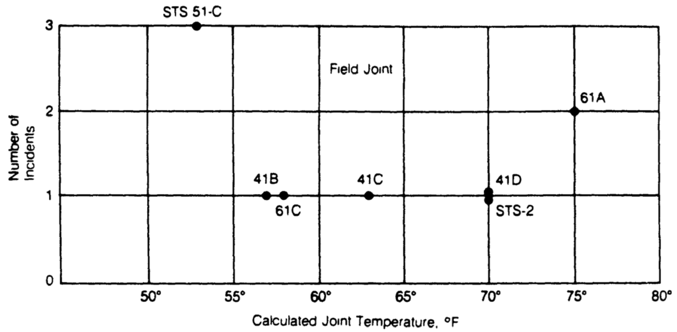


Figure 3. Solid Rocket Motor Cross Section: Tang, Clevis, and O-Rings. Picture: Dalal, Fowlkess and Hoadley 1989

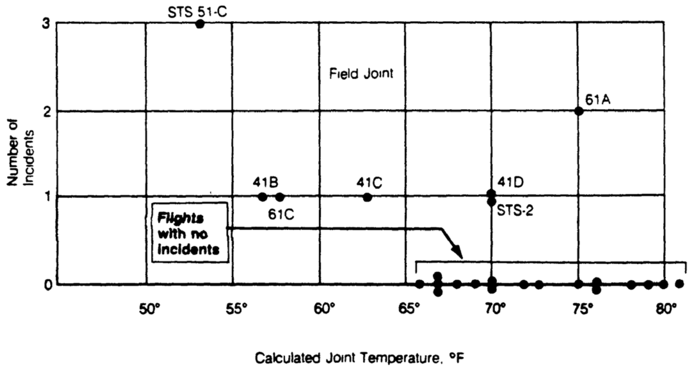
- The Shuttles each have **6 primary 0-rings**: 2 motors, each with 3 field joints.
- The 0-rings seal the field joints of the rocket motors. Cold 0-rings take longer time to seal, whick may cause them to fail. If not sealed appropriately, hot gas may escape through a gap between the tang and the secondary 0-ring.
- The (at that time unverified) temperature effect on 0-ring efficiency is labeled *thermal distress*.
- The evening before the launch, a meeting was held to consider potential thermal distress for the morning launch (predicted temperature: $31°F$), based on previous distress reportings of 0-ring distress:



- Message to NASA: ”Temperature data [are] not conclusive on predicting primary 0-ring blowby“, recommending launch as scheduled.

O-ring data

- **What they <span>should</span> have looked at, were the following data:**



- **Data are evaluated as follows:**
- Data documenting the presence or absence of one or more primary O-ring thermal distress in the 23 shuttle launches preceding the Challenger mission were collected.

Focus of these data is to determine if there is a relationship between the temperature at launch time and the number of o-rings with thermal distress.

The variables in the data set are the flight number (FLT), the temperature at launch (TEMP), and an indicator variable for whether or not there was thermal distress during the launch (TD) (***0=no distress, 1=distress***).

Flight number	Temp. at launch	O-ring Thermal Distress
1	66	0
2	70	1
3	69	0
4	68	0
5	67	0
6	72	0
7	73	0
8	70	0
9	57	1
10	63	1
11	70	1
12	78	0
13	67	0
14	53	1
15	67	0
16	75	0
17	70	0
18	81	0
19	76	0
20	79	0
21	75	1
22	76	0
23	58	1

Analysis in R

```
Challenger <- data.frame(flt = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23),
                        temp = c(66,70,69,68,67,72,73,70,57,63,70,78,67,53,67,75,70,81,76,79,75,76,58),
                          td = c(0,1,0,0,0,0,0,0,1,1,1,0,0,1,0,0,0,0,0,0,1,0,1))
Challenger <- Challenger[order(Challenger$td),]; row.names(Challenger)<-1:23
# OR, more compact: Challenger<-read.csv2(”Challenger.csv”)
t.test(Challenger$temp[Challenger$td == 0], Challenger$temp[Challenger$td == 1])
t.test(Challenger$temp[Challenger$td == 0], Challenger$temp[Challenger$td == 1], var.eq = TRUE)
# OR, more compact: with(Challenger, t.test(temp~td));
# formula form does not support var.equal=T; with(Challenger, t.test(temp[td==0],temp[td==1],var.eq=T))
```

We create the ***dataframe*** *Challenger* by specifying the values of the variables *flt, temp, and td* one variable at a time; Or load them directly from the data file `Challenger.csv` on DTU Learn…

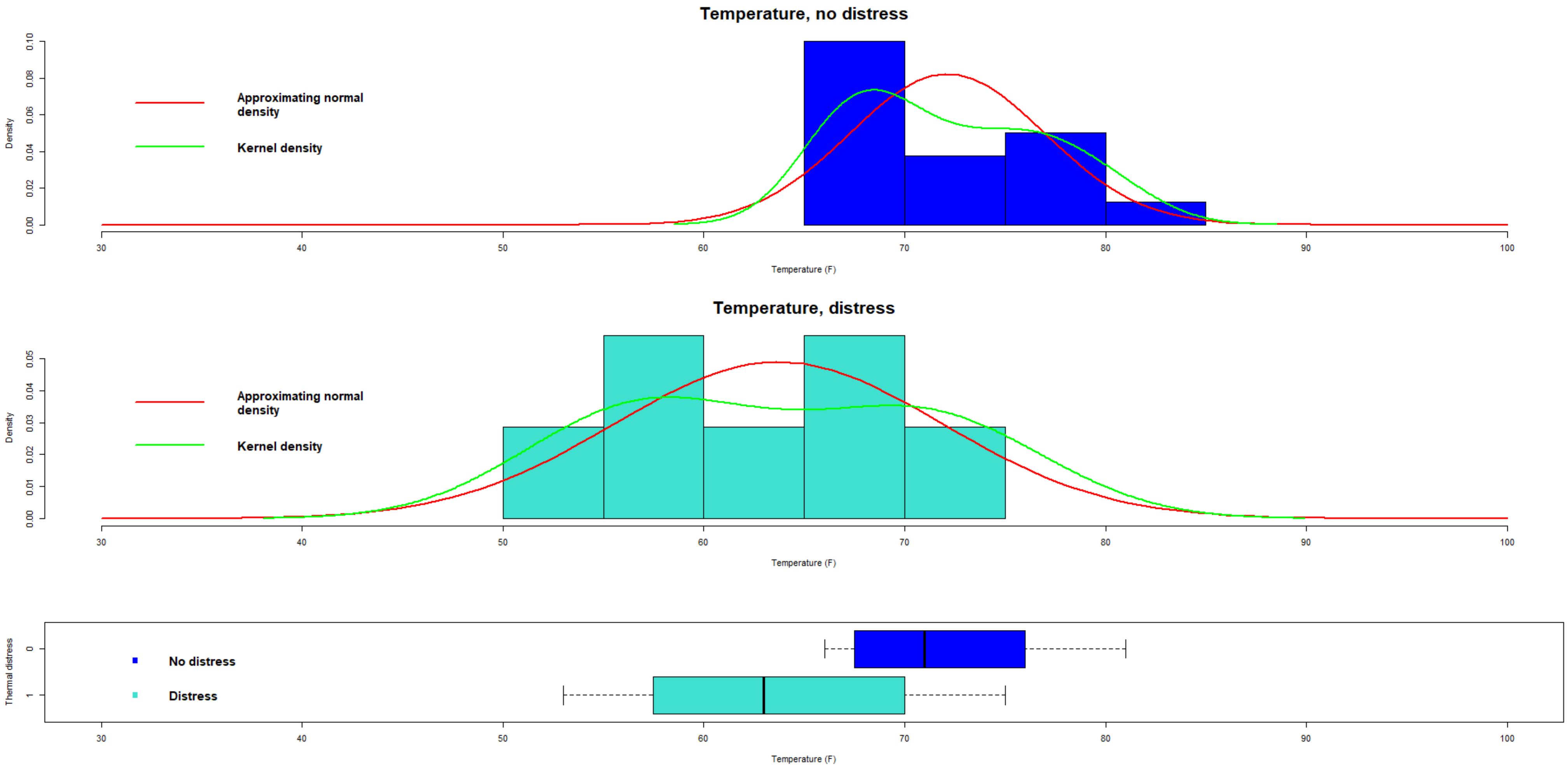
Then the data are ***sorted*** by *td, and row numbers adjusted.* This is just for view, and not necessary for the analysis.

Finally, we perform a ***t-test*** to see whether the temperature distribution means are the same for the two ’*td*’ values. **Note that we may choose to specify the kind of t-test (equal variances)**

The data

```
   flt temp td
1    1   66  0
2    3   69  0
3    4   68  0
4    5   67  0
5    6   72  0
6    7   73  0
7    8   70  0
8   12   78  0
9   13   67  0
10  15   67  0
11  16   75  0
12  17   70  0
13  18   81  0
14  19   76  0
15  20   79  0
16  22   76  0
17   2   70  1
18   9   57  1
19  10   63  1
20  11   70  1
21  14   53  1
22  21   75  1
23  23   58  1
```

Histogram and fitted frequency functions



T-test

- Two groups of sizes $n_1, n_2$with means μ_1, μ_2 and common variance σ^2:

$$H_0: \mu_1 = \mu_2$$

- Estimate μ_1, μ_2 by the empirical means, and σ^2 by the empirical variance. Then construct

$$T = \frac{\widehat{\mu_1} - \widehat{\mu_2}}{\sqrt{\frac{n_1 + n_2}{n_1 n_2} \widehat{\sigma^2}}}$$

T will then follow a t-distribution with $n_1 + n_2 - 2$ degrees of freedom, assuming that H_0 is correct.

With no common variance, Sattertwaithe’s approximation to the number of df is often used (see R output from `t.test`).

Estimates and test statistics

Data summary:

```
>with(Challenger,by(temp,td,summary))
td: 0
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  66.00   67.75   71.00   72.12   76.00   81.00 
------------------------------------------------------------
td: 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  53.00   57.50   63.00   63.71   70.00   75.00 
```

T-test:

```
 td  n  mean  sem  lower upper t.equal.var    t  p.equal.var   p
1  0 16 72.12 1.21 70.00 74.25          NA   NA          NA    NA
2  1  7 63.71 3.08 57.72 69.71        3.10 2.54       0.005**  0.04*
```

Levene’s test for equal variances:

```
>library(car);leveneTest(temp~as.factor(td),data=Challenger)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)  
group  1  3.6628 0.06937 .
      21                  
```

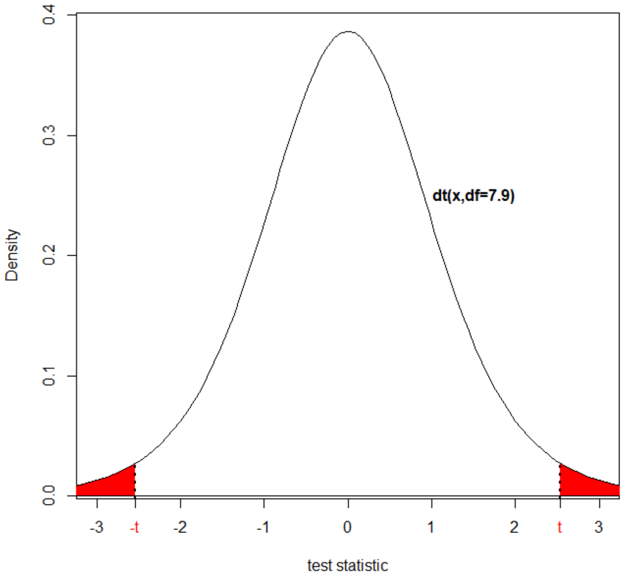
Statistical significance

- We perform a statistical test – we have a null-hypothesis. Here:

H_0: *temperature* for the two values of *td* are the same on average.

- We get an observed test statistic (`3.10` when assuming equal variances `2.54` when not), and we know the test-distribution and the degrees of freedom (t-test, 21 df for equal variances, 7.9 (Sattertwaithe approximation) when not).
- Then we calculate the **p-value**, the probability of getting a more extreme value, as:

`2*pt(-3.10,df=21),` respectively `2*pt(-2.54,df=7.9)` (use of non-rounded statistics are recommended)



- The significance level is simply the threshold we set beforehand. Typically, 0.05.
- If the p-value is **below** the significance level, the test is declared **significant**.
- Result: p=0.005,p=0.04, respectively.
- The **strength** of the evidence from the p-value is signified by stars/.:
- $. \quad : 0.1 > p \geq 0.05$
- $* \quad : 0.05 > p \geq 0.01$
- $** \ : 0.01 > p \geq 0.001$
- $***: 0.001 > p$

Overview: Logistic regression

- Let X be the number of 0-rings exhibiting thermal distress in a rocket launch.
- Assuming that each 0-ring acts independently of the others, given the temperature, we note that

$$X \sim Bin(6, p)$$

where p is the probability that an individual 0-ring experiences thermal distress. The data measures

$$Z = 1_{\{X>0\}}$$

Distribution:

$$E[Z|temp = x] = P(X > 0|temp = x) = p^*(x)$$

Relationship between p and p^*:

$$p^* = 1 - (1-p)^6, \qquad ie. \qquad p = 1 - (1-p^*)^{1/6}$$

What is a p-value?

The probability of a null hypothesis being true 0%

The probability of a null hypothesis being false 0%

The probability of getting the observed (or a more extreme) test value, given that the null hypothesis is true 0%

The probability of getting the observed (or a more extreme) test value, given that the null hypothesis is false 0%

Don't know 0%

Overview: Logistic regression

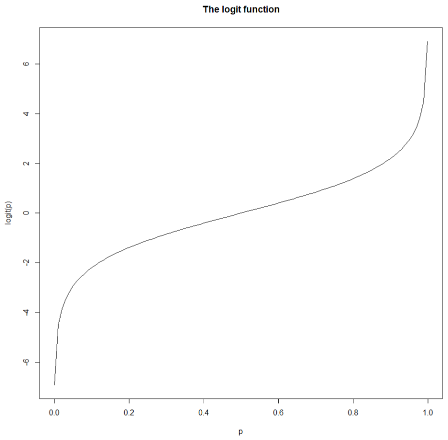
With Z having binary response, we model p^* through logistic regression:

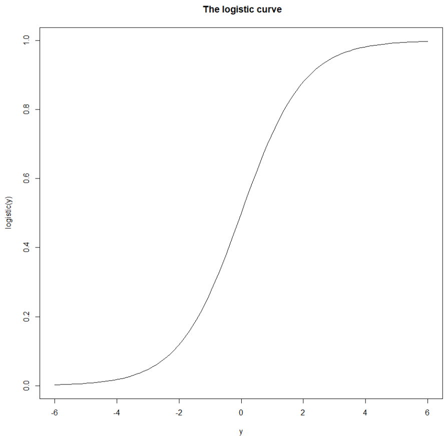
$$logit\big(p^*(x)\big) = log\left(\frac{p^*(x)}{1 - p^*(x)}\right) = \alpha + \beta x$$

Thus

$$E[Y] = \pi(x) = P(Y = 1|X = x)$$

$$p^*(x) = logit^{-1}(\alpha + \beta x) = logistic(\alpha + \beta x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$





Logistic regression on O-ring data

```
model <- glm(td ~ temp, family=binomial(link='logit'), data=Challenger)
summary(model)
```

Partial output:

```
Call:
glm(formula = td ~ temp, family = binomial(link = "logit"), data = Challenger)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  15.0429     7.3786   2.039   0.0415 *
temp         -0.2322     0.1082  -2.145   0.0320 *
```

The Z-tests above are just pointers. Use the Likelihood Ratio test:

```
drop1(model,test="Chisq")
```

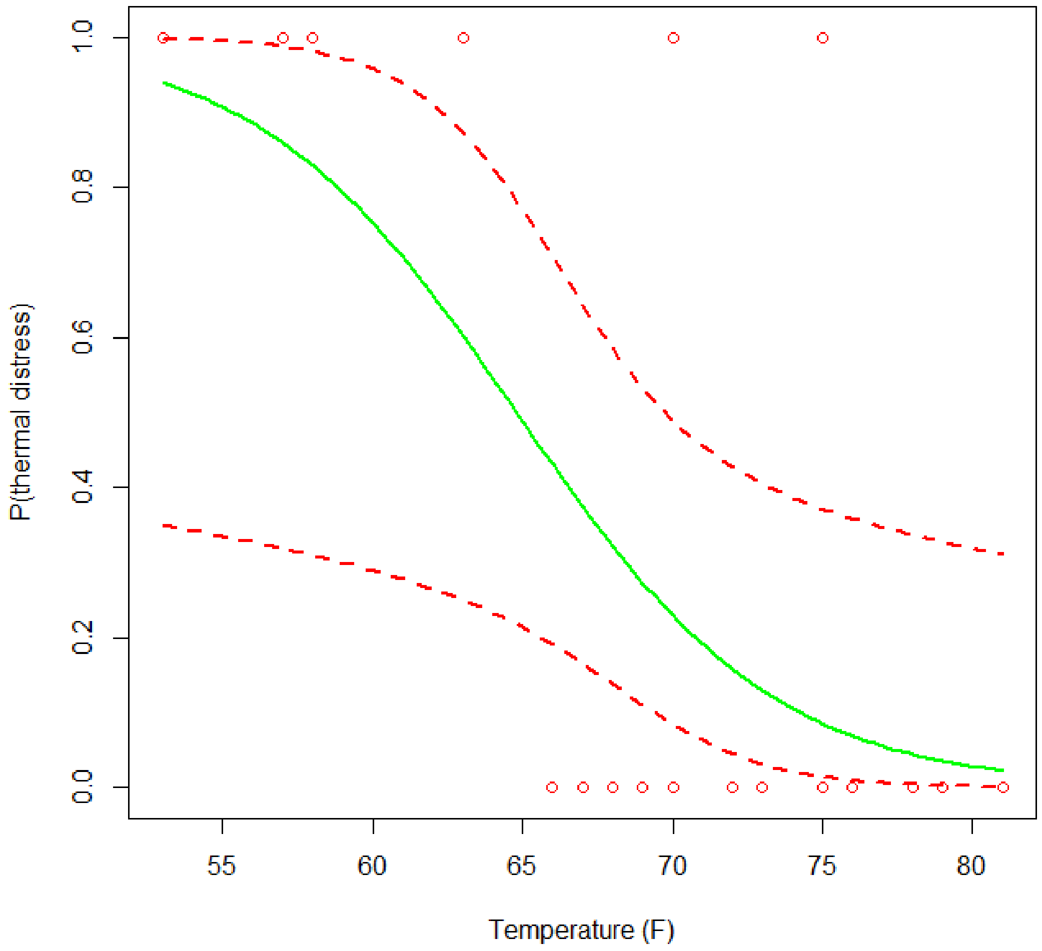
Output:

```
       Df Deviance    AIC   LRT Pr(>Chi)
<none>      20.315 24.315
temp    1   28.267 30.267 7.952 0.004804 **
```

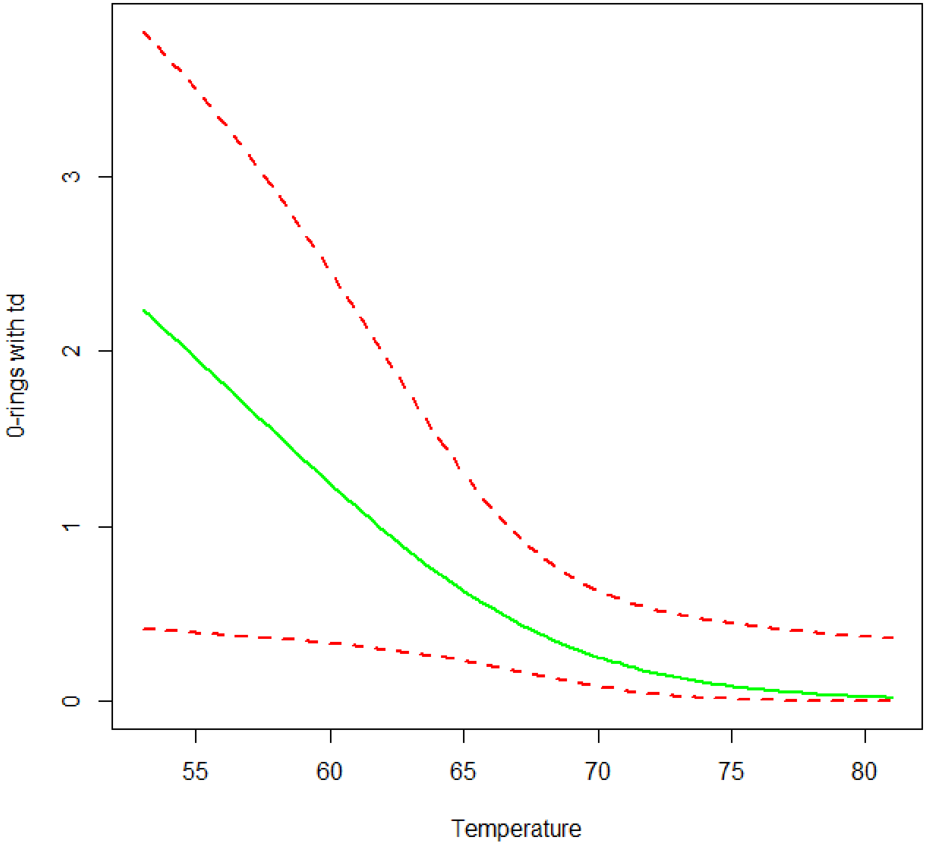
Very different strength of evidence – but both significant at the 5% test level

Visualizing the dependency of temperature

```
newdat <- data.frame(temp=seq(min(Challenger$temp),
                     max(Challenger$temp),length=100))
newdat$fit = predict(model, newdata=newdat)
newdat$se<-predict(model, newdata=newdat,se.fit=T)$se.fit
plot(td ~ temp, data=Challenger,
     col="red",xlab="Temperature (F)",
     ylab="P(thermal distress)")
lines(logistic(fit)~ temp, data=newdat, col="green",
      lwd=2)
lines(logistic(fit+1.96*se)~ temp, data=newdat,
      col="red", lty=2,lwd=2)
lines(logistic(fit-1.96*se)~ temp, data=newdat,
      col="red", lty=2,lwd=2)
```



Number of 0-rings with thermal distress



Prediction at 29°F:

```
newdat <- data.frame(temp=29)
newdat$fit = predict(model, newdata=newdat)
newdat$se<-predict(model, newdata=newdat,se.fit=T)$se.fit
logistic(newdat$fit)
[1] 0.9997541

p.temp<-logistic(newdat$fit)
6*(1-(1-p.temp)^(1/6))
[1] 4.498164
```

Thus, one would expect **4.5 out of 6** 0-rings to experience thermal distress for the Challenger (large uncertainty though).

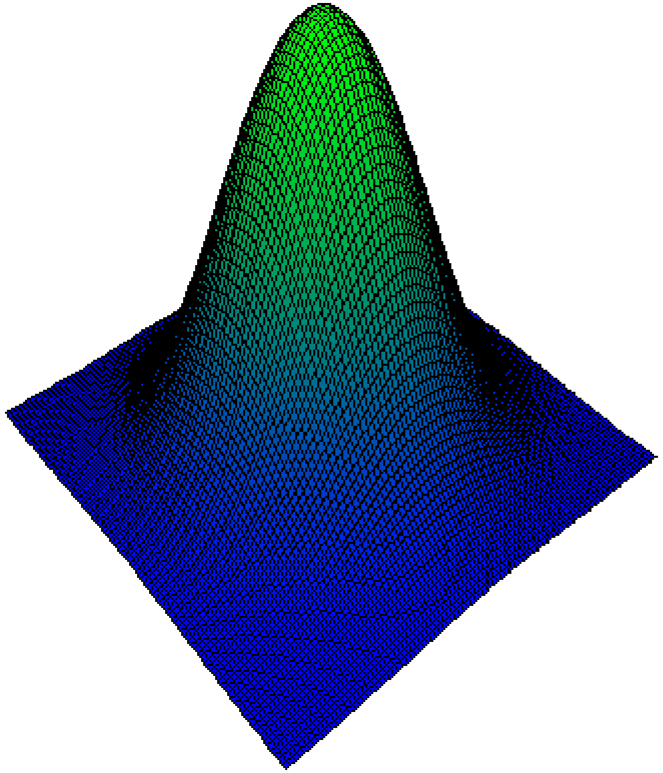
The Rogers commission concluded that the cause of the accident was a gas leak through a joint in one of the booster rockets, which was supposed to be sealed by an 0-ring, and that based on the above, the risk of a catastrophy from failing 0-rings was at least 13.2%, more than 600% higher compared to if the launch had been postponed to when the temperature had risen to 60F.

Ethics of Whistle-Blowing

- **Richard P. Feynman:**
- It appears that there are enormous differences of opinion as to the probability of a failure with loss of vehicle and of human life. The estimates range from roughly 1 in 100 to 1 in 100,000**. *The higher figures come from the working engineers, and the very low figures from management.*** What are the causes and consequences of this lack of agreement? Since 1 part in 100,000 would imply that one could put a Shuttle up each day for 300 years expecting to lose only one, we could properly ask ***"What is the cause of management's fantastic faith in the machinery?"***

- **For a successful technology, reality must take precedence over public relations, for nature cannot be fooled.**

The multivariate normal distribution



One-dimensional continuously distributed random variables

The **probability density function** f(x) (the **pdf**) of a random variable X determines the outcomes of X and is given by

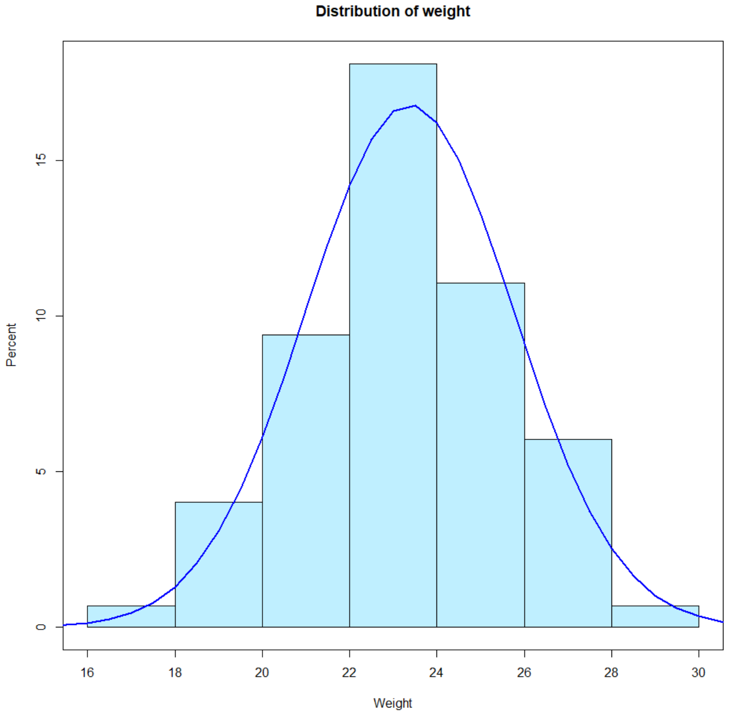
$$P\{X \in [x, x + dx]\} = \int_{x}^{x+dx} f(t)dt \sim f(x)dx$$

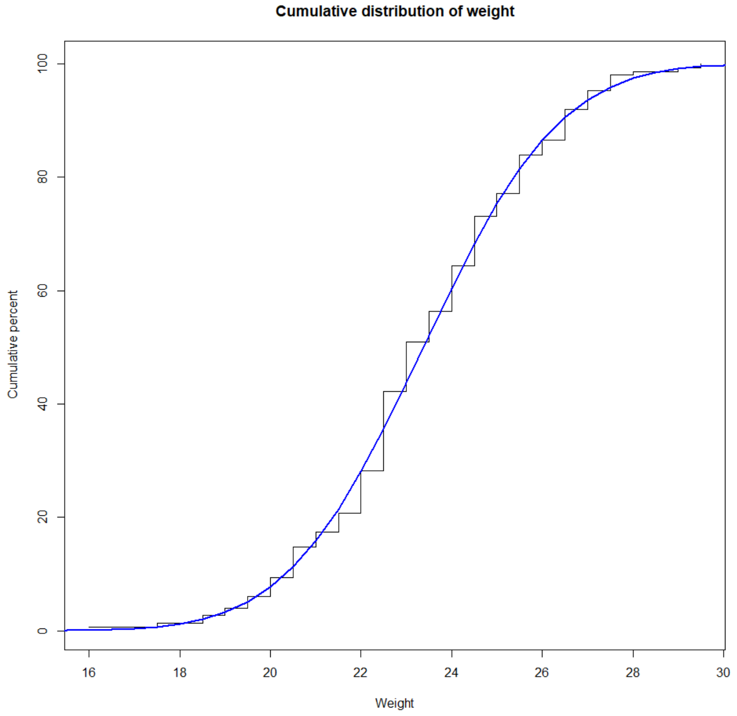
The **cumulative distribution function** F(x) (the **cdf**) of a random variable X is another assessment of the probabilities of outcomes of X and is given by

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)dt$$

Conversely we have

- $f(x) = F'(x)$


Distribution of weight


Cumulative distribution of weight

DETOUR TO MOMENTS OF STOCHASTIC VARIABLES

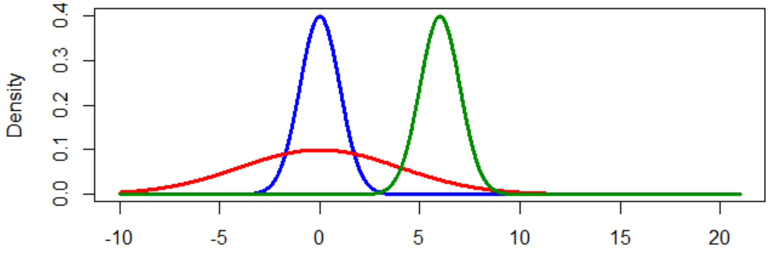
Moments of a random variable

The **mean** is the center of mass in a probability function:

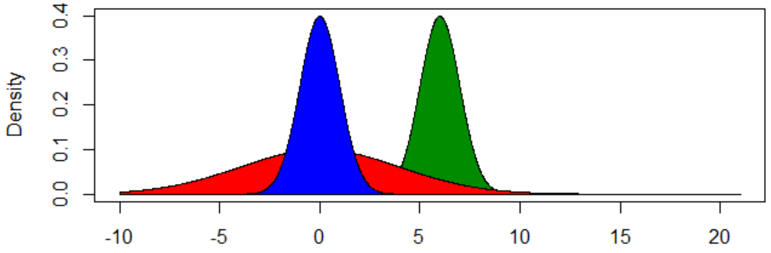
$$E(X) = \mu = \int_{-\infty}^{\infty} t f(t) dt$$

The **variance** measures the degree of variablity around the mean

$$V(X) = \ \sigma^2 = E\left([X-\mu]^2\right) = E\left(X^2\right) - \mu^2$$



$\boldsymbol{\mu_1 = 0}$ $\boldsymbol{\mu_2 = 0}$ $\boldsymbol{\mu_3 = 6}$



$\boldsymbol{\sigma_1^2 = 1}$ $\boldsymbol{\sigma_2^2 = 16}$ $\boldsymbol{\sigma_3^2 = 1}$

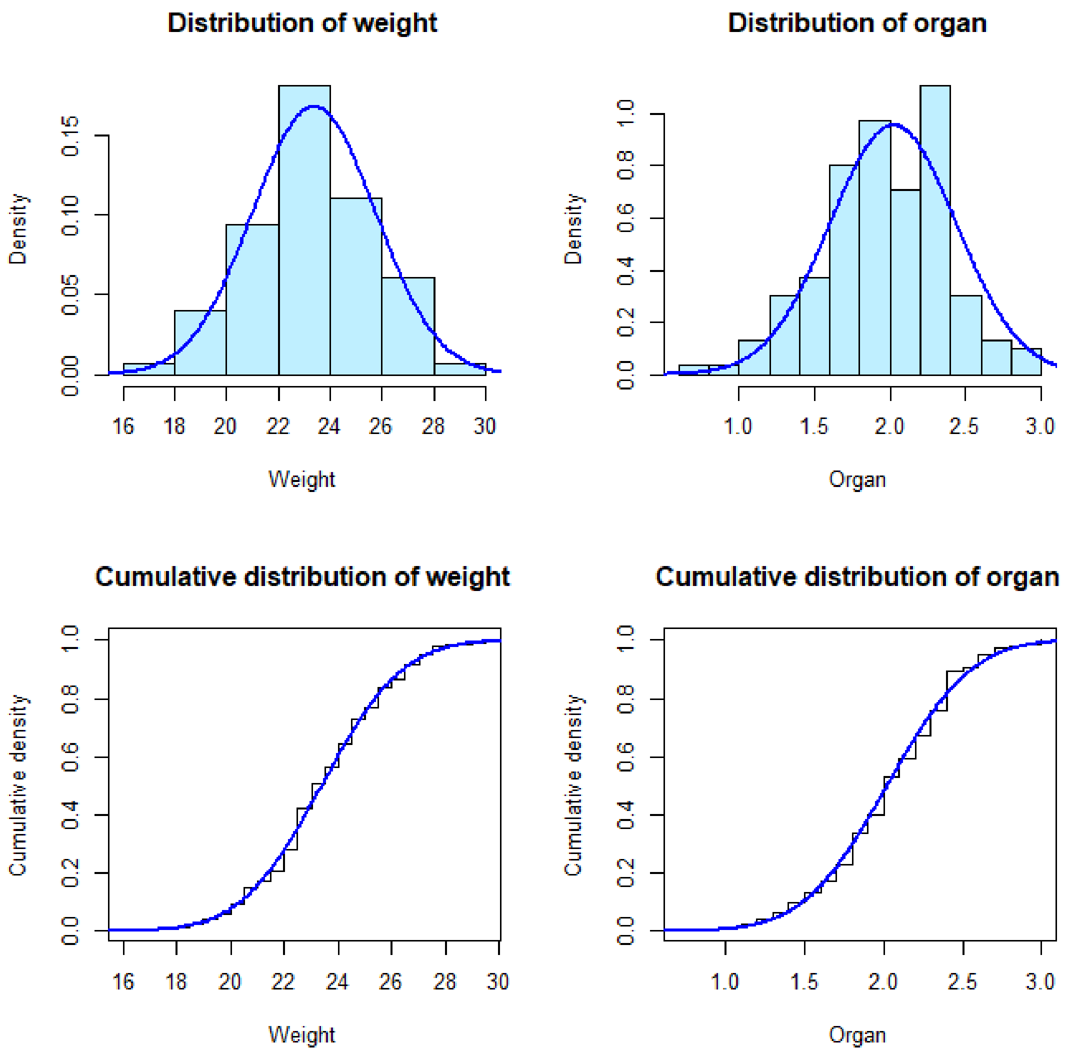
Covariance and correlation

- Example: Weight of a male mouse and the weight of its reproductive organs
- in 149 male Apodemus caught between 28 March and 13 April 1937, 1938 and 1939.
- Data from Hacker and Pearson, Biometrika 33, 2 (1944), pp. 136-162.

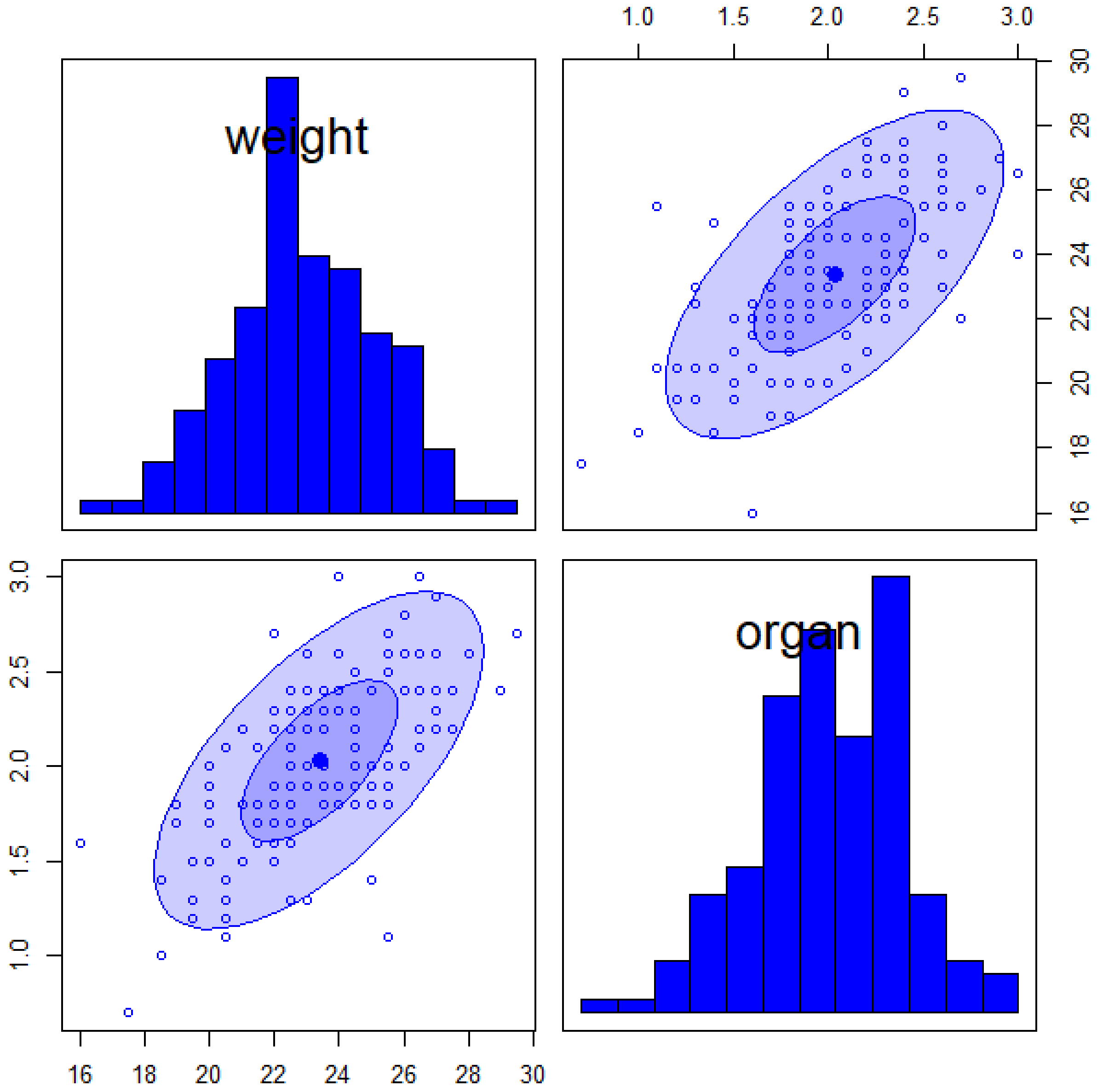
```
organ<-read.csv2("Data/organ.csv")
head(organ,n=20)
```

obs	Weight	organ
1	16.0	1.6
2	17.5	0.7
3	18.5	1.0
4	18.5	1.4
5	19.0	1.7
6	19.0	1.8
7	19.5	1.2
8	19.5	1.3
9	19.5	1.5
10	20.0	1.5
11	20.0	1.7
12	20.0	1.8
13	20.0	1.9
14	20.0	2.0
15	20.5	1.2
16	20.5	1.2
17	20.5	1.3
18	20.5	1.4
19	20.5	1.4
20	20.5	1.4

Univariate histograms/cdf and fitted Gaussian distribution



Matrix of 2D-scatterplots



The plot reveals that the univariate plots does not tell the whole story.

There is information in the pairwise behavior of weight and organ.

Two-dimensional random variable

A two-dimensional random variable is a vector $(X,Y)^T$ where each component is a random variable. The two-dimensional, or **simultaneous pdf** $f(x,y)$ of $(X,Y)^T$ satisfies:

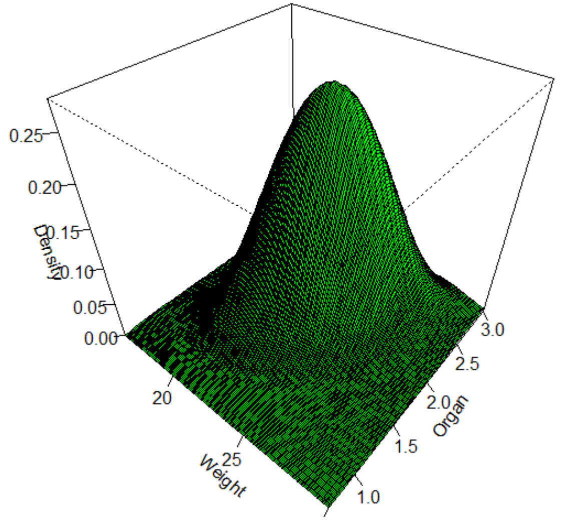
$$P\left\{\begin{bmatrix} X \\ Y \end{bmatrix} \in [x, x+dx] \times [y, y+dy]\right\} = \int_x^{x+dx} \int_y^{y+dy} f(s,t)dsdt \sim f(x,y)dxdy$$

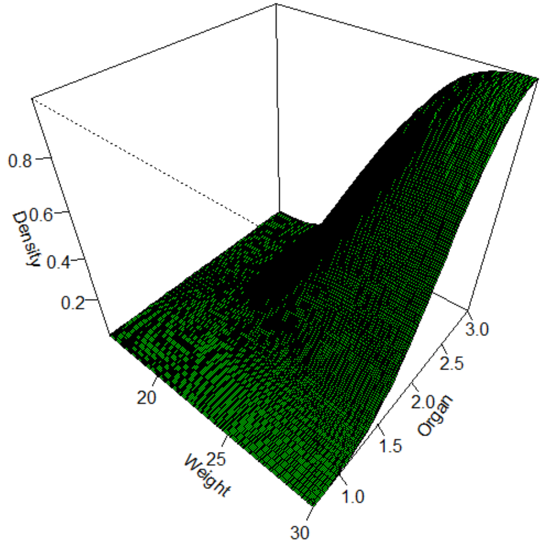
The **simultaneous cdf** $F(x,y)$ is given by:

$$F(x,y) = P\{X \le x\,, Y \le y\,\} = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)dsdt$$

And we have

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$$





Marginal distributions and independence

The **(marginal) pdf** of say Y is found by integrating f over all x-values, i.e

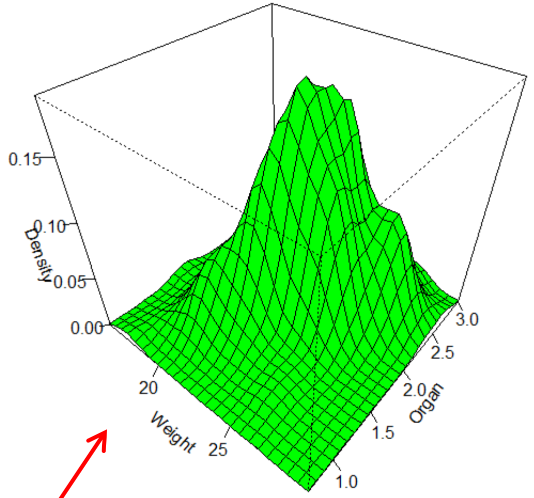
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

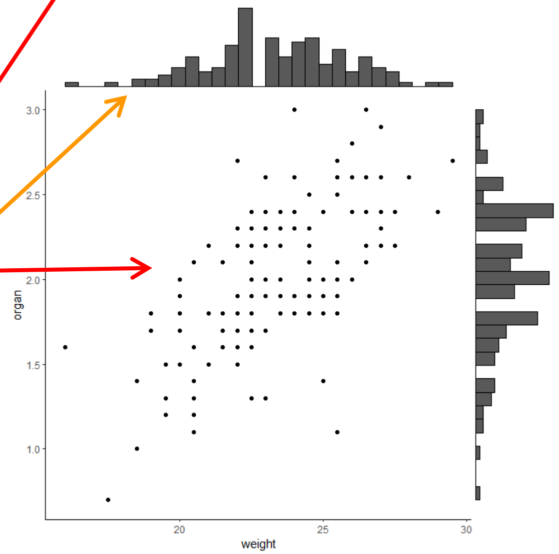
The **(marginal) cdf** $F_Y(y)$ may be found directly or found similarly:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = \int_{-\infty}^{\infty}\int_{-\infty}^{y} f(x,t)dxdt$$

We say that X and Y are **independent** iff

$$f(x,y) = f_X(x)f_Y(y) \quad \text{or} \quad F(x,y) = F_X(x)F_Y(y)$$





Smoothed and scatter simultaneous distribution of weight and organ

Marginal distribution of weight

Covariance and correlation

- The **covariance** is an expression for the relationship between two random variables:

$$Cov(X, Y) = \sigma_{xy} = E\left[(X - \mu_x)\left(Y - \mu_y\right)\right] = E[XY] - \mu_x \mu_y$$

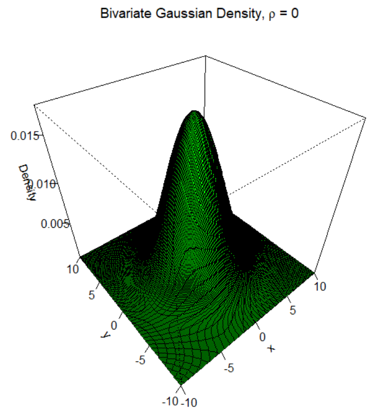
- The **correlation** measures the degree of relationship between the two variables:

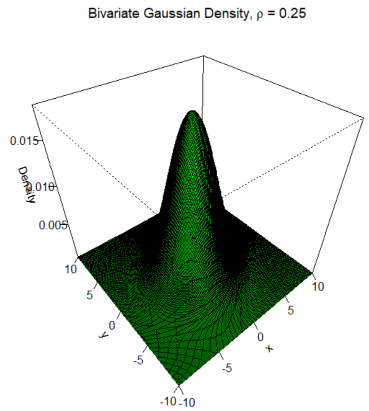
$$Cor(X, Y) = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

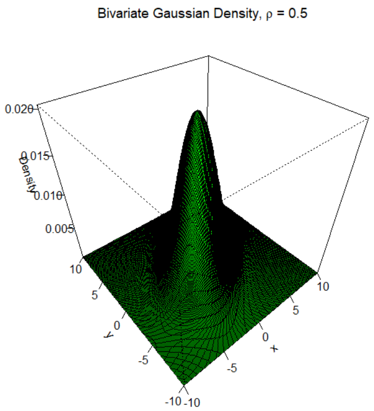
Are mouse weight and organ weight independent?

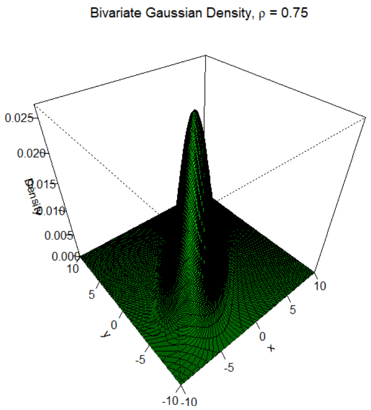
A Yes, they are independent	0%
B No, they are negatively correlated	0%
C No, they are positively correlated	0%
D Don't know	0%

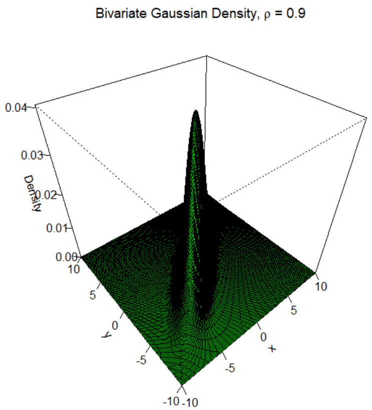
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

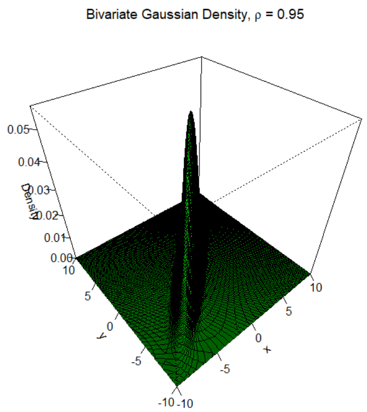

Bivariate Gaussian Density, ρ = 0

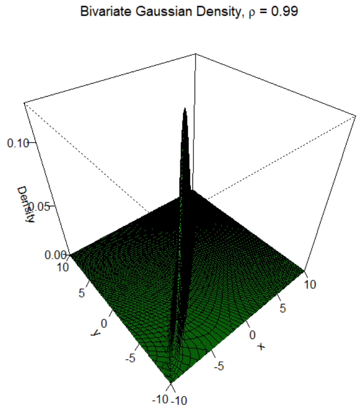

Bivariate Gaussian Density, ρ = 0.25


Bivariate Gaussian Density, ρ = 0.5


Bivariate Gaussian Density, ρ = 0.75


Bivariate Gaussian Density, ρ = 0.9

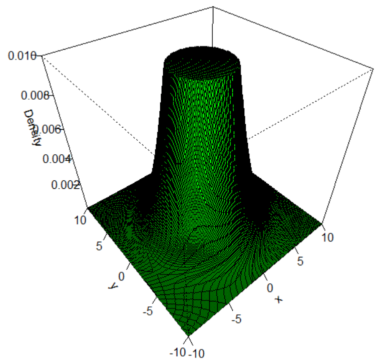

Bivariate Gaussian Density, ρ = 0.95


Bivariate Gaussian Density, ρ = 0.99

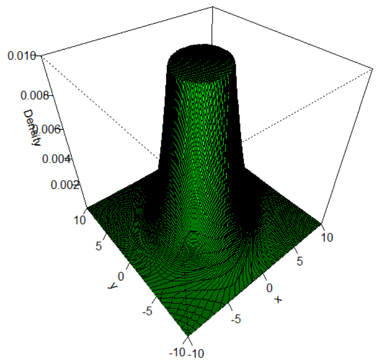
Probability Density Functions for bivariate Gaussians with correlation coefficients $\rho = 0, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99$

Distribution: $N_2\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}9 & 9\rho\\ 9\rho & 9\end{bmatrix}\right)$

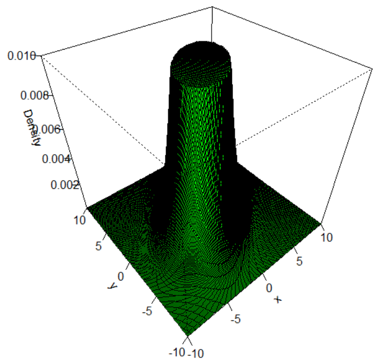
Bivariate Gaussian Density, ρ = 0



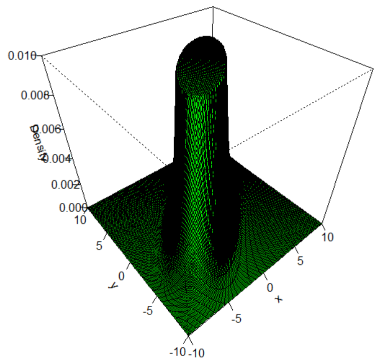
Bivariate Gaussian Density, ρ = 0.25



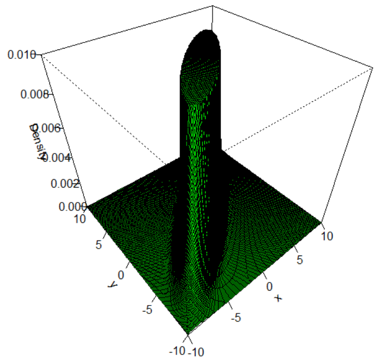
Bivariate Gaussian Density, ρ = 0.75



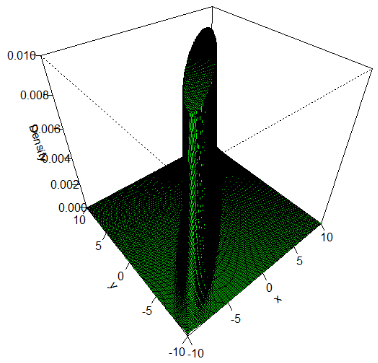
Bivariate Gaussian Density, ρ = 0.75



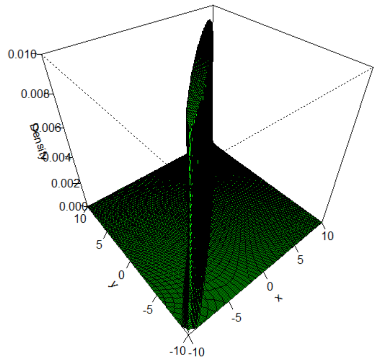
Bivariate Gaussian Density, ρ = 0.9



Bivariate Gaussian Density, ρ = 0.95



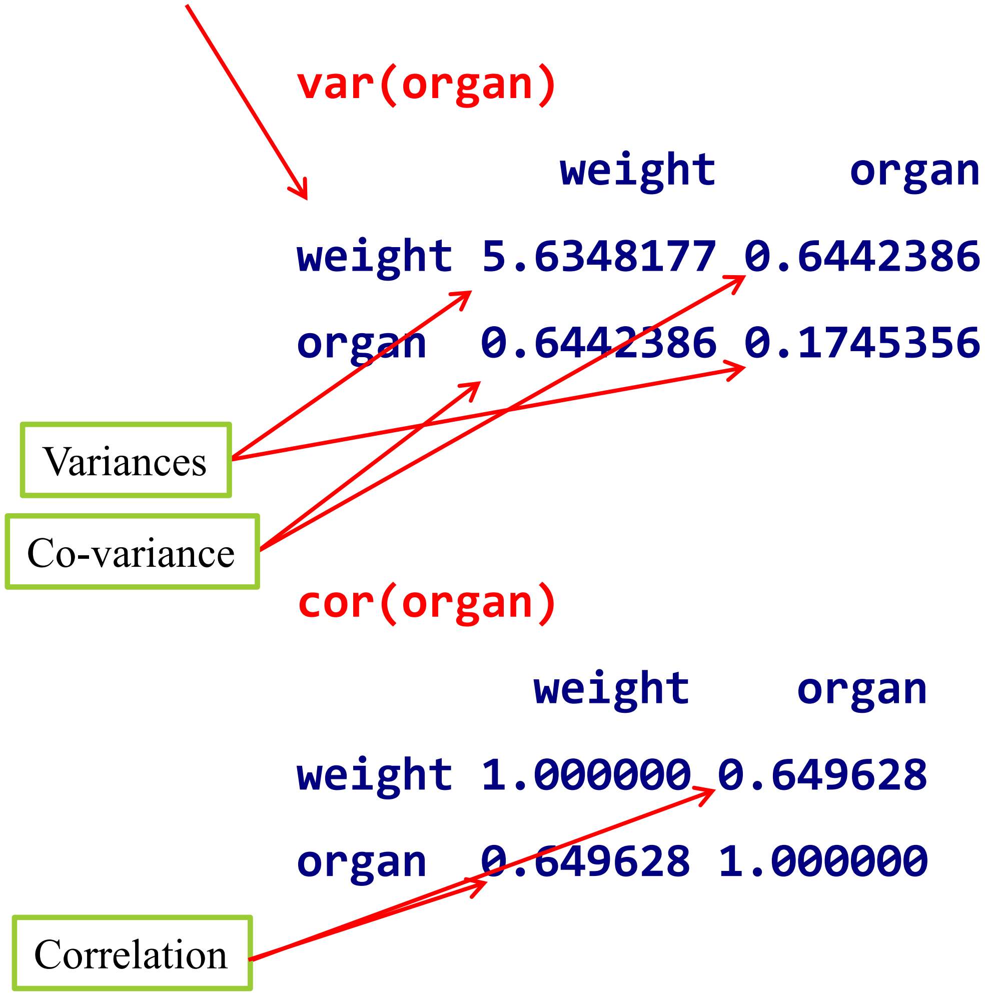
Bivariate Gaussian Density, ρ = 0.99



Densities from previous slide truncated at z=0.01

Covariance and correlation

*What we call the **Dispersion matrix**; or*
***Variance-Covariance matrix**; or*
Variance matrix



```
var(organ)
          weight     organ
weight 5.6348177 0.6442386
organ  0.6442386 0.1745356
```

Variances

Co-variance

```
cor(organ)
         weight    organ
weight 1.000000 0.649628
organ  0.649628 1.000000
```

Correlation

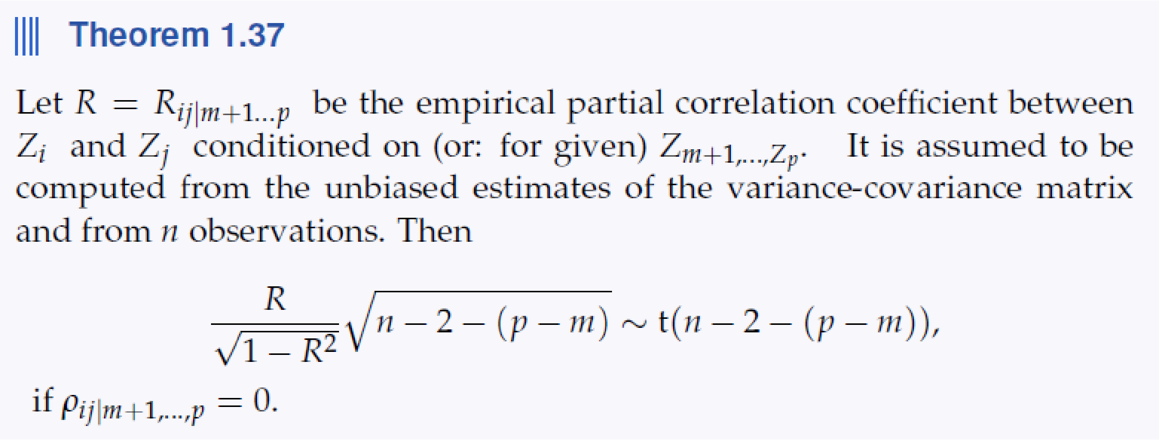
Better (for display purposes):

```
round(var(organ),digits=2)
       weight organ
weight   5.63  0.64
organ    0.64  0.17
```

```
round(cor(organ),digits=2)
       weight organ
weight   1.00  0.65
organ    0.65  1.00
```

Covariance and correlation

- Testing for correlation:



Theorem 1.37

Let $R = R_{ij|m+1\ldots p}$ be the empirical partial correlation coefficient between Z_i and Z_j conditioned on (or: for given) $Z_{m+1,\ldots,Z_p}$. It is assumed to be computed from the unbiased estimates of the variance-covariance matrix and from n observations. Then

$$\frac{R}{\sqrt{1-R^2}}\sqrt{n-2-(p-m)} \sim \mathrm{t}(n-2-(p-m)),$$

if $\rho_{ij|m+1,\ldots,p} = 0$.

- In our case (p-m=0, n=149);

$$\sqrt{147} * \frac{0.649628}{\sqrt{1-0.649628^2}} = 10.36$$

Test for that data are generated from a $t(147)$ distribution, of weight and organ are truly uncorrelated:

$$H_0\colon \rho = 0$$

P-value: `2*pt(-my.T,df=147)`

```
[1] 3.175071e-19
```

Covariance and correlation

- **Handled in one go by R:**

```
cor.test(organ[,1],organ[,2])

        Pearson's product-moment correlation

data:  organ[, 1] and organ[, 2]
t = 10.36, df = 147, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5458474 0.7337775
sample estimates:
     cor
0.649628
```

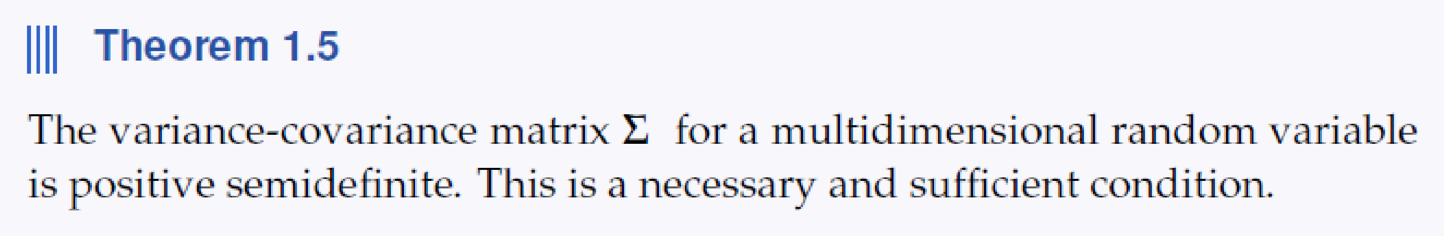
Moments of multivariate random variables:

Mean, **dispersion** and **covariance**

$$E(\boldsymbol{X}) = \boldsymbol{\mu} = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}$$

$$\mathrm{V}(\boldsymbol{X}) = \boldsymbol{\Sigma} = E\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\boldsymbol{T}}\right] = \begin{bmatrix} V(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_k) \\ Cov(X_2,X_1) & V(X_2) & \cdots & Cov(X_2,X_k) \\ \vdots & \vdots & & \vdots \\ Cov(X_k,X_1) & Cov(X_k,X_2) & & V(X_k) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_k^2 \end{bmatrix}$$

$V(X) \geq 0$ corresponds to $\mathrm{V}(X)$ symmetric and positive semidefinite



Theorem 1.5

The variance-covariance matrix $\boldsymbol{\Sigma}$ for a multidimensional random variable is positive semidefinite. This is a necessary and sufficient condition.

$$C(\boldsymbol{X},\boldsymbol{Y}) = E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\nu})^T\,] = \begin{bmatrix} Cov(X_1,Y_1) & \cdots & Cov(X_1,Y_q) \\ \vdots & & \vdots \\ Cov(X_p,Y_1) & \cdots & Cov(X_p,Y_q) \end{bmatrix}$$

Multivariate Normal or Gaussian Distribution

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \in N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

p-dimensional random variable

$$\boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_1 \end{bmatrix}$$

Mean value or expectation

$$\boldsymbol{\Sigma} = D(\boldsymbol{X}) = \begin{bmatrix} V(X_1) & \cdots & Cov(X_1, X_p) \\ \vdots & \ddots & \vdots \\ Cov(X_p, X_p) & \dots & V(X_p) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_p^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_1 \sigma_p \varrho_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_p \sigma_1 \varrho_{p1} & \dots & \sigma_p^2 \end{bmatrix}$$

Dispersion or

variance matrix

It it is 2 dimension it is a elipse
If it is more than 2 dimension it is a elipsoid

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}^p} \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

Frequency or density function

$$f(\boldsymbol{x}) = c \Leftrightarrow (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathrm{k}$$

Contour ’lines’ are ellipsoids

with main axes given by

eigenvectors of the variance matrix

Estimation of parameters I

The i’th

Observation

$$\boldsymbol{X}_i = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{ip} \end{bmatrix}$$

The mean

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

The empirical

Variance matrix

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T = \frac{1}{n-1}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T - \frac{n}{n-1}\bar{\boldsymbol{X}}\,\bar{\boldsymbol{X}}^T.$$

Observations collected

in data matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1^T \\ \vdots \\ \boldsymbol{X}_n^T \end{bmatrix} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}$$

Other formulas

for mean and dispersion

$$\bar{\boldsymbol{X}} = \frac{1}{n}\boldsymbol{X}^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{n}\boldsymbol{X}^T\mathbf{1}$$

$$(n-1)\mathbf{S} = \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T = \mathbf{X}^T\mathbf{X} - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T = \mathbf{X}^T\mathbf{X} - \frac{1}{n}\mathbf{X}^T\mathbf{1}\mathbf{1}^T X.$$

Estimation of parameters II

- We sample the following, where – *in this case* - each column is an observation and each row a variable:

$$X = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 5 \\ 1 & 5 & 9 \end{bmatrix}$$

- Determine the mean, the dispersion matrix and the correlation matrix

$$\boldsymbol{X}_i = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{ip} \end{bmatrix}$$

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T = \frac{1}{n-1}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T - \frac{n}{n-1}\bar{\boldsymbol{X}}\,\bar{\boldsymbol{X}}^T.$$

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

The mean of the shown data X is?

$$X = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 5 \\ 1 & 5 & 9 \end{bmatrix}$$

- $\begin{bmatrix} 3 \\ 12 \\ 15 \end{bmatrix}$ 0%
- $\begin{bmatrix} 1.5 \\ 6 \\ 7.5 \end{bmatrix}$ 0%
- $\begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$ 0%
- $\begin{bmatrix} \frac{5}{3} & 3 & \frac{16}{3} \end{bmatrix}$ 0%

Estimation of parameter: Mean

- We sample the following, where each column is an observation and each row a variable:

$$X = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 5 \\ 1 & 5 & 9 \end{bmatrix}$$

The mean $$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

is given by $\bar{X} = \frac{1}{3}\left\{\begin{bmatrix}1\\3\\1\end{bmatrix} + \begin{bmatrix}0\\4\\5\end{bmatrix} + \begin{bmatrix}2\\5\\9\end{bmatrix}\right\} = \frac{1}{3}\begin{bmatrix}3\\12\\15\end{bmatrix} = \begin{bmatrix}1\\4\\5\end{bmatrix}$ **Option C**

Estimation of parameters: Dispersion

- We sample the following, where each column is an observation and each row a variable:

$$X = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 5 \\ 1 & 5 & 9 \end{bmatrix} \qquad \bar{X} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$$

The empirical variance matrix

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T = \frac{1}{n-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T - \frac{n}{n-1}\bar{\mathbf{X}}\,\bar{\mathbf{X}}^T.$$

is given by

$$S = \frac{1}{3-1}\left\{\left(\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}\right)\left(\begin{bmatrix} 1 & 3 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}\right) + \left(\begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}\right)\left(\begin{bmatrix} 0 & 4 & 5 \end{bmatrix} - \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}\right) + \left(\begin{bmatrix} 2 \\ 5 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}\right)\left(\begin{bmatrix} 2 & 5 & 9 \end{bmatrix} - \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}\right)\right\} =$$

$$\frac{1}{2}\left\{\begin{bmatrix} 0 \\ -1 \\ -4 \end{bmatrix}\begin{bmatrix} 0 & -1 & -4 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}\begin{bmatrix} 1 & 1 & 4 \end{bmatrix}\right\} =$$

$$\frac{1}{2}\left\{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 4 & 16 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 4 \\ 1 & 4 & 16 \end{bmatrix}\right\} = \frac{1}{2}\begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 8 \\ 4 & 8 & 32 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 2 \\ 0.5 & 1 & 4 \\ 2 & 4 & 16 \end{bmatrix}$$

Estimation of parameters II

- We sample the following, where each column is an observation and each row a variable:

$$X = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 5 \\ 1 & 5 & 9 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 0.5 & 2 \\ 0.5 & 1 & 4 \\ 2 & 4 & 16 \end{bmatrix}$$

The correlation matrix $Cor(X,Y)$ is given by

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

With the variance matrix V(X) being the following, what is the correlation matrix?

$$V(X) = \begin{bmatrix} 1 & 0.5 & 2 \\ 0.5 & 1 & 4 \\ 2 & 4 & 16 \end{bmatrix}$$

$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ 0%

$\begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 1 \\ 0.5 & 1 & 1 \end{bmatrix}$ 0%

$\xcancel{\begin{bmatrix} 2 & 1 & 4 \\ 1 & 2 & 8 \\ 4 & 8 & 32 \end{bmatrix}}$ 0%

$\begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}$ 0%

Don't know 0%

Covariance do always have 1 in the diagonal

Estimation of parameters II

- We sample the following, where each column is an observation and each row a variable:

$$X = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 5 \\ 1 & 5 & 9 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 0.5 & 2 \\ 0.5 & 1 & 4 \\ 2 & 4 & 16 \end{bmatrix}$$

The correlation matrix is given by

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

$$S = \begin{bmatrix} 1 & 0.5 & 2 \\ 0.5 & 1 & 4 \\ 2 & 4 & 16 \end{bmatrix} \qquad \rho_{12} = \frac{0.5}{\sqrt{1 \cdot 1}} = 0.5 \qquad \rho_{13} = \frac{2}{\sqrt{1 \cdot 16}} = 0.5 \qquad \rho_{23} = \frac{4}{\sqrt{1 \cdot 16}} = 1$$

$$\mathrm{Corr} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 1 \\ 0.5 & 1 & 1 \end{bmatrix}$$ **Option B**

1st and 2nd order Moments - summary of calculation rules (section 1.1.4)

- **Univariate:**

$E[a + bX] = a + bE[X]$

$E[X + Y] = E[X] + E[Y]$

$V[a + bX] = b^2 V[X]$

$V[X + Y] \quad = V[X] + V[Y] + 2Cov[X, Y]$

$(= V[X] + V[Y] \text{ if } X, Y \text{ are independent})$

$Cov[X, X] = V[X]$

$Cov[X, Y] = Cov[Y, X]$

$Cov[aX + bY] = abCov[X, Y]$

$Cov[X, Y + Z] = Cov[X, Y] + Cov[X, Z]$

- **Multivariate:**

$E[A + X] = A + E[X]$

$E[AX] = AE[X]$

$E[XB] = E[X]B$

$E[X + Y] = E[X] + E[Y]$

$V[A + BX] = BV[X]B^T$

$V[AX] = AV[X]A^T$

$V[X + Y] = V[X] + V[Y] + Cov[X, Y] + Cov[X, Y]^T$

$(= V[X] + V[Y] \text{ if } X, Y \text{ are independent})$

$Cov[X, X] = V[X]$

$Cov[X, Y] = Cov[Y, X]^T$

$Cov[AX, BY] = ACov[X, Y]\mathrm{B}^{\mathrm{T}}$

$Cov[X, Y + Z] = Cov[X, Y] + Cov[X, Z]$

Example of computing moments - I

Rule 1: $E[A+X] = A + E[X]$

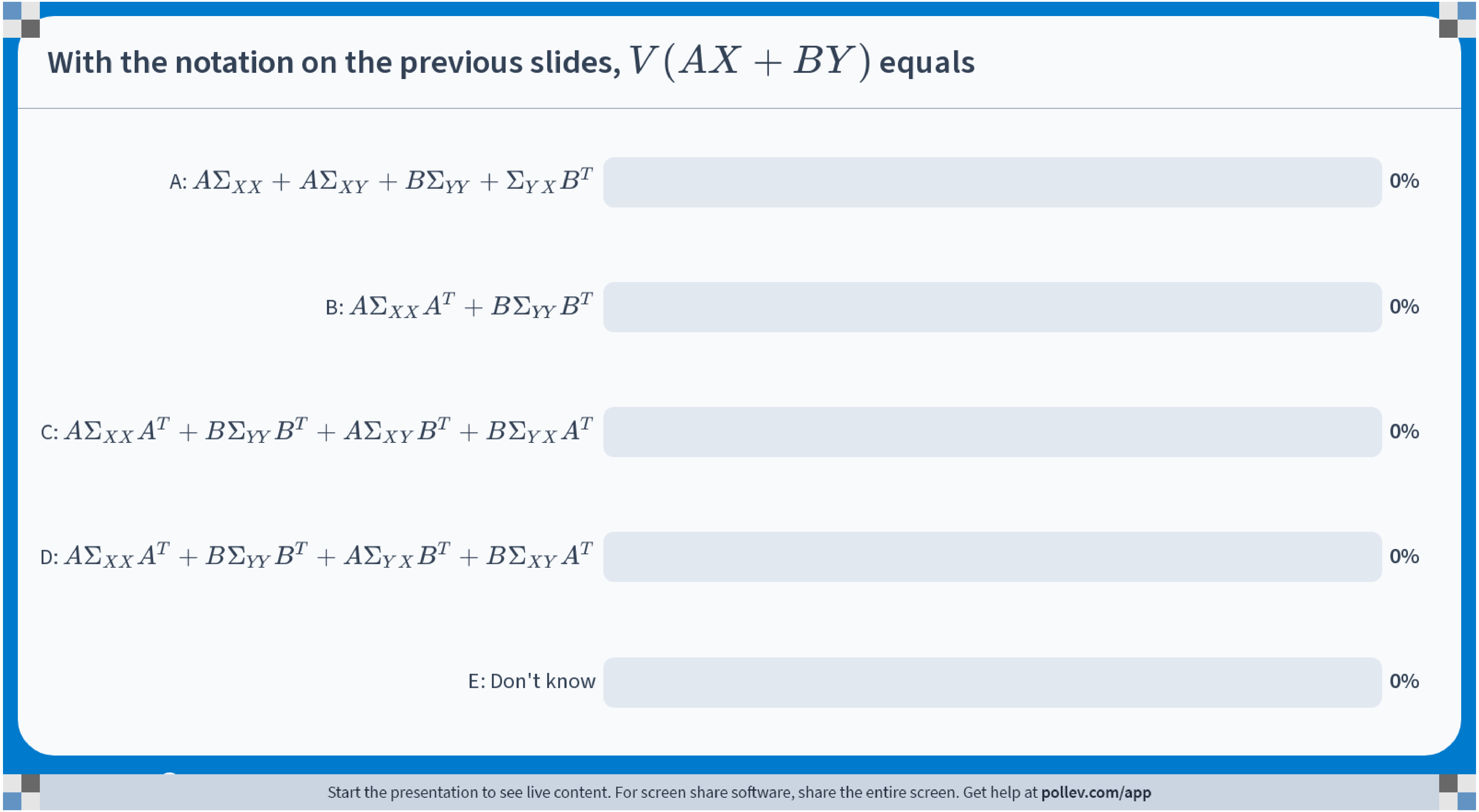
Rule 2: $E[AX] = AE[X]$

Rule 3: $E[XB] = E[X]B$

Rule 4: $E[X+Y] = E[X] + E[Y]$

- Consider the two random vectors X and Y. Assume that
- $E[X] = \mu_x,\ V(X) = \Sigma_{XX},\ E[Y] = \mu_Y,\ V[Y] = \Sigma_{YY},\ Cov[X,Y] = \Sigma_{XY}, Cov[Y,X] = \Sigma_{YX}$.
- Further, assume that the constant c and matrices A, B all have dimensions so that the involved matrix-vector operations are well defined.
- Let us work out $E[AX + BY + c]$:

$$\begin{aligned} E[AX+BY+c] &= E[AX+BY] + c && \text{Rule 1} \\ &= E[AX] + E[BY] + c && \text{Rule 4} \\ &= AE[X] + BE[Y] + c && \text{Rule 2} \\ &= A\,\mu_x + B\mu_y + c \end{aligned}$$



With the notation on the previous slides, $V(AX+BY)$ equals

A: $A\Sigma_{XX} + A\Sigma_{XY} + B\Sigma_{YY} + \Sigma_{YX}B^T$ 0%

B: $A\Sigma_{XX}A^T + B\Sigma_{YY}B^T$ 0%

C: $A\Sigma_{XX}A^T + B\Sigma_{YY}B^T + A\Sigma_{XY}B^T + B\Sigma_{YX}A^T$ 0%

D: $A\Sigma_{XX}A^T + B\Sigma_{YY}B^T + A\Sigma_{YX}B^T + B\Sigma_{XY}A^T$ 0%

E: Don't know 0%

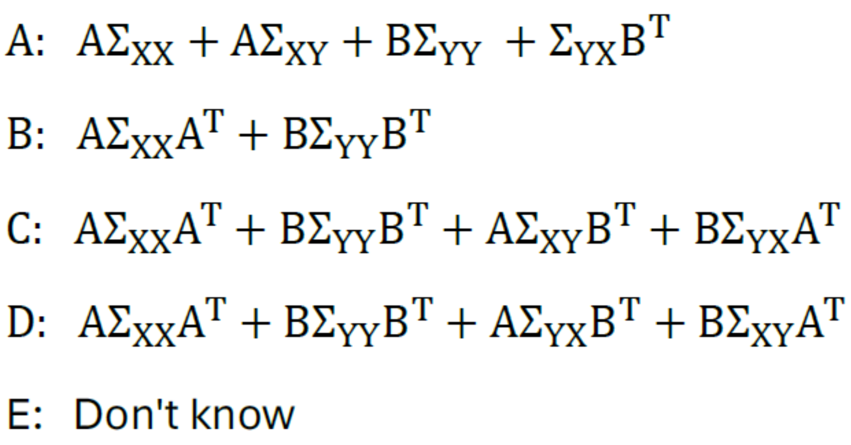
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Example of computing moments - II

- Let us work out $V[AX + BY]$:

$$\begin{aligned} V[AX+BY] &= V[AX] + V[BY] + Cov[AX,BY] + Cov[AX,BY]^T && \textit{Rule } 3 \\ &= AV[X]A^T + BV[Y]B^T + Cov[AX,BY] + Cov[BY,AX] && \textit{Rule } 2, \textit{Rule } 5 \\ &= AV[X]A^T + BV[Y]B^T + ACov[X,Y]B^T + BCov[Y,X]A^T && \textit{Rule } 6 \\ &= A\Sigma_{XX}A^T + B\Sigma_{YY}B^T + A\Sigma_{XY}B^T + B\Sigma_{YX}A^T \end{aligned}$$

Thus **option C**.



A: $A\Sigma_{XX} + A\Sigma_{XY} + B\Sigma_{YY} + \Sigma_{YX}B^T$

B: $A\Sigma_{XX}A^T + B\Sigma_{YY}B^T$

C: $A\Sigma_{XX}A^T + B\Sigma_{YY}B^T + A\Sigma_{XY}B^T + B\Sigma_{YX}A^T$

D: $A\Sigma_{XX}A^T + B\Sigma_{YY}B^T + A\Sigma_{YX}B^T + B\Sigma_{XY}A^T$

E: Don't know

*Rule*1: $V[A + BX] = BV[X]B^T$

*Rule*2: $V[AX] = AV[X]A^T$

*Rule*3: $V[X + Y] = V[X] + V[Y] + Cov[X,Y] + Cov[X,Y]^T$

($= V[X] + V[Y]$ if X, Y are independent)

Rule 4: $Cov[X,X] = V[X]$

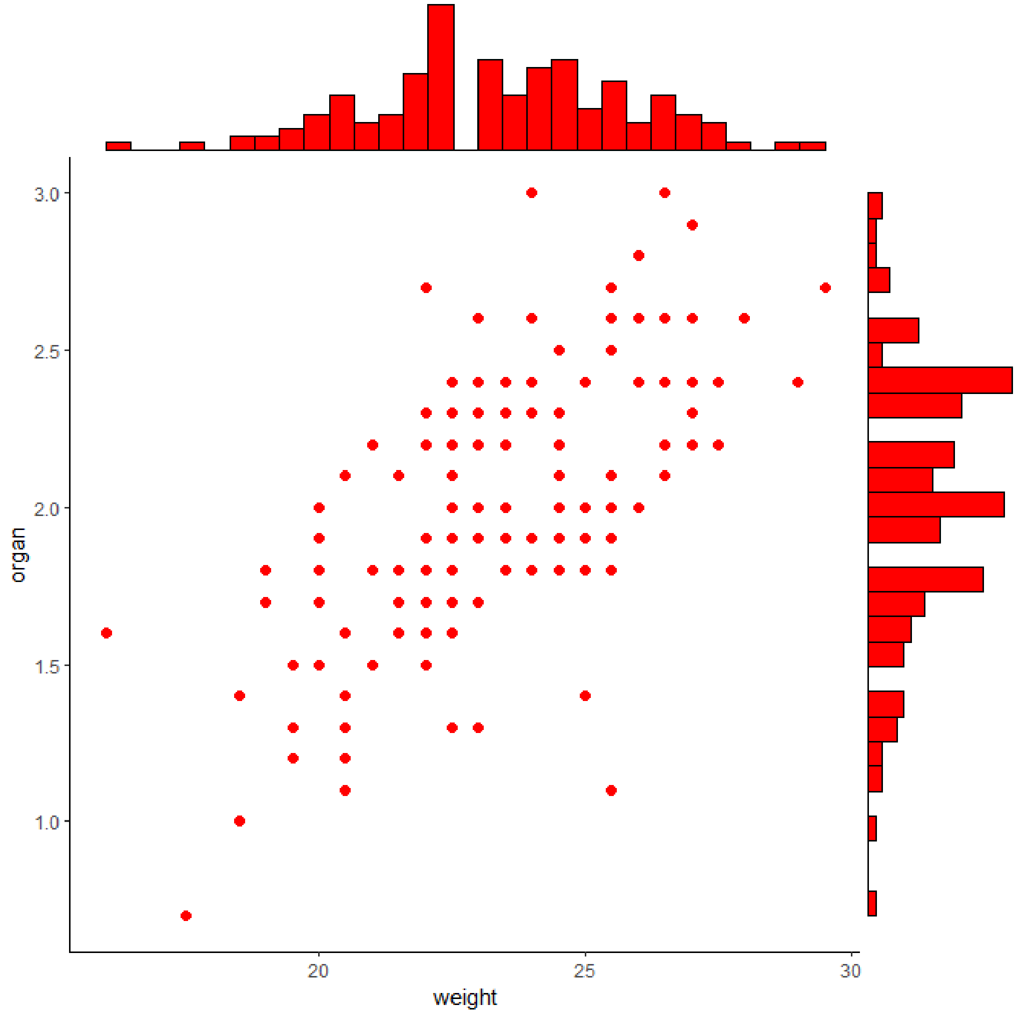
Rule 5: $Cov[X,Y] = Cov[Y,X]^T$

Rule 6: $Cov[AX,BY] = ACov[X,Y]B^T$

Rule 7: $Cov[X,Y+Z] = Cov[X,Y] + Cov[X,Z]$

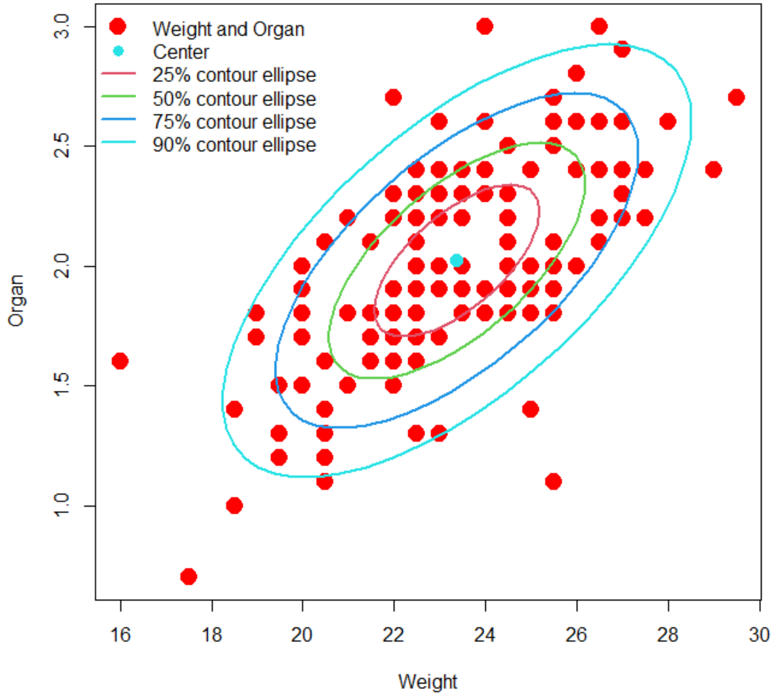
RETURNING TO ORGAN DATA

Scatterplot and marginal histograms for Apodemus Data



Weight of mouse and weight of its reproductive organs I

	N	Mean	SD	Min	Max
weight	149	23.3758	2.2738	16	23.3758
organ	149	2.0215	0.4178	0.7	2.0214



Covariance matrix:

```
              weight     organ
weight 5.6348177 0.6442386
organ  0.6442386 0.1745356
```

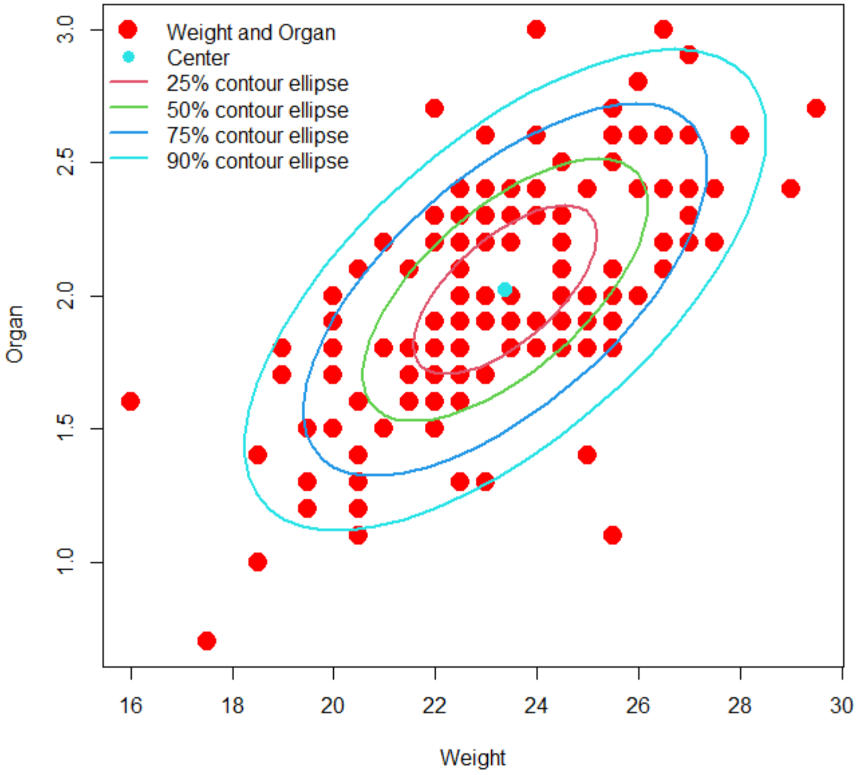
Correlation matrix:

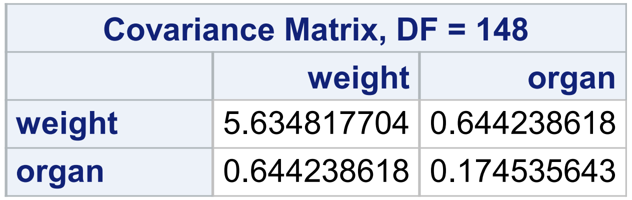
```
               weight    organ
  weight 1.000000 0.649628
  organ  0.649628 1.000000
```

Test:

```
        Pearson's product-moment correlation
data:  organ$weight and organ$organ
t = 10.36, df = 147, p-value < 2.2e-16
```

Weight of mouse and weight of its reproductive organs II





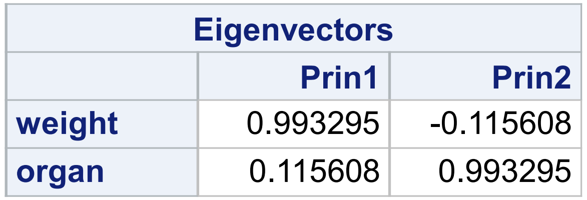
Covariance Matrix, DF = 148		
	weight	organ
weight	5.634817704	0.644238618
organ	0.644238618	0.174535643

$$f(x) = c \Leftrightarrow (x-\mu)'\Sigma^{-1}(x-\mu) = k$$

Contour 'lines' are ellipsoids with main axes given by eigenvectors of the variance matrix

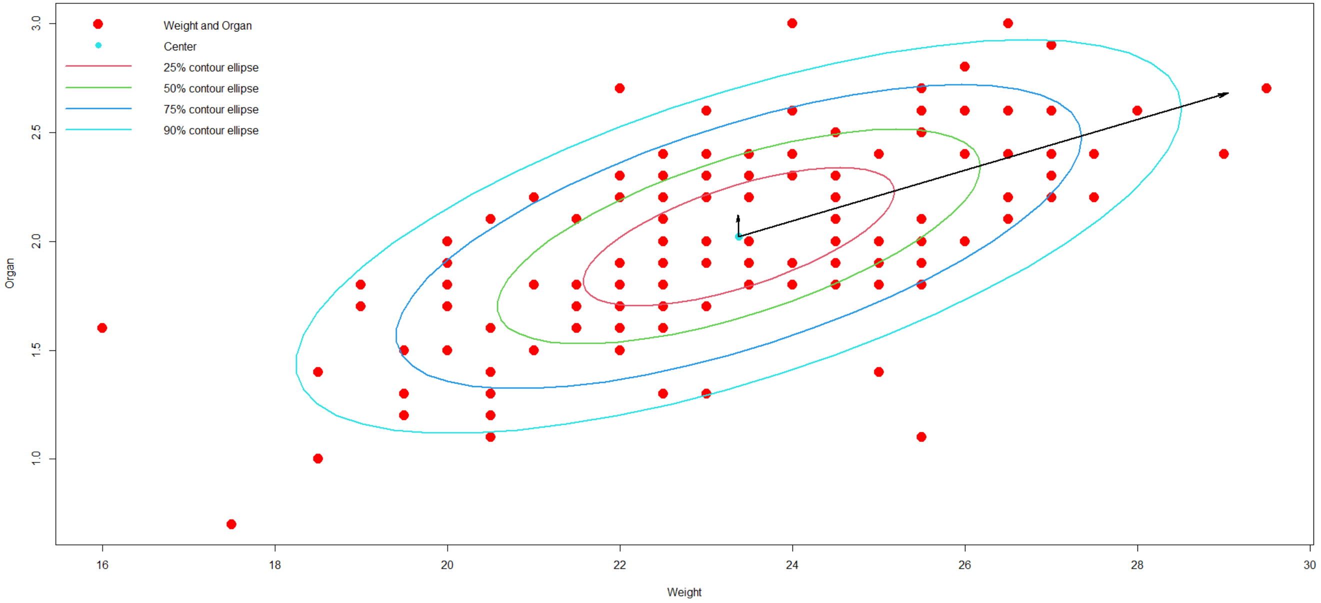
Eigenvectors:

```
eigen(var(organ))$values
# [1] 5.70979939 0.09955396
my.eigenvectors<--eigen(var(organ))$vectors
```



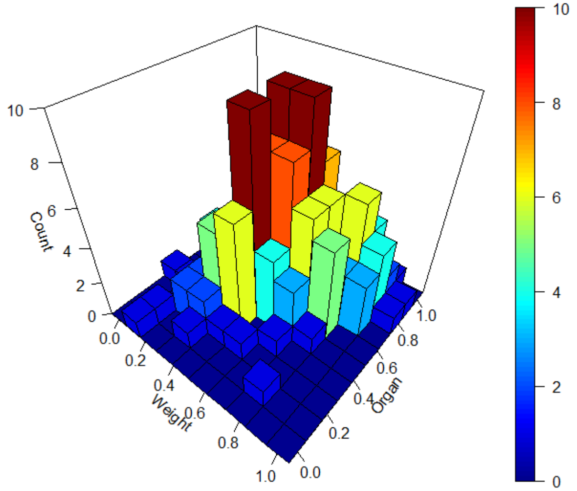
Eigenvectors		
	Prin1	Prin2
weight	0.993295	-0.115608
organ	0.115608	0.993295

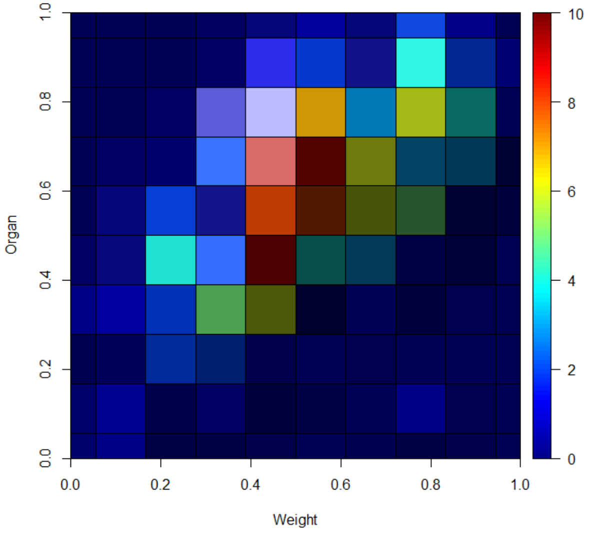
Weight of mouse and weight of its reproductive organs III

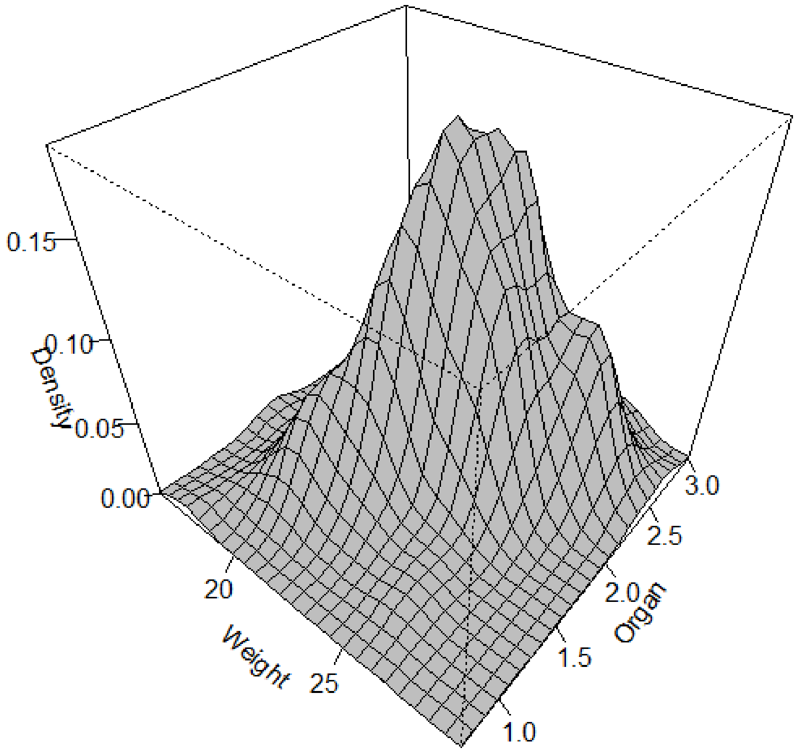


- Eigenvectors do not look perpendicular; but they are; axes are not scaled similarly.

Bivariate histogram, heatmap and kernel pdf

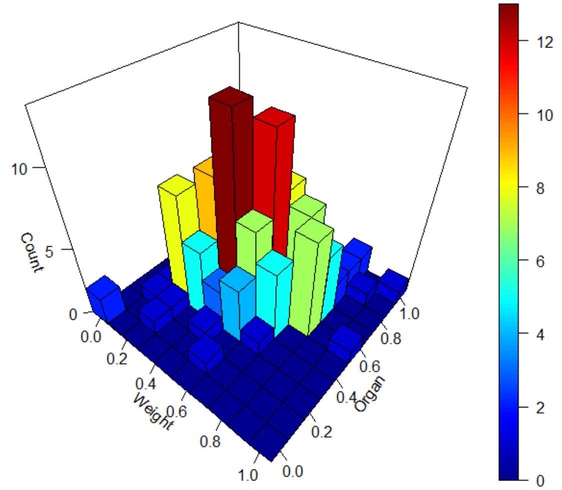


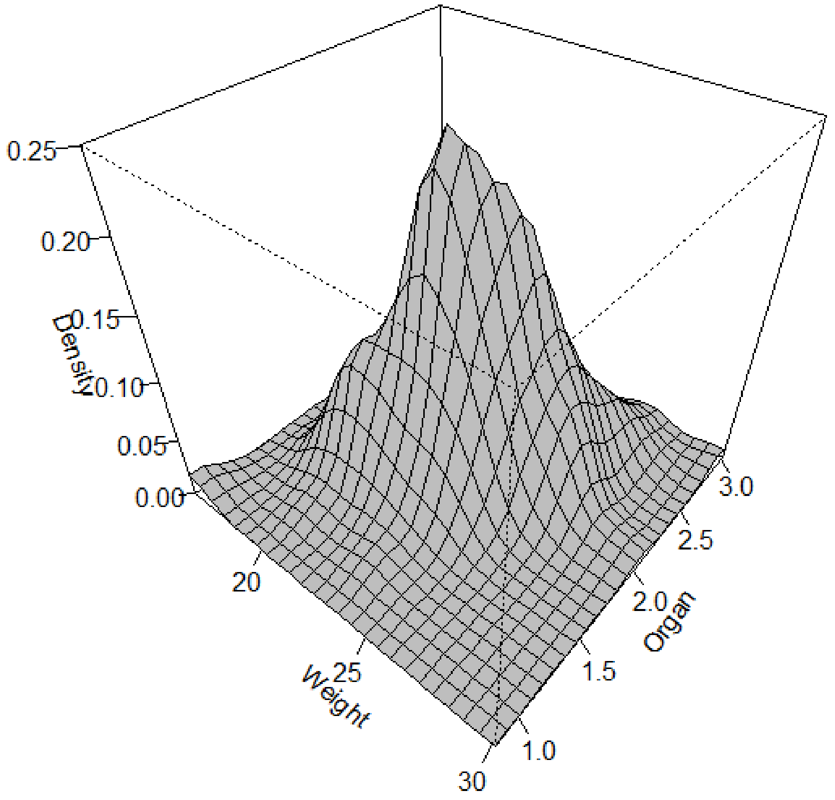


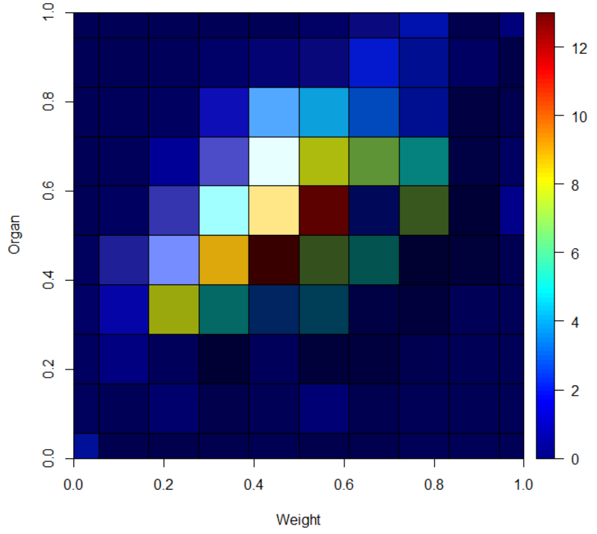


Bivariate histogram, heatmap and kernel pdf

Simulated normal data







25,000 records of human heights and weights

Source: Statistics Online Computational Ressource

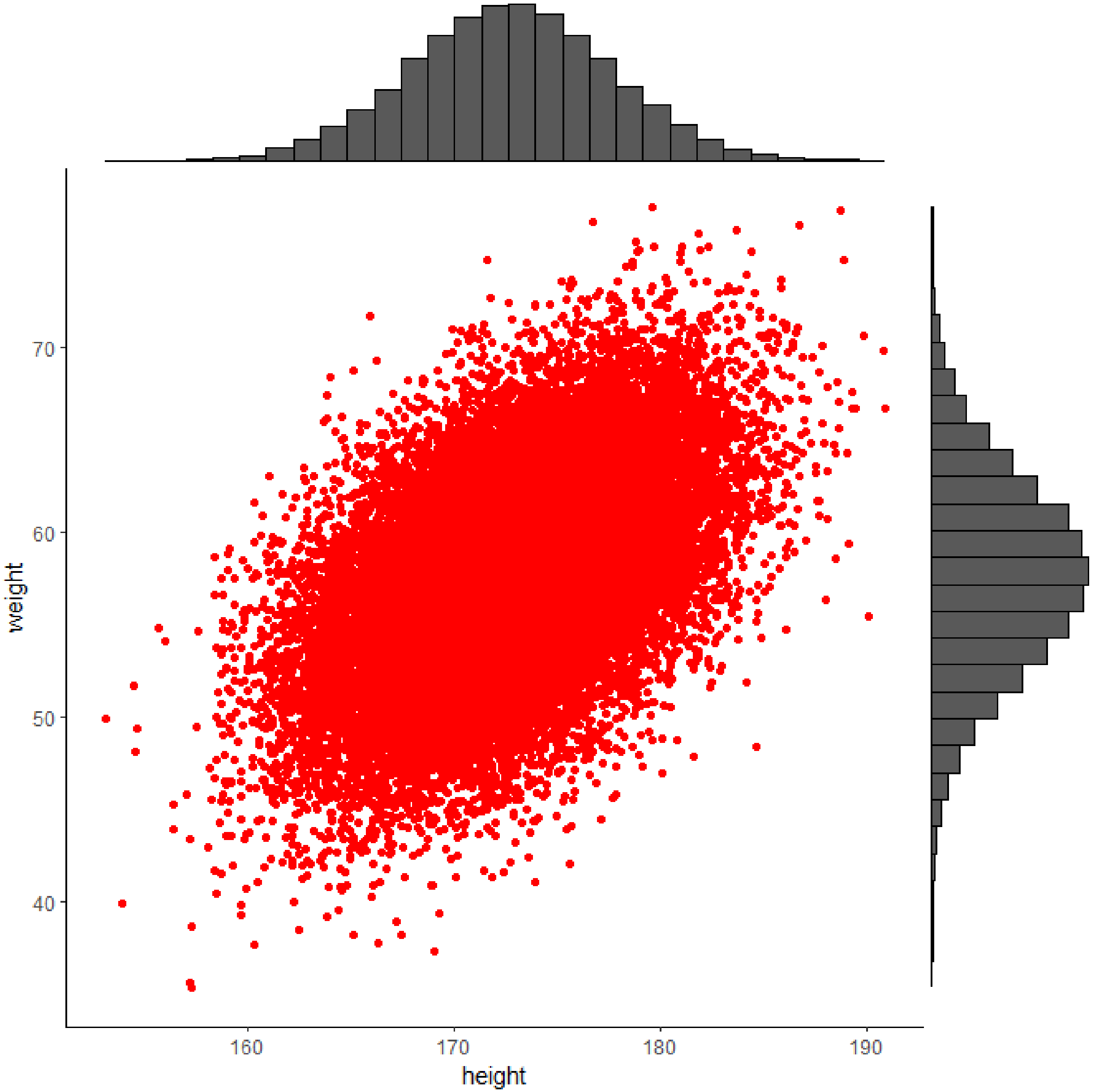
http://socr.ucla.edu/docs/resources/SOCR_Data/SOCR_Data_Dinov_020108_HeightsWeights.html

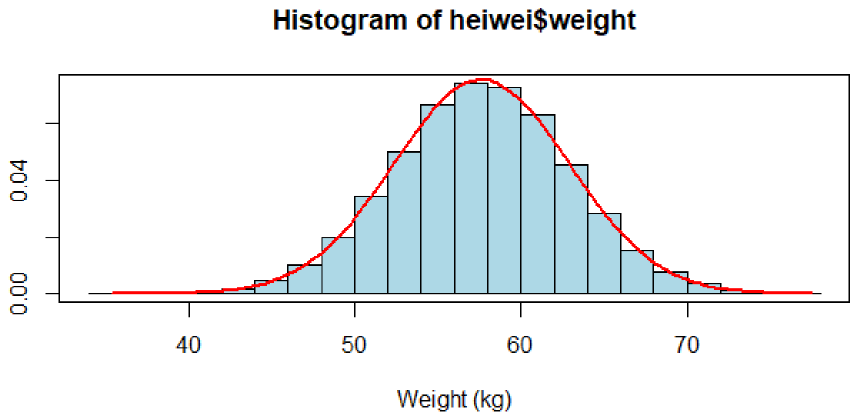
The dataset below contains 25,000 records of human heights and weights. These data were obtained in 1993 by a Growth Survey of 25,000 children from birth to 18 years of age recruited from Maternal and Child Health Centres (MCHC) and schools and were used to develop Hong Kong's current growth charts for weight, height, weight-for-age, weight-for-height and body mass index (BMI).

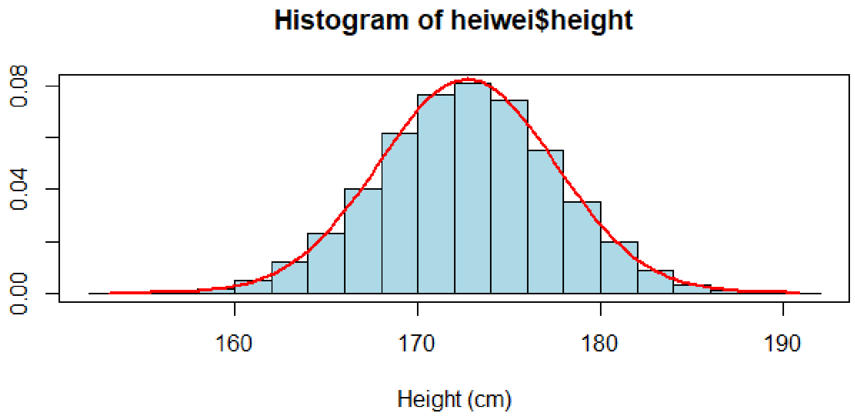
```
head(heiwei)
    height   weight
1 167.0896 51.25249
2 181.6486 61.90955
3 176.2728 69.41178
4 173.2702 64.56220
5 172.1810 65.45201
6 174.4925 55.92898
```

- Such a large dataset is less sensitive to individual measurements and the structure shows more clearly.

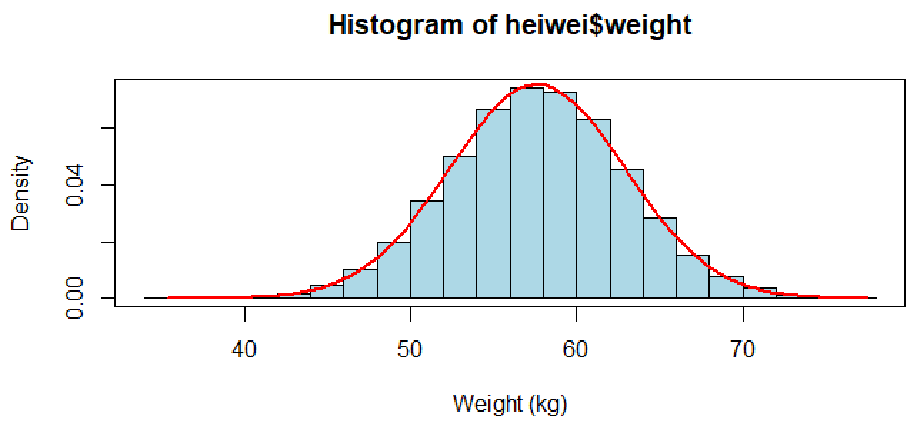
Univariate Histograms and pdf estimates

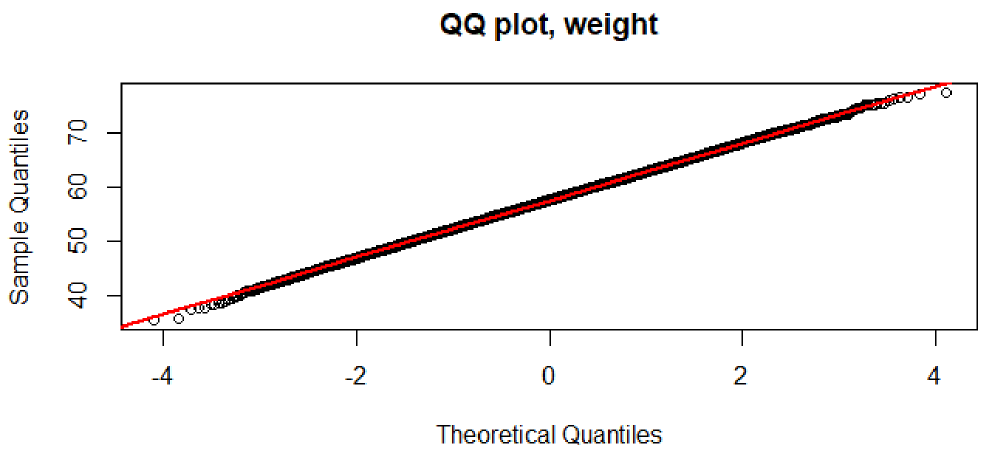


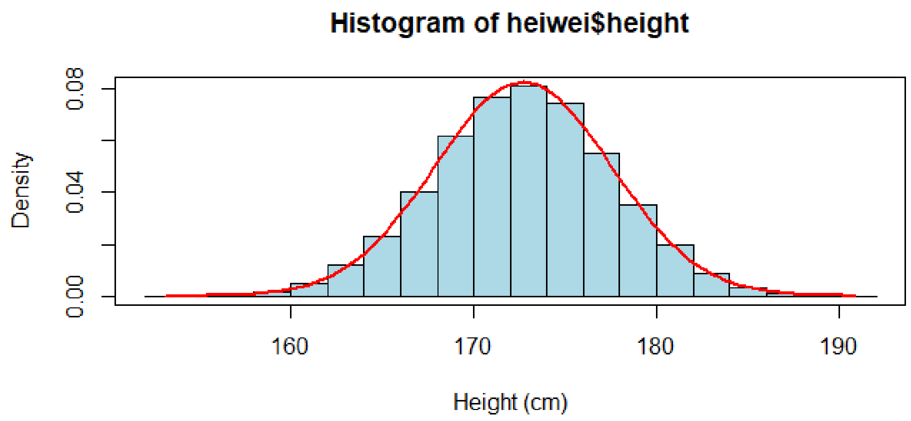


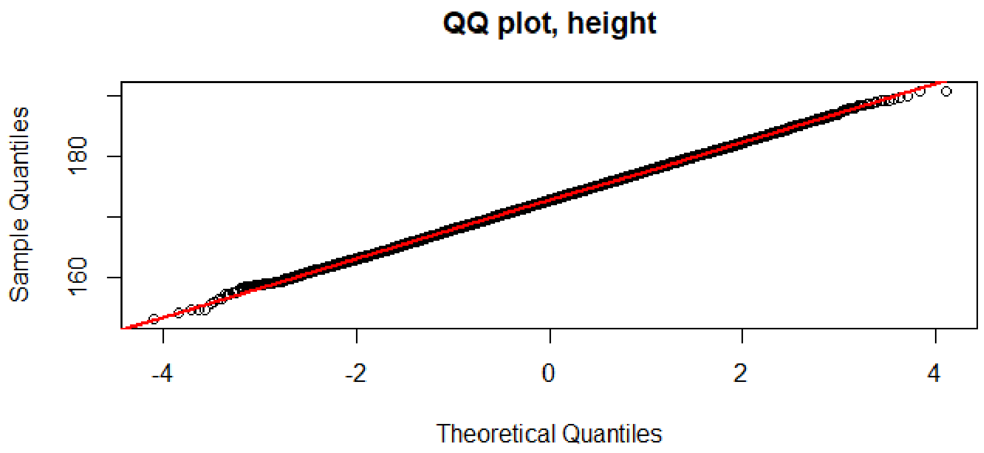


Univariate Histograms and QQ plots



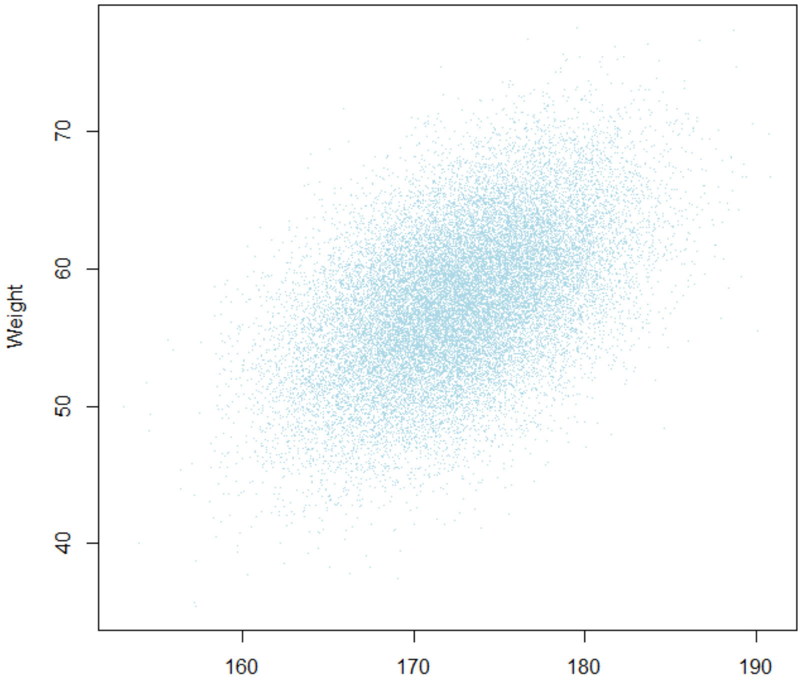


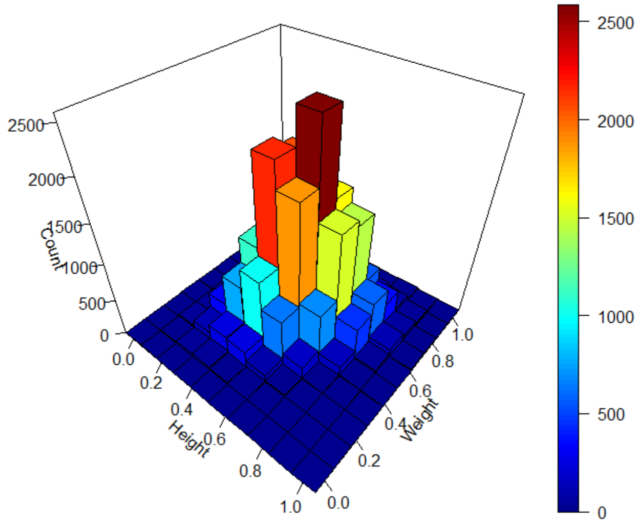


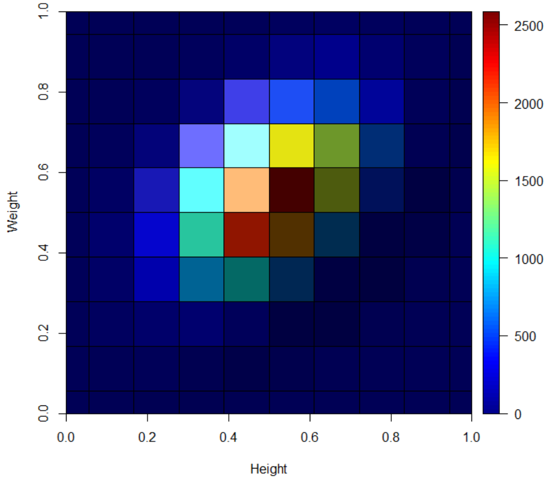


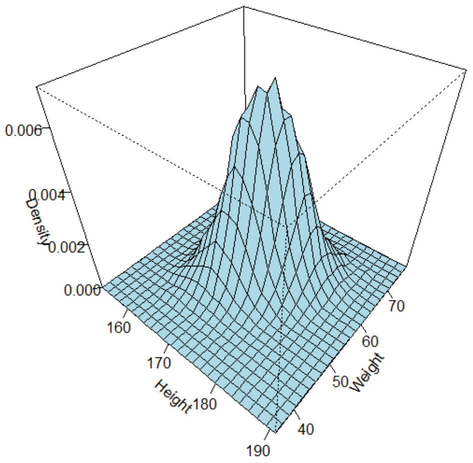
- Heiwei data are **MARGINALLY** NORMAL

Bivariate Histogram, heatmap and Kernel pdf









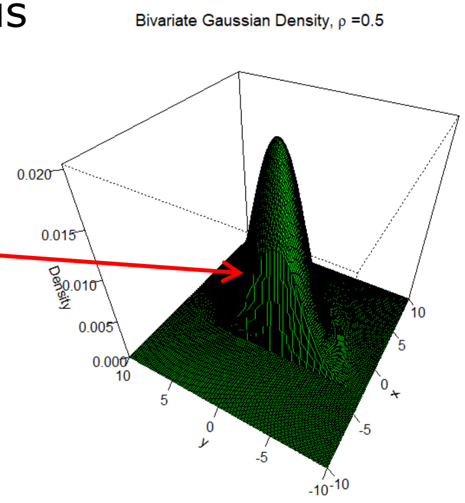
Conditional distributions

In probability theory, the **conditional probability** of an event A given another event B has occurred, is

$$P(A|B) \coloneqq \frac{P(A \cap B)}{P(B)}$$

Transferring this result to random variables gives the concept of the conditional distribution of one variable X given another variable Y. If (X, Y) has simultaneous density $f_{XY}(x, y)$, then the conditional density for Y given $X = x$ is

$$f(y|x) \coloneqq \frac{f_{XY}(x,y)}{f_X(x)}$$



Conditional distributions

The distribution of a variable X, when another variable Y is considered known.

We shall exemplify the concept with the normal distribution.

Note that

$$P(X \in A, Y \in B) = P(X \in A | Y \in B) P(Y \in B)$$

In particular,

$$P(X \in x + dx, Y y + dy) = P(X \in x + dx | Y \in y + dy) | P(Y \in y + dy) |$$

Divide by $|dx||dy|$, and go to the limit to obtain densities:

$$f_{XY}(x, y) = f_{X|Y=y}(x) f_Y(y)$$

ie.

$$f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Conditional distributions

If (X_1, X_2) is multivariate normal, the conditional distributions $\left(\mathcal{L}(X_1|X_2 = x_2)\right)$ are all normal for all values of x_2. Mean and variance are given by (Theorem 1.27):

$$X \sim N_p(\mu, \Sigma), \qquad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}:$$

$$E(X_1|\, X_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
$$V(X_1|\, X_2 = x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

The variance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is *independent* of x_2. Important feature for the normal distribution

Let

Conditional regression

The variance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is *independent* of x_2:

If (X_1, X_2) is multivariate normal, the residuals (ε_i) in the (potentially multivariate) regression

$$X_{1i} = \alpha + \beta x_{2i} + \varepsilon_i, i = 1, \dots, n$$

are independent and identically normal distributed. The converse is also true because

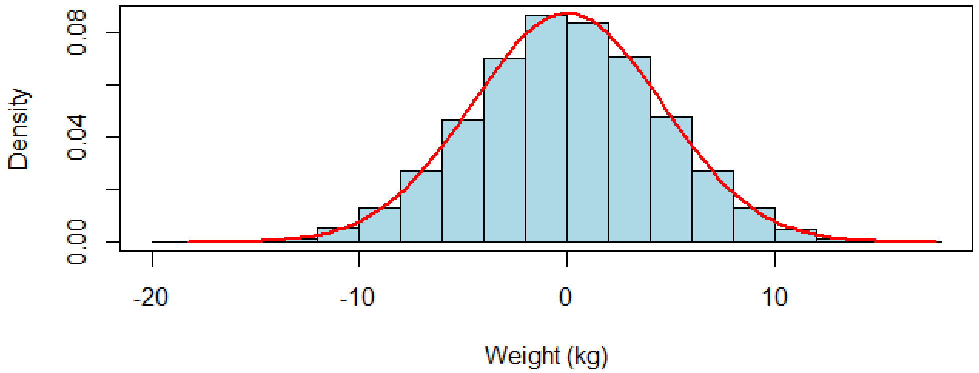
$$f_{XY}(x, y) = f_{X|Y=y}(x) f_Y(y)$$

THUS:

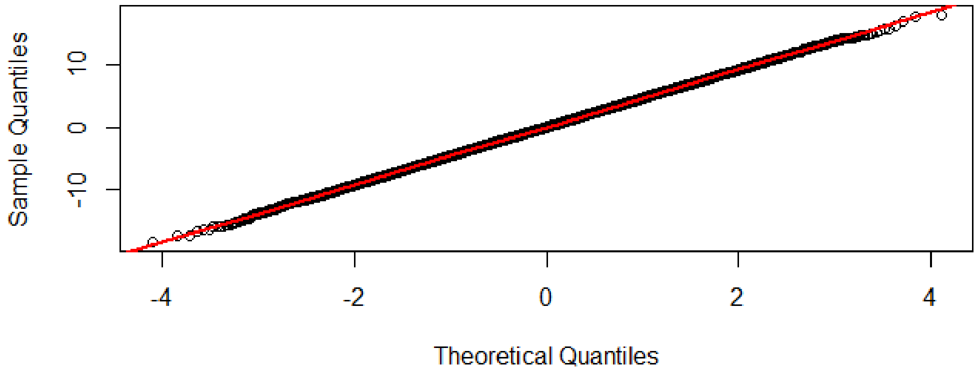
$(\boldsymbol{X_1}, \boldsymbol{X_2})$ is multivariate normal iff $(\boldsymbol{X_1}, \boldsymbol{X_2})$ is marginally normal, and the conditional distributions are normal!

Univariate Histograms and QQ plots

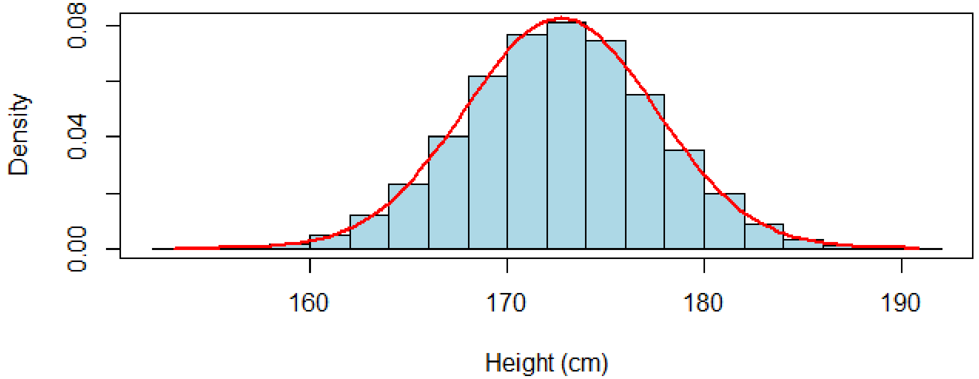
Histogram of Estimated Residuals
Regressing Weight on Height



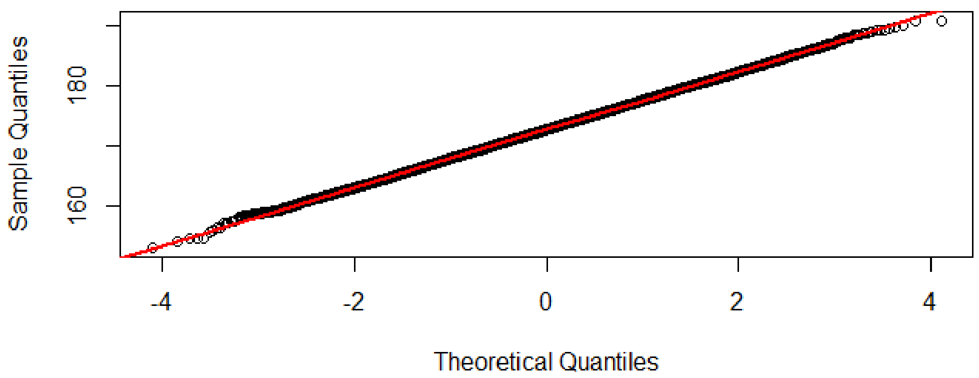
QQ plot, Estimated Residuals
Regressing Weight on Height



Histogram of Height



QQ plot of Height



- Heiwei data are **SIMULTANEOUSLY** NORMAL

Todays exercises

- make sure you have R up and running

- Exercise 1.1
  - Brush up on linear algebra
  - Properties of variance and correlation matrices

- Exercise 1.2 (Exam 2013, problem 2)
  - Calculation of multivariate mean, variance matrices and conditional mean

- Exercise E1: Exercise on the impact of correlation.

- FEEL FREE TO USE MAPLE / SYMBOLIC MATH-SOFTWARE!