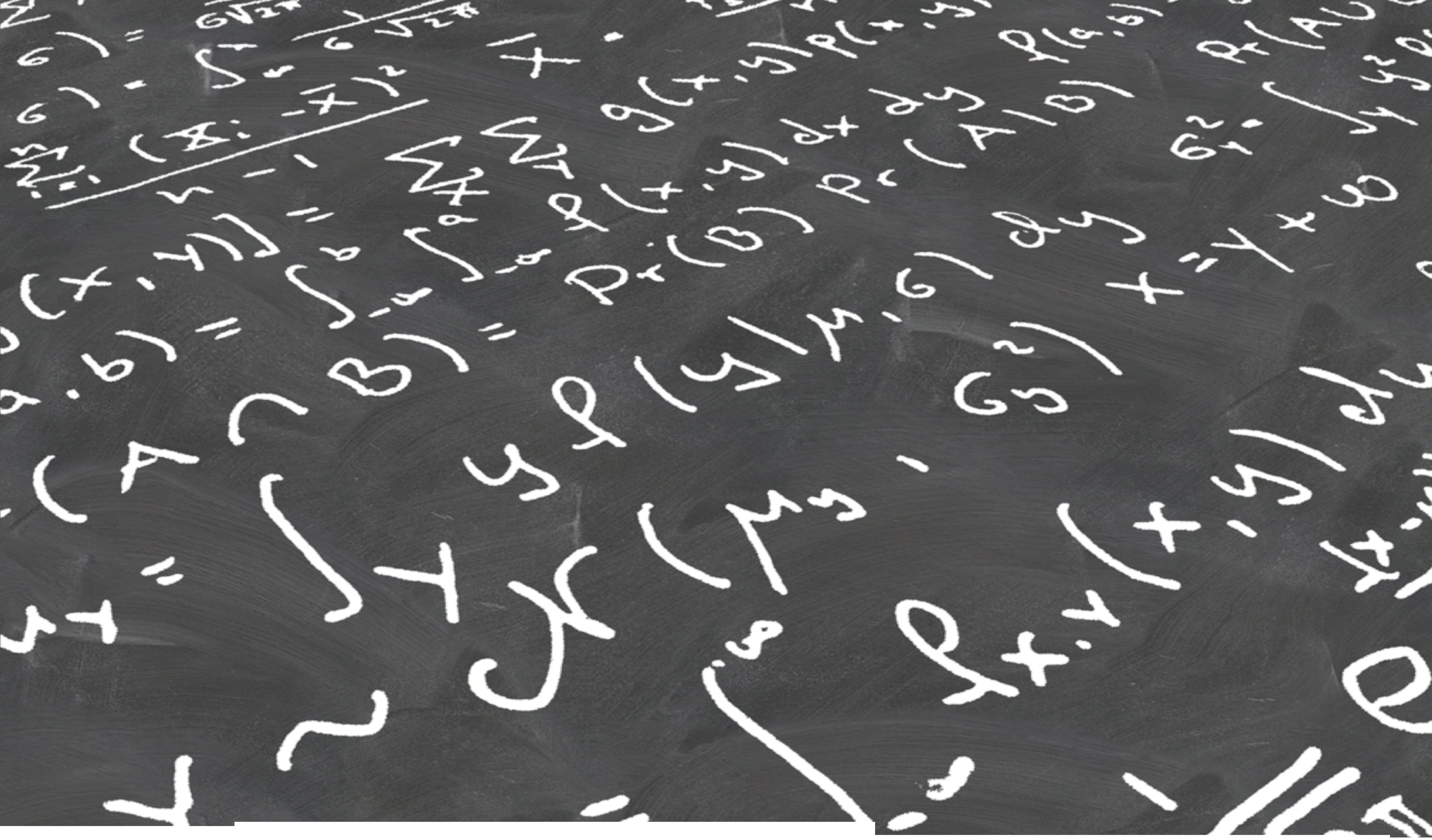
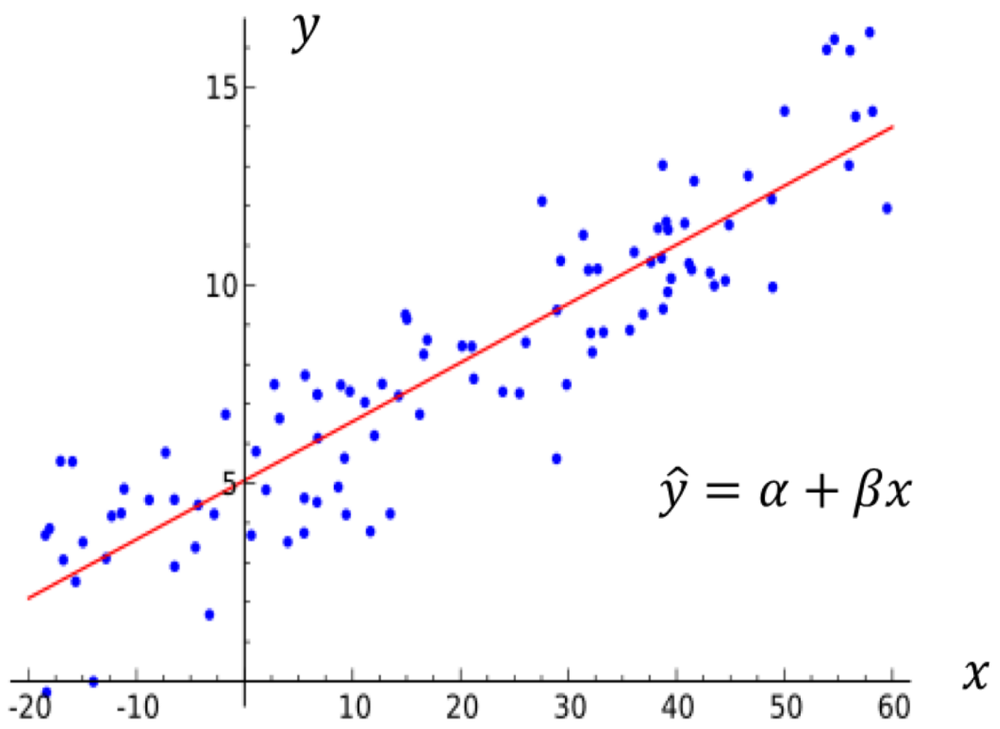


13. Linear Regression Models

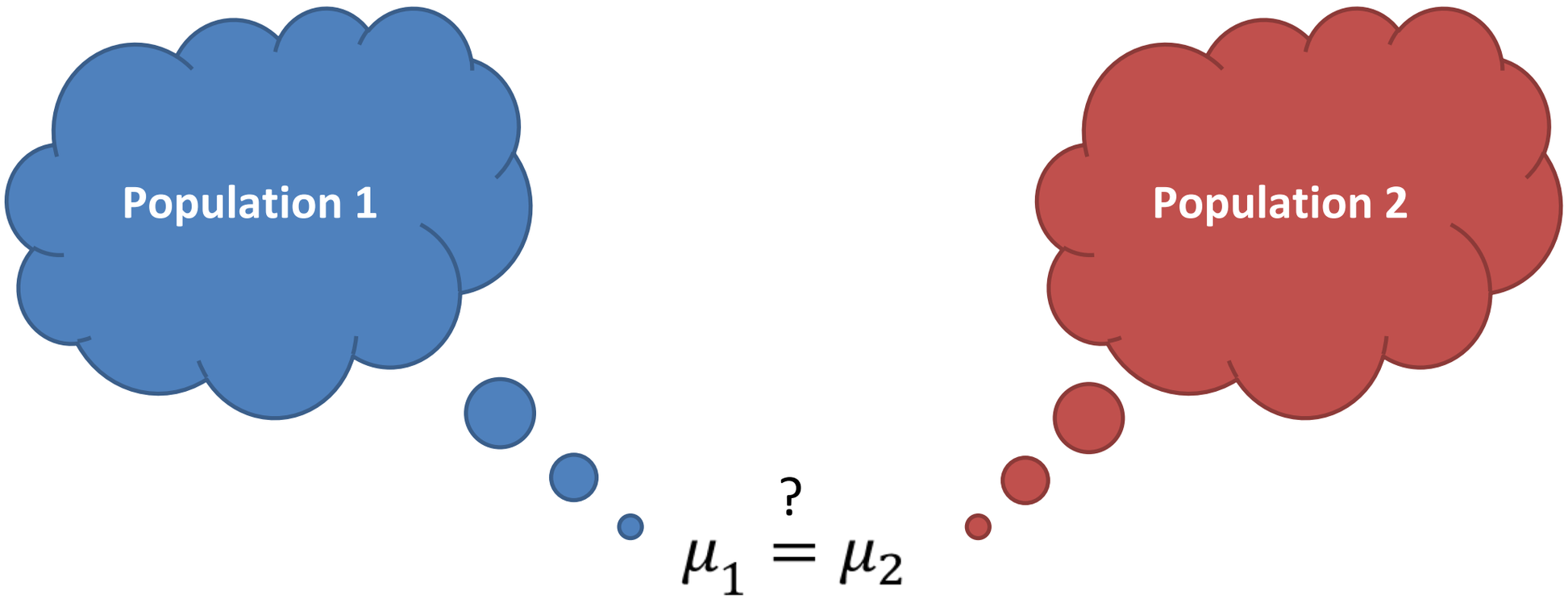
Gunvor Elisabeth Kirkelund
Lars Mandrup
Slides and material provided in parts by
Henrik Pedersen

Todays Content

- Repetition from last time
- Linear regression
- RANSAC



Comparing two population means



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

Test catalog for Comparing Two Means (known variance)

Statistical model:

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

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

Test Catalog for Comparing Two Means (unknown variance)

Statistical model:

- $X_{1i} \sim N\left(\mu_1, \sigma_1^2\right), i = 1,2,\dots,n_1$ and $X_{2i} \sim N\left(\mu_2, \sigma_2^2\right)\ i = 1,2,\dots,n_2$
- 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 $\dot{x_1} - \dot{x_2}$ = '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: $t = \frac{(\dot{x_1}-\dot{x_2})}{s\sqrt{1/n_1+1/n_2}} = \sim t(n_1+n_2-2)$
- Approximate p-value: $2 \cdot \left(1 - t_{cdf}(|t|, n_1+n_2-2)\right)$

95% confidence interval:

- $\delta_- = (\dot{x_1} - \dot{x_2}) - t_0 \cdot s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
- $\delta_+ = (\dot{x_1} - \dot{x_2}) + t_0 \cdot s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
- 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*

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:

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

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

- Where the observation is $\bar{d}$ = '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: $t = \frac{\bar{d}-\delta_0}{s_d/\sqrt{\mathrm{n}}} = \sim t(\mathrm{n}-1)$
- Approximate p-value: $2 \cdot \left(1 - t_{cdf}(|t|, \mathrm{n}-1)\right)$

95% confidence interval:

- $\delta_- = \bar{d} - t_0 \cdot \frac{s_d}{\sqrt{n}}$
- $\delta_+ = \bar{d} + t_0 \cdot \frac{s_d}{\sqrt{n}}$

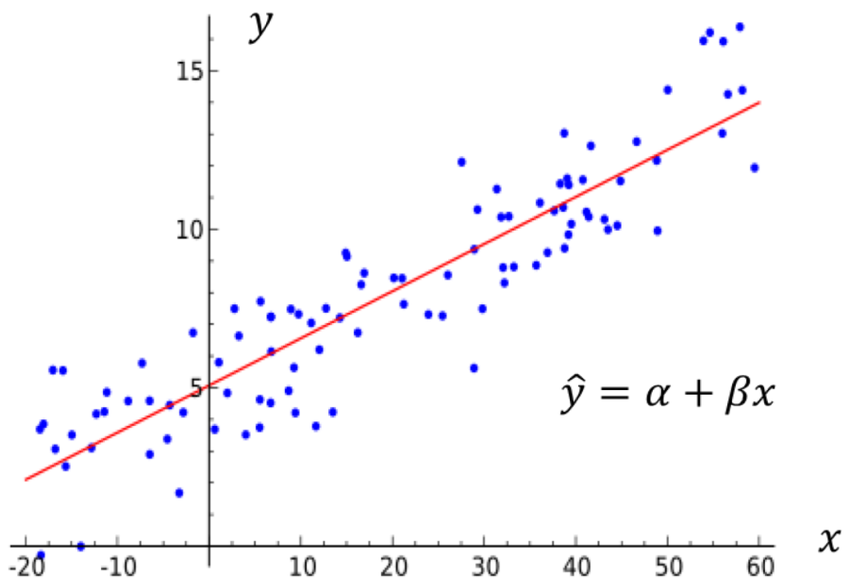
- where t0 = tinv(1-0.05/2,n-1)

Paired test (vs. unpaired test):

- *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_o hopothesis*

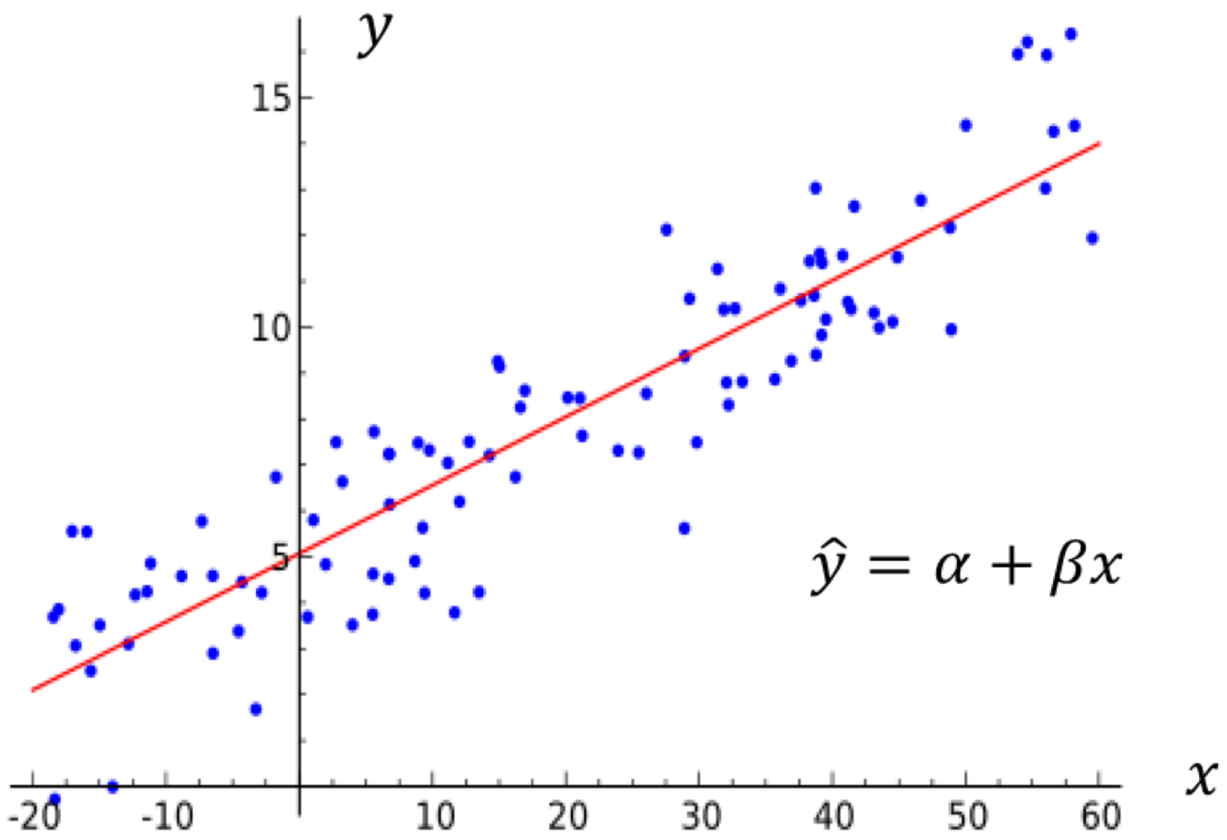
Linear Models – When/Why?

- Can be used when the mean changes over time.
- The variance should not change over time
- The mean is connected to time linearly!
- The simplest model – more advanced models not necesarily the best → start simpel (linear regression)



Linear Regression

- Fits a straight line through the set of n points (x_i, y_i)
- Make the sum of squared residuals ($\epsilon^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \sim \chi^2$ - the vertical distances between til points of data and the fitted line) of the model as small as possible



Statistical Model

- In linear regression, the data come in pairs

$$(x_i, y_i), \qquad \text{for } i = 1,2, \dots, n$$

fx. time t

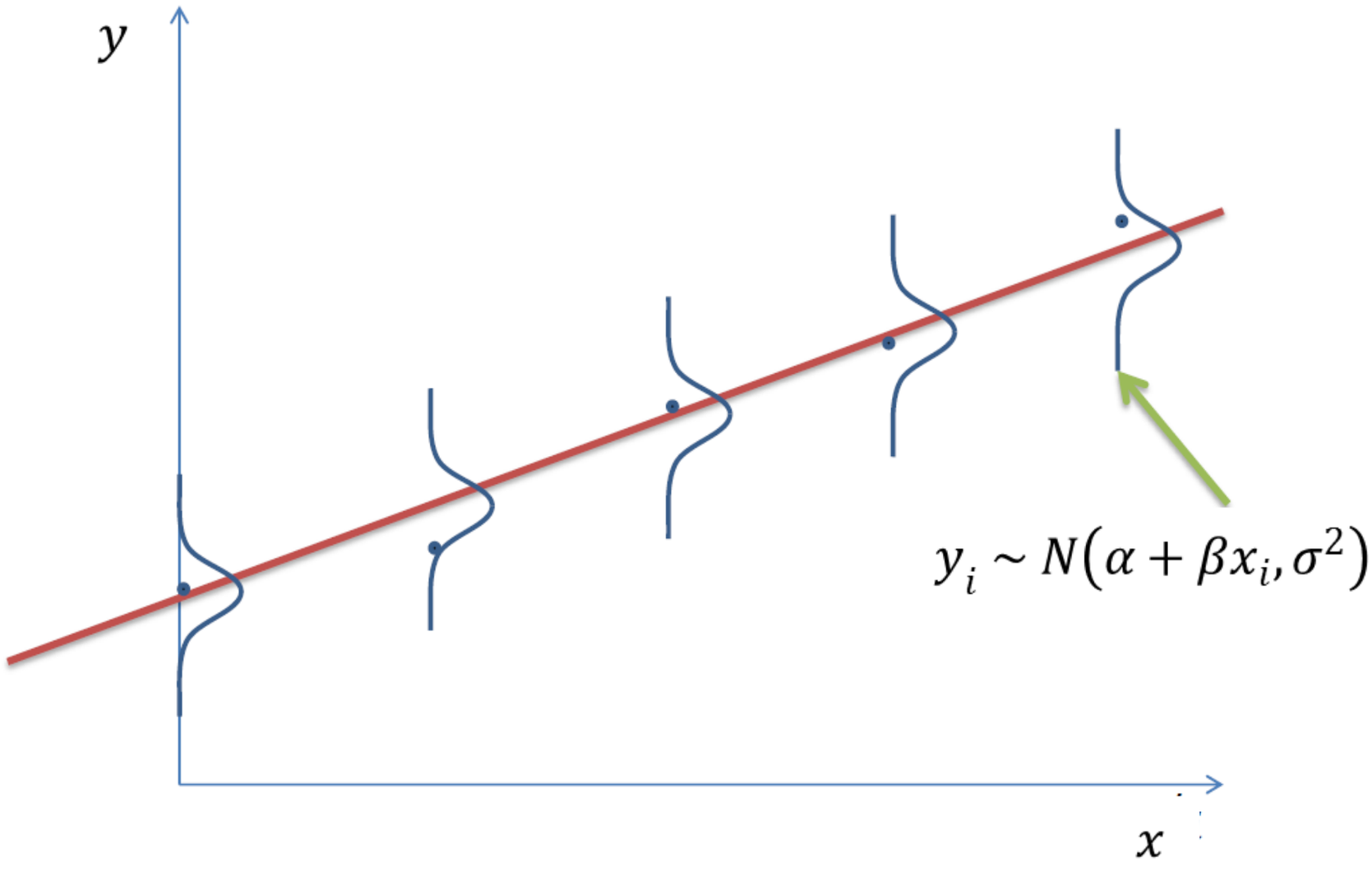
- where x is the independent variable and y is the dependent (or response) variable.
- Statistical model

$$y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

- where β is the slope of the straight line and α it's intercept with the y-axis.

OBS! In "Random Signals": $\alpha \rightarrow b_0$ and $\beta \rightarrow b_1$

Statistical Model



Residual

- A residual is the difference between the measured and predicted data.
- The residual of the *i*'th sample (y_i) for a given choice of α and β is denoted ε_i and is given by

$$\epsilon_i = y_i - (\alpha + \beta x_i), \qquad \text{for } i = 1,2,\dots,n$$

measured

estimated ($\hat{y}_i$)

Empirical Variance

- Recall that $y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma^2)$
- The unbiased estimator of the variance is

$$s_r^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - (\hat{\alpha} + \hat{\beta}x_i))^2 = \frac{1}{n-2}\sum_{i=1}^{n}{\epsilon_i}^2$$

Two constrains $(\hat{\alpha}, \hat{\beta})$ → ÷ two degrees of freedom

- Statistical interence in linear regression concerns the parameter estimates: $\hat{\alpha}, \hat{\beta}$ and s_r^2.

Model Fitting

- The goal of linear regression is to determine the choice of slope (β) and intercept (α) that minimizes the sum of squared residuals of the model.

$$R(\alpha, \beta) = \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 ,$$

ϵ_i ^{2}

- The parameter estimates that minimize *R* are

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

- where $\bar{x}$ is the average of $x_1, x_2, \ldots, x_n$ and $\bar{y}$ is the average of $y_1, y_2, \ldots, y_n$.

Derivation of the Intercept Parameter

- Partial derivative w.r.t. α and setting to zero:

$$\frac{\partial R(\alpha,\beta)}{\partial \alpha} = \frac{\partial}{\partial \alpha}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 = -2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0$$

- It follows that:

$$2n\alpha = 2\sum_{i=1}^{n} y_i - 2\beta\sum_{i=1}^{n} x_i = 2n\bar{y} - 2\beta n\bar{x}$$

$$\Downarrow$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Derivation of the Slope Parameter

- Partial derivative w.r.t. β and setting to zero:

$$\frac{\partial R(\alpha,\beta)}{\partial \beta} = \frac{\partial}{\partial \beta}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 = -2\sum_{i=1}^{n} x_i(y_i - \alpha - \beta x_i)$$
$$= -2\sum_{i=1}^{n}\left(x_i y_i - \alpha x_i - \beta x_i^2\right) = -2\sum_{i=1}^{n} x_i y_i + 2\alpha\sum_{i=1}^{n} x_i + 2\beta\sum_{i=1}^{n} x_i^2 = 0$$

- Inserting the result $\alpha = \bar{y} - \beta\bar{x}$ we get:

$$-2\sum_{i=1}^{n} x_i y_i + 2(\bar{y} - \beta\bar{x})\sum_{i=1}^{n} x_i + 2\beta\sum_{i=1}^{n} x_i^2$$
$$= -2\sum_{i=1}^{n} x_i(y_i - \bar{y}) + 2\beta\sum_{i=1}^{n} x_i(x_i - \bar{x})$$
$$= -2\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + 2\beta\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0$$

- It follows that:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i y_i - \bar{x}\,\bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}^2}{s_x^2}$$

- Where: $s_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i y_i - \bar{x}\,\bar{y})$ and $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Sample covariance

Sample variance

Example - Hubble's Law

- Hubble's law is the name for the observation in physical cosmology that objects observed in deep space are found to have a relative velocity away from the Earth that is approximately proportional to their distance from the Earth: $v = H \cdot x$
- Edwin Hubble's original measurements for 24 distant galaxies were (in Matlab notation)

```
Distance = [ 0.032 0.034 0.214 0.263 0.275 0.275 ...
             0.450 0.500 0.500 0.630 0.800 0.900 ...
             0.900 0.900 0.900 1.000 1.100 1.100 ...
             1.400 1.700 2.000 2.000 2.000 2.000 ];

Speed = [ 170 290 -130 -70 -185 -220 200  290 ...
          270 200  300 -30  650  150 500  920 ...
          450 500  500 960  500  850 800 1090 ];
```

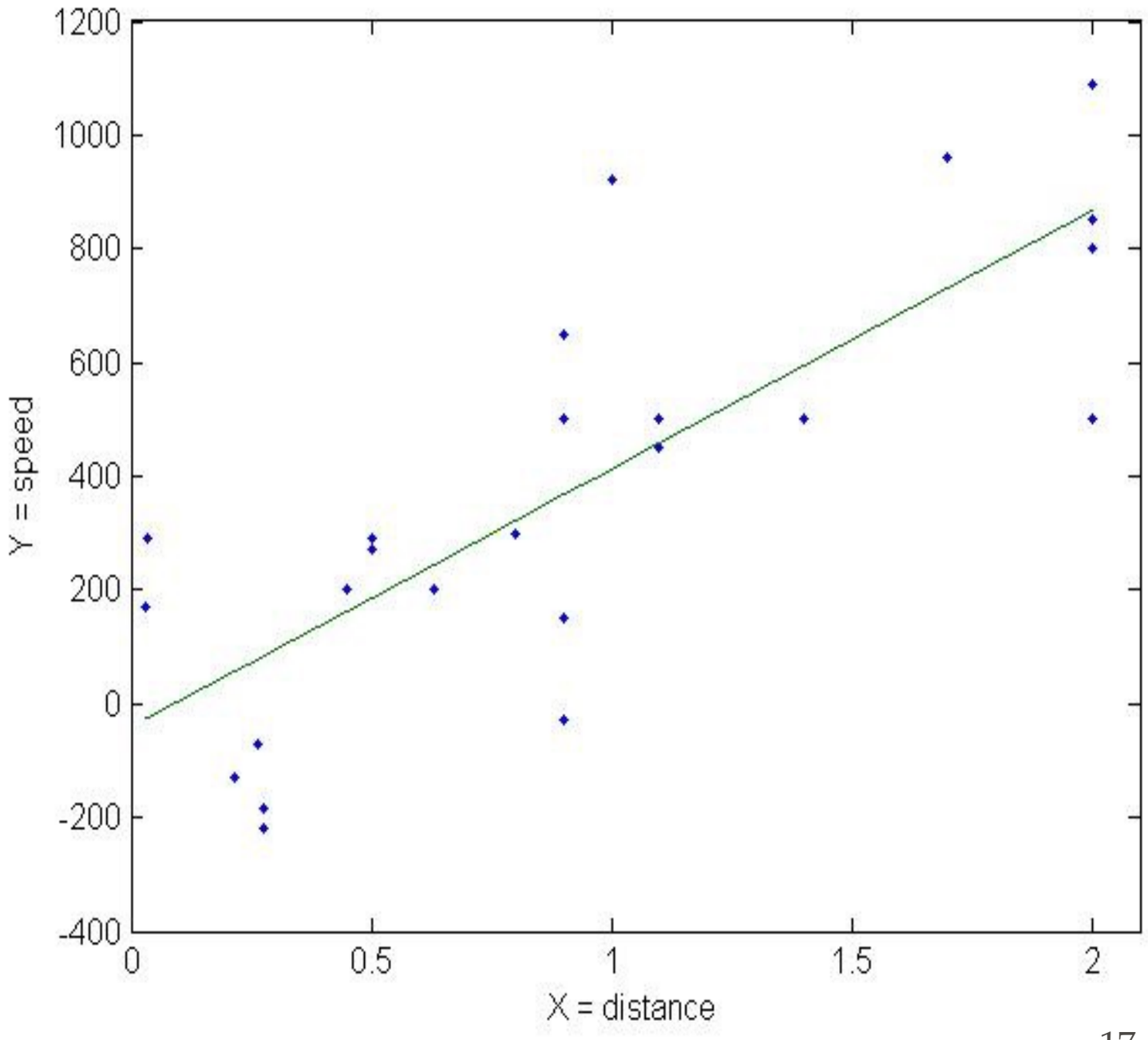
Example - Hubble's Law

- Choosing
  - x = Distance;
  - y = Speed;
- Slope estimate

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 454.1584$$

- Intercept estimate

$$\widehat{\alpha} = \bar{y} - \widehat{\beta} \cdot \bar{x} = -40.7836$$



Statistical Inference on the Regression Slope

- In general, the null hypothesis about the slope that we wish to test takes the following form

$$H_0: \beta = \beta_0$$

- It can be shown that the estimator of the slope is normally distributed with mean β and variance

$$\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- where σ^2 is the variance used in the statistical model, $y_i \sim N(\alpha + \beta x_i, \sigma^2)$.
- Using the estimated variance, s_r^2, instead of the population variance, the appropriate test statistic for $\widehat{\beta}$ is

$$t = \frac{\widehat{\beta} - \beta_0}{s_r \sqrt{1/\sum_{i=1}^{n}(x_i - \bar{x})^2}} \sim t(n-2)$$

- The p-value is

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

Example - Hubble's Law

- Let us test whether the regression slope deviates significantly from zero.
- Null hypothesis

$$H_0: \beta = 0$$

- Parameter estimates:

$$\hat{\beta} = 454.1584 \quad \text{and} \quad \hat{\alpha} = -40.7836$$

- Empirical variance/s.d.

$$s_r^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \left(\hat{\alpha} + \hat{\beta}x_i\right)\right)^2 = 54247 \text{ and } s_r = \sqrt{54247} = 232.91$$

- Test size:

$$t = \frac{\hat{\beta} - 0}{s_r\sqrt{1/\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 6.0364$$

- p-value:

$$2 \cdot \left(1 - t_{cdf}(|t|, n-2)\right) \approx 0$$

- Since p<0.05, we reject the null hypothesis that $\beta = 0$. In other words, the data suggest that the regression slope deviates significantly from zero.

Statistical Inference on the Regression Slope

- The 95% confidence interval for the slope is

$$\beta_{-} = \hat{\beta} - t_0 \cdot \frac{s_r}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \hat{\beta} - t_0 \cdot \frac{s_r}{s_x} \cdot \frac{1}{\sqrt{n-1}}$$

$$\beta_{+} = \hat{\beta} + t_0 \cdot \frac{s_r}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \hat{\beta} + t_0 \cdot \frac{s_r}{s_x} \cdot \frac{1}{\sqrt{n-1}}$$

- where

$$t_0 = tinv\left(1 - \frac{0.05}{2}, n-2\right) = tinv(0.975, n-2)$$

Example - Hubble's law

- 95% confidence interval:

$$t_0 = tinv(0.975,22) = 2.0739$$

$$\beta_- = \hat{\beta} - t_0 \cdot \frac{s_r}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 298.12$$

$$\beta_+ = \hat{\beta} + t_0 \cdot \frac{s_r}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 610.19$$

- The NULL hypothesis H_0: $\beta = 0$ is not within the 95% confidence interval, so we reject the NULL hypothesis

Statistical Inference on the Regression Intercept

- In general, the null hypothesis that we wish to test takes the following form

$$H_0: \alpha = \alpha_0$$

- It can be shown that the estimator of the intercept is normally distributed with mean α and variance

$$\sigma^2 \cdot \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$$

- where σ^2 is the variance used in the statistical model, $y_i \sim N(\alpha + \beta x_i, \sigma^2)$.
- The appropriate test statistic for $\widehat{\alpha}$ is

$$t = \frac{\widehat{\alpha} - \alpha_0}{s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}} \sim t(n-2)$$

- The p-value is

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

Example - Hubble's Law

- Let us test whether the regression intercept deviates significantly from zero.
- Null hypothesis

$$H_0\colon\ \alpha = 0$$

- Parameter estimates are the same as above:
- Test size:

$$t = \frac{\hat{\alpha} - 0}{s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} = -0.4888$$

- p-value:

$$2 \cdot \left(1 - t_{cdf}\left(|t|, n-2\right)\right) = 0.6298$$

- Since p>0.05, we fail to reject the null hypothesis that $\alpha = 0$. In other words, the data suggest that the regression intercept does not deviate significantly from zero.

Statistical Inference on the Regression Intercept

- The 95% confidence interval for the intercept α is:

$$\alpha_{-} = \hat{\alpha} - t_0 \cdot s_r \cdot \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

$$\alpha_{+} = \hat{\alpha} + t_0 \cdot s_r \cdot \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

- where

$$t_0 = tinv\left(1 - \frac{0.05}{2}, n - 2\right) = tinv(0.975, n - 2)$$

Example - Hubble's law

- 95% confidence interval:

$$\alpha_{-} = \hat{\alpha} - t_0 \cdot s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = -124.2$$

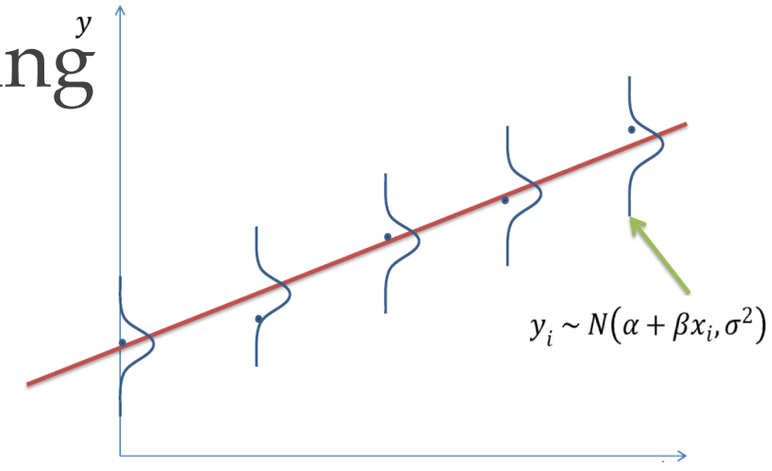
$$\alpha_{+} = \hat{\alpha} + t_0 \cdot s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 42.6$$

- The NULL hypothesis H_0: $\alpha = 0$ is within the 95% confidence interval, so we fail to reject the NULL hypothesis.

Checking for Normality

- Recalling that the statistical model underlying linear regression is

$$y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma^2)$$



- the residual of the i'th sample should be normally distributed with zero mean and variance σ^2

$$\epsilon_i = y_i - (\alpha + \beta x_i) \sim \mathcal{N}(0, \sigma^2)$$

- Hence, a good way to check whether the assumption of linearity between x and y holds is to first fit the linear model and subsequently check that the residuals ϵ_i are normally distributed using a Q-Q plot.

Checking for Normality Using Q-Q plot (Hubble's law)

- The residuals in Hubble's law example are

```
res=y-alpha-beta*x
```

- The resulting Q-Q plot

```
qqplot(res)
```

- shows that the residuals are approximately normally distributed, because the data points lie approximately on a straight line.

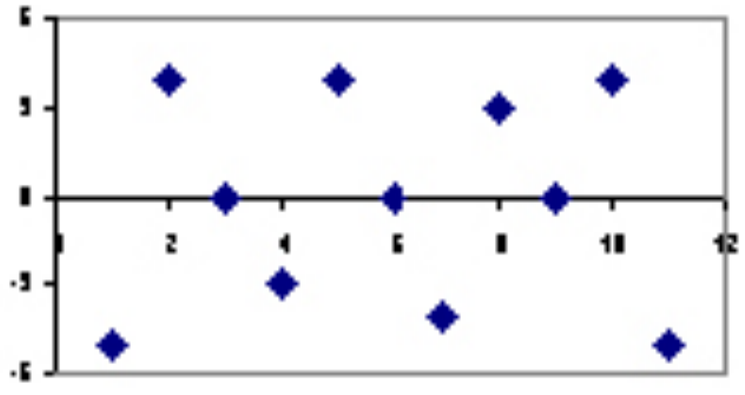
- Hence, it is safe to use simple linear regression to find the relation between the Speed and Distance of galaxies.

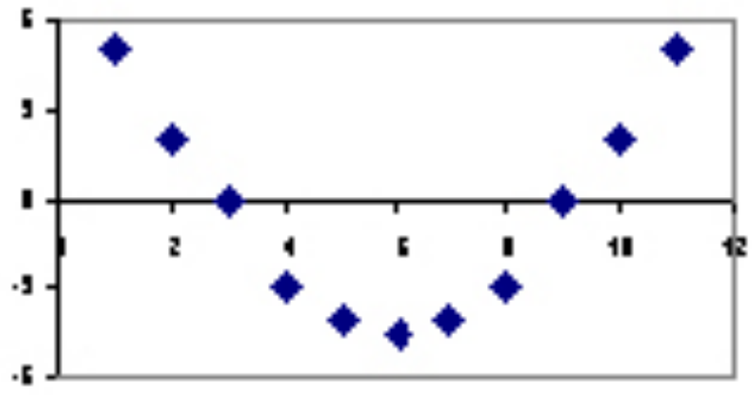
Residual Plots

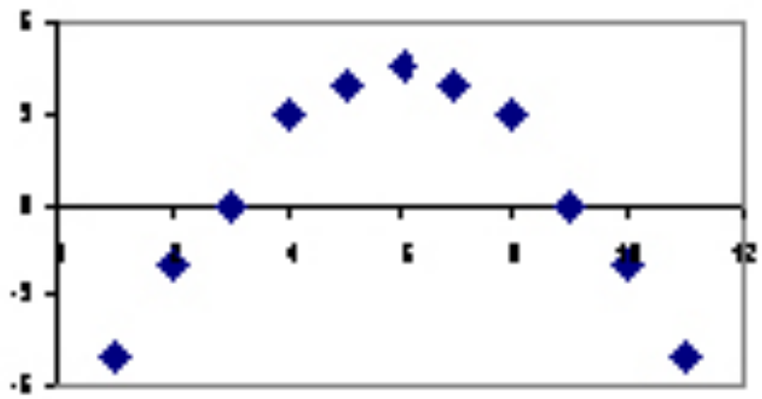
- Another way to check the normality assumption is to make a so-called *residual plot*.
- A residual plot is a graph that shows the residuals on the vertical axis and the independent variable (x) on the horizontal axis.
- If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.

Residual Plots

- Below, the residual plots show three typical patterns.
- The first plot shows a random pattern, indicating a good fit for a linear model.
- The other plot patterns are non-random (U-shaped and inverted U), suggesting a better fit for a non-linear model.

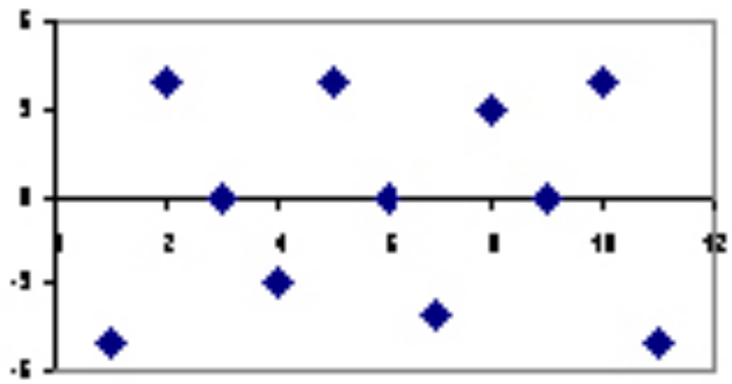


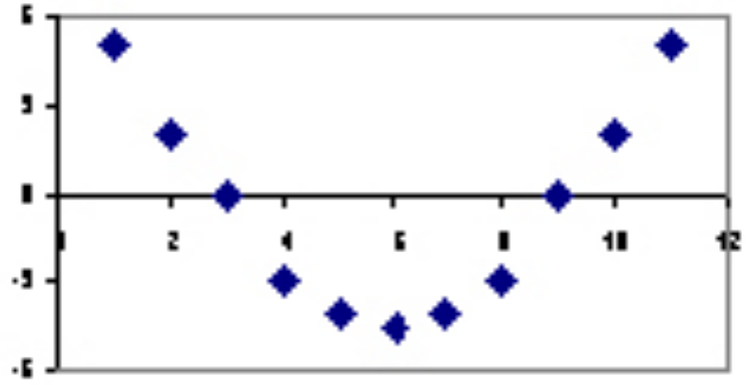


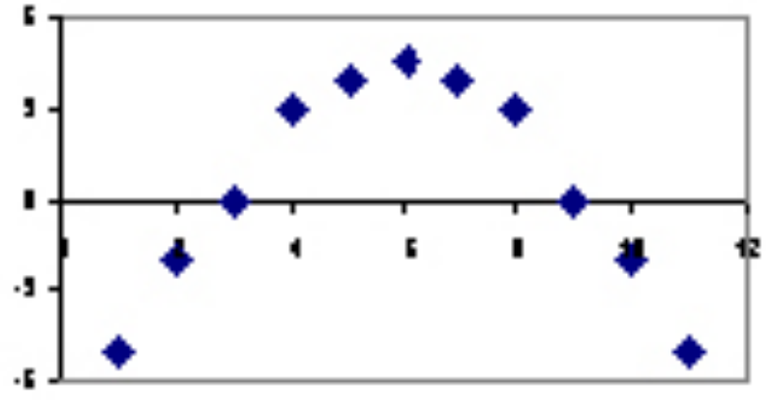


Residual Plots

- Formally, you must check the following two conditions:
  - The value of the residuals $\epsilon_i = y_i - (\alpha + \beta x_i)$ must not depend on x_i, but should lie randomly distributed around zero.
  - The variance of the residuals must not depend on x_i either.







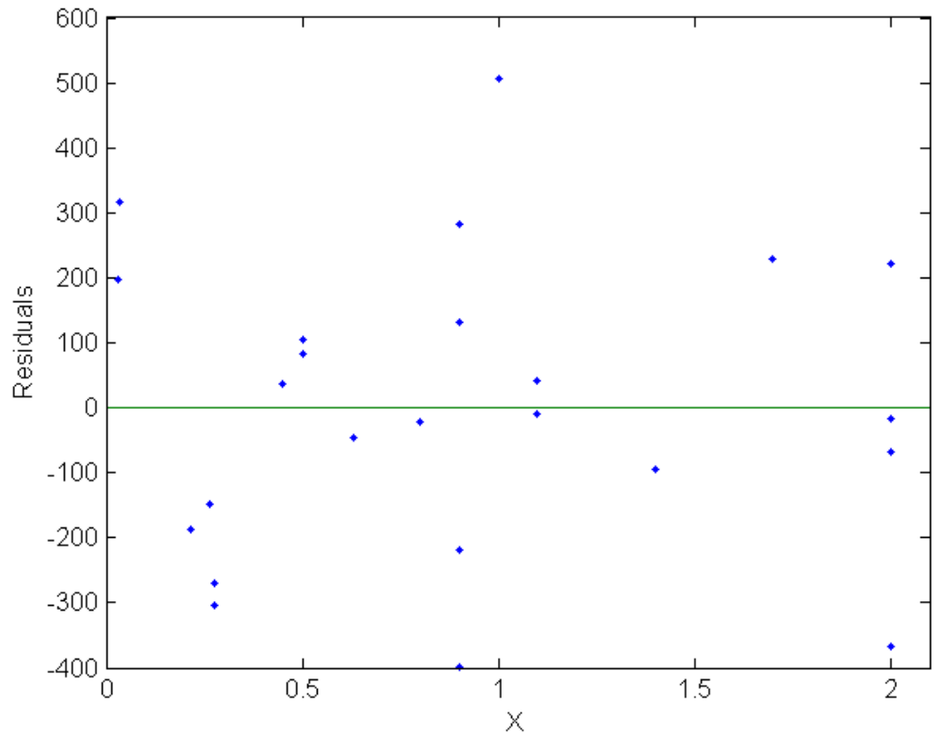
Checking for Normality Using Residual Plot (Hubble's law)

- The residuals in Hubble's law example are

```
res=y-alpha-beta*x
```

- The resulting residual plot

```
plot(x,res,'.',...
          [0 2.1],[0 0])
axis([0 2.1 -400 600])
xlabel('X')
ylabel('Residuals')
```

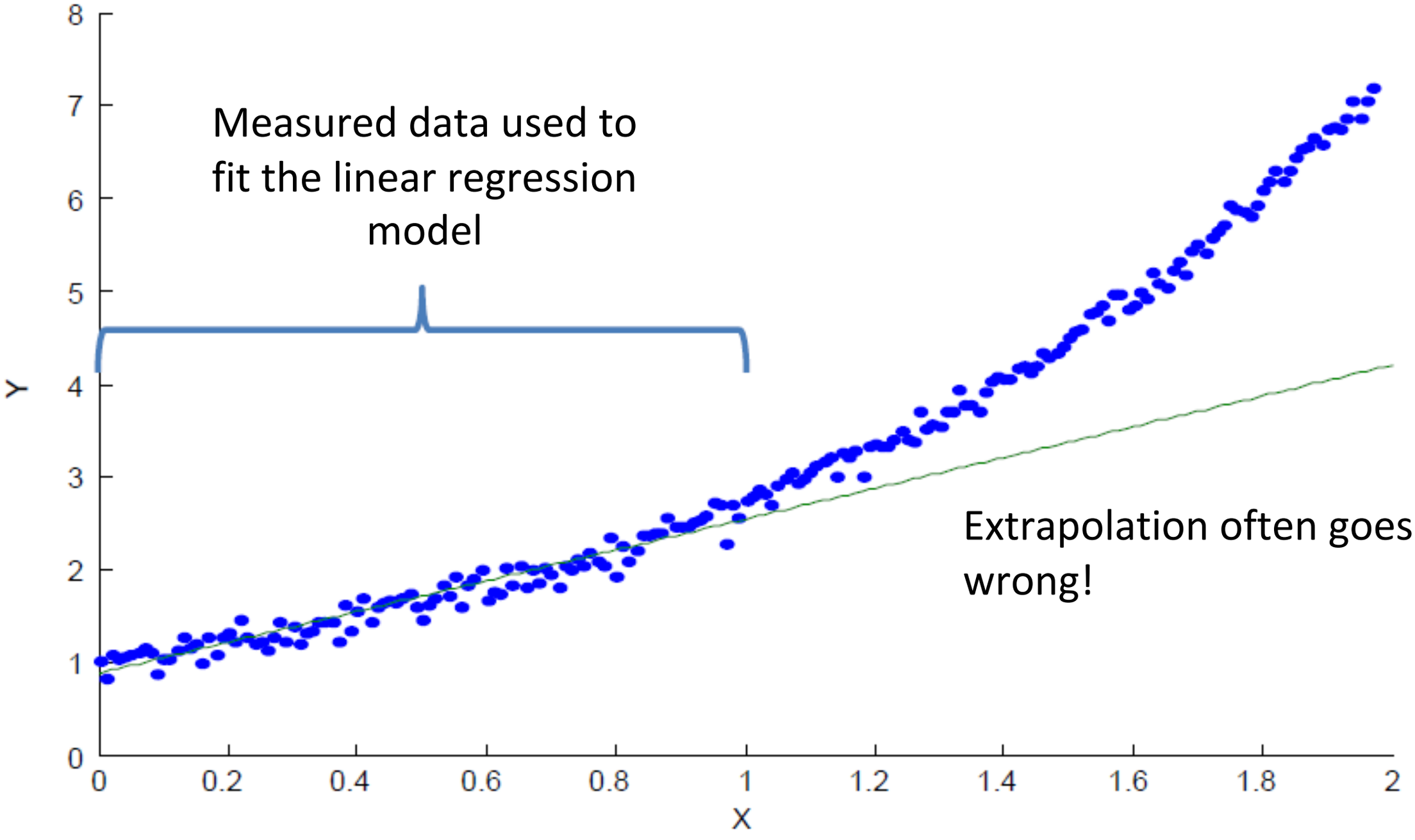


- Shows that the residuals are randomly distributed around zero and do not depend on x.
- Also, it appears that the variance of the residuals is independent of x.

Usage of Linear Regression

- Linear regression is often used for prediction.
- Suppose, for instance, that the relationship between daily energy consumption of a power plant and the outside temperature is linear.
- Then, given the temperature of tomorrow (from a weather forecast), we can give an estimate of tomorrows energy consumption of the power plant based on a linear model.
- When you use a regression equation, do not use values for the independent variable that are outside the range of values used to create the equation.
- That is called **extrapolation**, and it can produce unreasonable estimates.

Extrapolation



Sample Correlation Coefficient

- If we wish to quantify the strength of a linear relation, we can use the sample correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{{s_{xy}}^2}{s_x \cdot s_y} = \frac{Cov(x,y)}{\sqrt{Var(x) \cdot Var(y)}}$$

- where s_x and s_y are the empitical standard deviations of x and y.
- As we saw in ”Random Signals”, chap. 2.3.3, the correlaltions coefficient (ϱ) takes on values from -1 to 1.
- It can be shown that the estimate of the regression slope is lineary related to the correlation coefficient:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{{s_{xy}}^2}{{s_x}^2} = r\frac{s_x}{s_y}$$

Large β → large r → strong correlation

Coefficient of Determination

- In simple linear regression, the *coefficient of determination*

$$R^2 = r^2,$$

- indicates how well the data fit the linear model.
- The coefficient of determination ranges from 0 to 1 with value close to 1 suggesting a strong linear relationship, and values close to 0 suggesting no linear relationship.
- The coefficient of determination in the example with Hubble's law is $R^2 = 0.6235$.
- To calculate the sample correlation coefficient between *x* and *y* in Matlab, use the command `corr2(x,y)`.

Outliers

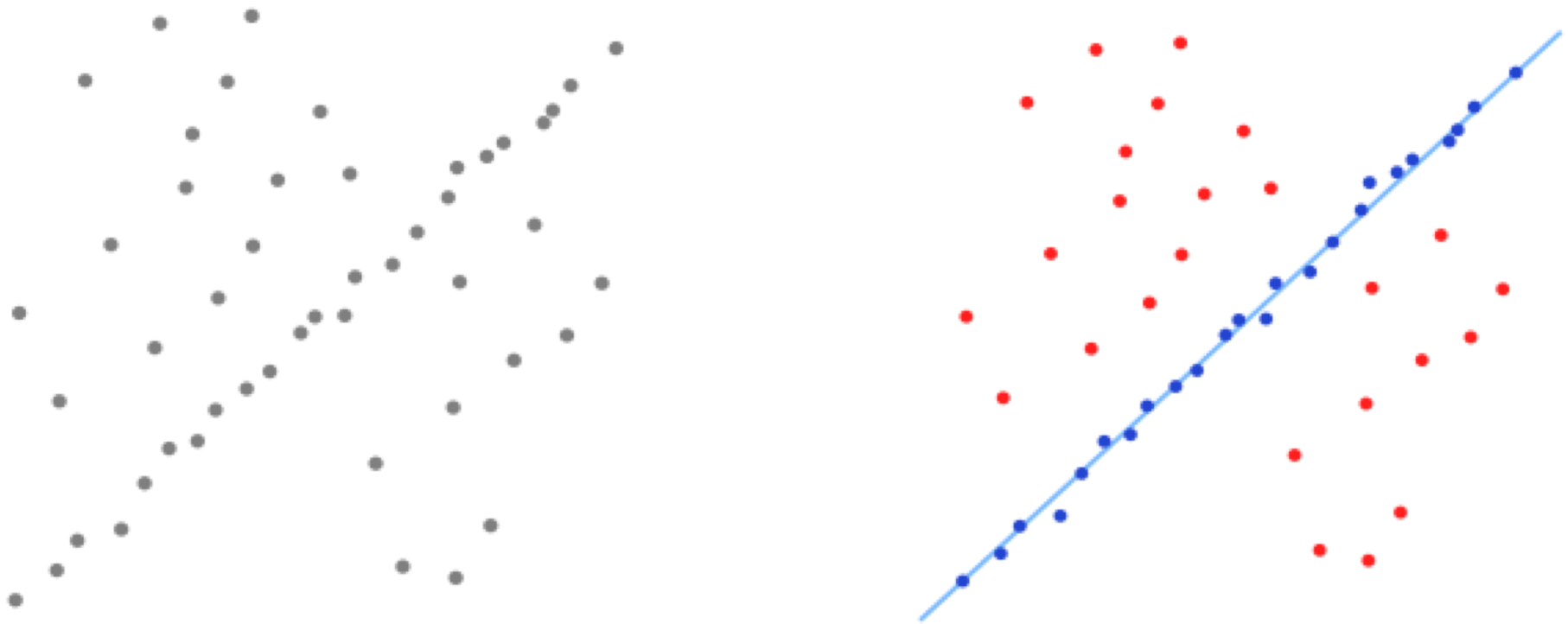
- Outliers are data points that are separated from the rest of the data and potentially influential for the regression analysis.
- Outliers can have a dramatic on the sample correlation coefficient (and therefore the slope).
- Recalling the definition of the sample correlation coefficient,

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- an outlier is a point (x_i, y_i), such that either $(x_i - \bar{x})$ or $(y_i - \bar{y})$, or both, is large.
- The extent of influence of any point can be judged in part by computing the correlation coefficient with and without that point.

RANSAC

Random Sample Consensus



RANSAC is an iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers. It is a non-deterministic algorithm in the sense that it produces a reasonable result only with a certain probability, with this probability increasing as more iterations are allowed.

RANSAC

- Step 1: Select a subset of the data.
- Step 2: Find the best fitted model to that subset.
- Step 3: Determine the dataset of the data that fits with the model (inliers).
- Step 4: Repeat set 1-3, and if the new model has more inliers than the previous one, replaced the model with the new.
- Step 5: After a number of iterations, reject all datapoints that are not inliers. This is the outliers.
- Step 6: Re-estimate the model based on all the inliers.

➢ i.e. Find the model (linear regression) that gives the largest number of inliers (smallest number of outliers)

Linear Models – When/Why?

- Can be used when the mean changes over time.
- The variance should not change over time
- The mean is connected to time linearly!

- Outliers must not be omitted from a conclusion
  - Fx. may new medication damage indiviual patients (allergy)
- Outliers can – with justification – be omitted from a linear regression

Words and Concepts to Know

Linear Regression

Random Sample Consensus

Linear Model

Slope parameter

Sample Correlation Coefficient

Regression Intercept

Intercept parameter

Extrapolation

Predicted data

Outliers

Model Fitting

Slope parameter

Residual

Empirical Variance

Residual plot

Inliers

RANSAC

Measured data

Coefficient of Determinaton