Statistical Inference

Anders Stockmarr
Course developers: Anders Stockmarr, Helle Rootzen

DTU Department of Applied Mathematics and Computer Science
Section for Statistics and Data Analysis
Technical University of Denmark
anst@dtu.dk

January 6th, 2025

DTU Compute
Department of Applied Mathematics and Computer Science

Plan for this week

- **Monday** Statistical inference, and the t-test
- **Tuesday** Simple and Multiple regression
- **Wednesday** ANOVA, ANCOVA, and linear models
- **Thursday** Categorical data, statistical report writing, logistic Regression
- **Friday** Introduction to repeated measures , Principal Component Analysis

Outline

- Introduction
- Background
- Summary Statistics
- Statistical Modelling
- Estimation
- Test
- Power

An Introductory Experiment

- How much time did you spend on social media yesterday? Guess?

Example: Low Birth Weight

```
BWTdata <- read.csv2("Data/lowbwt.txt")
head(BWTdata)
```

	ID	LOW	AGE	LWT	RACE	SMOKE	PTL	HT	UI	FTV	BWT
1	85	0	19	182	2	0	0	0	1	0	2523
2	86	0	33	155	3	0	0	0	0	3	2551
3	87	0	20	105	1	1	0	0	0	1	2557
4	88	0	21	108	1	1	0	0	1	2	2594
5	89	0	18	107	1	1	0	0	1	0	2600
6	91	0	21	124	3	0	0	0	0	0	2622

- Rows are *observations*
- Columns are *variables*

Hosmer & Lemeshow data 2000

Example: Low Birth Weight - Statistical Analysis

The variable that we want to analyse is **BWT**. When we look at data, what can we say about the underlying statistical model that gave rise to these data?

- A normal model with parameters mean μ and variance σ^2?
- A log-normal model with similar parameters?
- something else?

Once we settle upon a (reasonable) model, we can proceed to conduct *statistical Inference*

Example: Low Birth Weight - Statistical Analysis

- Estimation: With these 189 Birth weights, what can we say about the two unknown parameters μ and σ^2?
  Guesses on values, precision on guesses, possible dependencies on variables.
- Test: Does the mean μ depend on Age? Does the variance σ^2 depend on Race?
- Prediction: Given values of Age and Race, what expectation will we have to the birthweight? What will be the uncertainty of our expetation?

Statistical Inference - Study planning

When planning a study, the planned statistical inference is a key element:

- Formulate a scientific question that you wish to answer with your study
  - Is the birth weight different for smoking and non-smoking mothers?
  - If so, how big is the difference?
  - These questions should be answered through statistical inference.

Statistical Inference - Study planning

When planning a study, the planned statistical inference is a key element:

- Formulate a scientific question that you wish to answer with your study
  - Is the birth weight different for smoking and non-smoking mothers?
  - If so, how big is the difference?
  - These questions should be answered through statistical inference.
- Choosing the subjects
  - Who? Inclusion and Exclusion criteria. Availability?
  - How many?
  - these choices need to be made to enable the study to provide relevant results that may be used for inference on the scientific question.

Statistical Inference - Study planning

When planning a study, the planned statistical inference is a key element:

- Formulate a scientific question that you wish to answer with your study
  - Is the birth weight different for smoking and non-smoking mothers?
  - If so, how big is the difference?
  - These questions should be answered through statistical inference.
- Choosing the subjects
  - Who? Inclusion and Exclusion criteria. Availability?
  - How many?
  - these choices need to be made to enable the study to provide relevant results that may be used for inference on the scientific question.
- What information are we going to collect? *A suitable response variable (birth weight). Primary explanatory variables (smoking) Variables that may affect the outcome variable and/or cloud effects of the main explanatory variable (Age, gender,region etc).*
- *All of these choices are made to be able to conduct relevant statistical inference*

Background

- Probability - outcomes and sample spaces.
- Frequency
- Subjective probability
- Continuous variables

Background - Probability

A *probability* is relative to an event. It is a number between 0 and 1, indication uncertainty about if that event occurs or not.

- Probability of 0.5 - large uncertainty (toss of a coin)
- Probability close to 0 or 1 - small uncertainty (winning in lotto; surviving until tomorrow)
- *Outcomes* - elements of the *sample space*
- Coin toss: sample space $\{Heads, Tails\}$
- Birth weight in grams - sample space $(0; \infty)$

Example: Toss of a Die

Tossing a die: Sample space and outcome probability

Outcome	1	2	3	4	5	6	Landing on the Edge
Probability	1/6	1/6	1/6	1/6	1/6	1/6	0

Altering tools of the experiment: Put lead opposite the six eyes, a common trick for sharpers:

Outcome	1	2	3	4	5	6	Landing on the Edge
Probability	2/15	2/15	2/15	2/15	2/15	1/3	0

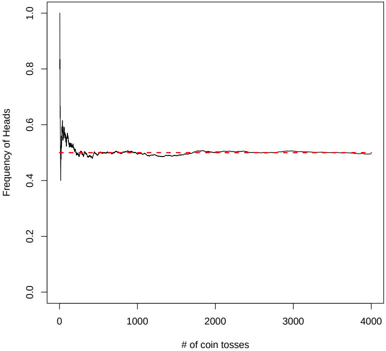
Altering circumstances of the experiment: Throw on a garden table:

Outcome	1	2	3	4	5	6	landing on the Edge
Probability	1/9	1/9	1/9	1/9	1/9	1/9	1/3

Frequency

- The frequency is the relative proportion of an event.
- In contrast to a probability, a frequency contains an element of randomness.
- The **Law of Large Numbers**: If an experiment is repeated many times without changing the circumstances, the frequency of an event will converge to the probability.
- The **Frequency Interpretation** of probabilities: Limit values of frequences

Example: Frequency of Heads in Coin Tosses



- Frequency converges to 0.5.

Subjective probabilities

- Reflect persons (subjects) individual assessment of probabilities from the own impression of circumstances and effects.
- Should NOT be confused with true probability.
- Example: Gut feeling, rule of thumb.
- Example: I think that the chance that FC Copenhagen will win the Danish soccer league is 80%, as they have been playing really well lately (when writing slides, FC Copenhagen was 1^{st} in the Danish soccer league).

Subjective probabilities

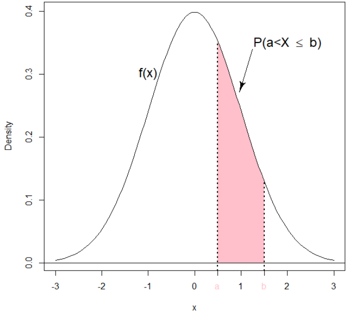
- Reflect persons (subjects) individual assessment of probabilities from the own impression of circumstances and effects.
- Should NOT be confused with true probability.
- Example: Gut feeling, rule of thumb.
- Example: I think that the chance that FC Copenhagen will win the Danish soccer league is 80%, as they have been playing really well lately (when writing slides, FC Copenhagen was 1^{st} in the Danish soccer league).
- Typically, subjective probabilities contain an element of *subject bias*.

Probabilities and Continuous Variables

- What does a probability mean when we are talking about a continuous variable X, with no upper limit to the number of possible values, like $X =$birth weight?
- The sample space is $(0; \infty)$. Each possible outcome, say 3545 g, has probability 0 of occurring.
- In this case we represent the probability distribution by a **density function**, and discuss the probability of an interval;

$$P(a < X \leq b)$$

Probabilities and Continuous Variables

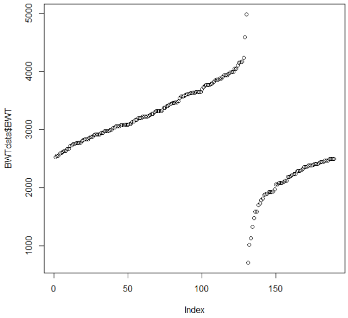


- Note that $\lim_{\varepsilon \to 0} P(X \in [x-\varepsilon; x+\varepsilon])/2\varepsilon = f(x)$
- Thus the density function is a limit of (normalized) probabilities.

Summary Statistics

- the first thing that we want to do is to plot the data:

```
plot(BWTdta$BWT)
```



Data are ordered according to low birthweigth (< 2500 g), and birthweight.

Summary Statistics - Location

After having plotted our data, we may want to get a further overview by calculating some simple statistics. Assume that we have observations $y_1, \ldots, y_n$ of a continuous variable Y.

The **location**, or **centre** of our data:

- Empirical mean: $\overline{y} = \frac{1}{n}(y_1 + \ldots + y_n)$
- Median: The middle observation when data are sorted and n is odd. The average of the two central observations when n is even.

Summary Statistics - Measures of Variation

The empirical variance (or standard deviation) is a measure of how much the observations are spread out

- Empirical Variance s^2:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

- Standard deviation (SD) s

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

- Percentiles: the median is the 50% percentile.

Summary Statistics - Measures of Variation

Percentile (and quartiles): Sort the data from smallest to largest:

2.5% percentile: First observation for which at least 2.5% of the observations are smaller or equal to, and at most 97.5% are larger.
25% percentile: First observation for which at least 25% of the observations are smaller or equal to, and at most 75% are larger.
50% percentile: First observation for which at least 50% of the observations are smaller or equal to, and at most 50% are larger.
75% percentile: First observation for which at least 75% of the observations are smaller or equal to, and at most 25% are larger.
97.5% percentile: First observation for which at least 97.5% of the observations are smaller or equal to, and at most 2.5% are larger.

Summary Statistics - Measures of Variation

Empirical percentiles (and quartiles): Sort the data from smallest to largest: *p*-percentile q_p: First observation for which at least the fraction p of the observations are smaller, and at most the fraction $1 - p$ are larger

- The quartiles are $q_{0.25}, q_{0.5}, q_{0.75}$. $q_{0.5}$ is usually equated with the median.
- Inter quartile range **IQR**: $q_{0.75} - q_{0.25}$. This is the size of the box in a box plot.
- $q_{0.025}$ and $q_{0.975}$ spans an interval where **95% of the observations** lie within.

Summary Statistics in R

```
mean(BWTdata$BWT)
```

```
[1] 2944.656
```

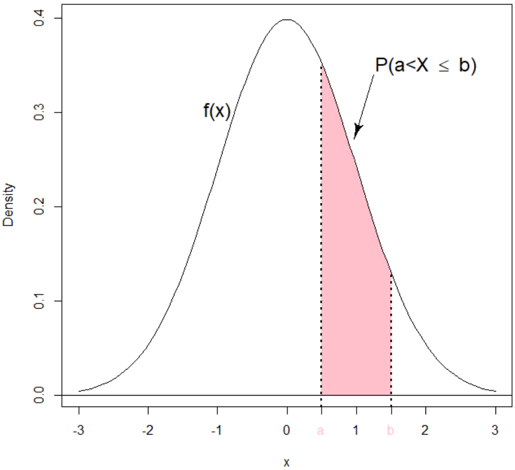
```
summary(BWTdata$BWT)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    709    2414    2977    2945    3475    4990
```

```
quantile(BWTdata$BWT,probs=seq(0,1,by=0.1),type=2)
```

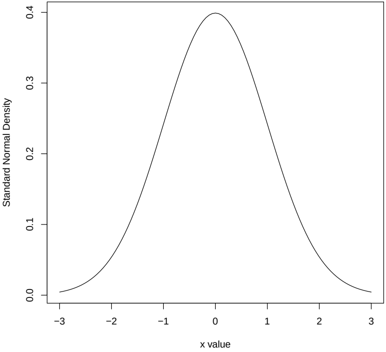
```
  0%  10%  20%  30%  40%  50%  60%  70%  80%  90% 100%
 709 1970 2325 2495 2778 2977 3175 3374 3629 3884 4990
```

Statistical Models from Probability Densities



The Standard Normal Density

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, with mean$= 0$ and SD$= 1$.



The Importance of the Normal Distribution

- **The Central Limit Theorem** : Averages of a large numbers of observations are approximately normally distributed, irrespectively of the (common) distribution that you start out with.

The Importance of the Normal Distribution

- **The Central Limit Theorem** : Averages of a large numbers of observations are approximately normally distributed, irrespectively of the (common) distribution that you start out with.
- Because of this, it turns out that the Normal distribution is often a good approximation to real life distributions (perhaps after a transformation with the log, the square root or other...).

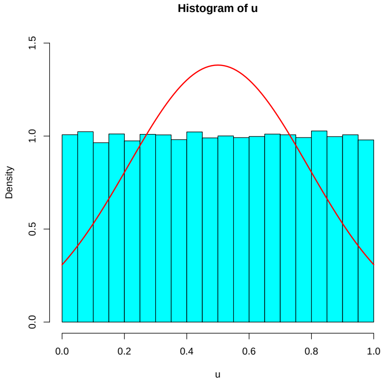
The Importance of the Normal Distribution

- **The Central Limit Theorem** : Averages of a large numbers of observations are approximately normally distributed, irrespectively of the (common) distribution that you start out with.
- Because of this, it turns out that the Normal distribution is often a good approximation to real life distributions (perhaps after a transformation with the log, the square root or other...).
- The structure of the Normal distribution is mathematically tractable, and software has been developed for a lot of situations.

A Non-Normal Distribution

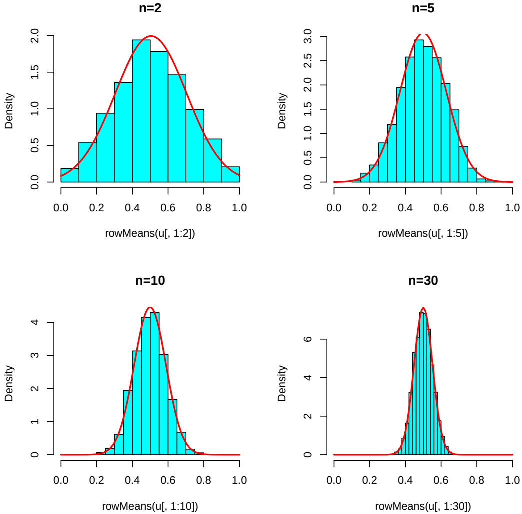
A non-normal distribution could be the uniform distribution between 0 and 1; $f(x) = 1, 0 < x < 1$. Histogram of u:

```
u<-runif(2500)
hist(u, col="cyan",probability=T,breaks=20,ylim=c(0,1.5))
curve(dnorm(x,mean=mean(u),sd=sd(u)),0,1,add=T,col="red",lwd=2)
```



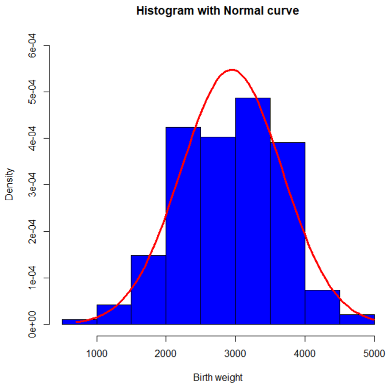
A Non-Normal Distribution

Histograms for means of uniformly distributed variables:



Normality Check

With normality, a data histogram should resemble normality. Birth weight data:

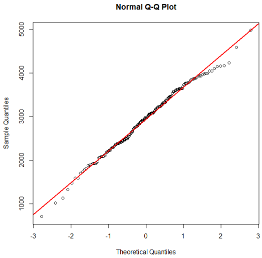


May be difficult to detect deviations from normality from a histogram.

Normality Check

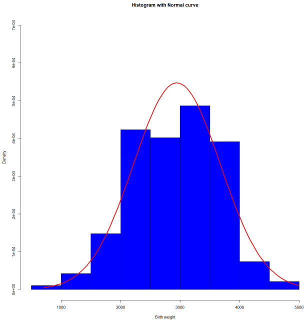
- Much better to use a *quantile-quantile plot* (qq-plot). Here, observed percentiles are plotted against percentiles from the normal distribution.
- If the empirical distribution is normal, a more or less straight line.
- If the data are non-normal, som deviation from a straight line should occur.

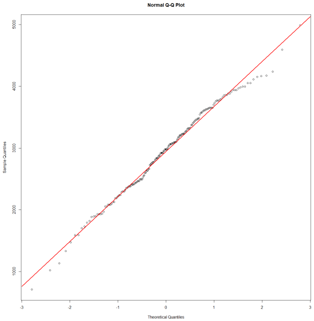
```
qqnorm(BWTdata$BWT)
lines((-3):3,((-3):3)*sd(BWTdata$BWT)+mean(BWTdata$BWT),
  type="l",col="red",lwd=2)
```



Statistical Model for the Birthweight data

Let $Y_1, \ldots, Y_{189}$ be the 189 registered birth weights in our dataset. We will assume that the 189 variables are **independent and normally distributed** with a common mean μ and common variance σ^2.





Statistical Models in General

In general, a statistical model consists of

- A family of distributions. In our example the 1-dimensional normal distributions (but there are others);
- the parameters that typically parametrizes the family of distributions. Eg. our example: The mean and variance in a normal distribution;

$$(\mu, \sigma^2) \in \mathbf{R} \times (0; \infty)$$

  or the probability of 'Heads' in a coin toss: $p \in [0; 1]$.

Parameter interpretation: Effects/associations; e.g. the decrease in mpg per lbs/1000 weight.

Exercise 1

- Set your working directory to where you keep your data for today.
- Load the cars dataset mtcars.txt.
- Describe the data
- Make plots of the variable miles per gallon, ”mpg”.
- Calculate summary statistics for mpg.
- Where would we expect most of the observations to be found?
- Calculate IQR and 0.025, 0.975 percentiles.

Exercise 2

- Load the cars data set mtcars.txt.
- can we assume that miles per gallon are norrmally distributed?
- The variable “am” is 0 for cars with automatic transmission and 1 for cars with manual. Make a boxplot for the two levels of am.
- can we assume that miles per gallon are normally distributed for each level of am?

Estimation

- Based on observations $X_1, \dots, X_n$, independent and each having a density function $f_\theta(x)$, we want to choose the parameters that fits our data the best.
- The density function measures the limiting probability of the data; The simultaneous (limiting) data probability is given by

$$f_\theta(X_1, \dots, X_n) := \prod_{i=1}^{n} f_\theta(X_i)$$

- Let's view this as a function of the parameter θ, rather than the data X:

$$L_X(\theta) := f_\theta(X_1, \dots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$$

We choose the parameter that **maximizes the probability of the data**; ie. we maximize $L_X(\theta)$ (or $\ell_X(\theta) = log(L_X(\theta))$).

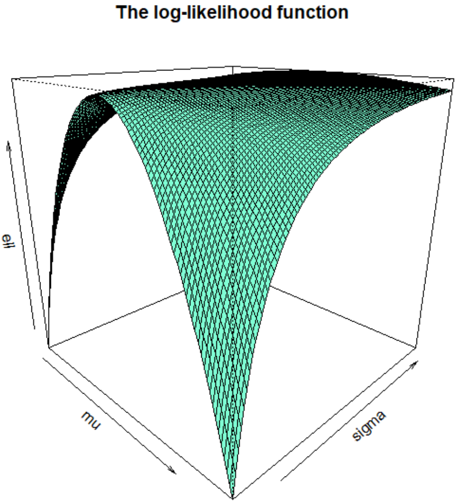
Estimation in the Birth Weights Example

This is also called the one-sample problem (will be treated in detail tomorrow). 189 samples assumed from the same normal distribution:

$$Y_i \sim N\left(\mu, \sigma^2\right)$$

$$\ell(\mu, \sigma^2) = -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Y - \mu\|^2$$

Birth Weights Example: The Log Likelihood Function



Estimation in the Birth Weights Example

$$\begin{aligned}\ell(\mu,\sigma^2) = & -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Y-\mu\|^2 \\ = & -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(n\overline{Y^2} - 2n\overline{Y}\mu + n\mu^2)\end{aligned}$$

For fixed σ^2 this is a 2nd order polynomial in μ - easy to maximize!

- maximization point:

$$(\hat{\mu}, \hat{\sigma}^2) = (\overline{Y}, Var(Y))$$

- $Var(Y) = \frac{1}{n}\sum(Y_i - \overline{Y})^2$ is downwards biased as an estimator of σ^2; the mean is not σ^2. We use the slightly bigger estimator $s^2 = \frac{n}{n-1}Var(Y)$ defined on slide 19, to arrive at an unbiased estimate. Thus, theoretical mean and variance is estimated by empirical mean and variance.
- values from R:

```
>Y<-BWTdata$BWT; mean(Y);var(Y)
[1] 2944.656
[1] 531473.7
```

Uncertainty of an Estimate

- We have $\hat{\mu} = 2944.7g$; this is our best guess on the parameter, but (likely) not the true value. How uncertain is our estimate?
- Just as Y_i has a distribution, so does $\hat{\mu}$, as it is nothing but a constructed random variable.
- We can use the standard deviation of this distribution to constuct measures of uncertainty. We call this the standard error of the estimate (or the Standard Error of the Mean (SEM)).

The Standard Error of the Mean (SEM):

The Standard Error of the Mean is calculated as

$$SEM = SD\left(\overline{Y}\right) = \frac{SD(Y_1)}{\sqrt{n}}$$

SEM gets smaller as $\sqrt{n}$ when n increases.

```
> SEM<-sd(BWTdata$BWT)/sqrt(length(BWTdata$BWT))
> SEM
[1] 53.02858
```

Confidence Intervals

The interval

$$[\hat{\mu} - 1.97SEM; \hat{\mu} + 1.97SEM]$$

is a stochastic interval that (in this case) has a 95% probability of containing the true value μ.

More on this later.

Statistical Tests

Could we use a simpler model? Could one or more parameters in the chosen model be of a known value (often 0)? We want to attempt to simplify our model, through a statistical test of hypotheses, where we decide if a given hypothesis is supported by the data at hand.
Examples of hypotheses:

- Is the mean birth weight 3000g ($\mu = 3000$)?
- Is the mean birth weight the same for smokers and non-smokers ($\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$)?
- Are the miles per gallon independent of the weight of the car (slope $\beta = 0$)?

Tests Statistics

- We test a hypothesis through a **test statistic**. A test statistic measures the discrepancy between a hypothesis and an already accepted model (say, a normality assumption).
- An extreme test statistic says that the hypothesis fits the data badly.
- We will be studying whether an observed test statistic is more extreme than what could be expected by chance, assuming that the hypothesis is correct.

A Test Statistic in the Birth Weight Example

We would like to test
H_0: The mean of a birth weight is 3000g.

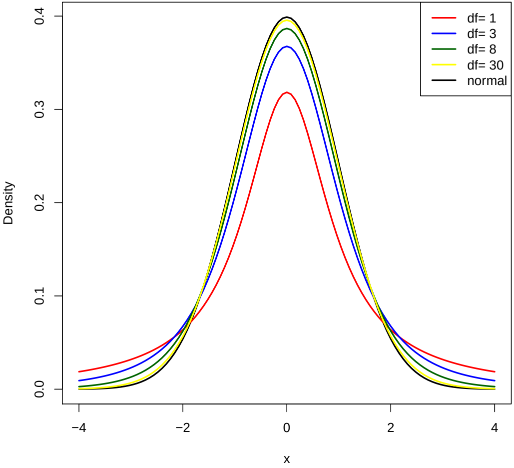
An obvious idea would be to assess whether our mean estimate, $\hat{\mu} = \overline{Y}$ is close to 3000, ie. whether $\overline{Y} - 3000$ is close to 0, relative to its uncertainty. Under H_0 we have:

$$\overline{Y} \sim N\left(3000, \frac{\sigma^2}{n}\right), \quad ie. \frac{\overline{Y} - 3000}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

We do not know σ^2, but we have an estimate $s^2 = 531473.7$. Replacing σ^2 with its estimate s^2 increases the uncertainty:

$$T = \frac{\overline{Y} - 3000}{\sqrt{\frac{s^2}{n}}} = \frac{\overline{Y} - 3000}{SEM} \sim t_{n-1}$$

Densities from t-distributions



A Test Statistic in the Birth Weight Example

We now have the test statistic

$$t = \frac{2944.7 - 3000}{53.03} = -1.0437$$

Is the value -1.0437 extreme in a t-distribution with 188 degrees of freedom? We should calculate the probability

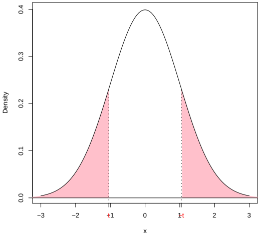
$$P\left(|t| > 1.0437\right)$$

This is straight forward in R:

```
> 2*pt(-1.0437,188)
[1] 0.2979644
```

p value

- The **p-value** is the probability of observing what we have seen, or something worse (more extreme) if H_0 is true.
- If what we have seen is very unlikely ($p<0.05$), We will not put faith in the hypothesis and **reject** H_0.
- We will not reject H_0 for the Birth weight data ($p=0.30$). *Statistical Inference*: The data *conforms* with a mean of 3000g at the 5% test level. But that doesn't mean that 3000g is the correct mean, only that the data can't support a rejection of it.



Significance Level

- If the p-value is below 0.05, we say that the test is *significant at the 5% level*, and we reject the hypothesis H_0
- If the p-value is above 0.05, we say that the test is *insignificant at the 5% level*, and we accept the hypothesis H_0.
- The choice of the threshold 5%, the *significance level*, is a standard, but not a convention that is applicable in all settings. Both the p-value and the significance level should normally be reported.

Confidence Interval

The interval reported earlier:

$$[\hat{\mu} - 1.97SEM; \hat{\mu} + 1.97SEM] \approx [2840.05; 3049.264]$$

is a **standard confidence interval**: $\hat{\mu} \pm qSEM$ where q is the 0.975 percentile in the t-distrubtion with 188 degrees of freedom. For high degrees of freedom, q will converge to 1.96. The 1.97 in the formula is rounded; the interval to the right is for the exact q value.

- Out best guess on μ is thus 2944.66g, but we cannot reject that it may be between 2840.05 and 3049.264.
- For normally distributed data, the probability that this stochastic interval contains the true parameter μ is 0.95 if the model is correct.
- For non-normal data, the probability is approximately 95% (The Central Limit Theorem), barring wierd cases where the CLT do not apply (Cauchy distributed data etc.).

A Test Statistic in the Birth Weight Example

The whole process in one go:

```
> t.test(BWTdata$BWT,mu=3000)

        One Sample t-test

data:  BWTdata$BWT
t = -1.0437, df = 188, p-value = 0.298
alternative hypothesis: true mean is not equal to 3000
95 percent confidence interval:
 2840.049 3049.264
sample estimates:
mean of x
 2944.656
```

'Exact' Confidence Intervals

- So-called Exact confidence intervals consists of the parameter values that a given test will accept as parameter value.
- A different but not necessarily more precise concept. Very handy in some situations (binomal distributions, more on this on Friday).
- The drawback is that the concept is dependent on a specific test.
- However, for the normal distribution, a standard CI and an 'exact' CI from the t-test **coincides**:

```
t.test(BWTdata$BWT,mu=2840.05)
t.test(BWTdata$BWT,mu=3049.264)
```

  both give borderline significant p-values EXACTLY equal to 0.05.

Exercise 3

- Load the mtcars data.
- Calculate the mean and SD for miles per gallon (mpg).
- Calculate a 95% CI for mean mpg.
- In which interval will you expect the true mean to be found?

Exercise 4

- Load the mtcars data.
- Test the hypothesis that the mean mpg is 22.
- Which values would be acceptable at a 1% test level?

Type 1 and Type 2 Error

When we are testing hypotheses, we can make (one of) two different types of errors:

Type I: Reject H_0 when H_0 is true.
Type II: Fail to reject H_0 when the alternative H_1 is true.

Standard notation:

$$\begin{aligned} P(\text{Type I error}) &= \alpha \\ P(\text{Type II error}) &= \beta \end{aligned}$$

Type 1 and Type 2 Errors

	Reject H_0	Fail to reject H_0
H_0 is true	Type 1 error (α)	Correct acceptance of H_0 ($1-\alpha$)
H_0 is false	Correct rejection of H_0 ($1-\beta$)	Type II error (β)

$1-\beta$ is called the power, and is the probability of rejecting a false hypothesis.

Difficulty: H_0 **can be wrong in many ways!**

Usually one looks at different possible scenarios (what if μ really was $4000g$, with what probability could I detect that? How about $3500g$ etc.).

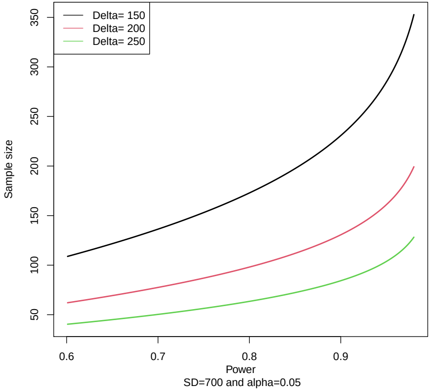
Planning a Study: The Power

Suppose that you have decided on a test statistic. One need to assign values to four of the following five quantities to be able to calculate the last:

- The sample size n.
- The significance level α of the test.
- A change in mean that you would want to detect $(\mu_0 - \mu_1)$.
- The population standard deviation σ.
- The power $1 - \beta$.

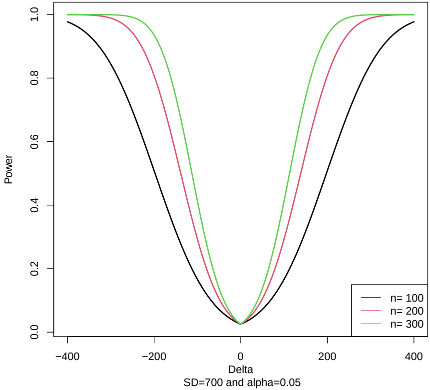
The Power Function

The sample size depends on the power, and the difference one wants to detect:



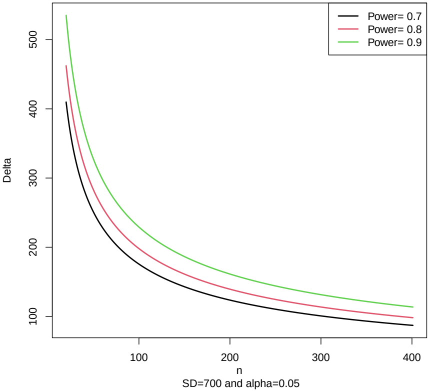
The Power Function

The power depends on the sample size, and the difference one wants to detect:



The Power Function

The difference that you can expect to detect depends on the sample size, and the power for your detection:



Planning a Study: The Power

Important note:

- The power is calculated before the study is carried out, in the planning phase. The main reason is to find the study sample size (should it be 10, 100 or 10000....)
- Once the study has been completed, report confidence intervals.
- The power calculations presented here are for the simple t-test. In general, power calculations follow the same principles, but may be **much** more complicated. Powers, sample sizes and minimum differences may be best found through simulations of statistical models.

Power Calculations for the t-test in R

If you wish to plan a study with a power of 0.8, $\alpha = 0.05$ to detect a difference of 250, where you expect the $sd = 750$, then you will need $n = 73$ subjects:

```
>power.t.test(power=0.8,delta=250,sd=750 , type=''one.sample'')

     One-sample t test power calculation

              n = 72.58407
          delta = 250
             sd = 750
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

Power Calculations for the t-test in R

If you wish to plan a study with 150 subjects, $\alpha = 0.05$ to detect a difference of 100, where you expect the $sd = 750$, then the power will be 37%:

```
>power.t.test(n=150,delta=100,sd=750 , type=''one.sample'')
```

```
     One-sample t test power calculation

              n = 150
          delta = 100
             sd = 750
      sig.level = 0.05
          power = 0.3678721
    alternative = two.sided
```

Power Calculations for the t-test in R

If you wish to plan a study with 150 subjects and a power of 0.8, $\alpha = 0.05$, where you expect the $sd = 750$, then the minimum difference that you can detect with such power will be $delta = 173$:

```
>power.t.test(n=150, power=0.8, sd=750 , type=''one.sample'')

     One-sample t test power calculation

              n = 150
          delta = 172.677
             sd = 750
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

How to Set Up the Analysis

- Explore the form of the data: Make plots of the following types:
  - Histogram
  - Box plots
  - Scatter plots
- Find preliminary values of centes and deviations etc: Descriptive Statistics like
  - Tables
  - Summary Statistics
- Then: Well prepared, proceed to **Analyses** (main focus for the rest of the course):
  - Select model
  - Estimation
  - Test

Steps in a Statistical Analysis

- Estimation: Which parameter values fit the observations best? How certain are we of our estimates?
- Model check : Are the assumptions on the underlying model fulfilled? Logically this should come first, but for practical reasons it comes after estimation.
- Simplifying the model (test): Is there a more simple model that fits the data nearly as well?

In practice, one can move back and forth between the first two steps a number of times, before a satisfying model is found.

Exercise 5

In a one-sample setting with $\alpha = 0.05$:

- Calculate the sample size to get a power of 80% when trying to detect a difference of 2, when SD=6 is expected.
- Calculate the power in a study planned to include 40 subjects, if we want to detect a difference of 3 and expect SD=6.
- What difference can we detect in a study with power of 80% , 45 subjects and SD=4?