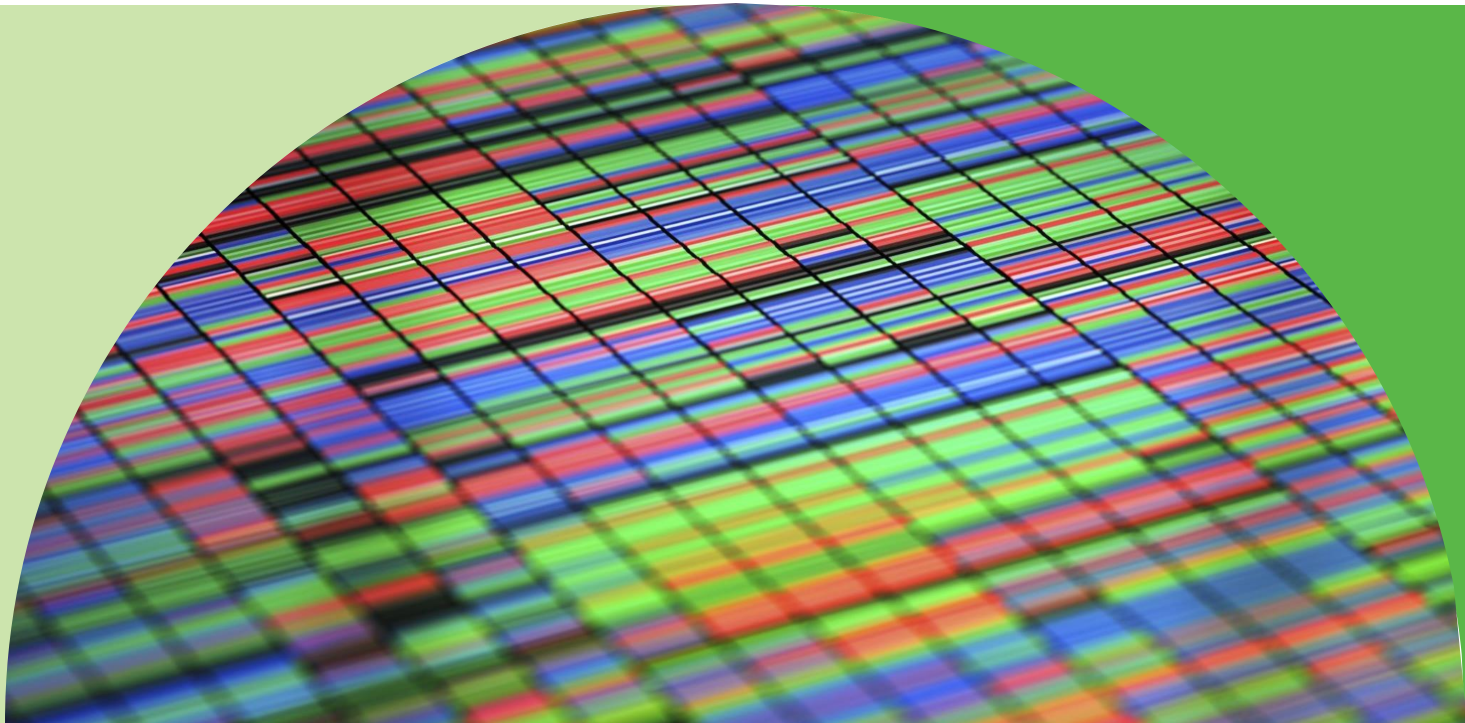
Bayesian Statistics

BRSM



Intuitive probability statements

- I'm carrying an umbrella, do you think it will rain?
- Data (d) = I'm carrying an umbrella.
- Hypotheses (h)
- 1) It rains today 2) It does not rain today

What are your prior beliefs about rain (hypotheses)?

- Say from historical records, we know that the chance of rain in April in Hyderabad is low (15%).

Hypothesis	Degree of Belief
Rainy day	0.15
Dry day	0.85

Likelihood: Theories about the data (me carrying an umbrella)

- Assumption: I'm not an idiot who randomly carries umbrellas but I can be very forgetful.. So I forget to carry an umbrella on about 70% of rainy days, and might carry an umbrella on 5% of dry days:

Likelihood = P(d|h)

	Data	
Hypothesis	Umbrella	No umbrella
Rainy day	0.30	0.70
Dry day	0.05	0.95

A reminder about basic probability rules

English	Notation		Formula
not A	$P(\neg A)$	$=$	$1 - P(A)$
A or B	$P(A \cup B)$	$=$	$P(A) + P(B) - P(A \cap B)$
A and B	$P(A \cap B)$	$=$	$P(A\|B)P(B)$

Joint probability of the hypothesis h and the data d
Probability that it is a rainy day AND I'm carrying an umbrella

$$P(d,h) = P(d|h)P(h)$$

$$\begin{aligned} P(\text{rainy}, \text{umbrella}) &= P(\text{umbrella}|\text{rainy}) \times P(\text{rainy}) \\ &= 0.30 \times 0.15 \\ &= 0.045 \end{aligned}$$

Repeat the exercise for all possibilities

	Umbrella	No-umbrella
Rainy	0.045	0.105
Dry	0.0425	0.8075

Repeat the exercise for all possibilities

	Umbrella	No-umbrella	Total
Rainy	0.0450	0.1050	0.15
Dry	0.0425	0.8075	0.85
Total	0.0875	0.9125	1

	Umbrella	No-umbrella	
Rainy	$P(\text{Umbrella, Rainy})$	$P(\text{No-umbrella, Rainy})$	$P(\text{Rainy})$
Dry	$P(\text{Umbrella, Dry})$	$P(\text{No-umbrella, Dry})$	$P(\text{Dry})$
	$P(\text{Umbrella})$	$P(\text{No-umbrella})$	

Repeat the exercise for all possibilities

	Umbrella	No-umbrella	Total
Rainy	0.0450	0.1050	0.15
Dry	0.0425	0.8075	0.85
Total	0.0875	0.9125	1

	Umbrella	No-umbrella	
Rainy	P(Umbrella, Rainy)	P(No-umbrella, Rainy)	P(Rainy)
Dry	P(Umbrella, Dry)	P(No-umbrella, Dry)	P(Dry)
	P(Umbrella)	P(No-umbrella)	

	d_1	d_2	
h_1	$P(h_1, d_1)$	$P(h_1, d_2)$	$P(h_1)$
h_2	$P(h_2, d_1)$	$P(h_2, d_2)$	$P(h_2)$
	$P(d_1)$	$P(d_2)$	

So far: we have calculated our prior beliefs before any data was given, in terms of many joint probabilities!

- We know how how confident we are about each of the different possibilities before we observed any data..

Now we are given data about the umbrella

	Umbrella	No-umbrella
Rainy		0
Dry		0
Total	1	0

Posterior Probability

Prior Joint Probabilities before observing the data

	Umbrella	No-umbrella	Total
Rainy	0.0450	0.1050	0.15
Dry	0.0425	0.8075	0.85
Total	0.0875	0.9125	1

Posterior Probability after observing the data (that I'm carrying an umbrella)

	Umbrella	No-umbrella
Rainy	0.514	0
Dry	0.486	0
Total	1	0

P(rain and umbrella) / P(umbrella)
= 0.045/0.0875 = 0.514

Posterior Joint probability

$$P(h|d) = P(d,h)/P(d)$$

Marginal probability

Arriving at Bayes' Rule

$P(d,h) = P(d|h) \times P(h)$ from our earlier basic probability rules

$P(h|d) = P(d,h)/P(d) = P(d|h) \times P(h)/P(d)$

So, to update your beliefs given data, you have to go from a prior probability via a likelihood function to a posterior probability

Posterior ~ Likelihood x Prior

$P(h|d) = P(d|h) \times P(h)/P(d)$

Now that you know Bayes' rule, you can do Bayesian Hypothesis Tests!

$$P(h_0|d) = \frac{P(d|h_0)P(h_0)}{P(d)}$$

$$P(h_1|d) = \frac{P(d|h_1)P(h_1)}{P(d)}$$

The Bayes Factor

- Posterior Odds = ratio of posterior probabilities for the hypotheses

$$\frac{P(h_1|d)}{P(h_0|d)} = \frac{0.75}{0.25} = 3$$

$$\frac{P(h_1|d)}{P(h_0|d)} = \frac{P(d|h_1)}{P(d|h_0)} \times \frac{P(h_1)}{P(h_0)}$$

Bayes Factor

The Bayes Factor or BF is used like the p value, to quantify the strength of evidence provided by the data.

Why not directly report the posterior odds?

- Because the prior beliefs may vary from researcher to researcher!!
- So the polite thing to do is to report BF and anyone can use the priors to calculate the posterior odds from the BF.
- Convention: to assume prior odds = 1, I.e., the null and the alternative hypotheses are equally likely.

Interpreting Bayes Factors

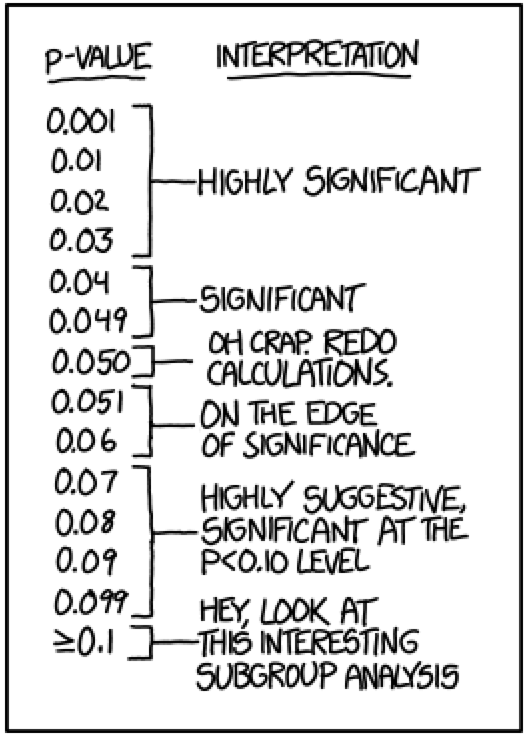
Bayes factor	Interpretation
1 - 3	Negligible evidence
3 - 20	Positive evidence
20 - 150	Strong evidence
$>$150	Very strong evidence

Why Bayesian Stats?

- Define p value?
- Define confidence interval?
- Wouldn't it be nice if you could say that the parameter lies in this range with 95% probability?
- All you need to do here is to be honest about what prior beliefs you brought to the table before you ran your study and then how your beliefs were updated..
- The dispute seems to be in the flexibility of prior beliefs, but there may be ways to justify these prior beliefs, specify them before you collect data, etc so that people are convinced you did you not use that flexibility to confirm your own biases.

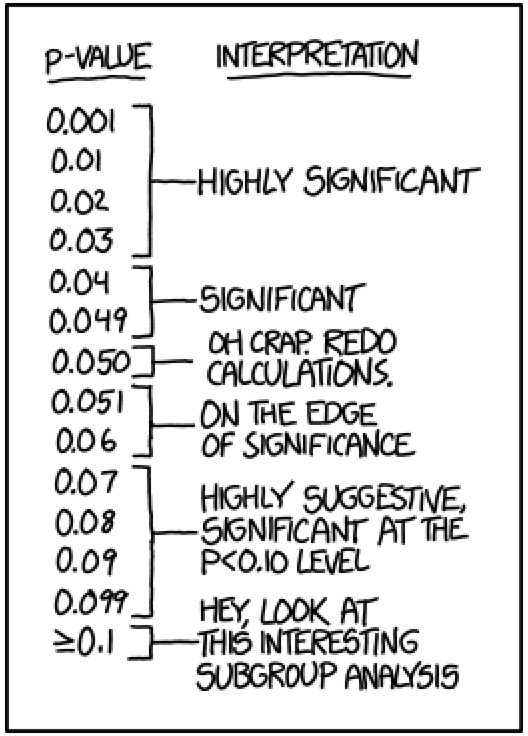
P-values

https://www.nature.com/articles/s41562-018-0311-x - our paper arguing why we need to do more work to justify the alpha level we choose for our domain



https://imgs.xkcd.com/comics/p_values.png

P-values



What happens if you peek at the data and say: "oh crap, p = 0.06, better collect some more data"??

https://imgs.xkcd.com/comics/p_values.png

p = 0.07, what do you do?

- You conclude that there is no effect, and try to publish it as a null result
- You guess that there might be an effect, and try to publish it as a "borderline significant" result
- You give up and try a new study
- You collect some more data to see if the p value goes up or (preferably!) drops below the magic criterion of p of 0.05

The fourth option?

- Messes up all p-values! They will all be incorrect if you want to interpret them as p-values (controlling for type-1 errors..). Why?

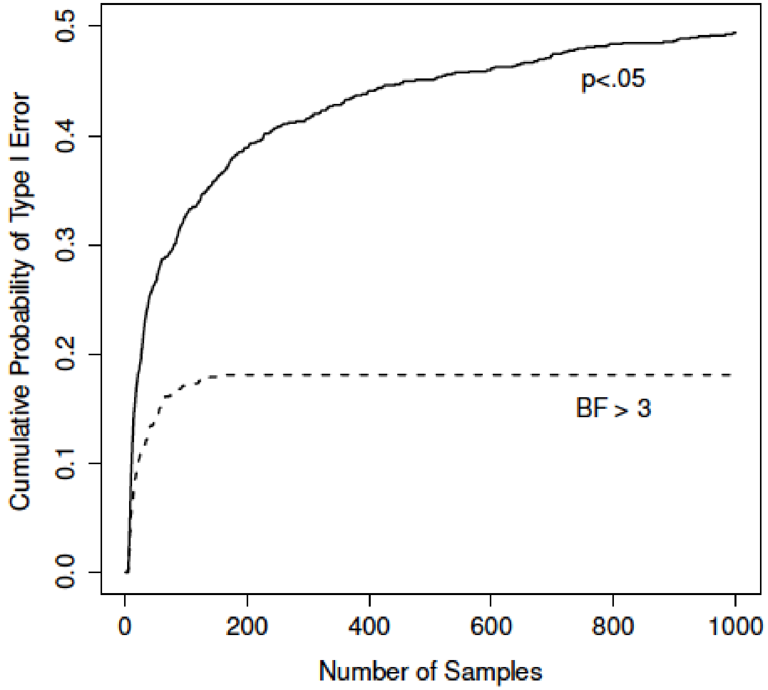
Outcome	Action
p less than .05	Reject the null
p greater than .05	Retain the null

Outcome	Action
p less than .05	Stop the experiment and reject the null
p between .05 and .1	Continue the experiment
p greater than .1	Stop the experiment and retain the null

Let's say you have a tight budget. Max 1000 subjects

- You peek at the data every time you collect an additional subject's worth of data
- Stop if $p<0.05$
- Let's say you keep doing it and never find an effect and you hit 1000 subjects.
- Assuming the null hypothesis is true, you should be able to get to this conclusion 95% of the times you run such an experiment. This is why you defined alpha = 0.05 at the outset.
- If we simulate this "optional stopping behavior", what is the true type 1 error?

Let's say you have a tight budget. Max 1000 subjects



Ok, but realistically, we don't peek 1000 times. How bad can it get?

- Assume your target is N = 80
- You collect N = 50. Your will power is exhausted. You decide to peek
- $p < 0.05$
- You decide to stop.
- The actual type 1 error instead of 5% now is 8%. So if you're honest, you have to say $p < 0.08$ in your reporting.
- So that is what happens with just ONE peek.
- So the rules are fairly strict with frequentist null hypothesis testing. No peeking.

Bayesian independent samples t-test

```
> load( "harpo.Rdata" )
> head(harpo)
  grade      tutor
1    65  Anastasia
2    72 Bernadette
3    66 Bernadette
4    74  Anastasia
5    73  Anastasia
6    71 Bernadette
```

Bayesian independent samples t-test

```
> load( "harpo.Rdata" )
> head(harpo)
  grade      tutor
1    65  Anastasia
2    72 Bernadette
3    66 Bernadette
4    74  Anastasia
5    73  Anastasia
6    71 Bernadette
```

```
Test results:
   t-statistic:  2.115
   degrees of freedom:  31
   p-value:  0.043
```

Bayesian independent samples t-test

```
> load( "harpo.Rdata" )
> head(harpo)
  grade      tutor
1    65  Anastasia
2    72 Bernadette
3    66 Bernadette
4    74  Anastasia
5    73  Anastasia
6    71 Bernadette
```

```
> ttestBF( formula = grade ~ tutor, data = harpo )
```

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 1.754927 ±0%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

Bayesian paired t-test

```
> load(chico)
> head(chico)
        id grade_test1 grade_test2
1 student1        42.9        44.6
2 student2        51.8        54.0
3 student3        71.7        72.3
4 student4        51.6        53.4
5 student5        63.5        63.8
6 student6        58.0        59.3
```

```
> ttestBF(
+     x = chico$grade_test1,
+     y = chico$grade_test2,
+     paired = TRUE
+ )
```

and here's the output:

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 5992.05 ±0%

Against denominator:
  Null, mu = 0
---
Bayes factor type: BFoneSample, JZS
```

Bayesian Regression

```
> load("parenthood.Rdata")
> head(parenthood)
  dan.sleep baby.sleep dan.grump day
1      7.59      10.18        56   1
2      7.91      11.66        60   2
3      5.14       7.92        82   3
4      7.71       9.61        55   4
5      6.68       9.75        67   5
6      5.99       5.04        72   6
```

Bayesian Regression

```
> model <- lm(
+    formula = dan.grump ~ dan.sleep + day + baby.sleep,
+    data = parenthood
+ )
```

```
> summary(model)

BLAH BLAH BLAH

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 126.278707   3.242492  38.945   <2e-16 ***
dan.sleep    -8.969319   0.560007 -16.016   <2e-16 ***
day          -0.004403   0.015262  -0.288    0.774
baby.sleep    0.015747   0.272955   0.058    0.954

BLAH BLAH BLAH
```

Bayesian Regression

```
> regressionBF(
+     formula = dan.grump ~ dan.sleep + day + baby.sleep,
+     data = parenthood
+ )
```

```
Bayes factor analysis
--------------
[1] dan.sleep                    : 1.622545e+34 ±0%
[2] day                          : 0.2724027    ±0%
[3] baby.sleep                   : 10018411     ±0%
[4] dan.sleep + day              : 1.016578e+33 ±0.01%
[5] dan.sleep + baby.sleep       : 9.770233e+32 ±0.01%
[6] day + baby.sleep             : 2340755      ±0%
[7] dan.sleep + day + baby.sleep : 7.835625e+31 ±0%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

Bayesian Regression

Not super helpful comparing everything to the intercept-only model which you typically don't care about much

Highest BF is easy enough to see here but in more complex situations can be harder

So you can use the head function to pick a few best models

However, this is still comparing models to the intercept-only model

```
> models <- regressionBF(
+    formula = dan.grump ~ dan.sleep + day + baby.sleep,
+    data = parenthood
+ )
```

```
> head( models, n = 3)

Bayes factor analysis
--------------
[1] dan.sleep                : 1.622545e+34 ±0%
[2] dan.sleep + day          : 1.016578e+33 ±0.01%
[3] dan.sleep + baby.sleep : 9.770233e+32 ±0.01%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

Bayesian Regression

```
> head( models/max(models), n = 3)

Bayes factor analysis
--------------
[1] dan.sleep               : 1          ±0%
[2] dan.sleep + day         : 0.06265328 ±0.01%
[3] dan.sleep + baby.sleep  : 0.06021549 ±0.01%

Against denominator:
  dan.grump ~ dan.sleep
---
Bayes factor type: BFlinearModel, JZS
```

```
> models[1] / models[4]

Bayes factor analysis
--------------
[1] dan.sleep : 15.96086 ±0.01%

Against denominator:
  dan.grump ~ dan.sleep + day
---
Bayes factor type: BFlinearModel, JZS
```

Bayesian Regression: individual coefficients

```
> regressionBF(
+   formula = dan.grump ~ dan.sleep + baby.sleep,
+   data = parenthood,
+   whichModels = "top"
+ )
```

```
Bayes factor top-down analysis
--------------
When effect is omitted from dan.sleep + baby.sleep , BF is...
[1] Omit baby.sleep : 16.60702     ±0.01%

[2] Omit dan.sleep  : 1.025401e-26 ±0.01%

Against denominator:
  dan.grump ~ dan.sleep + baby.sleep
---
Bayes factor type: BFlinearModel, JZS
```

Bayesian ANOVA

- Very similar to regression

```
> load("clinicaltrial.Rdata")
> head(clin.trial)
      drug    therapy mood.gain
1  placebo no.therapy       0.5
2  placebo no.therapy       0.3
3  placebo no.therapy       0.1
4 anxifree no.therapy       0.6
5 anxifree no.therapy       0.4
6 anxifree no.therapy       0.2
```

```
> models <- anovaBF(
+   formula = mood.gain ~ drug * therapy,
+   data = clin.trial
+ )
```

```
> models/max(models)

Bayes factor analysis
--------------
[1] drug                           : 0.3521273   ±0.96%
[2] therapy                        : 0.001047637 ±0.96%
[3] drug + therapy                 : 1           ±0%
[4] drug + therapy + drug:therapy  : 0.9856421   ±1.62%

Against denominator:
  mood.gain ~ drug + therapy
---
Bayes factor type: BFlinearModel, JZS
```

Summary

- BayesFactor package
- ttestBF()
- regressionBF()
- anovaBF()
- More interpretable Bayes Factors which can be converted to posterior odds ratios via the prior odds ratios.
- Combining p values and Bayes Factors can help you evaluate evidence better.

Project final presentations

- Use the weekend to do some serious analysis of your data for your projects.
- Next week – in-class help with projects
- Endsem – all inclusive, more emphasis on descriptive Qs, slightly more open-ended, will require you to write down your assumptions and justifications.