Report coursework assignment A - 2021

CS4125 Seminar Research Methodology for Data Science

Nikki Bouman (4597648), Anuj Singh (5305926), Gwennan Smitskamp (4349822)

20/04/2021

Contents

1 Introduction

We will be using the following packages:

```
library(ggplot2) # plotting
library(AICcmodavg) # aictab
library(pander) #for rendering output
library(rethinking) # for stan
```

```
## Loading required package: rstan

## Loading required package: StanHeaders

## rstan (Version 2.21.2, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
```

```
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)

## Do not specify '-march=native' in 'LOCAL_CPPFLAGS' or a Makevars file

## Loading required package: parallel

## Loading required package: dagitty

## rethinking (Version 2.01)

##
## Attaching package: 'rethinking'

## The following object is masked from 'package:AICcmodavg':
##
##     DIC

## The following object is masked from 'package:stats':
##
##     rstudent
```

2 Part 1 - Design and set-up of true experiment

2.1 The motivation for the planned research

The recent outbreak of the COVID-19 pandemic has changed our daily lives significantly. People are obligated to stay at home, disrupting their usual social interactions in both work and private life. The situation compels people to meet online. Typically, in such digital interactions, interlocutors can see each other by means of webcam streaming. However, this may not always be the case. Some or all interlocutors may not be visible during online dialogue, which could affect the quality of the conversation and the mutual understanding.

An important effect of the shift from face-to-face to online interaction can be revealed by studying laughter, as it is extremely contagious social behavior (Provine, 1992). Humans are very prone to unintentionally or unconsciously laugh as a social signal in any form; from a minor smile to laughing out loud. Additionally, laughing is one of the most important social signals for lubricating the flow of social interaction (Griffin et al., 2015).

2.2 The theory underlying the research

The effect of visibility on the use of gestures as a communicative function has been studied broadly (Alibali, Heath, & Myers, 2001; J. B. Bavelas, Chovil, Lawrie, & Wade, 1992; Cohen & Harrison, 1973; Cohen, 1977; Emmorey & Casey, 2001; Krauss, Dushay, Chen, & Rauscher, 1995; Rim´e, 1982). , J. Bavelas, Gerwing, Sutton, and Prevost (2008) provide a summary of previous experiments where rate and form of gestures were compared under two conditions: where the addressee could see the speaker and where the addressee could not see the speaker. These experiments show that speakers gestured at higher rate when they communicated with mutual visibility than without. J. Bavelas et al. (2008) extended these experiments by focusing on both visibility and dialogue as a variable, finding similar results. Furthermore, they found that speakers gestured at a significantly higher rate in a telephone dialogue than in a monologue to a tape recorder, confirming that visibility is not the only variable operating in telephone conversations. These experiments showed us that visibility plays a major role in the rate of gesturing, but that people also gesture when they are not visible to each other. As laughter can be seen as a form of gesturing, these findings are relevant for this study.

Laughing together is found to be essentially collaborative (Mehu & Dunbar, 2008; Coates, 2007). Joint laughter therefore serves important means to achieve effective team meetings (Ponton, Osbourne, Greenwood, & Thompson, 2018), considering that people who laugh on video are perceived with a higher likeability than people who do not (Reysen, 2006). This social function of joint laughter emphasises the relevance of studying the occurrence, now that the majority of meetings take place online.

2.3 Research questions

We will aim to answer the following research question:

What is the effect of webcam visibility during online dialogue on the frequency of joint laughter?

When recognizing laughter we do not focus on the reason why someone is laughing. We consider anything from an awkward laugh in a moment of silence to laughing out loud about a joke as a laughter episode regardless of the context.

2.4 The related conceptual model

The conceptual model related to the research question can be viewed in Figure 1. The main question is about the effect of mutual visibility on joint laughter. The mediating variable is familiarity, which we can define as the level of friendliness or intimacy between people. This can be caused when people know each other (which we aim to avoid), but also when people find similarities in their interests and behaviours. It is even possible that people of the same gender will feel more familiar with each other. A moderating variable is the duration of the experiment. The participants will do the experiment for about one hour, which can have a negative effect on the frequency of laughter, thus on the frequency of joint laughter.

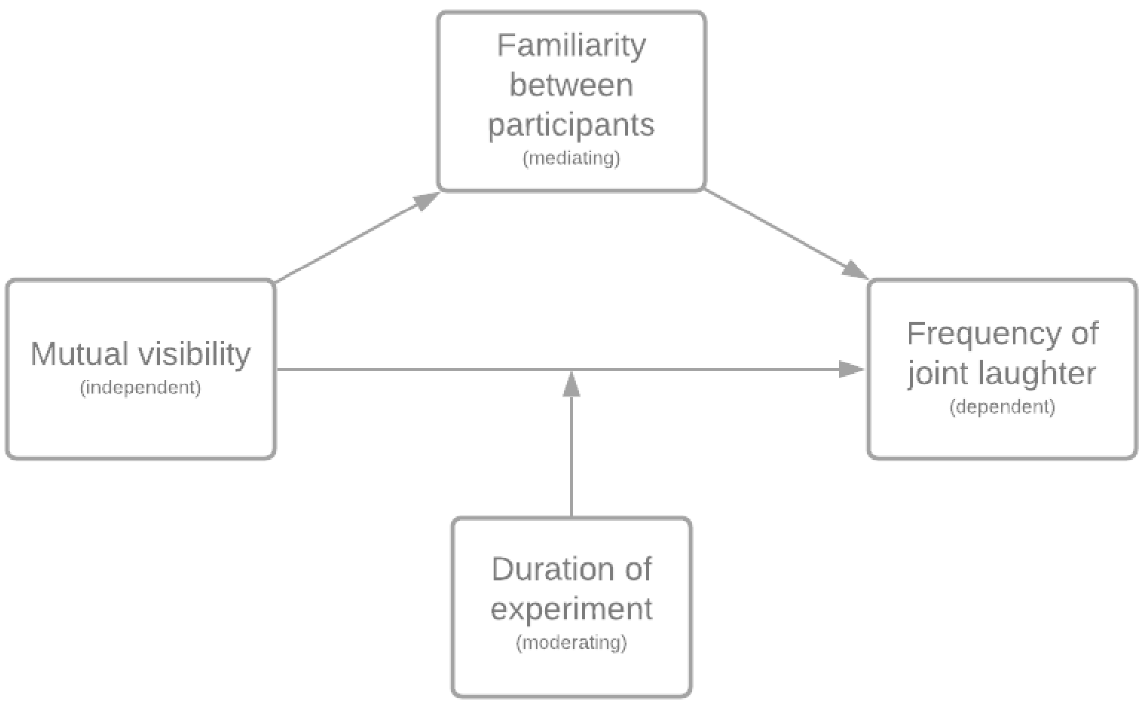


Figure 1: The conceptual model to test the effect of mutual webcam visibility on joint laughter

2.5 Experimental Design

One of the most important requirements for the setup of the experiment, was the creation of a comfortable and pleasant ambiance so that people would laugh. Therefore, it was decided that a game would comply, as the participants get the chance to interact with each other in a undemanding setting where the attention of the participants would be drawn to a task. It was reasoned that this would contribute to a reduction of awkwardness and give all the participants the option to speak and laugh. Additionally, the game needed to have a smooth flow that would automatically keep going to keep the interference of the researchers to a minimum.

The game that was chosen is called 30 Seconds. During the game, participants work together in teams (in the case of the experiment: two teams of two people) and gain points by guessing what the team member is describing. These descriptions include concepts such as famous persons, locations, movies and brands. Every team gets 30 seconds to guess as many concepts on the card as possible. Who is describing and who is guessing switches after every card.

2.6 Experimental procedure

The experiment will be set up in an online setting in a Zoom meeting. The host, one of us (not visible), will be able to send private messages containing the five words, share their screen and sound for a 30 second timer, and turn the participants' webcams on and off.

The following will repeat for every card of five words:

1. The host sends the words to the participant who has the turn to describe.
2. The host starts the 30 second timer.
3. The participant will try to describe as many words as possible, while his/her teammate will try to guess the words.
4. The timer rings, the host puts the score in the chat.

During the experiment, multiple things will happen. Each time all players have guessed and described a card (i.e. after four cards total), their webcams will switch on or off. After each player has guessed and described four cards (i.e. after sixteen cards total), the final score will be displayed and the teams will be rearranged. The previous will then repeat until every participant has been in a team with every other participant. An example of such an experiment is displayed in Figure 2. To counterbalance the experiments, half of the experiments will start with the webcams on, while the other half will start with the webcams off.

2.7 Measures

The data that has been collected includes audio and video of participants. The first step in data analysis involves annotating the signals. This will be done with a program in which we can manually select timesteps in which the participant is laughing. In the end we will have annotations for every person that contains the total amount of laughter (frequency), the amount of laughter with their webcam off, and the amount of laughter with their webcam on..

In a case where different people will annotate this data, we will first let every one of them annotate the same data sample. Then we can calculate the consistency in the annotations with for example Krippendorf's Alpha or Cohen's Kappa. When this is high enough, they can start annotating separate data.

2.8 Participants

The research is not specifically about a certain group of people, but more in general. However, we do want to aim for people with experience in an online setting and people who speak the same language. To be exact, we will conduct the experiment with people from the ages of 18 to 50 who speak dutch. The number of patricipants depends on the acceptable margin or error, which we do not know since we would have to go much more in-depth. However, since the population size basically includes more than 10.000 people and the independent variable is categorical, we should aim for around 385 participants (W.P. Brinkman, 2009). This results into around 96 experiments.

The experiments should be easy to conduct since the participants participate online; there is no need to travel. Moreover, the whole experiment should not take long. There will be 48 cards and for each card they have 30 seconds to explain. Taking some talking afterwards into account, the experiment should take 45 to 60 minutes.

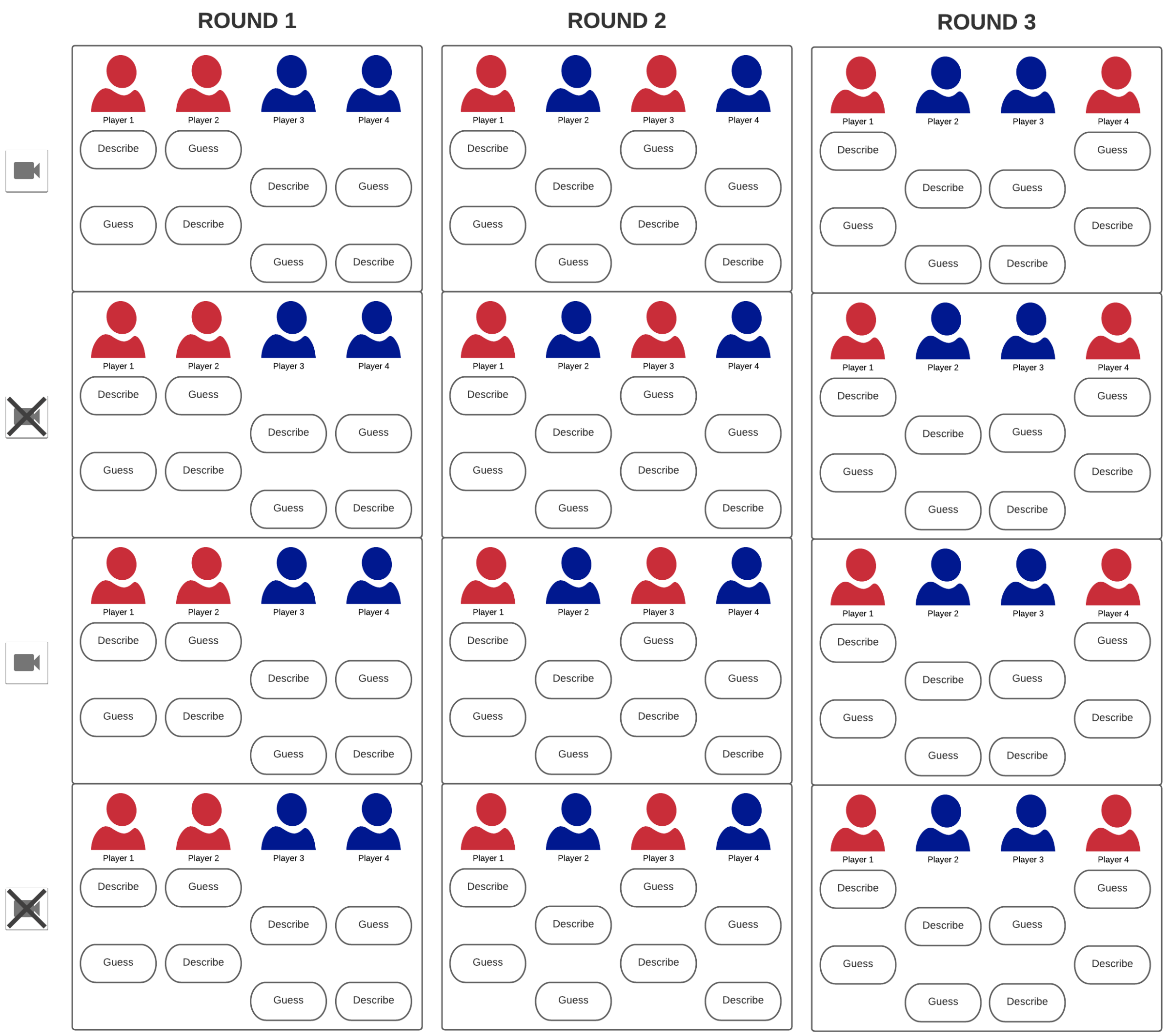


Figure 2: An example of an experimental setup.

To find the participants we can thus search online. Using a medium on which we state the experiment, we can find dutch people from all around the Netherlands of different age groups. They could for example sign up for a specific date, and if four people have signed up, the experiment can be held.

2.9 Suggested statistical analyses

To determine the significance of the results we will subject the data to statistical tests. To test the effect of mutual visibility (categorical, since it is either on or off) on the frequency (numerical) we will use the Wilcoxon signed rank test. The Wilcoxon signed rank test is a suitable statistical test when the measurements for a single variable are taken under two different conditions. It is similar to the paired t-test, however the paired t-test assumes that the data is normally distributed which we cannot assume for frequency of laughter. It tests the null hypothesis that the median difference of a two sets of observations is zero.

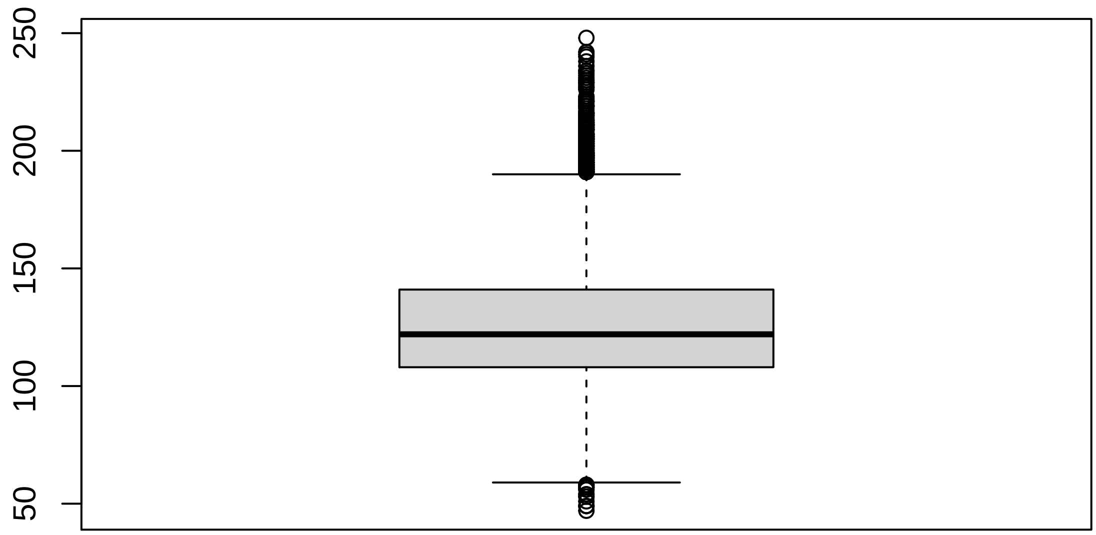
3 Part 3 - Multilevel model

3.1 Visual inspection

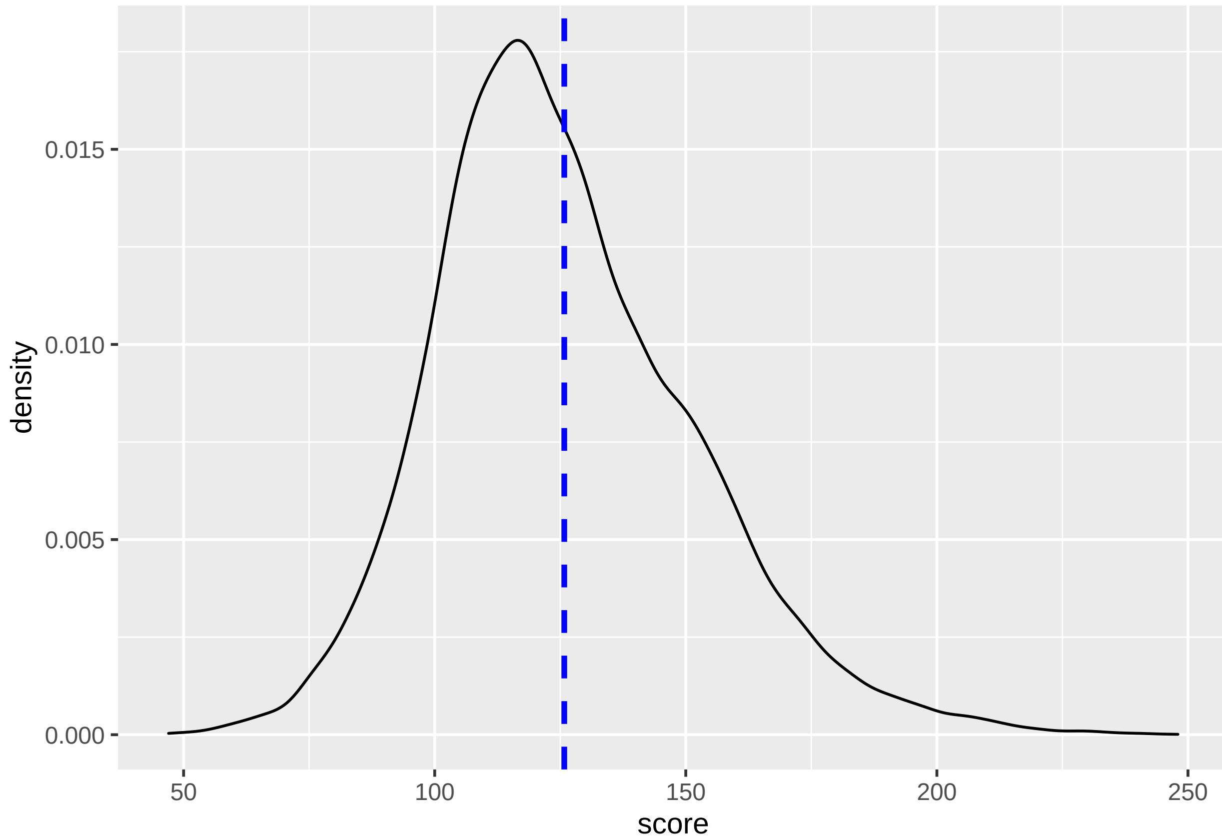
The boxplot and density plot show the distribution of the score. We can see that the mean score is 122 points. The minimum is set at 59, with outliers until 46, while the maximum is set at 190, with outliers until 248.

```
# Get data
filepath <- ("set0.csv")
ds <- read.csv(file=filepath, header=TRUE)
ds <- data.frame(ds)

# boxplot score overall distribution (session independent)
boxplot(ds$score)
```



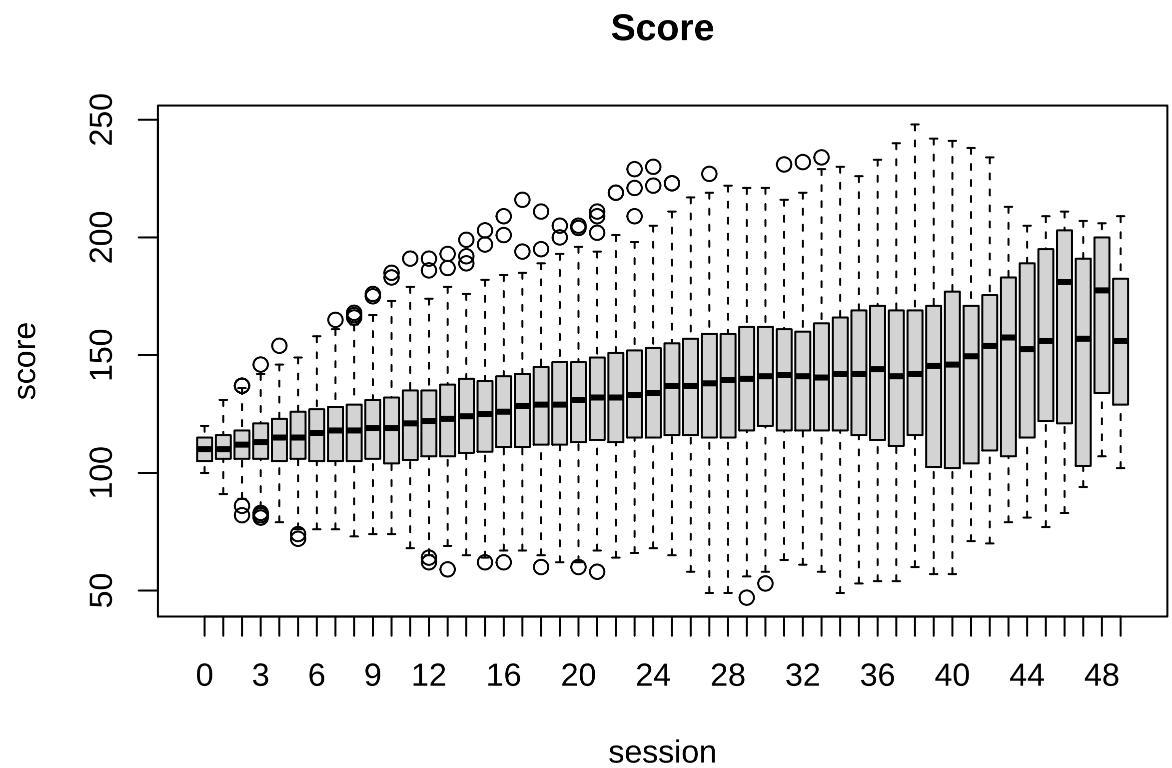

```r
# density score overall distribution (with mean line)
p <- ggplot(ds, aes(x=score)) + geom_density()
p + geom_vline(aes(xintercept=mean(score)), color="blue", linetype="dashed",size=1)
```



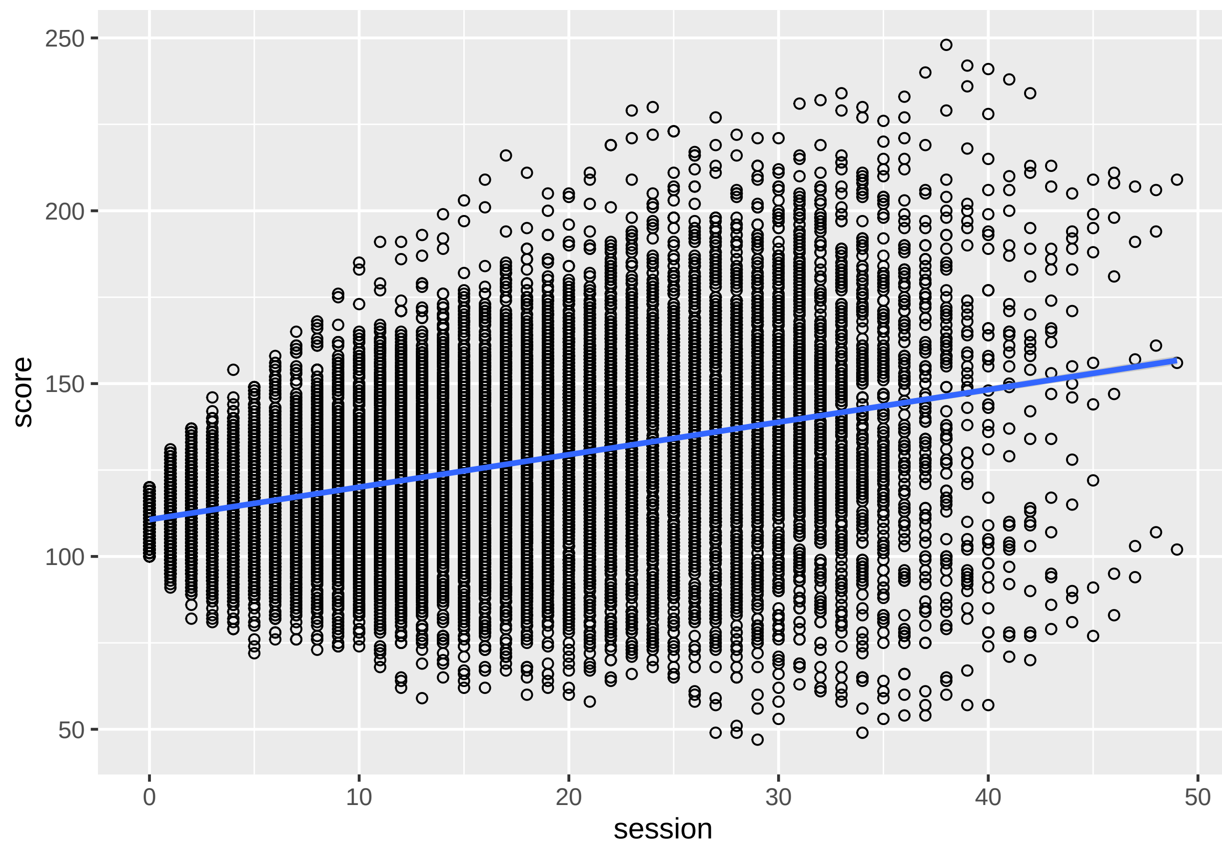
The relationship between the score and the session can be observed with the next two figures. The regression line (blue) in the scatterplot clearly shows how the score rises with the amount of sessions. This can also be observed in the box plot when looking at the mean (black line) for every box.

```
# set labels
ds$sessionF <- factor(ds$session, levels=c(0:49), labels=c(0:49))

# boxplot score per session
boxplot(score~sessionF, data=ds, main="Score", xlab="session", ylab="score")
```




```r
# ggplot score per session
hp <- ggplot(ds, aes(x=session, y=score)) + geom_point(shape=1) +
  geom_smooth(formula = y ~ x,method=lm)
hp
```



3.2 Frequentist approach

3.2.1 Multilevel analysis

We have conducted a multilevel analysis. We have an intercept only model (model0) which we compare to a model that includes a predictor parameter for the session (model1). By comparing these two models, we will know whether there is a difference in the score over the sessions.

```
# create models as given in slides lecture 4
model0 <- lm(formula=score~1, data=ds, na.action=na.exclude)
model1 <- lm(formula=score~sessionF, data=ds, na.action=na.exclude)

# analysis, see if predictor improves fitting
pander(anova(model0,model1))
```

Table 1: Analysis of Variance Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
16127	10713477	NA	NA	NA	NA
16078	9228641	49	1484836	52.79	0

```
pander(anova(model1))
```

Table 2: Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sessionF	49	1484836	30303	52.79	0
Residuals	16078	9228641	574	NA	NA

From this analysis we can see there is a significant variation between the sessions. We take a further look at the summary results.

```
pander(summary(model1))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	110.3	1.07	103	0
sessionF1	0.7166	1.514	0.4734	0.636
sessionF2	1.908	1.514	1.261	0.2075
sessionF3	2.96	1.514	1.955	0.05054
sessionF4	3.838	1.514	2.536	0.01123
sessionF5	5.004	1.514	3.306	0.0009494
sessionF6	5.972	1.514	3.945	8.005e-05
sessionF7	7.016	1.514	4.635	3.599e-06
sessionF8	7.637	1.514	5.045	4.585e-07
sessionF9	8.551	1.514	5.649	1.642e-08
sessionF10	9.097	1.515	6.004	1.969e-09
sessionF11	10.36	1.515	6.836	8.453e-12
sessionF12	11.22	1.515	7.402	1.41e-13
sessionF13	12.44	1.515	8.209	2.407e-16
sessionF14	13.6	1.515	8.974	3.161e-19
sessionF15	14.45	1.515	9.539	1.644e-21
sessionF16	15.63	1.515	10.31	7.277e-25
sessionF17	16.69	1.516	11.01	4.212e-28
sessionF18	18.07	1.518	11.9	1.62e-32
sessionF19	19.18	1.521	12.61	2.751e-36
sessionF20	19.8	1.521	13.02	1.539e-38
sessionF21	20.85	1.525	13.67	2.489e-42
sessionF22	21.35	1.529	13.96	5.131e-44
sessionF23	22.39	1.536	14.58	7.593e-48
sessionF24	23.32	1.542	15.12	2.555e-51
sessionF25	25.06	1.555	16.11	5.923e-58
sessionF26	26.11	1.574	16.59	2.818e-61
sessionF27	26.54	1.597	16.62	1.578e-61
sessionF28	27.37	1.64	16.69	5.045e-62
sessionF29	28.76	1.68	17.12	4.195e-65
sessionF30	29.52	1.73	17.07	9.765e-65
sessionF31	30.66	1.804	16.99	3.552e-64
sessionF32	30.02	1.908	15.73	2.406e-55
sessionF33	30.19	2.034	14.85	1.568e-49
sessionF34	30.08	2.213	13.59	7.575e-42
sessionF35	30.84	2.388	12.92	5.712e-38
sessionF36	31.17	2.571	12.12	1.116e-33
sessionF37	29.23	2.825	10.35	5.185e-25
sessionF38	32.42	3.076	10.54	6.942e-26
sessionF39	31.77	3.767	8.434	3.618e-17
sessionF40	32.78	4.031	8.131	4.553e-16

	Estimate	Std. Error	t value	Pr(>|t|)
sessionF41	32.06	4.503	7.119	1.13e-12
sessionF42	34.55	5.109	6.763	1.397e-11
sessionF43	37.34	5.748	6.496	8.474e-11
sessionF44	38.8	6.492	5.976	2.33e-09
sessionF45	43.17	8.057	5.358	8.536e-08
sessionF46	50.16	9.118	5.5	3.846e-08
sessionF47	40.13	10.77	3.727	0.0001948
sessionF48	56.73	12.03	4.717	2.417e-06
sessionF49	45.39	13.87	3.272	0.00107

Table 4: Fitting linear model: score ~ sessionF

Observations	Residual Std. Error	R^2	Adjusted R^2
16128	23.96	0.1386	0.136

The summary compares the first session (intercept) with the other sessions. Looking at the estimates, we can see that compared to the first session, the scores are higher every later session. Next, we will take a look at the Akaike Information Criterion (AIC) to compare the models on the goodness-of-fit concering the out-of-sample deviance.

```
#AIC
models <- list(model0, model1)
model.names <- c("model0", "model1")
pander(aictab(cand.set = models, modnames=model.names),
       caption="Model selection based on AICc.")
```

Table 5: Model selection based on AICc.

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
2	model1	51	148277	0	1	1	-74087	1
1	model0	2	150584	2308	0	0	-75290	1

Here we can see that model 1 has the best goodness-of-fit as it has the smallest AICc value. Lastly, we will obtain a 95% confidence interval of the estimates we have obtained earlier.

```
# gives CI95%
pander(confint(model1), caption="95% confidence interval of the estimates.")
```

Table 6: 95% confidence interval of the estimates.

	2.5 %	97.5 %
(Intercept)	108.2	112.4
sessionF1	-2.251	3.684
sessionF2	-1.059	4.875
sessionF3	-0.007004	5.927
sessionF4	0.8712	6.805
sessionF5	2.037	7.971
sessionF6	3.005	8.939
sessionF7	4.049	9.983

	2.5 %	97.5 %
sessionF8	4.67	10.6
sessionF9	5.584	11.52
sessionF10	6.127	12.07
sessionF11	7.388	13.33
sessionF12	8.245	14.19
sessionF13	9.468	15.41
sessionF14	10.63	16.57
sessionF15	11.48	17.42
sessionF16	12.66	18.6
sessionF17	13.72	19.67
sessionF18	15.09	21.04
sessionF19	16.2	22.16
sessionF20	16.82	22.79
sessionF21	17.86	23.84
sessionF22	18.35	24.34
sessionF23	19.38	25.41
sessionF24	20.3	26.34
sessionF25	22.01	28.11
sessionF26	23.02	29.19
sessionF27	23.41	29.67
sessionF28	24.15	30.58
sessionF29	25.47	32.06
sessionF30	26.13	32.91
sessionF31	27.12	34.19
sessionF32	26.28	33.76
sessionF33	26.2	34.18
sessionF34	25.74	34.42
sessionF35	26.16	35.52
sessionF36	26.13	36.21
sessionF37	23.69	34.76
sessionF38	26.39	38.45
sessionF39	24.39	39.16
sessionF40	24.88	40.68
sessionF41	23.23	40.89
sessionF42	24.54	44.57
sessionF43	26.07	48.6
sessionF44	26.07	51.52
sessionF45	27.38	58.96
sessionF46	32.28	68.03
sessionF47	19.02	61.23
sessionF48	33.15	80.3
sessionF49	18.2	72.59

Here we can again see an increase in score related to the sessions. From this we can conclude that the session has a positive effect on people's score. Also, it seems there is a significant variance between the participants in their score when the sessions increase.

3.2.2 Report section for a scientific publication

A Linear Model analysis was conducted to test the difference between sessions on the score. The results found a significant effect ($F(49,16078) = 52.793$, $p < .001$) for the sessions on the score. From the results we can conclude that over sessions the score per participant generally increases. Moreover, the variance between the participants increases when the sessions increase, which we think is caused by missing scores on later sessions.

3.3 Bayesian approach

3.3.1 Model description

For model 2, the model with session as a factor, we take as prior a normal distribution of N(125,30). This comes from the mean of the score, 125, and a bit more than the standard deviation, which is around 27. Our sigma is set at a uniform distribution of U(0.001,30).

$$score \sim Norm(\mu, \sigma)$$

$$\mu = \alpha$$

$$alpha = Norm(125, 30)$$

$$\sigma = Uniform(0.001, 30)$$

3.3.2 Model comparison

We will create and compare the three described models. From the results we can see that model 1, the model with the adaptive prior for subject id, has the best fit since it has the smallest WAIC value and largest Akaike weight.

```
ds <- ds[!(ds$Subject>99),] # select first 100 subjects
ds$Subject <- ds$Subject +1 # increase subject number with 1 to overcome Stan zero index problem

mean(ds$score) # check mean
```

```
## [1] 125.5142
```

```
sd(ds$score) # check standard deviation
```

```
## [1] 27.402
```

```
ds$sessionF <- factor(ds$session, levels=c(0:49), labels=c(0:49))
ds$subjF <- factor(ds$Subject, levels=c(1:100), labels=c(1:100))

da <- subset(ds, select=c(score, sessionF))
da1 <- subset(ds, select=c(score, sessionF, subjF))

# create model with fixed intercept (i)
m0 <- map2stan(
  alist(
    score ~ dnorm(mu, sigma),
    mu <- a,
    a ~dnorm(125,30), # mean and sd from what we found above
    sigma ~dunif(0.001,30)
  ), data = da, iter = 10000, chains = 4, cores = 4
)
```

```
## Computing WAIC
```

```r
# create model extended with an adaptive prior for subject id (ii)
m1 <- map2stan(
  alist(
    score ~ dnorm(mu, sigma),
    mu <- a + a_subj[subjF],
    a_subj[subjF] ~ dnorm(0,sigma_subj),
    sigma_subj ~ dcauchy(0,10),
    a ~ dnorm(125,30),
    sigma ~ dcauchy(0.001,30)
  ), data = da1, iter = 10000, chains = 4, cores = 4
)
```

```
## Computing WAIC
```

```r
# create model with session as a factor (iii)
m2 <- map2stan(
  alist(
    score ~ dnorm(mu, sigma),
    mu <- a[sessionF],
    a[sessionF] ~ dnorm(125,30),
    sigma ~dunif(0.001,30)
  ), data = da, iter = 10000, chains = 4, cores = 4
)
```

```
## Computing WAIC
```

```r
pander(compare(m0,m1,m2,func=WAIC))
```

	WAIC	SE	dWAIC	dSE	pWAIC	weight
m1	28322	96.4	0	NA	93.2	1
m2	30578	98.62	2256	120.9	67.92	0
m0	31039	98.25	2717	104.5	2.465	0

3.3.3 Estimates examination

From the previous comparison we could see that model 1 is the best fit model. We will further examine this model with 95% credible intervals of the parameters of this model.

```r
pander(precis(m1, depth=2, prob=.95))
```

	mean	sd	2.5%	97.5%	n_eff	Rhat4
a_subj[1]	-6.555	5.75	-17.78	4.733	5212	1.001
a_subj[2]	-19.44	3.581	-26.5	-12.5	1982	1.002
a_subj[3]	44.54	3.658	37.4	51.66	1958	1.003
a_subj[4]	14.25	3.506	7.372	21.1	1872	1.002
a_subj[5]	-5.815	3.603	-12.85	1.174	2042	1.002
a_subj[6]	-10.55	3.757	-17.79	-3.121	2007	1.002
a_subj[7]	-4.478	3.684	-11.7	2.775	1910	1.002
a_subj[8]	9.469	3.677	2.312	16.71	2016	1.002
a_subj[9]	-2.501	4.603	-11.53	6.64	3219	1.001
a_subj[10]	20.74	3.717	13.46	27.98	2200	1.002
a_subj[11]	13.28	3.245	6.911	19.54	1526	1.002
a_subj[12]	-19.62	3.702	-26.9	-12.38	1978	1.002
a_subj[13]	9.602	3.552	2.651	16.56	1773	1.003

	mean	sd	2.5%	97.5%	n_eff	Rhat4
a_subj[14]	10.35	3.341	3.798	16.92	1665	1.003
a_subj[15]	-25.6	3.469	-32.37	-18.75	1750	1.003
a_subj[16]	26.66	3.564	19.7	33.75	1901	1.003
a_subj[17]	2.465	3.417	-4.196	9.193	1754	1.003
a_subj[18]	-42.34	3.838	-49.96	-34.92	2214	1.002
a_subj[19]	-1.086	3.868	-8.622	6.506	2301	1.002
a_subj[20]	-2.402	3.838	-9.859	5.171	2253	1.002
a_subj[21]	-13.21	3.509	-20.11	-6.305	1825	1.002
a_subj[22]	10.95	4.331	2.507	19.49	2824	1.001
a_subj[23]	-6.51	3.48	-13.34	0.2933	1835	1.002
a_subj[24]	-6.145	4.123	-14.19	2.063	2576	1.001
a_subj[25]	-15.47	3.781	-22.93	-8.071	2225	1.001
a_subj[26]	-7.532	3.332	-14.04	-0.9554	1702	1.002
a_subj[27]	2.301	3.436	-4.378	9.147	1856	1.002
a_subj[28]	7.908	3.7	0.585	15.14	2060	1.002
a_subj[29]	-30.01	3.625	-37.11	-22.94	1977	1.002
a_subj[30]	23.88	3.712	16.68	31.16	2028	1.002
a_subj[31]	23.55	3.946	15.85	31.33	2338	1.001
a_subj[32]	-43.11	3.582	-50.18	-36.11	1868	1.003
a_subj[33]	34.27	3.47	27.47	41.12	1765	1.003
a_subj[34]	-28.95	3.818	-36.46	-21.48	2207	1.001
a_subj[35]	4.998	3.883	-2.583	12.56	2286	1.002
a_subj[36]	-32.07	3.322	-38.55	-25.53	1641	1.003
a_subj[37]	-4.731	3.673	-11.9	2.474	1948	1.002
a_subj[38]	20.38	3.415	13.67	27	1670	1.002
a_subj[39]	-12.5	3.729	-19.89	-5.264	2101	1.002
a_subj[40]	-10.76	3.858	-18.26	-3.164	2193	1.002
a_subj[41]	-24.52	3.856	-32.17	-17	2181	1.002
a_subj[42]	14.19	3.789	6.638	21.6	2226	1.002
a_subj[43]	-2.566	3.718	-9.799	4.749	1965	1.003
a_subj[44]	-29.37	3.741	-36.73	-21.95	2146	1.002
a_subj[45]	-16.94	3.593	-24.04	-9.9	1929	1.002
a_subj[46]	1.457	3.432	-5.349	8.155	1714	1.002
a_subj[47]	9.219	3.564	2.347	16.13	1901	1.002
a_subj[48]	-10.61	3.568	-17.62	-3.577	1874	1.003
a_subj[49]	38.8	3.513	31.91	45.69	1798	1.002
a_subj[50]	19.4	3.565	12.4	26.34	1926	1.002
a_subj[51]	1.341	3.666	-5.915	8.522	2099	1.001
a_subj[52]	-7.972	3.467	-14.87	-1.234	1839	1.002
a_subj[53]	17.13	3.743	9.849	24.38	2069	1.002
a_subj[54]	7.963	3.616	0.777	14.96	1981	1.002
a_subj[55]	-13.47	3.956	-21.23	-5.626	2508	1.002
a_subj[56]	-28.92	4.009	-36.69	-21.09	2311	1.001
a_subj[57]	10.6	3.658	3.281	17.69	1919	1.002
a_subj[58]	14.24	3.781	6.875	21.72	2227	1.002
a_subj[59]	-24.86	4.136	-32.99	-16.79	2604	1.002
a_subj[60]	21.39	3.447	14.58	28.23	1778	1.002
a_subj[61]	-17.88	3.5	-24.73	-11.01	1734	1.002
a_subj[62]	72.88	3.326	66.35	79.51	1609	1.003
a_subj[63]	-34.09	3.684	-41.21	-26.81	2029	1.003
a_subj[64]	-19.53	3.803	-27.09	-12.19	2280	1.002
a_subj[65]	-27.71	3.56	-34.67	-20.67	1939	1.002

	mean	sd	2.5%	97.5%	n_eff	Rhat4
a_subj[66]	-20.32	3.452	-27.17	-13.58	1697	1.003
a_subj[67]	17.17	3.815	9.726	24.55	2191	1.002
a_subj[68]	-15.56	3.511	-22.49	-8.782	1819	1.002
a_subj[69]	13.41	3.74	6.007	20.72	2065	1.002
a_subj[70]	33.68	3.61	26.53	40.7	1804	1.002
a_subj[71]	-27.72	3.706	-35	-20.54	2054	1.002
a_subj[72]	-16.43	3.678	-23.57	-9.219	2008	1.002
a_subj[73]	6.018	3.852	-1.513	13.56	2123	1.002
a_subj[74]	13.41	3.557	6.449	20.33	1905	1.002
a_subj[75]	-17.84	3.251	-24.19	-11.43	1604	1.002
a_subj[76]	12.98	3.6	5.931	20.14	1941	1.002
a_subj[77]	29.72	3.402	23.02	36.37	1710	1.002
a_subj[78]	-5.454	3.634	-12.65	1.683	1811	1.003
a_subj[79]	-15.92	3.845	-23.38	-8.41	2169	1.002
a_subj[80]	-3.509	3.445	-10.24	3.244	1780	1.002
a_subj[81]	6.939	3.598	-0.1488	14.03	1947	1.002
a_subj[82]	-9.651	3.525	-16.61	-2.733	1869	1.002
a_subj[83]	-3.863	3.814	-11.32	3.78	2071	1.002
a_subj[84]	-33.33	3.633	-40.48	-26.24	2044	1.002
a_subj[85]	-7.805	3.678	-14.98	-0.6448	2033	1.002
a_subj[86]	20.48	3.291	13.97	26.86	1686	1.002
a_subj[87]	7.577	3.753	0.2093	14.88	2005	1.002
a_subj[88]	21.34	3.894	13.7	28.97	2297	1.002
a_subj[89]	-1.193	3.45	-7.985	5.568	1774	1.002
a_subj[90]	19.21	3.713	11.91	26.52	2071	1.002
a_subj[91]	-0.7188	3.616	-7.844	6.376	1923	1.002
a_subj[92]	15.97	3.794	8.378	23.27	2038	1.002
a_subj[93]	3.501	3.632	-3.604	10.56	2025	1.003
a_subj[94]	3.075	3.508	-3.882	9.876	1830	1.002
a_subj[95]	-8.297	4.139	-16.44	-0.1126	2663	1.002
a_subj[96]	1.199	3.845	-6.376	8.779	2075	1.002
a_subj[97]	29.34	3.843	21.86	36.94	2119	1.002
a_subj[98]	24.36	3.479	17.43	31.09	1888	1.002
a_subj[99]	0.341	3.501	-6.557	7.077	1797	1.003
a_subj[100]	14.69	3.469	7.905	21.58	1817	1.002
sigma_subj	20.5	1.484	17.82	23.62	26305	1
a	124.9	1.98	121	128.8	596.9	1.007
sigma	17.86	0.2262	17.43	18.31	26257	0.9999

We can observe that the mean between the subjects has a high variance. This means that although the scores increase per session, the subjects have very different prior skills, achieving relatively higher or lower scores in their first session than average.