

Exercise day 3: AMH - solutions

Alessandra Meddis

Birth control pills and Anti-Mullerian hormone

In a clinical study researchers wished to investigate the degree of which taking birth control pills changed the level of Anti-Müllerian hormone (AMH) in the human body. AMH is expressed by granulosa cells of the ovary during the reproductive years, and limits the formation of primary follicles by inhibiting excessive follicular recruitment by follicle-stimulating hormone. It is thus expected that women taking birth control pills will show a significantly lower level of AMH on average. This is an observational study i.e., the participants were not randomized to either birth control or no birth control, but a random sample of 732 women were included in the study.

In this exercise we would like to create some descriptive analysis to check the researcher hypothesis **taking birth control pills lowers the level of AMH on average**.

The data set contains the following variables:

- *amh* : anti-mullerian hormone level
- *pill*: usage of birth control pills (no/yes)
- *age*: age of woman in years
- *BMI* : BMI of women
- *smoking* : if a smoker (1: yes, 0: no)

Question 0. First look at the data

0. Print a summary of the data and check if you have to transform some variables.

```
summary(amh)
```

```
##           X           id           age           bmi
## Min.      : 1.0   Min.      : 1.0   Min.      :21.90   Length:732
## 1st Qu.:183.8   1st Qu.:183.8   1st Qu.:29.20   Class :character
## Median :366.5   Median :366.5   Median :32.45   Mode  :character
## Mean    :366.5   Mean    :366.5   Mean    :32.57
## 3rd Qu.:549.2   3rd Qu.:549.2   3rd Qu.:35.80
## Max.    :732.0   Max.    :732.0   Max.    :41.80
##      smoking      pill           amh           afc
## Min.      :0.0000   Length:732   Min.      : 0.395   Min.      : 1.00
## 1st Qu.:0.0000   Class :character   1st Qu.: 10.700   1st Qu.:12.75
## Median :0.0000   Mode  :character   Median : 19.796   Median :19.00
## Mean    :0.2063           Mean    : 24.593   Mean    :21.08
## 3rd Qu.:0.0000           3rd Qu.: 32.030   3rd Qu.:27.25
## Max.    :1.0000           Max.    :126.304   Max.    :80.00
```

We can transform bmi and pill into factors, since they are categorical variables.

```
amh$bmi<-factor(amh$bmi)
amh$pill<-factor(amh$pill)
```

1. Imagine you have to create a "Table1: table of individuals characteristics". We want to study the AMH by contraceptive use, we can provide the proportion of contraceptive users, the median of age by contraceptive use and number of smokers and individuals in each BMI group by contraceptive use.

```
## proportion of users:
prop.table(table(amh$pill))
```

```
##
##      no      yes
## 0.6885246 0.3114754
```

```
# BMI group by contraceptive use:
table(amh$bmi, amh$pill)
```

```
##
##              no yes
## normal      385 180
## overweight  108  44
## underweight  11   4
```

```
# smokers by contraceptive use:
table(amh$smoking, amh$pill)
```

```
##
##      no yes
## 0 401 180
## 1 103  48
```

```
#proportion of smokers among users:
48/(180+48)
```

```
## [1] 0.2105263
```

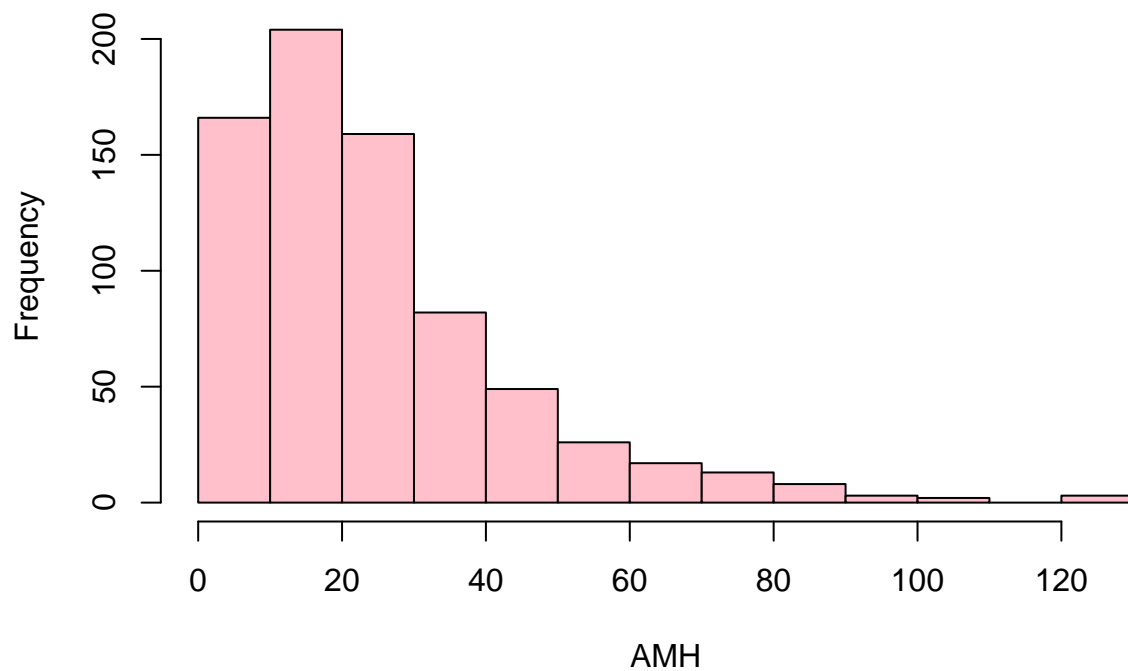
```
#proportion of smokers among non-users:
103/(103+401)
```

```
## [1] 0.2043651
```

2. Calculate the descriptive statistics you believe are relevant, explain the results.

Question 1. Distribution of AMH

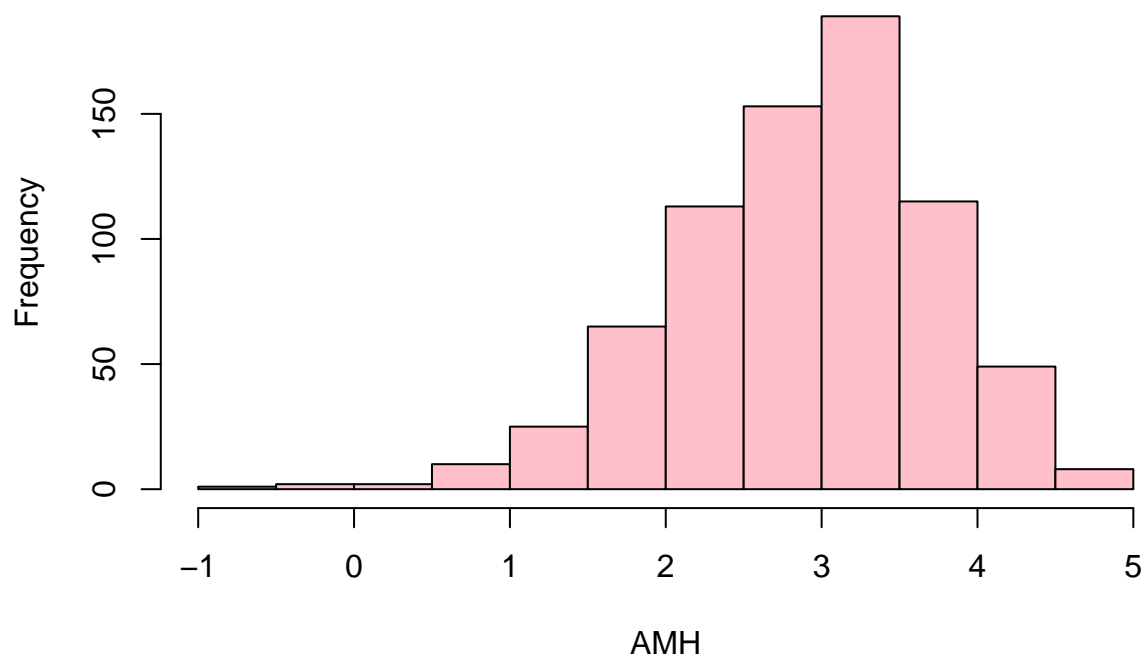
We start looking at the distribution of the AMH.



The histogram shows an heavy tail, so we might prefer to use the logarithmic scale of AMH for the analysis.

1. Add a new variable with logarithmic transformation for the AMH. Create the histogram for the log-AMH.

```
amh$logamh<-log(amh$amh)
hist(amh$logamh, col="pink", xlab="AMH", main=" ")
```



Question 2. Descriptive analysis: AMH by contraceptive use

We want to understand if taking birth control has an effect on the AMH level. What would you calculate to check whether the AMH level either increase or decrease by use of contraceptive?

1. Calculate the statistic you had in mind.
2. Create a plot (choose the one you prefer) to show the distribution of AMH level by contraceptive use.

What would be your conclusion on the hypothesis?

To assess the if AMH is impacted by the use of contraception we could calculate the difference in mean for the AMH level between contraceptive users and non-users.

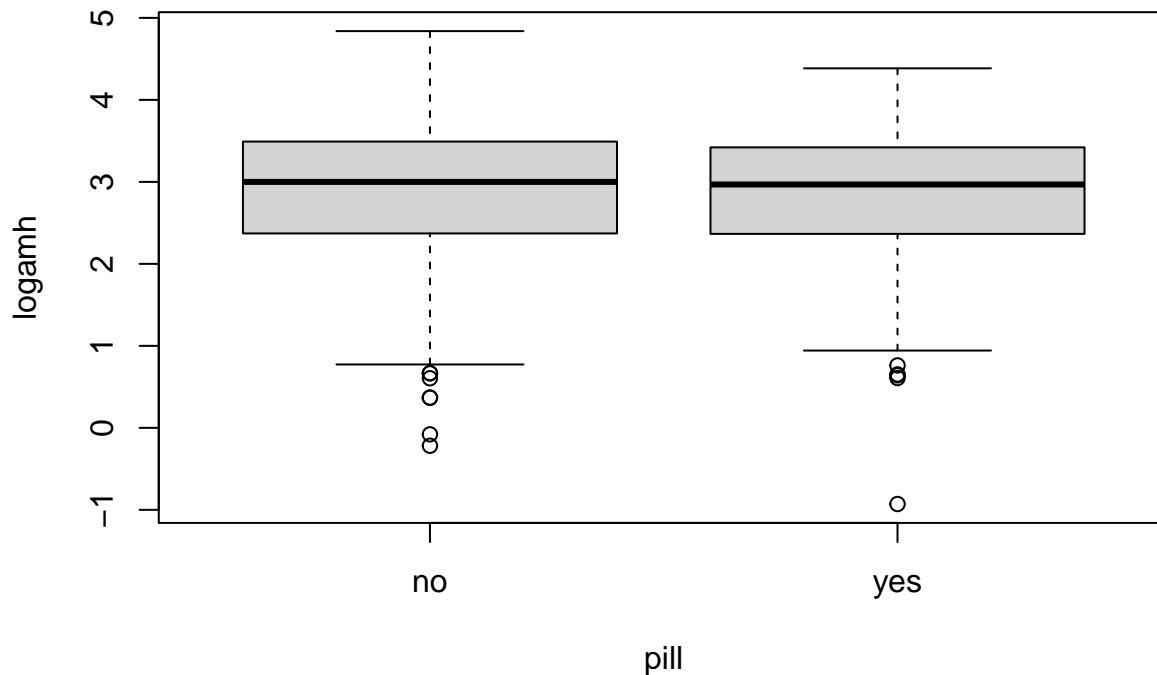
```
aggregate(logamh~pill, amh, FUN=mean)
```

```
## pill  logamh
## 1   no 2.911799
## 2   yes 2.853779
diff<-2.91 - 2.853
diff
```

```
## [1] 0.057
```

From the output we observe that the average AMH value in the non-users is 2.91 and among the users it is 2.85, with a difference of 0.058. It seems like the use of contraceptive is not decreasing the AMH level. We can also provide the boxplot to show the distribution of AMH by contraceptive use.

```
boxplot(logamh~pill, amh)
```



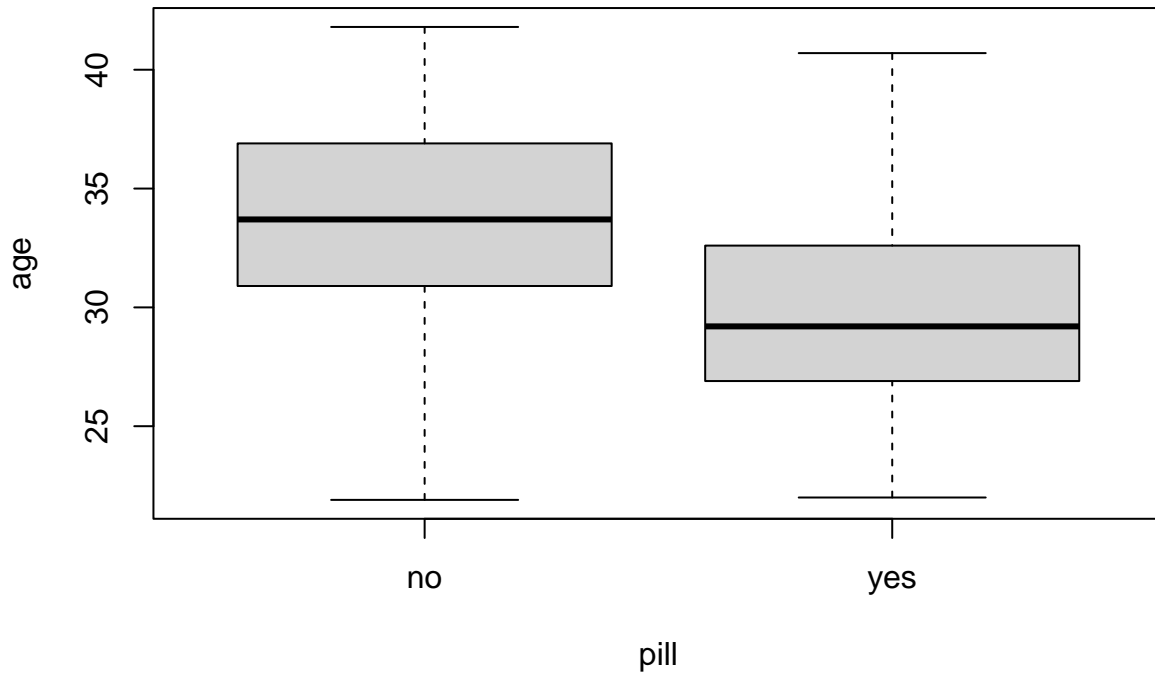
Question 3. Understanding data better:

This is an observational study, no randomization on the contraceptive use. Check at the results of the descriptive obtained in Question 0, in particular at the median age among users and non-users. What can you notice?

We can notice that users seems to be younger

1. Create a plot (choose the one you think it is most relevant) to show the distribution of age by contraceptive use.

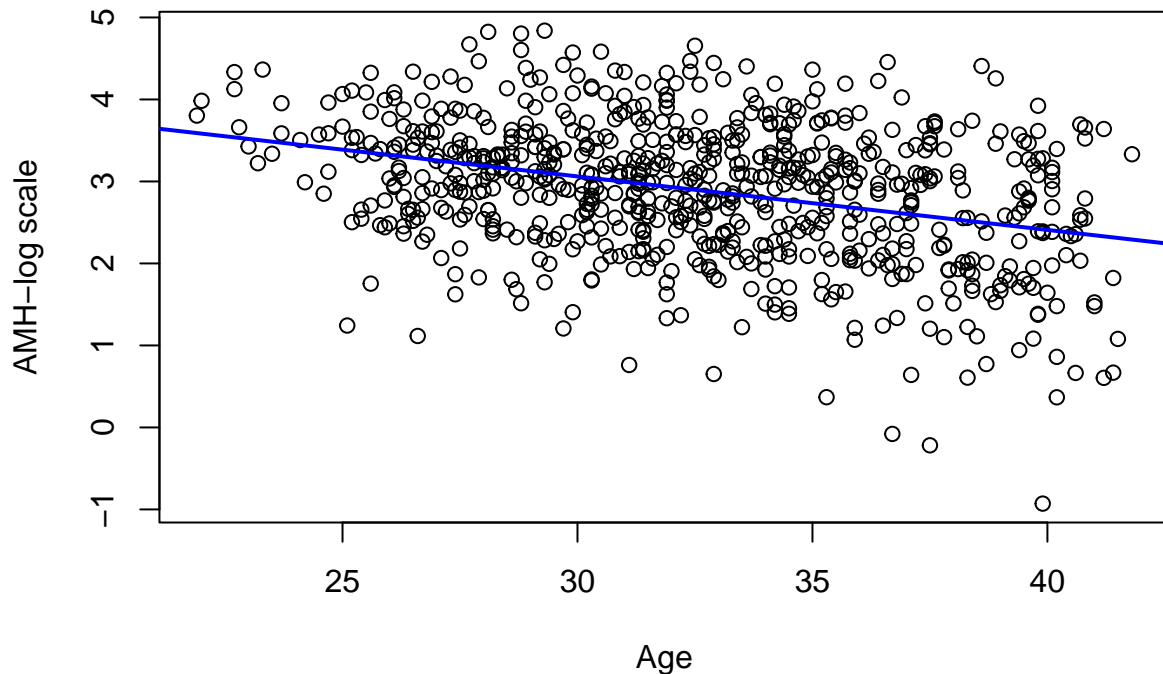
```
boxplot(age~pill, amh)
```



2. Create a scatterplot where the y-axis is the AMH level (log scale) and the x-axis is age. Add the regression line, namely the fitted line of the average AMH level by age (in year). You can add the line using the command `abline(lm(y-Variable ~ x-Variable, data), col="Choose a color")` (see slides).

What does this show?

```
plot(amh$age, amh$logamh, xlab="Age", ylab="AMH-log scale")  
abline(lm(logamh~age, amh), col="blue", lwd=2)
```



The plot shows that the AMH is decreasing with age

4. Would you change the analysis? Would you consider other variables when studying if AMH level is decreasing with contraceptive use?

We could re-do the analysis assessing the AMH level change among users and non-users given age. Looking at the other variables, it seems like their distribution is balanced among the two groups (users/non-users).

Question 4. Descriptive analysis: AMH and contraceptive use by age

So far, we believe that age might be relevant when assessing if AMH level is decreasing with contraceptive use. We can consider age as a categorical variable and have a stratified analysis, namely we re-run the descriptive analysis in each age-group.

1. Create a new variable of age group. You can choose the number of groups to use and the cut-off value(s).

```
# Create the categorical age group with two levels with the median as cut-off
median(amh$age)
```

```
## [1] 32.45
```

```
amh$age_cat <- cut(amh$age, breaks=c(-Inf,32,Inf), labels=c("<=32",">32"))
table(amh$age_cat)
```

```
##
```

```
## <=32 >32
```

```
## 344 388
```

2. Calculate the descriptive statistic (the same chosen in Question 1) by age group. What would be your conclusion now?

```
aggregate(logamh~ pill + age_cat, amh, FUN=mean)
```

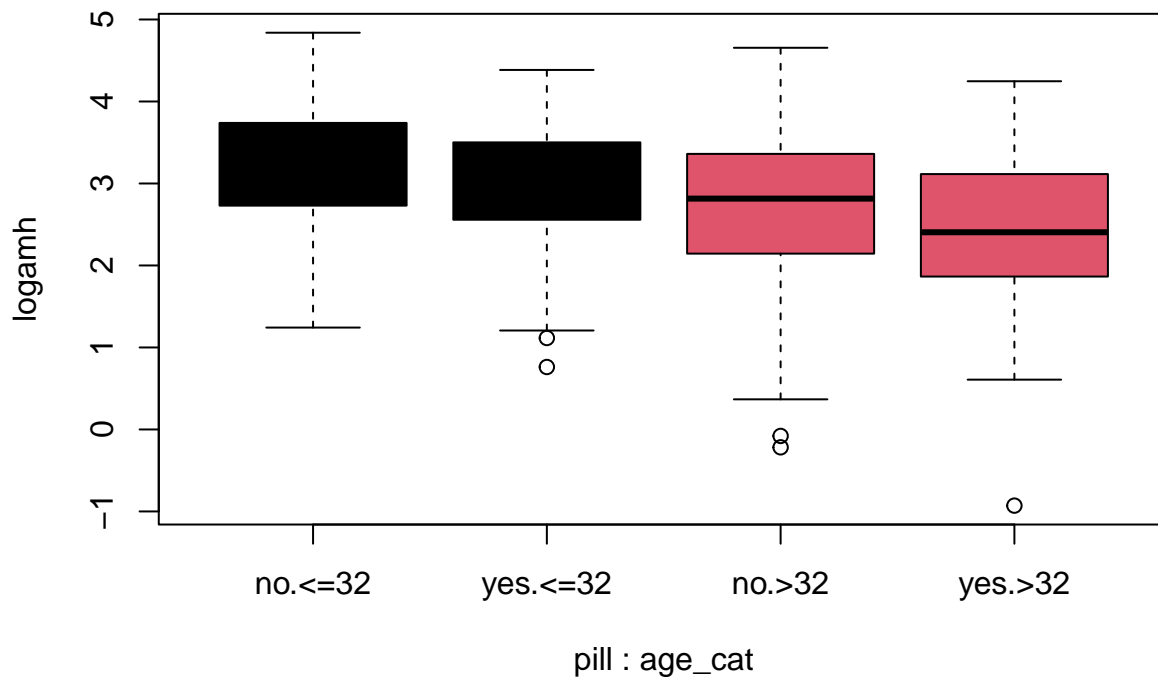
```
## pill age_cat logamh
## 1 no <=32 3.215594
## 2 yes <=32 3.030133
## 3 no >32 2.744478
```

```
## 4 yes >32 2.391900
```

The difference of AMH between non-users and users within the younger women is now $3.22 - 3.03 = 0.19$, while for older women is $2.74 - 2.39 = 0.35$. It seems like the AMH level is decreasing with the use of contraception, with a stronger effect among older women.

3. Create a plot you believe it would best explain the results. (You can choose to use either the continuous or categorical version of age)

```
## boxplot with categorical age_group:
boxplot(logamh~pill+age_cat, amh, col=amh$pill)
```



```
# using the regression line, with age in a continuous scale:
plot(amh$age, amh$logamh, xlab="Age", ylab="AMH-log scale")
abline(lm(logamh~age, data=subset(amh, pill=="yes")), col="blue", lwd=2)
abline(lm(logamh~age, data=subset(amh, pill=="no")), col="green", lwd=2)
legend("bottomleft", c("yes", "no"), col = c("blue", "green"), title="contraceptive", lty=1)
```

