Assignment 8 for Statistical Computing and Empirical Mathods (Answers)

Dr. Henry WJ Reeve

Teaching block 1 2021

Introduction

This document describes your eighth assignment for Statistical Computing and Empirical Methods (Unit EMATM0061) on the MSc in Data Science. Before starting the assignment it is recommended that you first watch video lectures 18 and 19.

Begin by creating an Rmarkdown document with html output. You are not expected to hand in this piece of work, but it is a good idea to get used to using Rmarkdown.

1 Obstacles to valid scientific inference

(Q) For each of the following give (A) an explanation of the concept and (B) an example of a situation (real or hypothetical) where they create a barrier to drawing scientific conclusions based on data. You are encouraged to discuss these concepts with your colleagues.

1. Measurement distortions
2. Selection bias
3. Confounding variables

(A)

1. **Measurement distortions**

Measurement distortions occur whenever there is a mismatch between the quantities recorded within the data and the true variable of interest.

Example: Suppose we are investigating the effect of a new type of feed on chickens. For the purpose of the experiment a large sample of chickens is split into two groups. One group is given feed type A, and the other feed type B. After some time has elapsed, the chicken's weights are all measured. Suppose the measuring equipment used to weight the chickens given feed type A had a downward bias. On the other hand, the measuring device used to weight the chickens given feed type B has no such bias. We could then observe a data set where the recorded weights for chickens receiving feed source B typically exceed those receiving feed type A. However, this difference is purely an artefact of the faulty measurement equipment.

2. **Selection bias**

Selection bias occurs whenever the sample included in the analysis misrepresents the underlying population of interest. Examples include:

a) Sample bias: When some members of the population are more likely to be selected than others.

b) Self selection bias: Occurs whenever people decide whether or not they should be assigned to a particular group.

c) Attrition bias: Occurs whenever the sample is distorted by people leaving the study.

d) Post-hoc selection: Occurs when a subset of the data is chosen based on the sample itself.

A classic historical example of sample bias is the Literary Digest poll of 1936. In this people were asked about their voting intentions. The prediction was for a 57% victory for Republican Landon over Roosevelt. In fact Roosevelt won the election. The data was collected based upon surveys. These suveys were carried out based upon lists from telephone books and club membership. At the time both telephone ownership and club membership were indicators of wealth, so the sample was biased towards people with higher levels of wealth. At the time, higher levels of wealth were slightly more likely to vote Republican than Democrat. Hence, the selection bias resulted in a misleading result. Subsequent Gallop polls created more accurate results with smaller sample sizes.

3. **Confounding variables**

Suppose that we are interested in understanding the causal effect of an independent variable X on a dependent variable Y. A confounding variable is a third causal factor Z which effects both X and Y. This makes it difficult to disentangle causal effects from purely correlative behavior.

As an example consider a scientific study into the causal effect of regular cardio-vascular exercise on longevity. Here a confounding variable could be someones overall interest in a healthy lifestyle. This is likely to effect someones via the causal effect of increased cardio-vascular exercise. However, it is likely that an interest in a healthy lifestyle will also effect the amount of fresh fruit and vegetables someone eats, for example. This may also have a causal effect on longevity. Hence, in light of the confounding variables it is difficult to distinguish a causal effect from a purely correlatative relationship.

2 An unpaired t test

In this question the goal is to create a function called `t_test_function` which implements an unpaired Student's t test, in order to test for a difference of population means between two unpaired samples from two distributions. Your function will play a similar role to the following standard R function:

```
t.test(body_mass_g~species, data=peng_AC,var.equal = TRUE)
```

Begin by creating a data frame called "peng_AC" which is a subset of the Palmer penguins data set consisting of all those penguins which belong to either the "Adelie" or the "Chinstrap" species of penguins.

```
library(palmerpenguins)

peng_AC<-penguins%>%
  drop_na(species,body_mass_g)%>%
  filter(species!="Gentoo")
```

(Q) Your function should take in the following arguments:

1. "data" - A data frame argument,
2. "val_col"- A string argument containing a column name for a continuous variable,
3. "group_col" - A string argument containing a column name for a binary variable.

The function should carry out an unpaired Student's t test based on the value of the continuous variable with column name "val_col". The function should begin by partitioning the sample into two groups based on the value of the binary variable named "group_col". Your function should then compute the sample mean, sample variance and sample size for each of these two groups, based upon the variable within the column named "var_col".

Your function should compute a test statistic for the Student's unpaired t-test. In addition, the function should compute the corresponding **p-value**. Finally, your function should compute an estimate for the effect size.

(A)

```
t_test_function<-function(data,val_col,group_col,var_equal=TRUE){


  stats<-data%>%
    rename(group=!!(group_col),val=!!(val_col))%>%
    group_by(group)%>%
    drop_na(val)%>%
    summarise(mn=mean(val),vr=var(val),n=n())

  pooled_sd<-sqrt(((stats$n[1]-1)*stats$vr[1]+(stats$n[2]-1)*stats$vr[2])/(stats$n[1]+stats$n[2]-2))


  if(var_equal){

    t_stat<-(stats$mn[1]-stats$mn[2])/(pooled_sd*sqrt(1/stats$n[1]+1/stats$n[2]))

    dof<-stats$n[1]+stats$n[2]-2

  }else{

    t_stat=(stats$mn[1]-stats$mn[2])/sqrt(stats$vr[1]/stats$n[1]+stats$vr[2]/stats$n[2])

    dof=(stats$vr[1]/stats$n[1]+stats$vr[2]/stats$n[2])^2/(
      (stats$vr[1]/stats$n[1])^2/(stats$n[1]-1)+
      (stats$vr[2]/stats$n[2])^2/(stats$n[2]-1)
    )


    }

  p_val<- 2*(1-pt(abs(t_stat),df=dof))

  return(data.frame(t_stat=t_stat,dof=dof,p_val=p_val))

}
```

Your function should have the following output:

```
t_test_function(data=peng_AC,val_col="body_mass_g",group_col="species")
```

```
##       t_stat dof     p_val
## 1 -0.5080869 217 0.6119085
```

As an additional challenge you can modify your function so that it takes a fourth argument called "var_equal" which takes a Boolean value. If the input of "var_equal" is the Boolean "TRUE" your function should compute the test-statistic and p-value for an unpaired Student's t test. If, on the other hand, the input of "var_equal" is the Boolean "FALSE" your function should compute the test-statistic and p-value for Welch's t-test. Your function should have the following output:

```
t_test_function(data=peng_AC,val_col="body_mass_g",group_col="species",var_equal=FALSE)
```

```
##       t_stat      dof     p_val
## 1 -0.5430902 152.4548 0.5878608
```

(Q) You can compare the output of your function with R's inbuilt **t.test()** function.

(A)

```
t.test(body_mass_g~species, data=peng_AC,var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  body_mass_g by species
## t = -0.54309, df = 152.45, p-value = 0.5879
## alternative hypothesis: true difference in means between group Adelie and group Chinstrap is not equa
## 95 percent confidence interval:
##  -150.38481   85.53284
## sample estimates:
##    mean in group Adelie mean in group Chinstrap
##                3700.662                3733.088
```

3 Statistical hypothesis testing

(Q) Explain the following concepts:

1. Null hypothesis
2. Alternative hypothesis
3. Test statistic
4. Type 1 error
5. Type 2 error
6. The size of a test
7. The power of a test
8. The significance level
9. The p-value

10. Effect size

(A)

1. The null hypothesis is our default position in a statistical hypothesis which typically declares the absence of some interesting phenomenon for example the equality of two statistical parameters.
2. The alternative hypothesis is a statistical hypothesis which contradicts the null hypothesis and typically declares the presence of some interesting phenomenon, often consistent with the research hypothesis a scientist is attempting to prove. For example, a difference in the values of two statistical parameters.
3. A test statistic is some function of the data used within a statistical hypothesis test. The test statistic must have a known distribution under the null hypothesis. In addition, the test statistic should emphasize differences between null and alternative hypothesis.
4. A type 1 error is a rejection of the null hypothesis in favor of the alternative hypothesis when the null hypothesis is true.
5. A type 2 error is a failure to reject the null hypothesis in favor of the alternative hypothesis when the alternative hypothesis holds.
6. The size of the test is the probability of a type 1 error under the null hypothesis.
7. The power of a test is one minus the probability of a type 2 error under an alternative hypothesis. That is, the probability of rejecting the null hypothesis under an alternative hypothesis. Typically this depends upon the particular alternative hypothesis.
8. The significance level is an upper bound on the size of the test. This should be chosen in advance of seeing the data. A typical value is $\alpha = 0.05$.
9. The p-value is probability under the null hypothesis that the test statistic will achieve a value as extreme or more extreme than the value which is actually observed. A very small p-value indicates that the observed data is sufficiently inconsistent with the null hypothesis that the null hypothesis can be reasonably rejected.
10. The effect size is a measure of the magnitude of the observed phenomenon which reflects the extent to which the null hypothesis is false.

(Q) Is the p-value the probability that the null hypothesis is true?

(A) No! The p-value is the probability, under the null hypothesis, of a test-statistic as extreme or more extreme than the observed numerical value.

(Q) If I conduct a statistical hypothesis test, and my p-value exceeds the significance level, do I have good evidence that the null hypothesis is true?

(A) No. This often occurs when the null hypothesis is false, but we have either a small sample size or a relatively small effect size.

4 Investigating test size for an unpaired Student's t-test

In this question we shall investigate the performance of an unpaired Student's t test from the perspective of test size. Recall a Type 1 error occurs when we reject the null hypothesis, when the null hypothesis is true. The size of a test is the probability of a Type 1 error. A key property of valid statistical hypothesis tests with a given significance level is that the size of the test does not exceed the significance level.

Note that we can apply unpaired Student's t-test with significance level `alpha` on a samples `sample_0`, `sample_1`:

```
t.test(sample_0,sample_1,var.equal = TRUE, conf.level = 1-alpha)
```

We can apply an unpaired Student's t-test and extract the p-value as follows:

```
t.test(sample_0,sample_1,var.equal = TRUE)$p.value
```

Notice that the significance level wasn't supplied as an argument. Is this a problem?

The following code checks the size of an unpaired Student's t-test with a significance level of $\alpha = 0.05$.

```
num_trials<-10000
sample_size<-30
mu_0<-1
mu_1<-1
sigma_0<-3
sigma_1<-3
alpha<-0.05

set.seed(0) # set random seed for reproducibility

single_alpha_test_size_simulation_df<-data.frame(trial=seq(num_trials))%>%
  mutate(sample_0=map(.x=trial,.f=~rnorm(n=sample_size,mean=mu_0,sd=sigma_0)),
         sample_1=map(.x=trial,.f=~rnorm(n=sample_size,mean=mu_1,sd=sigma_1)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap(.l=list(trial,sample_0,sample_1),
                      .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
  # generate p values
  mutate(type_1_error=p_value<alpha)

single_alpha_test_size_simulation_df%>%
  pull(type_1_error)%>%
  mean() # estimate probability of type I error i.e. the size of the test
```

Check that you understand the above code.

(Q) Modify the above code to explore how the size of the test varies as a function of the significance level α.

(A)

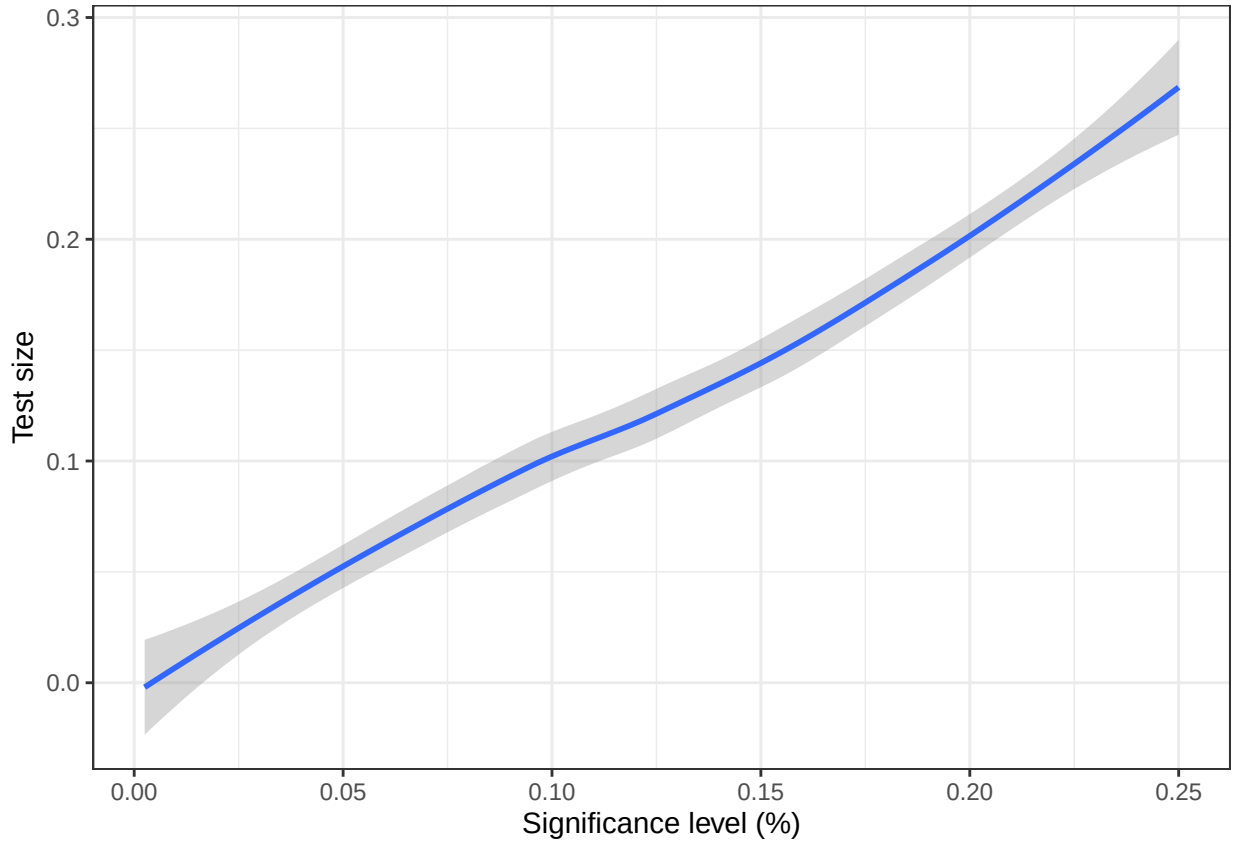
```
num_trials_per_alpha<-100
sample_size<-30
mu_0<-1
mu_1<-1
```

```r
sigma_0<-3
sigma_1<-3
alpha_min<-0.0025
alpha_max<-0.25
alpha_inc<-0.0025

set.seed(0) # set random seed for reproducibility

many_alpha_test_size_simulation_df<-crossing(trial=seq(num_trials_per_alpha),
                                             alpha=seq(alpha_min,alpha_max,alpha_inc)
                                             )%>%
  mutate(sample_0=map(.x=trial,.f=~rnorm(n=sample_size,mean=mu_0,sd=sigma_0)),
         sample_1=map(.x=trial,.f=~rnorm(n=sample_size,mean=mu_1,sd=sigma_1)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap(.l=list(trial,sample_0,sample_1),
                      .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
  # generate p values
  mutate(type_1_error=p_value<alpha)

many_alpha_test_size_simulation_df%>%
  group_by(alpha)%>%
  summarise(test_size=mean(type_1_error))%>%
  ggplot(aes(x=alpha,y=test_size))+
  geom_smooth()+xlab("Significance level (%)")+ylab("Test size")+
  theme_bw()
```



5 The power of an unpaired t-test

In this question we shall investigate the performance of an unpaired Student's t test from the perspective of statistical power. Recall that the statistical power of a test is the probability of correctly rejecting the null hypothesis when an alternative hypothesis holds.

Consider a setting in which we have two samples i.i.d with Gaussian distribution. The first sample consists of n_0 observations with population mean μ_0 and population variance σ_0^2. The second sample consists of n_1 observations with population mean μ_1 and population variance σ_1^2.

The following code checks the statistical power of an unpaired Student's t-test in a sample sizes $n_0 = n_1 = 30$, $\mu_0 = 3$, $\mu_1 = 4$, $\sigma_0 = \sigma_1 = 1$ and with a significance level of $\alpha = 0.05$.

```
num_trials<-10000
n_0<-30
n_1<-30
mu_0<-3
mu_1<-4
sigma_0<-2
sigma_1<-2
alpha<-0.05

set.seed(0) # set random seed for reproducibility

data.frame(trial=seq(num_trials))%>%
  mutate(sample_0=map(.x=trial,.f=~rnorm(n=n_0,mean=mu_0,sd=sigma_0)),
         sample_1=map(.x=trial,.f=~rnorm(n=n_1,mean=mu_1,sd=sigma_1)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap(.l=list(trial,sample_0,sample_1),
                      .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
  # generate p values
  mutate(reject_null=p_value<alpha)%>%
  pull(reject_null)%>%
  mean() # estimate of coverage probability
```

(Q) Conduct a simulation study to explore how the statistical power varies as a function of the significance level.

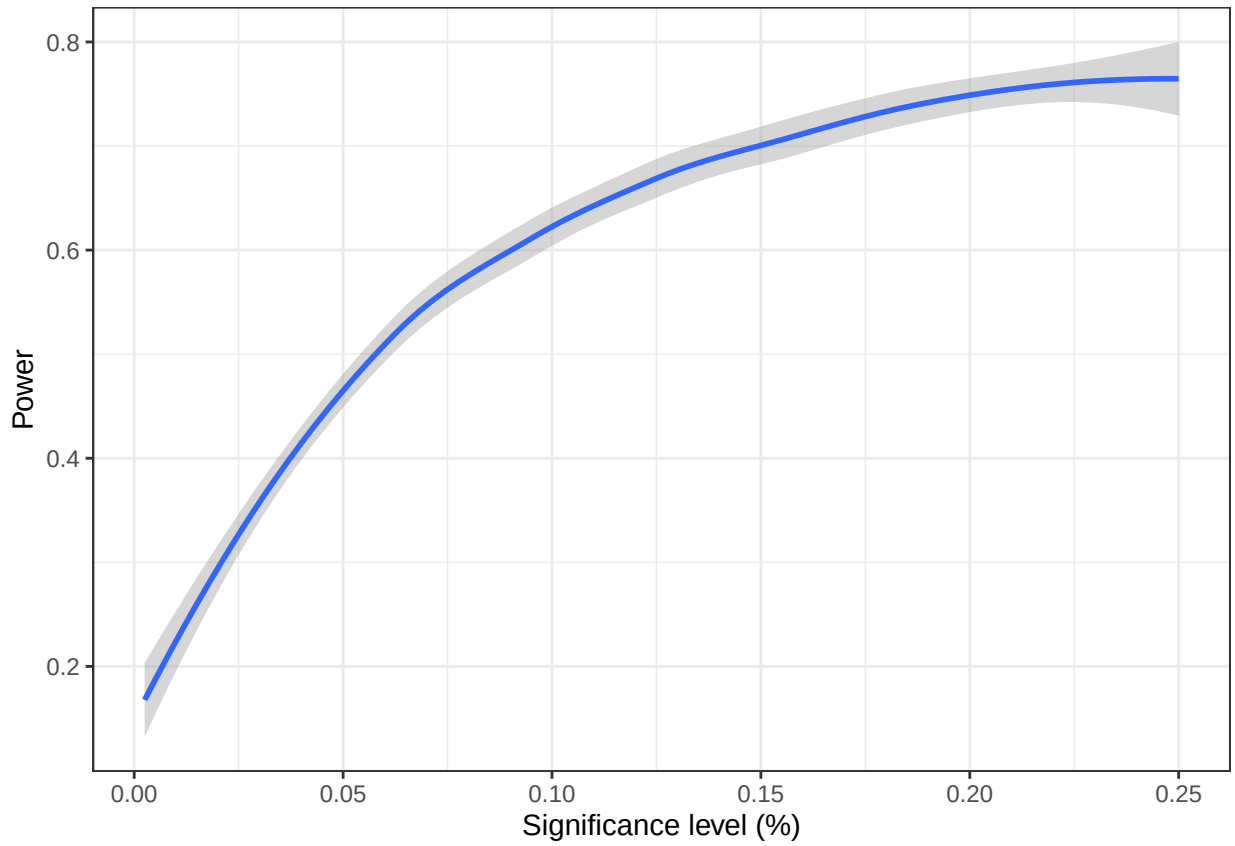
(A)

```
num_trials_per_scenario<-100
n_0<-30
n_1<-30
mu_0<-3
mu_1<-4
sigma_0<-2
sigma_1<-2

alpha_min<-0.0025
alpha_max<-0.25
alpha_inc<-0.0025

set.seed(0) # set random seed for reproducibility
```

```
crossing(trial=seq(num_trials_per_scenario),
         alpha=seq(alpha_min,alpha_max,alpha_inc))%>%
  mutate(sample_0=map(.x=trial,.f=~rnorm(n=n_0,mean=mu_0,sd=sigma_0)),
         sample_1=map(.x=trial,.f=~rnorm(n=n_1,mean=mu_1,sd=sigma_1)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap(.l=list(trial,sample_0,sample_1),
                      .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
  # generate p values
  mutate(reject_null=p_value<alpha)%>%
  group_by(alpha)%>%
  summarise(statistical_power=mean(reject_null))%>%
  ggplot(aes(x=alpha,y=statistical_power))+
  geom_smooth()+xlab("Significance level (%)")+ylab("Power")+
  theme_bw()
```



(Q) Conduct a simulation study to explore how the statistical power varies as a function of the difference in means $\mu_1 - \mu_0$.

(A)

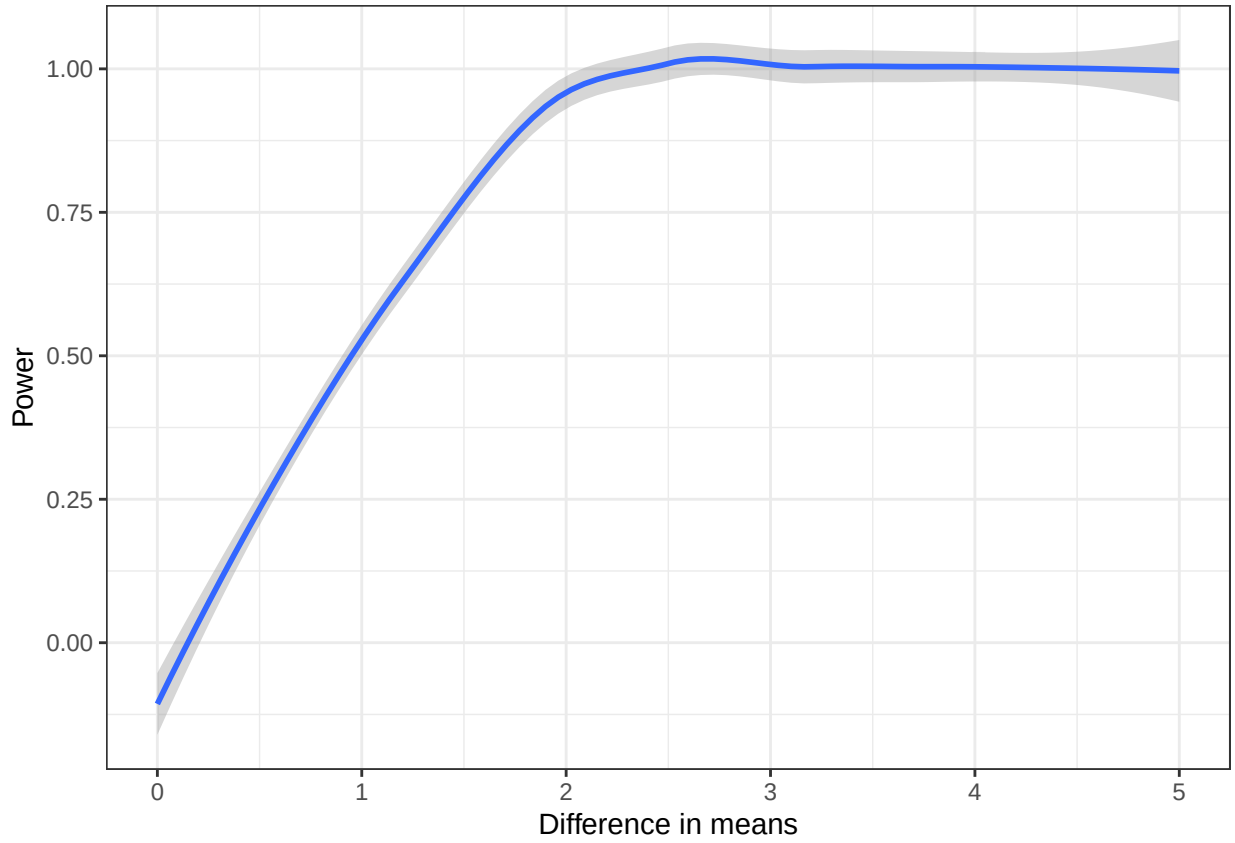
```
num_trials_per_scenario<-100
n_0<-30
n_1<-30
mu_0<-3
sigma_0<-2
```

```r
sigma_1<-2
alpha<-0.05

delta_min<-0
delta_max<-5
delta_inc<-0.1


set.seed(0) # set random seed for reproducibility

crossing(trial=seq(num_trials_per_scenario),
         delta=seq(delta_min,delta_max,delta_inc))%>%
  mutate(sample_0=map(.x=trial,.f=~rnorm(n=n_0,mean=mu_0,sd=sigma_0)),
         sample_1=map2(.x=trial,.y=delta,.f=~rnorm(n=n_1,mean=mu_0+.y,sd=sigma_1)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap(.l=list(trial,sample_0,sample_1),
                      .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
  # generate p values
  mutate(reject_null=p_value<alpha)%>%
  group_by(delta)%>%
  summarise(statistical_power=mean(reject_null))%>%
  ggplot(aes(x=delta,y=statistical_power))+
  geom_smooth()+xlab("Difference in means")+ylab("Power")+
  theme_bw()
```



(Q) Conduct a simulation study to explore how the statistical power varies as a function of the population standard deviation $\sigma = \sigma_0 = \sigma_1$.

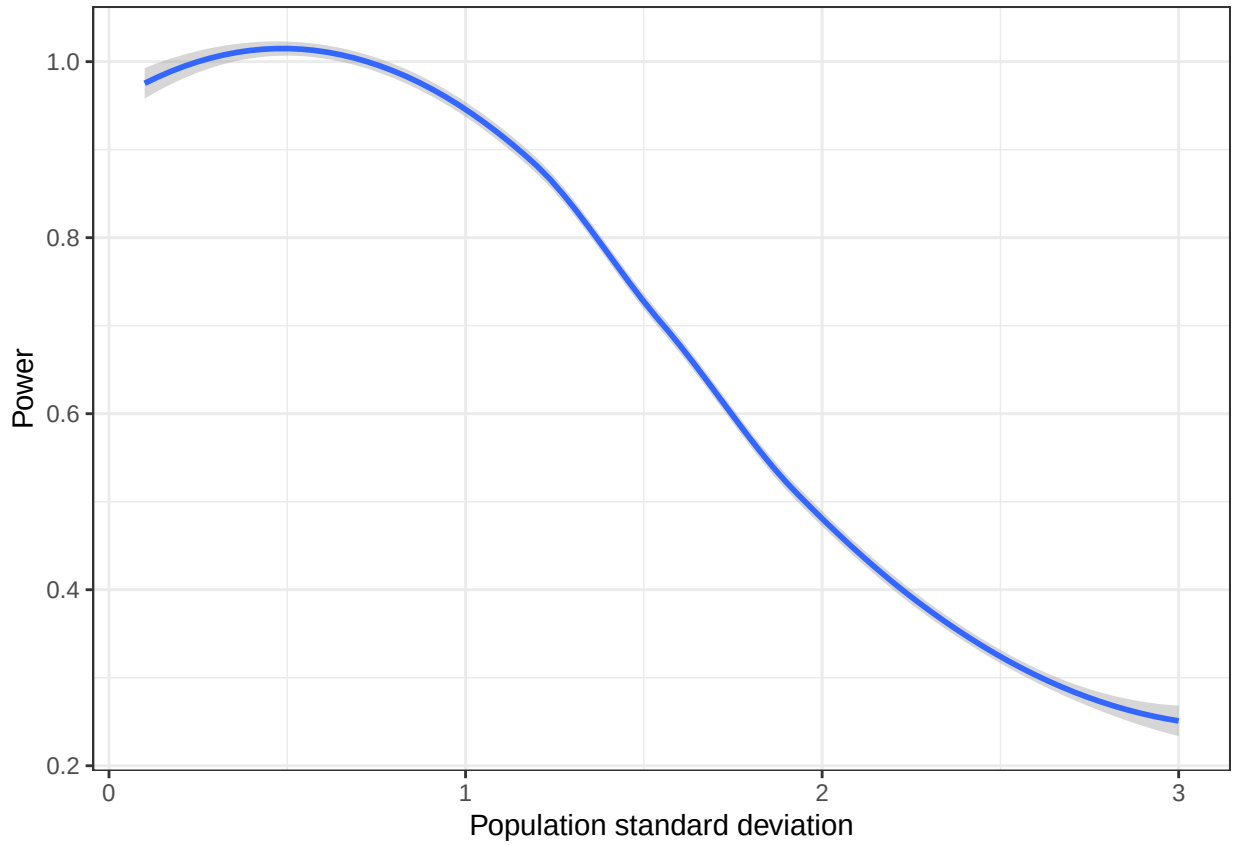
(A)

```
num_trials_per_scenario<-100
n_0<-30
n_1<-30
mu_0<-3
mu_1<-4
alpha<-0.05

sigma_min<-0.1
sigma_max<-3
sigma_inc<-0.01


set.seed(0) # set random seed for reproducibility

crossing(trial=seq(num_trials_per_scenario),
         sigma=seq(sigma_min,sigma_max,sigma_inc))%>%
  mutate(sample_0=map2(.x=trial,.y=sigma,.f=~rnorm(n=n_0,mean=mu_0,sd=.y)),
         sample_1=map2(.x=trial,.y=sigma,.f=~rnorm(n=n_1,mean=mu_1,sd=.y)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap_dbl(.l=list(trial,sample_0,sample_1),
                          .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
  # generate p values
  mutate(reject_null=p_value<alpha)%>%
  group_by(sigma)%>%
  summarise(statistical_power=mean(reject_null))%>%
  ggplot(aes(x=sigma,y=statistical_power))+
  geom_smooth()+xlab("Population standard deviation")+ylab("Power")+
  theme_bw()
```



(Q) Conduct a simulation study to explore how the statistical power varies as a function of the sample size $n = n_0 = n_1$.

(A)

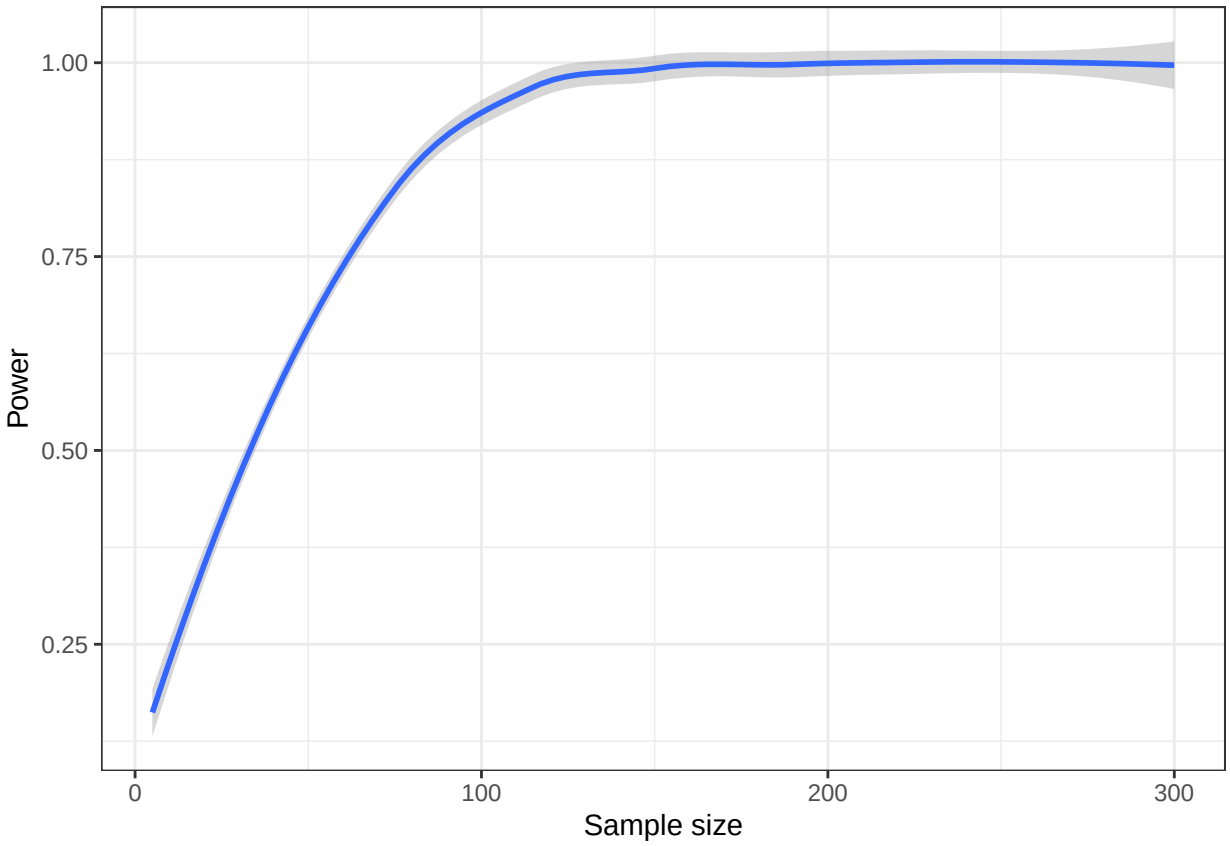
```
num_trials_per_scenario<-100
mu_0<-3
mu_1<-4
alpha<-0.05
sigma_0<-2
sigma_1<-2


n_min<-5
n_max<-300
n_inc<-5


set.seed(0) # set random seed for reproducibility

crossing(trial=seq(num_trials_per_scenario),
         sample_size=seq(n_min,n_max,n_inc))%>%
  mutate(sample_0=map2(.x=trial,.y=sample_size,.f=~rnorm(n=.y,mean=mu_0,sd=sigma_0)),
         sample_1=map2(.x=trial,.y=sample_size,.f=~rnorm(n=.y,mean=mu_1,sd=sigma_1)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap_dbl(.l=list(trial,sample_0,sample_1),
```

```
                    .f=~t.test(..2,..3,var.equal = TRUE)$p.value))%>%
# generate p values
mutate(reject_null=p_value<alpha)%>%
group_by(sample_size)%>%
summarise(statistical_power=mean(reject_null))%>%
ggplot(aes(x=sample_size,y=statistical_power))+
geom_smooth()+xlab("Sample size")+ylab("Power")+
theme_bw()
```



6 Comparing the paired and unpaired t-tests (Optional)

The aim of this question is to explore the benefits of using a paired test when a natural pairing is available. Consider a situation in which we have two i.i.d. samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$.

Suppose that $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and for each $i = 1, \ldots, n$, we have $Y_i = X_i + Z_i$ where $Z_1, \ldots, Z_n \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$ are independent and identically distributed random variables. It follows that $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ are independent and identically distributed with $\mu_Y = \mu_X + \mu_Z$ and $\sigma_Y^2 = \sigma_X^2 + \sigma_Z^2$.

In this situation we only observe the two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$. We are interested in performing a statistical hypothesis test to see if $\mu_X \neq \mu_Y$. We have two options here. We could either (1) use the pairing and apply a paired test or (2) ignore the pairing and use an unpaired test. In the console run `?t.test()` to see how to carry out an unpaired and a pared test within R.

(Q) Conduct a simulation study to explore the statistical power of these two approaches. You may want to consider a setting in which $n = 30$, $\mu_X = 10$, $\sigma_X = 5$, $\mu_Z = 1$ and $\sigma_Z = 1$. Consider a range of different significance levels α.

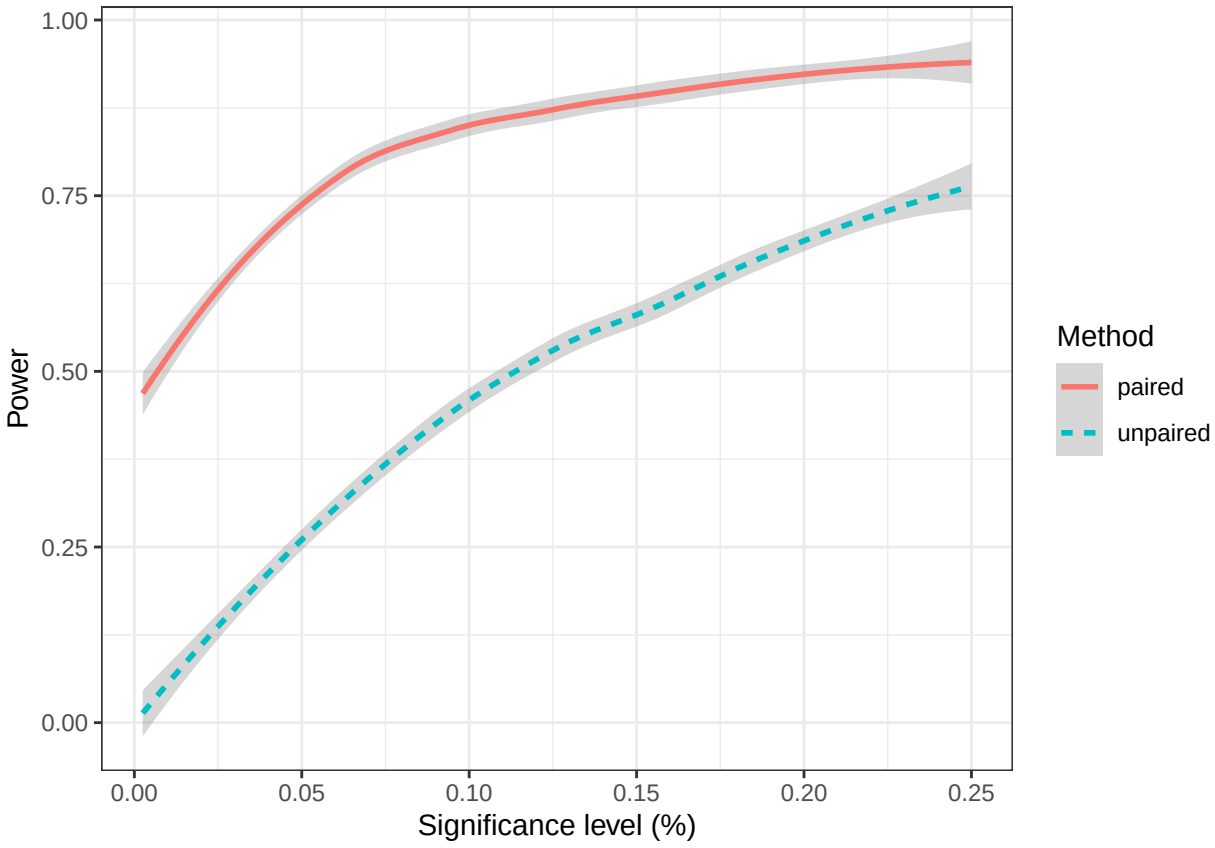
(A)

```
num_trials_per_scenario<-100
n<-30
mu_X<-3
mu_Z<-1
sigma_X<-2
sigma_Z<-2

alpha_min<-0.0025
alpha_max<-0.25
alpha_inc<-0.0025

set.seed(0) # set random seed for reproducibility

crossing(trial=seq(num_trials_per_scenario),
         alpha=seq(alpha_min,alpha_max,alpha_inc))%>%
  mutate(sample_X=map(.x=trial,.f=~rnorm(n=n,mean=mu_X,sd=sigma_X)),
         sample_Z=map(.x=trial,.f=~rnorm(n=n,mean=mu_Z,sd=sigma_Z)))%>%
  mutate(sample_Y=pmap(.l=list(sample_X,sample_Z),.f=~(..1+..2)))%>%
  # generate random Gaussian samples
  mutate(p_value_paired=pmap_dbl(.l=list(sample_X,sample_Y),
                      .f=~t.test(..1,..2,paired=TRUE)$p.value),
         p_value_unpaired=pmap_dbl(.l=list(sample_X,sample_Y),
                      .f=~t.test(..1,..2,paired=FALSE)$p.value))%>%
  # generate p values
  rename(paired=p_value_paired,unpaired=p_value_unpaired)%>%
  select(trial,alpha,paired,unpaired)%>%
  pivot_longer(cols=c(paired,unpaired),
               names_to="Method",values_to = "p_value")%>%
  mutate(reject_null=p_value<alpha)%>%
  group_by(alpha,Method)%>%
  summarise(statistical_power=mean(reject_null))%>%
  ggplot(aes(x=alpha,y=statistical_power,color=Method,linetype=Method))+
  geom_smooth()+xlab("Significance level (%)")+ylab("Power")+
  theme_bw()
```


1.00
0.75
0.50
0.25
0.00
Power
0.00
0.05
0.10
0.15
0.20
0.25
Significance level (%)
Method
paired
unpaired

7 A chi-squared test of population variance (Optional)

It is recommended that you watch lecture 20 before completing this question.

Suppose we have an i.i.d. sample $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ and a conjectured value for the population variance σ_0^2. We wish to test the null hypothesis that $\sigma^2 = \sigma_0^2$.

(Q) Implement a function called `chi_square_test_one_sample_var` which takes as input a sample `sample` and a null value for the variance `sigma_square_null`.

(A)

```
chi_square_test_one_sample_var<-function(sample,sigma_square_null){

  sample<-sample[!is.na(sample)]
  # remove any missing values
  n<-length(sample)
  # sample length
  chi_squared_statistic<-(n-1)*var(sample)/sigma_square_null
  # compute test statistic
  p_value<-2*min(pchisq(chi_squared_statistic,df=n-1),
                 1-pchisq(chi_squared_statistic,df=n-1))
  # compute the p-value

  return(p_value)
}
```

(Q) Conduct a simulation study to see how the size of the test varies as a function of the significance level.

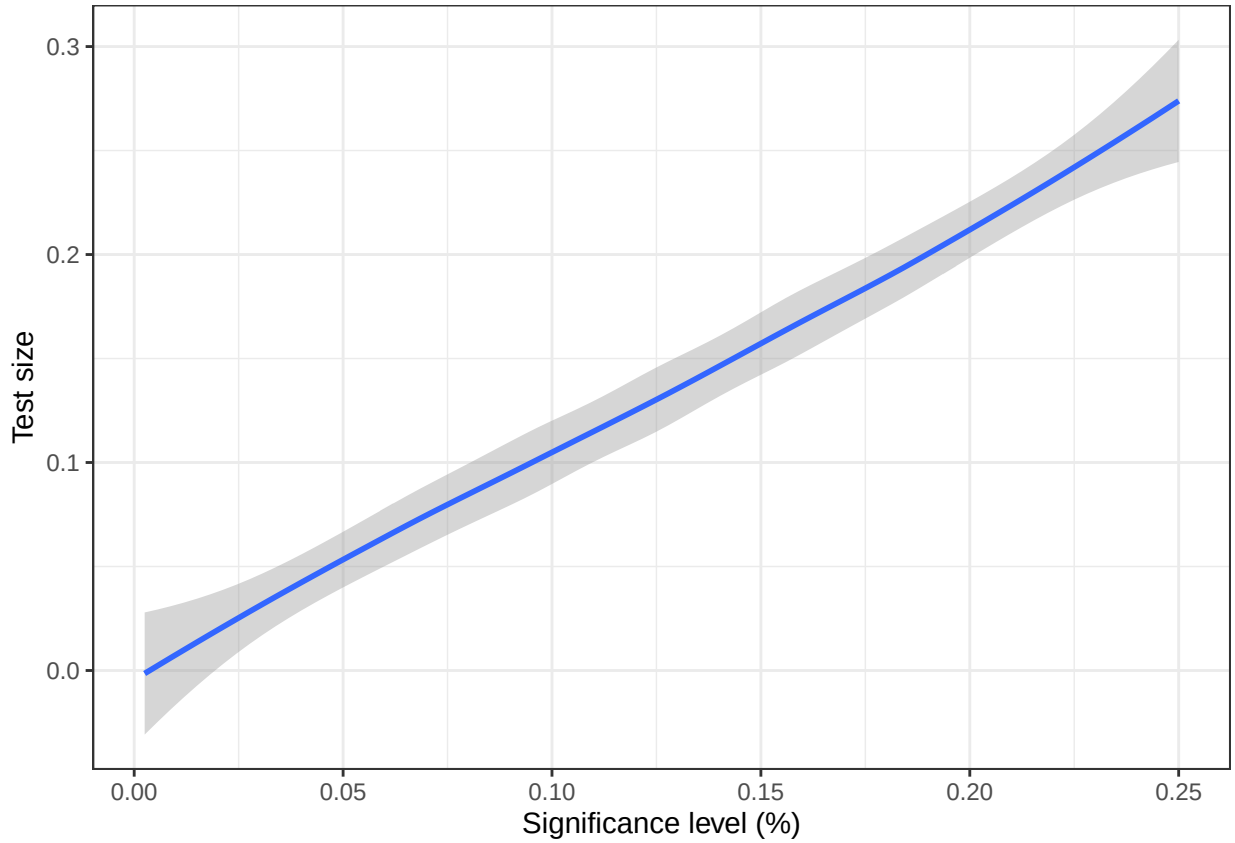
(A)

```
num_trials_per_scenario<-100
n<-30
mu<-1
sigma<-1
sigma_0<-1
alpha_min<-0.0025
alpha_max<-0.25
alpha_inc<-0.0025

set.seed(0) # set random seed for reproducibility

crossing(trial=seq(num_trials_per_scenario),
         alpha=seq(alpha_min,alpha_max,alpha_inc))%>%
  mutate(sample=map(.x=trial,.f=~rnorm(n=n,mean=mu,sd=sigma)))%>%
  # generate random Gaussian samples
  mutate(p_value=pmap(.l=list(trial,sample),
                      .f=~chi_square_test_one_sample_var(..2,sigma_0^2)))%>%
  # generate p values
  mutate(type_1_error=p_value<alpha)%>%
  group_by(alpha)%>%
  summarise(test_size=mean(type_1_error))%>%
  ggplot(aes(x=alpha,y=test_size))+
```

```
  geom_smooth()+xlab("Significance level (%)")+ylab("Test size")+
  theme_bw()
```



(Q) Load the "Palmer penguins" library and extract a vector called "bill_adelie" consisting of the bill lengths of the Adelie penguins belonging to the Adelie species.

(A)

```
library(palmerpenguins)
bill_adelie<-penguins%>%
  filter(species=="Adelie")%>%
  pull(bill_length_mm)%>%
  na.omit()
```

Suppose we model the sequence of bill lengths as a sample of independent and identically distributed Gaussian random variables $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ with population mean μ and population standard deviation σ.

(Q) Now apply your function `chi_square_test_one_sample_var` to test the null hypothesis that the population standard deviation is 3mm at a significance level of $\alpha = 10\%$.

(A)

```
chi_square_test_one_sample_var(bill_adelie,3^2)
```

```
## [1] 0.05196711
```