Understanding Data and Statistical Design (60117)

Lab 2: One factor, two level experiments

This lab is marked from 18.

Please submit in PDF format via Canvas.

Due by the conclusion of the lab class

In this week's lab we look at a range of T-tests.

Question 1 [12 marks]

In this question we analyse the effect of fertilizer on the yield of a type of grass. The variables we consider are summarised in the table below.

Name	Type	Description
yld	continuous numerical	grass yield in 100lbs per acre
$fert$	factor	fertilizer quantity: 1 (none), 2 (low)

The sample data consists of 12 observations of yld for each level of $fert$ (data available in lab2a.csv on Canvas). Not much information is available with this data set, so we will suppose that 24 plots were prepared in a homogenous manner. The 2 fertilizer concentrations were randomly assigned 12 times each across these 24 plots.

The statistical model for this experiment is

$$yld_{ij} = \mu_i + \epsilon_{ij}, \qquad i \in \{1,2\}, \qquad j \in \{1,2,\dots,12\},$$

where

- yld_{ij} is the yield from the j-th plot with $fert = i$
- μ_i is the population mean yield with $fert = i$
- ϵ_{ij} is the random effect from the j-th plot with $fert = i$.

The components of the experiment design:

- **factor** – fertiliser quantity (variable $fert$) with **levels** 1 (none) and 2 (low)
- **treatments** – same as the factor levels as there is only 1 factor
- **experimental units** – the 2 groups of 12 plots each allocated a treatment
- **measurement units** – the 24 plots used in the experiment
- **response variable** – grass yield (variable yld).

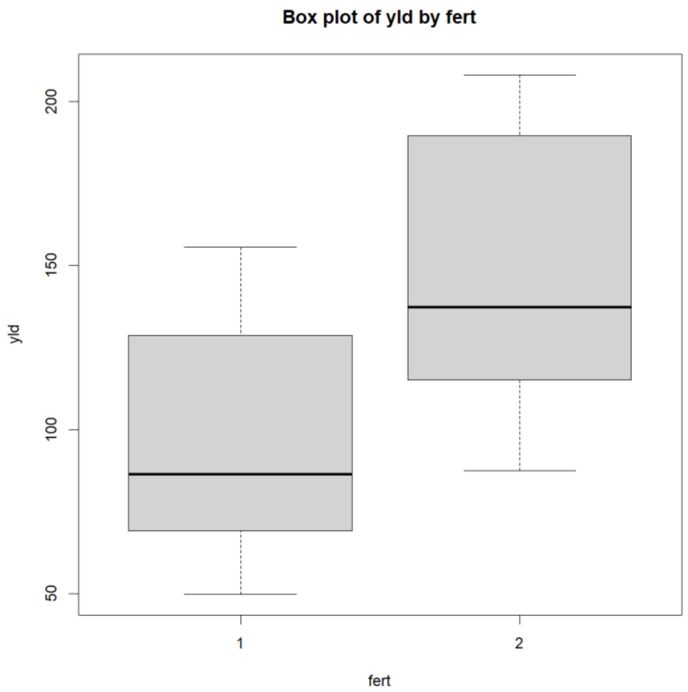
First, we compute sample means of yld for each level of $fert$ (R output copied below).

```
        1         2
96.08333 148.43333
```

Next, we compute sample standard deviations of yld for each level of $fert$ (R output copied below).

```
        1        2
35.92244 43.73857
```

(a) Construct a box plot for yld for each level of $fert$ and display in a single chart. Compare the location and scale of the two samples **[3 marks]**.



The two levels have notable differences between them, first and foremost, the level 2 median is superior compared to the level 1 median, as well as the distribution magnitude, the level 1 values start from 50 to 160 approx. while for the level 2 start from around 80 to 200 plus. Variation of both levels are similar, being the level 2 variance a bit higher.

The distributions of both levels are not symmetric, they both are skewed to the right, since both median are not in the middle, but near the q25.

One sample T-test

To begin we just consider the yld measurements where no fertilizer has been used ($fert = 1$).

(b) Using significance level $\alpha = 0.05$, perform a one sample T-test to determine if (population) mean yield when no fertilizer is used is less than 11500lbs per acre. Write down the hypotheses, the test statistic and p-value, the test decision (with reason) and conclusion (using a minimum of mathematical language) **[3 marks]**.

Hint. Remember that yld is measured in 100lbs per acre.

```
        One Sample t-test

data:  data2a[data2a$fert == 1, "yld"]
t = -1.8242, df = 11, p-value = 0.04769
alternative hypothesis: true mean is less than 115
95 percent confidence interval:
     -Inf 114.7065
sample estimates:
mean of x
 96.08333

> qt(p=0.05, df=11) #0.05 quantile from T(11)
[1] -1.795885
> pt(q=-1.8242, df=11) #recalculate p-value
[1] 0.04768916
```

Null hypothesis: the population mean is equal to 11500lbs per acre.
Alternative hypothesis: the population mean is less than 11500lbs per acre.
T statistic = -1.8242, p_value = 0.04769
Decision: given that p_value is less than alpha and the t statistic is less than the q005 of the T distribution (-1.795885), we reject the null hypothesis, and we have strong evidence that the population mean is less than 11500lbs per acre.

(c) Using significance level $\alpha = 0.05$, perform a one sample T-test to determine if (population) mean yield when no fertilizer is used is different to 8000lbs per acre. Write down the hypotheses, the test statistic and p-value, the test decision (with reason) and conclusion (using minimum of mathematical language) **[3 marks]**.

```
        One Sample t-test

data:  data2a[data2a$fert == 1, "yld"]
t = 1.551, df = 11, p-value = 0.1492
alternative hypothesis: true mean is not equal to 80
95 percent confidence interval:
  73.25931 118.90736
sample estimates:
mean of x
 96.08333


>
> qt(p=0.025, df=11) #0.025 quantile from T(11)
[1] -2.200985
> qt(p=0.975, df=11) #0.975 quantile from T(11)
[1] 2.200985
>
> pt(q=1.551, df=11) #recalculate p-value
[1] 0.9254105
```

Null hypothesis: the population mean is equal to 8000lbs per acre.
Alternative hypothesis: the population mean is different than 8000lbs per acre.
T statistic = 1.551, p_value = 0.1492
Decision: Given that p_value is greater than alpha and the t statistic is inside q0.025 and q0.975 (-2.200985, 2.200985), we don't have evidence to reject the null hypothesis. In conclusion, the population mean is not different than 8000lbs per acre.

Power analysis for one sample T-test

The R code file contains power analysis that is not assessed. The power of a hypothesis test is the probability of rejecting the null hypothesis when it is false. It is related to the probability of a Type II error (denoted β), which is the probability of retaining the null hypothesis when it is false.

The other type of error that can be made is called Type I, which is rejecting the null hypothesis when it is true. The probability of this is the significance level α.

When designing an experiment, the experimenter sets the probabilities of Type I and II errors (α and β) and then calculates the sample size necessary to achieve these.

To perform the analysis for the test just performed in (b), we are going to assume that $\mu_1 = 100$ (power analysis requires us to set such a value). R returned the following output.

```
     One-sample t test power calculation

              n = 12
          delta = 15
             sd = 35.92244
      sig.level = 0.05
          power = 0.3868396
    alternative = one.sided
```

The power of the test was

$$\text{Prob}(\text{Reject } H_0 | H_0 \text{ is false}) = 0.3868396$$

which is quite low.

This implies that the probability of a Type II error was

$$\begin{aligned} \beta = \text{Prob}(\text{Type II error}) &= \text{Prob}(\text{Retain } H_0 | H_0 \text{ is false}) \\ &= 1 - \text{Prob}(\text{Reject } H_0 | H_0 \text{ is false}) \\ &= 1 - 0.3868396 = 0.6131604 \end{aligned}$$

which is quite high.

In the second run of the analysis, we calculate the necessary sample size to obtain power of $1 - \beta = 0.8$ (a common level) in the test just performed in (b). Of course, changing the sample size would also change the test statistic.

```
     One-sample t test power calculation

              n = 36.8494
          delta = 15
             sd = 35.92244
      sig.level = 0.05
          power = 0.8
    alternative = one.sided
```

We see that a sample size of $n = 37$ (rounding up as sample size must be a whole number) would have been needed.

Two sample independent T-test

We now compare the yld measurements where no fertilizer has been used and where a low quantity of fertilizer has been used ($fert = 1$ against $fert = 2$).

(d) Using significance level $\alpha = 0.05$, perform a two sample independent upper tail T-test (assuming unequal variances) to determine if mean yield for low quantity fertiliser is more than 2000lbs per acre higher than for no fertiliser. Write down the hypotheses, the test statistic and p-value, the test decision (with reason) and conclusion (using a minimum of mathematical language) **[3 marks]**.

```
        Welch Two Sample t-test

data:  data2a[data2a$fert == 2, "yld"] and data2a[data2a$fert == 1, "yld"]
t = 1.9799, df = 21.199, p-value = 0.03042
alternative hypothesis: true difference in means is greater than 20
95 percent confidence interval:
 24.24727      Inf
sample estimates:
mean of x mean of y
148.43333  96.08333

>
> qt(p=0.95, df=21) #0.95 quantile from T(21)
[1] 1.720743
>
> pt(q=1.9799, df=21, lower.tail=FALSE) #recalculate p-value
[1] 0.03048495
```

Null hypothesis: The mean yield for low quantity fertiliser is not more than 2000lbs per acre higher than for no fertiliser.

Alternative hypothesis: The mean yield for low quantity fertiliser is more than 2000lbs per acre higher than for no fertiliser.

T statistic = 1.9799, p_value = 0.03042

Decision: Given that p_value is less than alpha, the t statistic is greater than the q095 of the T distribution (1.720743), we reject the null hypothesis, and we have strong evidence that the mean yield for low quantity fertiliser is more than 2000lbs per acre higher than for no fertiliser.

The tests just performed rely on the data samples being normally distributed, as the sample sizes are too small to rely on the sample means being normally distributed if the data itself is not normal. Next week we will look at how we can assess whether this assumption has been met.

QUESTION 2 [6 marks]

In this question we consider the time taken to perform a task under low and high noise conditions. The variables we consider are summarised in the table below.

Name	Type	Description
time	continuous numerical	time taken to perform task in seconds
noise	factor	background noise: 1 (low), 2 (high)

The sample data consists of 20 paired observations of $time$, one for each level of $noise$ (data available in lab2b.csv on Canvas). In the experiment, 20 randomly selected individuals were asked to perform a task twice: once with low background noise and again with high background noise.

The statistical model for the experiment is

$$time_{ij} = \mu_i + \epsilon_{ij}, \qquad i \in \{1,2\}, \qquad j =\in \{1,2, \dots ,20\},$$

where

- $time_{ij}$ is the time taken by the j-th individual with $noise = i$
- μ_i is the population mean time taken with $noise = i$
- ϵ_{ij} is the random effect from the j-th individual with $noise = i$.

The components of the experiment design:

- **factor** – background noise (variable $noise$) with **levels** 1 (low) and 2 (high)
- **treatments** – same as the factor levels as there is only 1 factor
- **experimental units** – the group of 20 individuals used twice
- **measurement units** – the 20 individuals used in the experiment
- **response variable** – time taken to perform task (variable $time$).

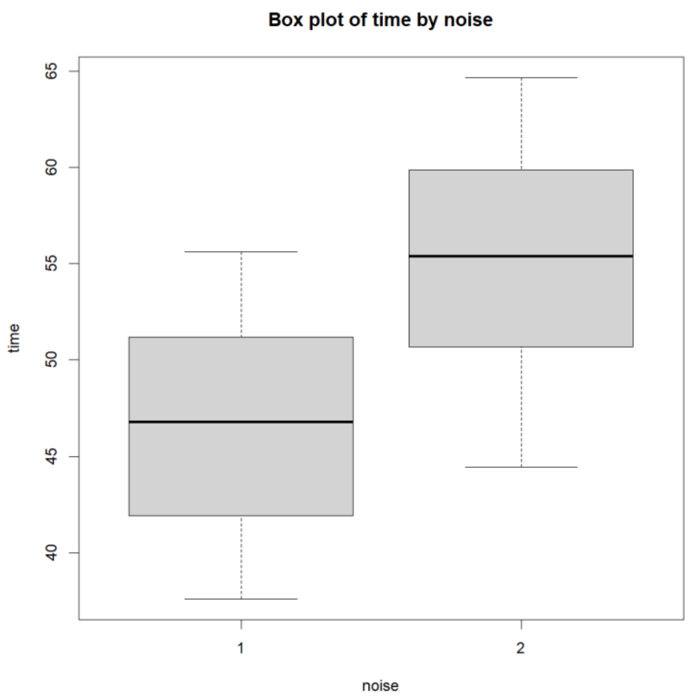
First, we compute sample means of $time$ for each level of $noise$ (R output copied below).

```
      1       2
46.4455 55.1035
```

Next, we compute sample standard deviations of $time$ for each level of $noise$ (R output copied below).

```
       1        2
5.196913 6.014407
```

(a) Construct a box plot for *time* for each level of *noise* and display in a single chart. Compare the location and scale of the two samples **[3 marks]**.



First and foremost, the scale of time for the high noise is higher than for the low noise. Both distribution seems to been symmetrical since both medians are located near the middle of the box. As for the variance, analysing the IQR of both factors, it appears that there are no significant differences between the variances.

Two sample paired T-test

(b) Using significance level $\alpha = 0.05$, perform a two sample paired upper tail T-test to determine if mean time taken for high background noise is more than 5 seconds higher than for low noise. Write down the hypotheses, the test statistic and p-value, the test decision (with reason) and conclusion (using a minimum of mathematical language) **[3 marks]**.

```
        Paired t-test

data:  data2b[data2b$noise == 2, "time"] and data2b[data2b$noise == 1, "time"]
t = 1.9707, df = 19, p-value = 0.03175
alternative hypothesis: true mean difference is greater than 5
95 percent confidence interval:
 5.448353      Inf
sample estimates:
mean difference 
          8.658 

> qt(p=0.95, df=21) #0.95 quantile from T(21)
[1] 1.720743
> 
> pt(q=1.9799, df=21, lower.tail=FALSE) #recalculate p-value
[1] 0.03048495
```

Null hypothesis: The mean time taken for high background noise is not more than 5 seconds higher than for low noise.
Alternative hypothesis: The mean time taken for high background noise is more than 5 seconds higher than for low noise.

T statistic = 1.9707, p_value = 0.03175

Decision: Given that p_value is less than alpha, the t statistic is greater than the q095 of the T distribution (1.720743), we reject the null hypothesis, and we have strong evidence that the mean time taken for high background noise is more than 5 seconds higher than for low noise.