MAST20005/MAST90058: Week 9 Lab

Goals: (i) Analysis of variance; (ii) Study the distribution of the test statistics and p-values through simulations; (iii) Compare tests by simulating power curves.

Data for Section 1: Corn data (`corn.txt`). The yield obtained, in bushels per acre, of 4 test plots for each of 4 different varieties of corn. The data file can be obtained from the shared folder in the computer labs, or from the LMS.

1 Analysis of variance (ANOVA)

Consider the corn dataset. Let μ_i be the average yield of variety i of corn, $i = 1, 2, 3, 4$. Test the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ at the 5% level of significance.

Variety	Yield			
1	68.82	76.99	74.30	78.73
2	86.84	75.69	77.87	76.18
3	90.16	78.84	80.65	83.58
4	61.58	73.51	74.57	70.75

1. Assuming the data file is in your working directory, the following commands will load the data into R, convert `Corn` to a 'factor' (categorical variable) and plot the data.

```
data  <- read.table("corn.txt", header = TRUE)
Corn  <- factor(data[, 1])
Yield <- data[, 2]
Corn
table(Corn)
Yield
tapply(Yield, list(Corn), mean)  # group means
boxplot(Yield ~ Corn)
```

2. Do an ANOVA:

```
m1 <- lm(Yield ~ Corn)
qqnorm(residuals(m1))
summary(m1)
anova(m1)
```

Hence complete the following table:

Source	df	SS	MS	F	p-value
Treatment					
Error					
Total					

3. Enter the command:

```
pairwise.t.test(Yield, Corn, pool.sd = FALSE, p.adjust.method = "none")
```

which gives the p-values for all possible paired t-tests. (Note for advanced users: this version does not adjust for multiple testing.) To confirm this table, try commands like:

```
t.test(Yield[Corn == 1], Yield[Corn == 3])
```

4. What are your conclusions? Can you interpret the coefficients in the `lm()` output?

5. What assumptions have you made? Do you think they are reasonable?

2 Simulate the distribution of test statistics and p-values

In this task we explore the null distribution of test statistics. To this end, it is useful to know how to extract various objects from the R test output. For example,

```
x <- rnorm(5)
y <- rnorm(5)
test.result <- t.test(x, y, var.equal = TRUE)
names(test.result)
```

To extract `statistic`, we use the dollar sign operator:

```
test.result$statistic
```

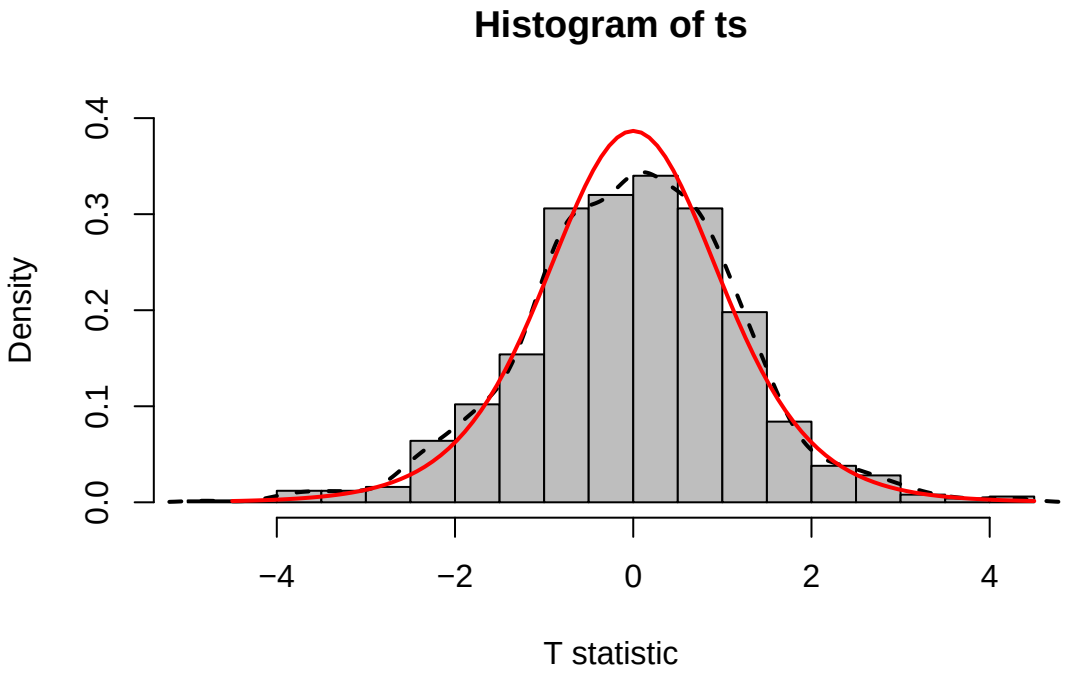
We cannot learn much by looking at a single value of the statistic. Thus, we generate many such statistics before we can look at their properties.

1. To generate 1000 t-statistics from testing two groups of 10 standard random normal numbers, we can use

```
ts <- replicate(1000,
                t.test(rnorm(5), rnorm(5), var.equal = TRUE)$statistic)
```

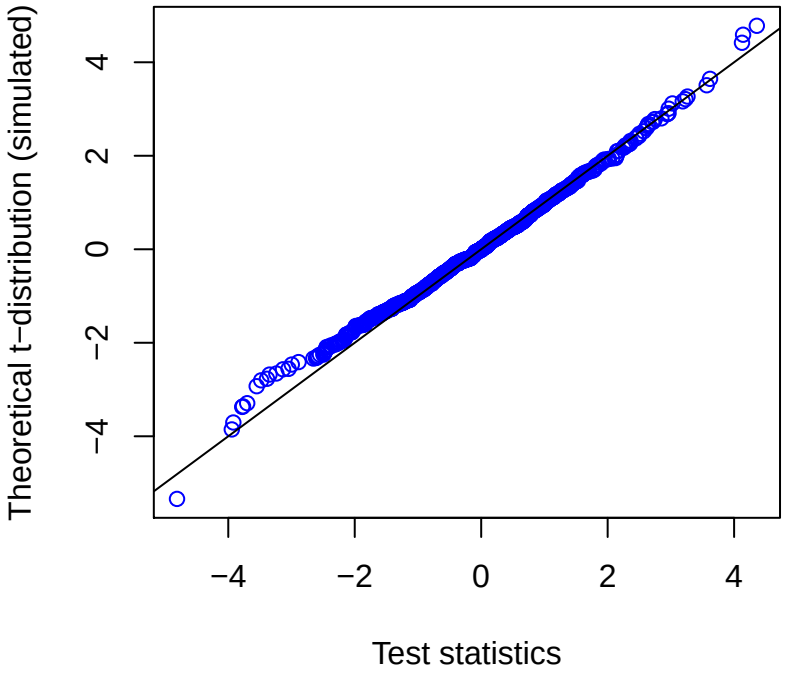
2. Under appropriate assumptions, the test statistic follows a t-distribution with 8 degrees of freedom. How can we test if that is true? One way to check this is to plot the theoretical density of the t-statistic we should be seeing, and superimposing the density of our simulated statistics on top of it.

```
hist(ts, freq = FALSE, nclass = 25, col = "grey", ylim = c(0, 0.4),
     xlab = "T statistic", ylab = "Density")  # histogram
lines(density(ts), lty = 2, lwd = 2)      # smooth density estimate
curve(dt(x, df = 8), from = -4.5, to = 4.5, add = TRUE,
      col = "red", type = "l", lwd = 2)  # theoretical density
```



3. Another way is to use a QQ plot:

```
qqplot(ts, rt(1000, df = 8), col = 4,
       xlab = "Test statistics",
       ylab = "Theoretical t-distribution (simulated)")
abline(0, 1)
```



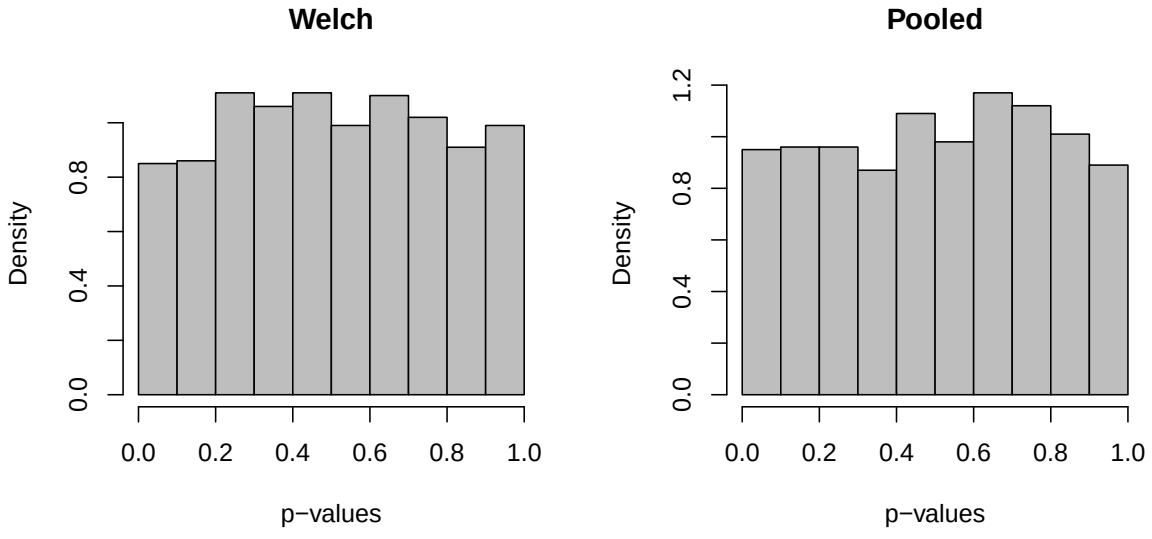
4. The central part of the graph seems to agree, but there are various discrepancies near the tails. The tails of a distribution are the most difficult part to measure. This is unfortunate, because those are often the values that interest us most since they provide us with enough evidence to reject a null hypothesis. Because the tails are so important, another way to test to see if a distribution of a sample follows some hypothesized distribution is to calculate the quantiles of some tail probabilities (using the quantile function) and compare them to the theoretical probabilities from the distribution.

```
probs <- c(0.9, 0.95, 0.99)
quantile(ts, probs)

##      90%      95%      99%
## 1.407738 1.849849 2.962914

qt(probs, df = 8)

## [1] 1.396815 1.859548 2.896459
```

The quantiles agree fairly well, especially at the 0.95 and 0.99 quantiles. Using a larger number of simulations runs (e.g. try 5000) or using a large sample size for the two groups would probably result in values closer to what we have theoretically predicted.

5. What is the behaviour of the p-values? While it may not be immediately apparent, p-values are actually random variables. Under the null hypothesis, the p-values for any statistical test should form a uniform distribution between 0 and 1 (see the Appendix). Consider Welch's t-test and the equal-variances t-test.

```
pvals.welch <- replicate(1000, t.test(rnorm(10), rnorm(10))$p.value)
pvals.ttest <- replicate(1000, t.test(rnorm(10), rnorm(10),
                                      var.equal = TRUE)$p.value)
par(mfrow = c(1, 2))
hist(pvals.welch, freq = FALSE, col = 8, xlab = "p-values", main = "Welch")
hist(pvals.ttest, freq = FALSE, col = 8, xlab = "p-values", main = "Pooled")
```



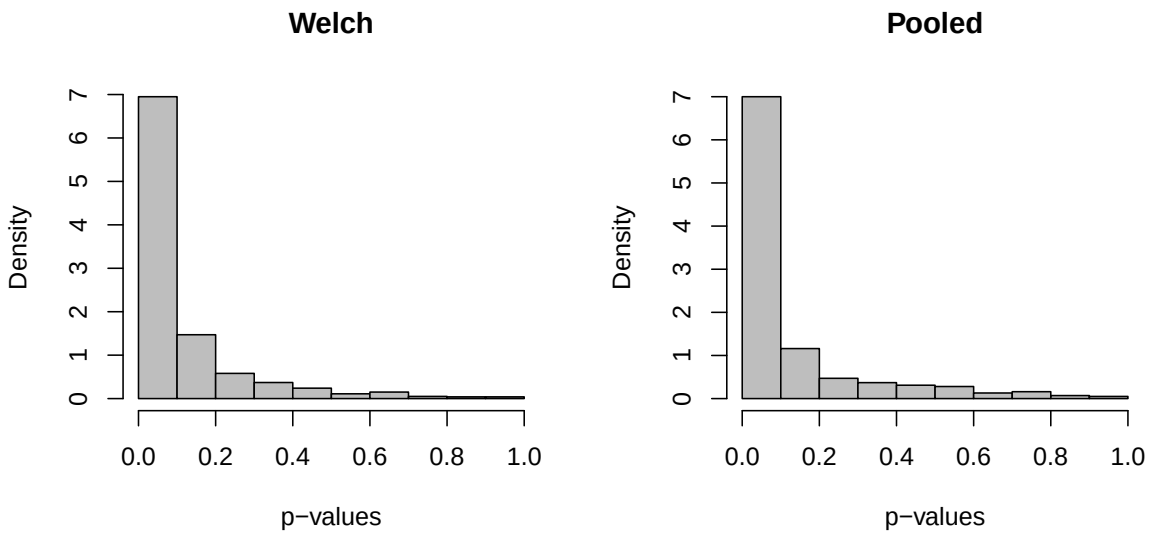
The idea that the p-values follow a uniform distribution seems reasonable. Now, let's look at some of the quantiles

```
probs <- c(0.5, 0.7, 0.9, 0.95, 0.99)
quantile(pvals.welch, probs)

##       50%       70%       90%       95%       99%
## 0.5013845 0.6961268 0.8985751 0.9497797 0.9890055

quantile(pvals.ttest, probs)

##       50%       70%       90%       95%       99%
## 0.5173447 0.7056634 0.8874657 0.9422718 0.9844300
```

There is not that much of a difference between theoretical and simulated quantiles for both tests.

6. What happens to the distribution of the p-values under the alternative hypothesis? For example, try $\mu_X - \mu_Y = 1$.

```
pvals.welch <- replicate(1000, t.test(rnorm(10), rnorm(10, 1))$p.value)
pvals.ttest <- replicate(1000, t.test(rnorm(10), rnorm(10, 1),
                                      var.equal = TRUE)$p.value)
par(mfrow = c(1, 2))
hist(pvals.welch, freq = FALSE, col = 8, xlab = "p-values", main = "Welch")
hist(pvals.ttest, freq = FALSE, col = 8, xlab = "p-values", main = "Pooled")
```



```
quantile(pvals.welch, probs)

##        50%        70%        90%        95%        99%
## 0.03692803 0.10248368 0.29723121 0.43562115 0.73201027

quantile(pvals.ttest, probs)

##        50%        70%        90%        95%        99%
## 0.03474991 0.09975027 0.39903759 0.57613280 0.81187650
```

The distribution is not uniform. Thus, the probability that the p-value is smaller than $\alpha = 0.05$ under the alternative hypothesis is higher than under the null hypothesis, and this effect is more pronounced as the true difference $\mu_X - \mu_Y$ moves away from zero.

3 Power comparisons by simulation

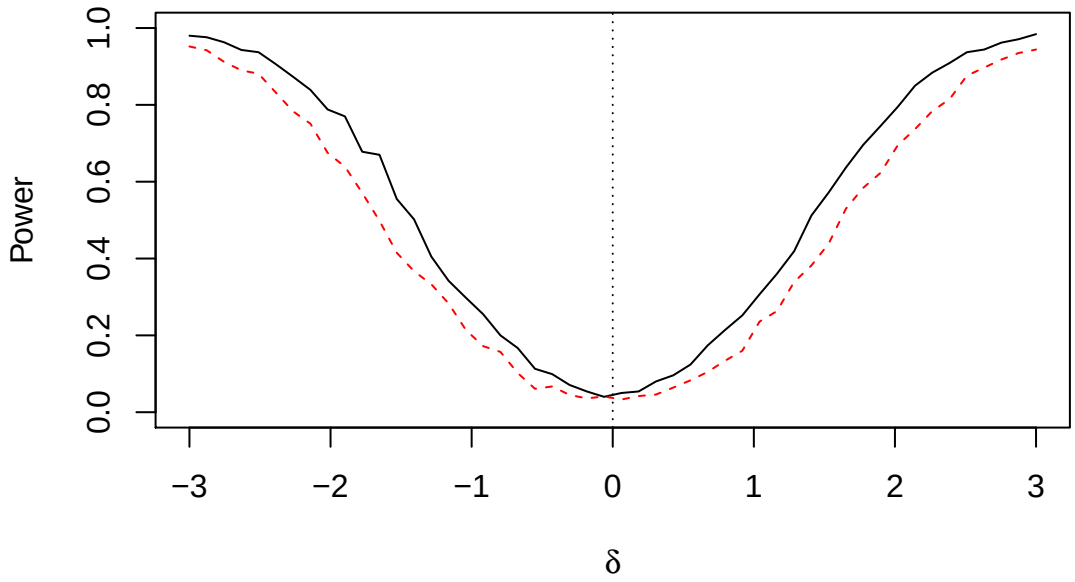
We carry out a a sequence of simulations to compare the power curves of two tests. We generate samples of size $n = n_X = n_Y = 5$ from $X \sim \mathrm{N}(\mu_X, 1)$ and $Y \sim \mathrm{N}(\mu_Y, 1)$ for various values of $\delta = \mu_X - \mu_Y$. We consider $B = 1000$ simulation runs and for each run we carry out the t-test (equal variances) and the Wilcoxon rank-sum test.

```
B <- 1000                                # number of simulation runs
R <- 50                                  # number of power values
n <- 5                                   # sample sizes
delta.seq <- seq(-3, 3, length = R)      # sequence of true differences
power.t <- numeric(R)                    # initialize power vectors
power.w <- numeric(R)

for (i in 1:R) {
  delta <- delta.seq[i]

  # Simulate B p-values for each test.
  pvals.t <- replicate(B, t.test(rnorm(n), rnorm(n, delta),
                                 var.equal = TRUE)$p.value)
  pvals.w <- replicate(B, wilcox.test(rnorm(n), rnorm(n, delta),
                                      exact = FALSE)$p.value)

  # Record the estimated power (proportion of rejections).
  power.t[i] <- mean(pvals.t < 0.05)
  power.w[i] <- mean(pvals.w < 0.05)
}
```

Now the vectors `power.t` and `power.w` contain simulated power values for a range of values of δ. This allows us to draw the (estimated) power curves:

```
# Plot simulated power for t- and Wilcoxon tests.
plot(delta.seq, power.t, type = "l", ylim = c(0, 1),
     ylab = "Power", xlab = expression(delta))
```

```r
lines(delta.seq, power.w, lty = 2, col = 2)
abline(v = 0, lty = 3)
```



This will take a few seconds since you are running 1000 simulations times 50 values for δ. The simulated power curve for the t-test is uniformly better than that of the Wilcoxon test for any $|\delta| > 0$. How could you explain this behaviour? What is the meaning of the points where power curves intersect the vertical axis in the plot?

Appendix: p-values and their distribution under H_0

The p-value, denoted here by U, can be seen as a random variable depending on the sample. Suppose we are interested in testing $H_0\colon \mu = \mu_0$ vs $H_1\colon \mu > \mu_0$ by a one-sample t-test. Then the p-value for the given observed statistic, t, is

$$u = \Pr(T \geqslant t \mid H_0) = 1 - F_0(t),$$

where F_0 denotes the cdf of T under the null hypothesis. In other words, the p-value is just a statistic (a function of the data) and therefore has a sampling distribution. In particular, the above formula shows a realisation of $U = 1 - F_0(T)$, which is the random-variable version of the p-value.

(The p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Importantly, note that the p-value does not correspond to the probability that H_0 is true given the data we actually observed. Rather, it is a probability that relates to the data under hypothetical repetitions of the experiment.)

Next, we show that the p-value, U, follows a uniform distribution on the unit interval, $[0, 1]$. We will use a trick called probability integral transform.

Thanks to the fact that the cdf $F_0(t)$ is monotonic, increasing and (left-)continuous, we can write:

$$\Pr(T \geqslant t \mid H_0) = \Pr(F_0(T) \geqslant F_0(t)) = 1 - \Pr(F_0(T) \leqslant F_0(t)).$$

This shows $\Pr(F_0(T) \leqslant F_0(t)) = F_0(t)$. But recall that the random variable X follows the uniform distribution $X \sim \text{Unif}(0,1)$ if and only if its cdf is $P(X \leqslant x) = x$. Hence, $F_0(T) \sim \text{Unif}(0,1)$ and also $U = 1 - F_0(T) \sim \text{Unif}(0,1)$.

Exercises

Note: some of the R commands you'll need to complete these exercises were covered in the lectures but not in these particular lab notes.

1. Consider the `InsectSprays` dataset, which comes provided by default with R. It shows the counts of insects in an agricultural experiment that compares different insecticides (you can read more about it on its help page by running `?InsectSprays`).

   (a) Do a one-way ANOVA to show that the experiment provides strong evidence that the insect sprays that were tested vary in their effectiveness.

   (b) Inspect the data with a boxplot. Is there an assumption we make in the ANOVA that you suspect might not hold?

   (c) A square root transformation is often applied to count data to make them more compatible with the ANOVA assumptions. Take the square root of the counts and repeat the ANOVA. Have your conclusions changed?

2. Consider the `ToothGrowth` dataset, also provided with R, which describes an experiment to measure the effect of vitamin C on tooth growth (see the help page for more info).

   (a) Do a two-way ANOVA (do not include interaction terms) to assess the impact of vitamin C dose and delivery method on the length of the odontoblasts.

   (b) Is there evidence for an interaction? Do an ANOVA with interaction terms and assess the output.

   (c) Draw an interaction plot for these data and interpret it.

3. Consider the same scenario as in Section 3 but this time with $n_X = 5$ and $n_Y = 10$. Use simulations to compare the power of Welch's t-test and the equal-variances t-test under the following conditions:

   (a) The variances of both groups are equal to 1.

   (b) The variance of the X group is 1 and the Y group is 4.

   What do you notice?