HW7 - Lucas Fellmeth, Sven Bergmann

2023-12-14

Chapter 15

Problem 3

Refer to insect waiting times for the female Western White Clematis in Table 10.1. Use the percentile method to find a 90 confidence interval for $F(30)$, the probability that the waiting time is less than or equal to 30 minutes.

Table 10.1

Waiting times for insects to visit flowers.

Male flowers			Female flowers		
1	9	27	1	19	57
1	9	27	2	23	59
2	9	30	4	23	67
2	11	31	4	26	71
4	11	35	5	28	75
4	14	36	6	29	75
5	14	40	7	29	78
5	14	43	7	29	81
6	16	54	8	30	90
6	16	61	8	32	94
6	17	68	8	35	96
7	17	69	9	35	96
7	18	70	14	37	100
8	19	83	15	39	102
8	19	95	18	43	105
8	19	102	18	56	
		104			

```r
library(survival)
female <- c(1, 2, 4, 4, 5, 6, 7, 7, 8, 8, 8, 9, 14, 15, 18, 18, 19, 23, 23, 26, 28,
    29, 29, 29, 30, 32, 35, 35, 37, 39, 43, 56, 57, 59, 67, 71, 75, 75, 78, 81, 90,
    94, 96, 96, 100, 102, 105)
female.event <- c(rep(1, 32), 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)

T_i <- replicate(10000, {
    idx_boot <- sample(seq_along(female), size = length(female), replace = T)
    x_boot <- female[idx_boot]
    female.event.boot <- female.event[idx_boot]
```

```
    kmest <- (1 - survfit(Surv(x_boot, event = female.event.boot, type = "right") ~
        1, type = "kaplan-meier")$surv)
    kmest[sum(30 >= x_boot)]
})
quantile(T_i, probs = c(0.05, 0.95), na.rm = T)
```

```
##        5%        95%
## 0.6595745 0.9106383
```

Problem 7

Seven patients each underwent three different methods of kidney dialysis. The following values were obtained for weight change in kilograms between dialysis sessions:

Patient	Treatment 1	Treatment 2	Treatment 3
1	2.90	2.97	2.67
2	2.56	2.45	2.62
3	2.88	2.76	1.84
4	2.73	2.20	2.33
5	2.50	2.16	1.27
6	3.18	2.89	2.39
7	2.83	2.87	2.39

Test the null hypothesis that there is no difference in mean weight change among treatments. Use properly designed permutation test.

```
perm_pair <- function(vector_1, vector_2, reps = 10000) {
    T_obs <- mean(vector_1) - mean(vector_2)
    T_perm <- replicate(reps, {
        sample_perm <- sample(c(vector_1, vector_2), size = length(vector_1) + length(vector_2),
            replace = F)
        vector_2_perm <- sample_perm[(length(vector_1) + 1):(length(vector_1) + length(vector_2))]
        vector_1_perm <- sample_perm[seq_along(vector_1)]
        mean(vector_1_perm) - mean(vector_2_perm)
    })
    return(list(T_obs = T_obs, T_perm = T_perm, p_value = mean(T_perm > T_obs)))
}
```

```
treatment_1 <- c(2.9, 2.56, 2.88, 2.73, 2.5, 3.18, 2.83)
treatment_2 <- c(2.97, 2.45, 2.76, 2.2, 2.16, 2.89, 2.87)
treatment_3 <- c(2.67, 2.62, 1.84, 2.33, 1.27, 2.39, 2.39)
```

```
plot(seq_along(treatment_1), treatment_1, type = "l", col = "red", ylim = c(1, 3.5),
    xlab = "patient number", ylab = "weight change")
abline(h = mean(treatment_1), col = "red", lty = 2)
lines(treatment_2, col = "blue")
abline(h = mean(treatment_2), col = "blue", lty = 2)
lines(treatment_3, col = "green")
abline(h = mean(treatment_3), col = "green", lty = 2)
legend("bottomleft", legend = c("Treatment 1", "Treatment 1 mean", "Treatment 2",
```

```
    "Treatment 2 mean", "Treatment 3", "Treatment 3 mean"), col = c("red", "red",
    "blue", "blue", "green", "green"), lty = 1:2, cex = 0.6)
```
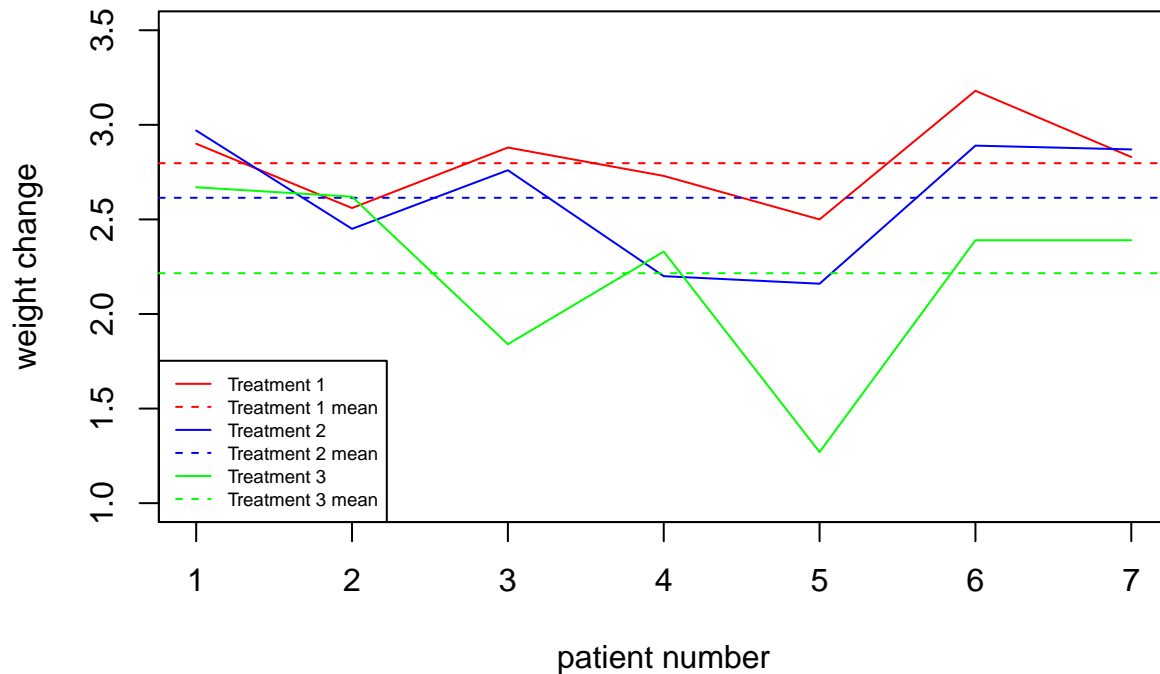


```
p_value_1_2 <- perm_pair(treatment_1, treatment_2)$p_value
p_value_1_3 <- perm_pair(treatment_1, treatment_3)$p_value
p_value_2_3 <- perm_pair(treatment_2, treatment_3)$p_value

cat("p-value for perm test between treatment_1 and treatment_2:", p_value_1_2, "\n",
    "p-value for perm test between treatment_1 and treatment_3:", p_value_1_3, "\n",
    "p-value for perm test between treatment_2 and treatment_3:", p_value_2_3, "\n")
```

```
## p-value for perm test between treatment_1 and treatment_2: 0.1335
##  p-value for perm test between treatment_1 and treatment_3: 0.0021
##  p-value for perm test between treatment_2 and treatment_3: 0.0505
```

```
p_values <- c(p_value_1_2, p_value_2_3, p_value_1_3)
threshhold <- 0.05

if (any(p_values < threshhold)) {
    cat("Since at least one of the $p$-values is smaller than", threshhold, ", we reject $H_0$.")
} else {
    cat("Since all $p$-values are greater or equal to", threshhold, ", we do not reject $H_0$.")
}
```

```
## Since at least one of the $p$-values is smaller than 0.05 , we reject $H_0$.
```

Problem 10

Return to Example 7.5 in which tread wear for tires is measured using weight loss and groove wear. Construct
a randomization test to compute the significance of the Pearson correlation coefficient. How does this value
compare with the p-value in the R function `cor.test`?

Example 7.5

Stichler, Richey, and Mandel (1953) list tread wear for tires (see table below), each tire measured by two methods based on (i) weight loss and (ii) groove wear. In R, the function `cor(x,y,method="spearman")` computes the Spearman coefficient. For this example, $\hat{\rho} = 0.9265$. Note that if we opt for the parametric measure of correlation, the Pearson coefficient is 0.948:

Weight	Groove	Weight	Groove
45.9	35.7	41.9	39.2
37.5	31.1	33.4	28.1
31.0	24.0	30.5	28.7
30.9	25.9	31.9	23.3
30.4	23.1	27.3	23.7
20.4	20.9	24.5	16.1
20.9	19.9	18.9	15.2
13.7	11.5	11.4	11.2

```r
weight <- c(45.9, 37.5, 31, 30.9, 30.4, 20.4, 20.9, 13.7, 41.9, 33.4, 30.5, 31.9,
    27.3, 24.5, 18.9, 11.4)

groove <- c(35.7, 31.1, 24, 25.9, 23.1, 20.9, 19.9, 11.5, 39.2, 28.1, 28.7, 23.3,
    23.7, 16.1, 15.2, 11.2)

T_obs <- cor(weight, groove)

X <- weight
T_i <- replicate(10000, {
    Y <- sample(groove, length(groove), replace = FALSE)
    cor(X, Y)
})

p_value <- mean(T_i > T_obs)

p_value_of_r_test <- cor.test(x = weight, y = groove)$p.value

cat("Difference between our p_value, which is:", p_value, "\n")
```

```
## Difference between our p_value, which is: 0
```

```r
cat("and the p_value from the test in R, which is:", p_value_of_r_test, "\n")
```

```
## and the p_value from the test in R, which is: 2.459998e-08
```

```r
cat("is:", abs(p_value - p_value_of_r_test))
```

```
## is: 2.459998e-08
```