Inference about Multivariate Mean and Multiple Testing

Bartosz Makaruś

2025-04-12

Problem 1

Let's consider a sequence of independent variables $X_1, X_2 \ldots$, each following the standard normal distribution. Using it, we can define a chi-squared distributed variable as

$$\chi^2 = X_1^2 + \ldots X_p^2,$$

where p, number of degrees of freedom, is a parameter of the distribution (we denote $\chi^2 \sim \chi^2_p$). Now say we are given two chi-squared distributed variables $U \sim \chi^2_{d_1}, \ V \sim \chi^2_{d_2}$. Then a variable

$$F = \frac{\frac{U}{d_1}}{\frac{V}{d_2}},$$

follows the F distribution (parametrised by d_1, d_2 - degrees of freedom, $F \sim F(d_1, d_2)$).
Let us focus on the distribution $F(p, n-p)$ which we'll encounter later. Say that $p = 4$ and $n = 1000$. Since n is large, we might want to approximate the true distribution with an asymptotic case.
Recall that the variable under consideration is of the form $\frac{U/p}{V/(n-p)}$, where $U \sim \chi^2_p, \ V \sim \chi^2_{n-p}$. Under the law of large numbers, the denominator converges to the expectation of a χ^2_1 variable, which is known to be 1. From there, using Slutsky's theorem, we gather that

$$F = \frac{U \,/\, p}{V \,/\, (n-p)} \xrightarrow[n \to \infty]{D} \frac{U}{p}.$$

Consider the statistic $n(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is sample mean vector from sample $X_1, \ldots, X_n \sim N_p(\mu, \Sigma)$. Since the covariance matrix Σ is positive semidefinite (so there exists a square root $\Sigma^{\frac{1}{2}}$) we could write

$$n(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu) = n(\bar{X} - \mu)^T (\Sigma^{\frac{1}{2}})^{-1} (\Sigma^{\frac{1}{2}})^{-1} (\bar{X} - \mu),$$

Σ is symmetric, therefore its square root is symmetric, and then the square root's inverse is also symmetric, so the statistic can be rewritten as

$$n((\Sigma^{\frac{1}{2}})^{-1} (\bar{X} - \mu))^T (\Sigma^{\frac{1}{2}})^{-1} (\bar{X} - \mu) = \sqrt{n}((\Sigma^{\frac{1}{2}})^{-1} (\bar{X} - \mu))^T \sqrt{n}(\Sigma^{\frac{1}{2}})^{-1} (\bar{X} - \mu),$$

which is a dot product of a transformed normal random vector - $\bar{X}$ follows distribution $N_p(\mu, \frac{\Sigma}{n})$, so $\bar{X} - \mu \sim N_p(0, \frac{\Sigma}{n})$, and finally $\sqrt{n}(\Sigma^{\frac{1}{2}})^{-1}(\bar{X} - \mu) \sim N_p(0, I)$. That means the dot product is a sum of squares of independent standard normal variables, i.e. the statistic is chi-squared distributed.

When Σ is unknown, to verify hypothesis $H_0 : \mu = \mu_0$ we will operate the T^2 statistic: $n(\bar{X} - \mu_0)^T S^{-1} (\bar{X} - \mu_0)$. When properly scaled, T^2 follows the F distribution if we assume the null hypothesis:

$$\frac{n(n-p)}{(n-1)p} (\bar{X} - \mu_0)^T S^{-1} (\bar{X} - \mu_0) \sim F(p, n-p),$$

which allows us to write down a test at α significance level. We reject H_0 when

$$T^2 > \frac{(n-1)p}{n-p} F_{p,n-p}(1-\alpha),$$

where $F_{p,n-p}(1-\alpha)$ is a $1-\alpha$ level quantile.
As n approaches ∞, the right side is approximately $\chi^2_p(1-\alpha)$. Still under the assumption of null hypothesis, the left side is asymptotically χ^2_p distributed. It means that the probability of rejecting H_0 converges to α. If we assume the alternative is true - $\mu \neq \mu_0$, then $T^2 = n(\bar{X}-\mu_0)^T S^{-1}(\bar{X}-\mu_0)$ diverges to ∞ and so the probability of rejecting the null approches 1 (the test is consistent).

Problem 2

Now we consider the multiple testing problem for the mean of $X \sim N_{10}(\mu, I)$. The null hypotheses are is $H_{0i} : \mu_i = 0$, the alternatives are two-tailed. We set $\alpha = 0.05$. We are given a realisation $x = (1.7, 1.6, 3.3, 2.7, -0.04, 0.35, -0.5, 1.0, 0.7, 0.8)$. Let's apply two multiple testing procedures: Bonferroni and Benjamini-Hochberg. Values x correspond to p-values

$$p = (p_1, \ldots, p_{10}) = (0.0891, 0.1096, 0.001, 0.0069, 0.9681, 0.7263, 0.6171, 0.3173, 0.4839, 0.4237)$$

in standard z-tests.
The Bonferroni correction means we compare the p-values not with α, but with $\frac{\alpha}{n} = 0.005$. This way we reject only the third null.

Table 1: Bonferroni correction results

p-values	0.0891	0.1096	0.001	0.0069	0.9681	0.7263	0.6171	0.3173	0.4839	0.4237
threshold	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
corresponding H_0	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	H_{06}	H_{07}	H_{08}	H_{09}	H_{010}

In Benjamini-Hochberg procedure, we first sort the p-values,

$$p_s = (p_{(1)}, \ldots, p_{(10)}) = (0.001, 0.0069, 0.0891, 0.1096, 0.3173, 0.4237, 0.4839, 0.6171, 0.7263, 0.9681)$$

then we find the largest i such that $p_{(i)} < \frac{i\alpha}{10}$, and ultimately reject the hypotheses corresponding to all the smaller p-values. In our case the procedure rejects nulls H_{03} and H_{04}.

Table 2: Benjamini-Hochberg procedure results

sorted p-values	0.001	0.0069	0.0891	0.1096	0.3173	0.4237	0.4839	0.6171	0.7263	0.9681
threshold	0.005	0.01	0.015	0.02	0.025	0.03	0.035	0.04	0.045	0.05
corresponding H_0	H_{03}	H_{04}	H_{01}	H_{02}	H_{08}	H_{010}	H_{09}	H_{07}	H_{06}	H_{05}

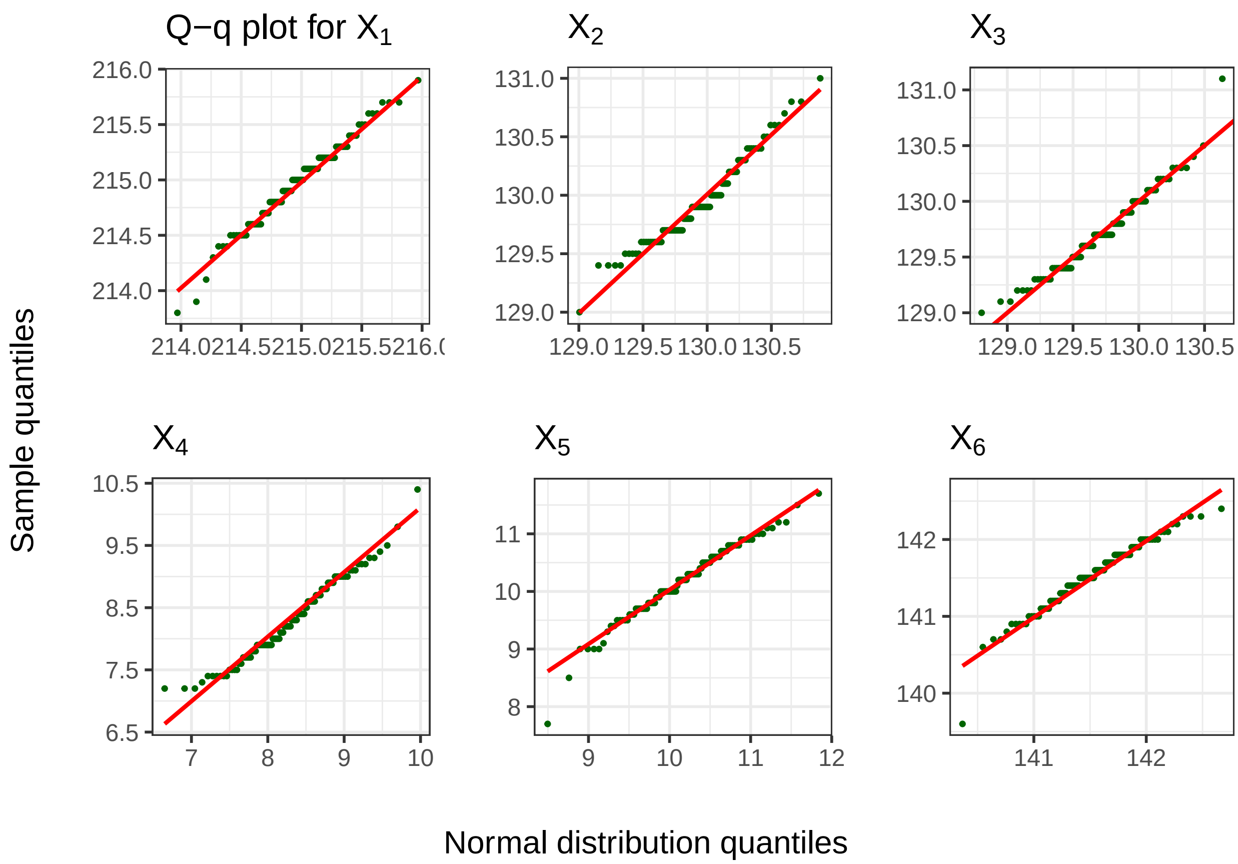
Now assuming that the alternative is true for the first 3 components of X, we see that the False Discovery Proportion, defined as $FDP = \frac{\#\{\text{false discoveries}\}}{\#\{\text{total discoveries}\}}$, is 0 for the Bonferroni procedure and 0.5 for Benjamini-Hochberg.

Project 2

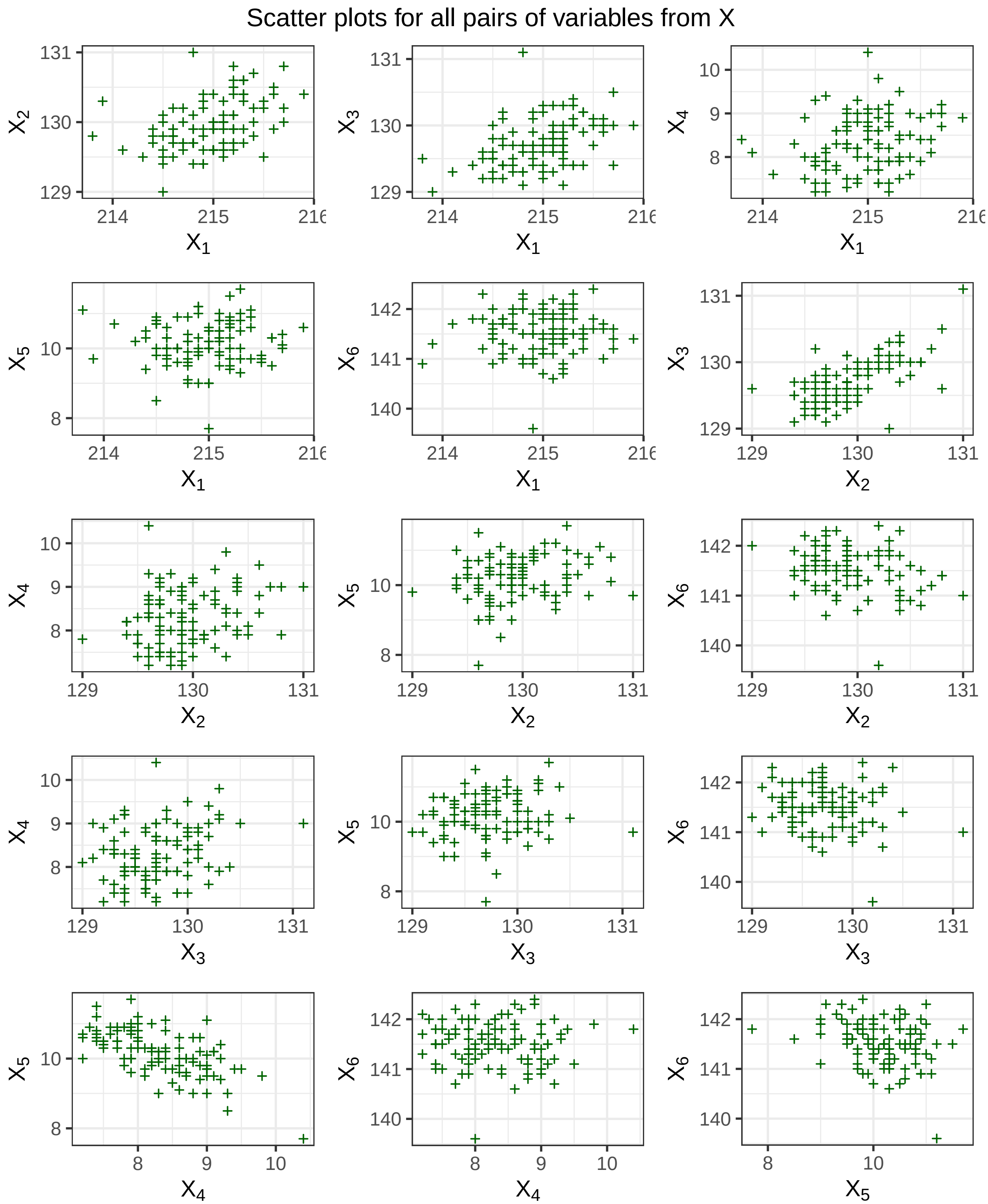
We will consider the dataset `BankGenuine`, which contains 100 observations of six characteistics of bank notes. The variables are

- X_1 - length of the note
- X_2 - height on the left side
- X_3 - height on the left side
- X_4 - distance from inner frame to the top
- X_5 - distance from inner frame to the bottom
- X_6 - length of the diagonal

First, we want to check for any reasons to discard the assumption that $X = (X_1, \ldots, X_6)^T$ follows multidimensional normal distribution. Let us draw quantile-quantile plots for the components of X.



The sample quantiles fit the normal distribution quantiles rather well for all $X_1, \ldots, X_6$, so we can assume the marginal distributions are normal. For pairs of variables, we take a look at scatter plots of observations:



The clouds of points look in line with how points from bivariate normal distributions would look - they are concentrated in ellipse-like shapes centred on the mean.

We now work with the assumption that $X = (X_1, \ldots, X_6)^T$ has distribution $N_6(\mu, \Sigma)$, where we estimate the unknown parameters by a vector of sample means $\bar{x}$ and a sample covariance matrix S:

$$\bar{x} = \begin{bmatrix} 214.969 \\ 129.943 \\ 129.72 \\ 8.305 \\ 10.168 \\ 141.517 \end{bmatrix}, \quad S = \begin{bmatrix} 0.1502 & 0.058 & 0.0573 & 0.0571 & 0.0145 & 0.0055 \\ 0.058 & 0.1326 & 0.0859 & 0.0567 & 0.0491 & -0.0431 \\ 0.0573 & 0.0859 & 0.1263 & 0.0582 & 0.0306 & -0.0238 \\ 0.0571 & 0.0567 & 0.0582 & 0.4132 & -0.2635 & -0.0002 \\ 0.0145 & 0.0491 & 0.0306 & -0.2635 & 0.4212 & -0.0753 \\ 0.0055 & -0.0431 & -0.0238 & -0.0002 & -0.0753 & 0.1998 \end{bmatrix}$$

Let's consider confidence regions (ellipsoids) for the parameter μ.
We know the statistic $\frac{(n-1)p}{n-p}T^2$, where $T^2 = n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu)$, follows the distribution $F(p, n-p)$. Hence, the confidence region for μ on significance level α is given by

$$n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) < \frac{n-p}{(n-1)p} F_{p,\, n-p}(\alpha),$$

where $F_{p,\, n-p}(\alpha)$ is a $1-\alpha$ quantile.

Now let's say we receive a new batch of bank notes and we wish to check whether their dimensions measure up to previous standards, i.e. if the mean of the new sample falls into the 95% confidence region given by T^2 (of course we use estimates and n, p from previous sample).
The new mean vector is `214.97 130 129.67 8.3 10.16 141.52` and upon checking the inequality we find that the mean falls outside the confidence interval.

A different way to construct a confidence region for the mean vector, is to calculate confidence interval for each component separately, ignoring correlation between the variables. For each X_i, $i = 1, \ldots, 6$, the interval is

$$\bar{X}_i - t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{S_i^2}{n}} < \mu_i < \bar{X}_i + t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{S_i^2}{n}}.$$

Of course the vector falls in the confidence region iff each of the variables in it falls in respective intervals. We check that the new mean vector (`214.97 130 129.67 8.3 10.16 141.52`) falls into the confidence region.

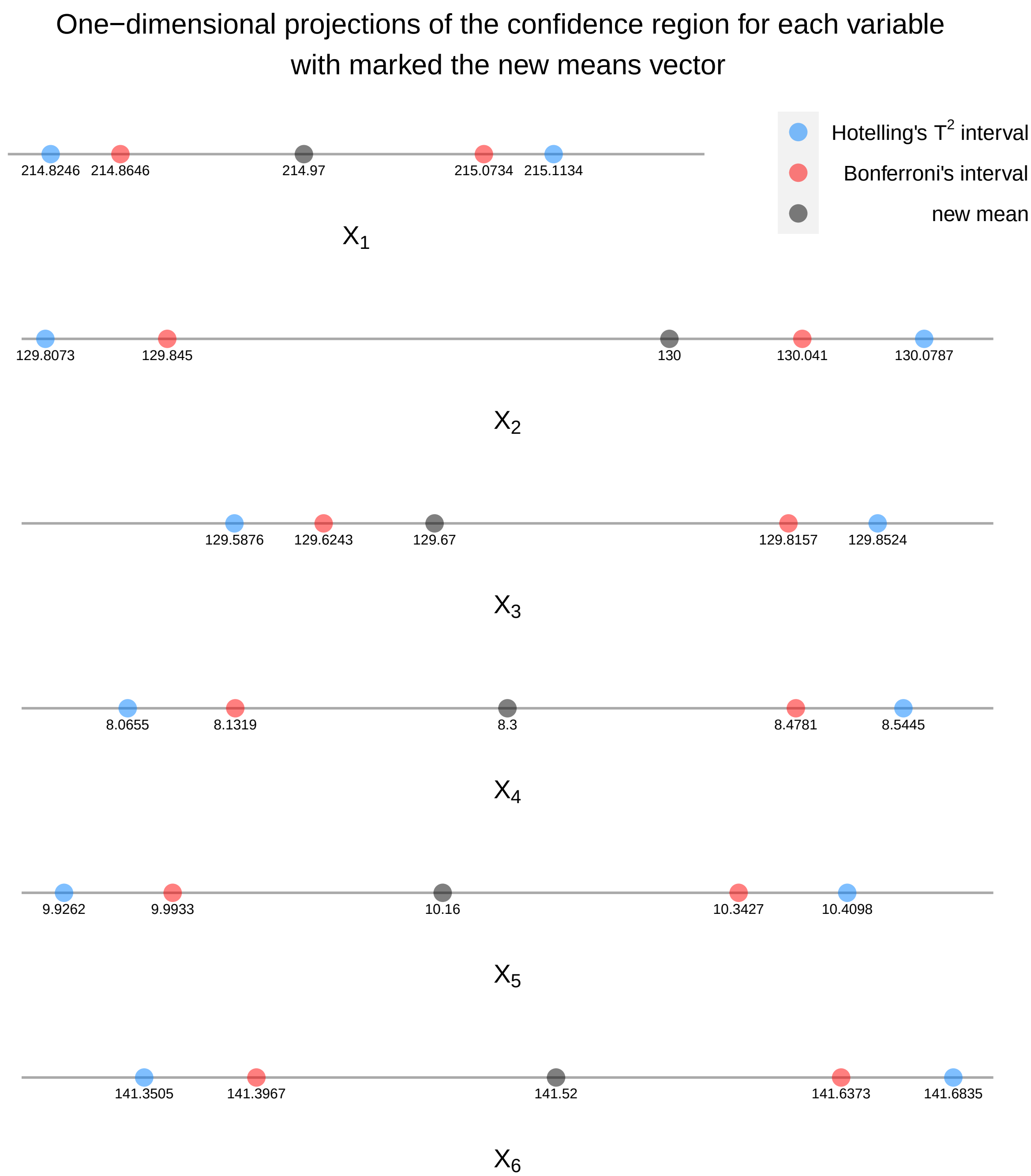
Table 3: Confidence intervals for $X_1, \ldots, X_6$ and new mean values

	X_1	X_2	X_3	X_4	X_5	X_6
Lower bound	214.8646	129.845	129.6243	8.131935	9.993272	141.3967
New mean	214.9700	130.000	129.6700	8.300000	10.160000	141.5200
Upper bound	215.0734	130.041	129.8157	8.478065	10.342728	141.6373

We can plot one-dimensional projections of both types of confidence regions on each axis (variable). In the case of Bonferroni region, these are simply the confidence intervals based on the t statistic. Regarding the Hotelling's T^2 confidence region, from maximisation lemma we get that

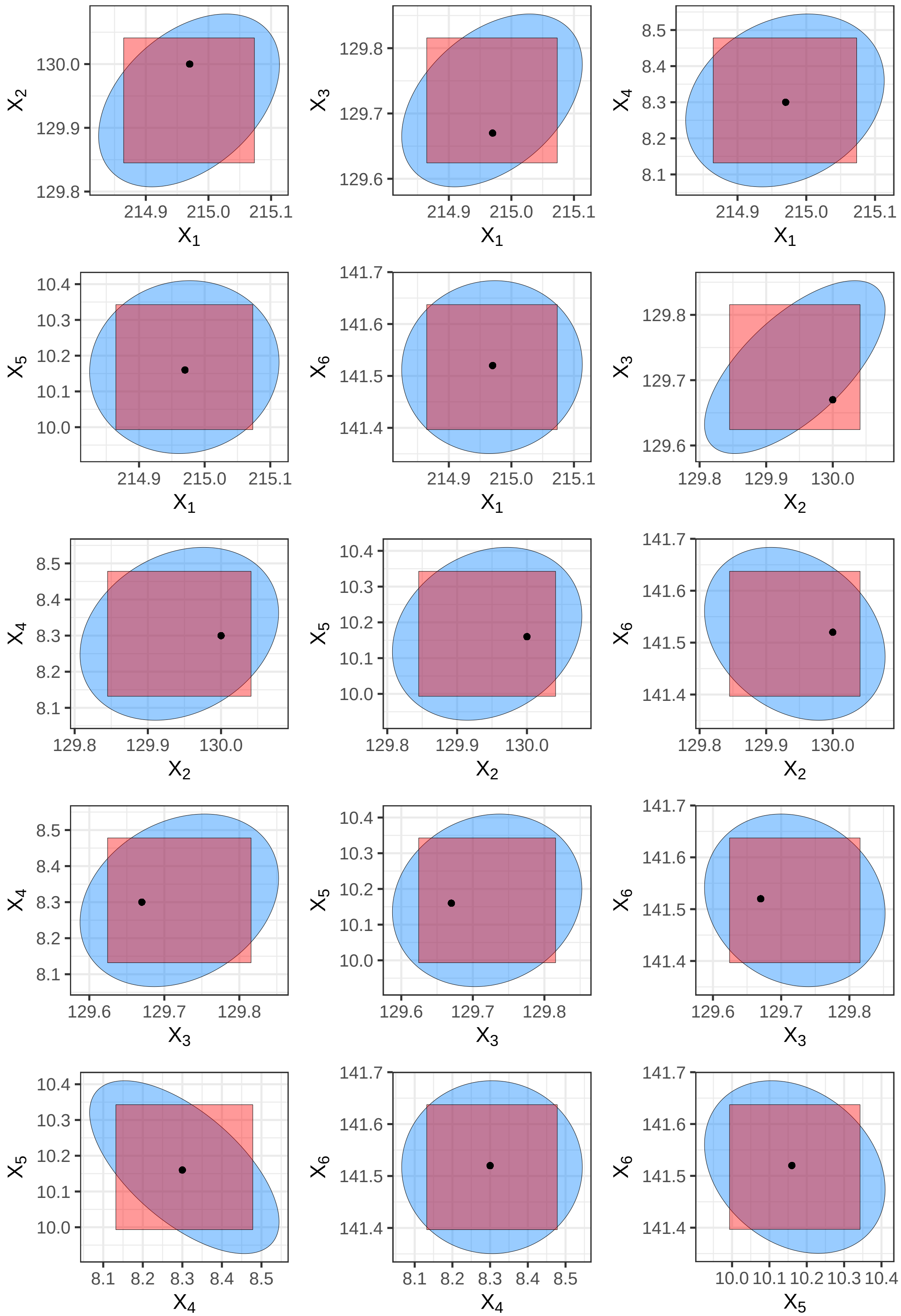
$$a^t\bar{X} \pm \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha) a^T S a}$$

with probability $1-\alpha$, so we use this result with a that has terms 1 for a single given variable and 0 elsewhere.



We see that the projections of T^2 confidence regions are wider than the Bonferroni intervals. Moreover, all projections of the mean we were testing fall inside both types of intervals.
Similarly we take a look at two-dimensional projections of the regions, that is we project them on a plane for each pair X_i, X_j, $i \neq j$.

Two−dimensional projections of the confidence regions for each pair of variables
with marked projection of the new means vector



The projections for Bonferroni method are always rectangles, while for T^2 we get ellipses. Notice how the slopes of the ellipses' axes correspond to correlation between components.
We can also see that the new mean falls inside all the projections, despite not being in the T^2-based confidence region - it's due to the shape of the region which is an ellipsoid, so there is no equivalence between belonging to the region and lying in all the projections of it. The Bonferroni region however, being a (hyper-)cuboid, with walls peripendicular to the projection planes, has that property.

Let's say that the results were deemed not up to standards and so the process was tuned, and new mean values have been obtained: `m1 = 214.99 129.95 129.73 8.51 9.96 141.55`.
Same as before, we check if the vector falls inside the confidence regions for the mean. It turns out, that the new values lie inside the Hotelling's T^2 region, but outside the Bonferroni region. Even though we consider Bonferroni's method as generally more liberal, such a case is not impossible, as can be seen from the projections plotted earlier.

Say the production process was tuned some more, and another mean vector was calculated: `m2 = 214.9473 129.9243 129.6709 8.3254 10.0389 141.4954`.
We verify if the mean belongs to the confidence regions and this time the values fall inside both the T^2-based region and the Bonferroni method region. It can be considered as an indication that no further improvements in production are neccesary as of now.

Simulation 1

To compare results of two approaches to multiple testing - Bonferroni and Benjamini-Hochberg procedures, we will (repeatedly) generate $p = 5000$ realisations of of an independent sequance of normally distributed variables $X_1, \ldots, X_p$, where the first p_1 variables have mean $\sqrt{2 \log p}$, the rest have mean 0, and the variance is 1 for all of them. We consider two cases - for $p_1 = 10$, and $p_1 = 500$. For each variable X_i we test the hypothesis H_{i0} that $\mu_i = 0$ (versus $H_{1i} : \mu_i \neq 0$). We repeat the simulation $n = 10000$ times and with that sample of testing results, for each of the methods we estimate two measures for controlling false discoveries:

$$FWER = P(V > 0), \quad FDR = \mathbb{E}[\frac{V}{R \vee 1}],$$

where V is the number of false discoveries and R - total number of rejections, as well as the power of the test (probability of rejecting H_{0i} when it's false). We set $\alpha = 0.05$. The simulation results are as follows.

Table 4: Estimates of $FWER$, FDR and power for multiple testing procedures in two cases

	$FWER$	FDR	Power
Bonferroni ($p_1 = 10$ case)	0.0491	0.0116	0.3854
Bonferroni ($p_1 = 500$ case)	0.0437	0.0002	0.3860
Benjamini-Hochberg ($p_1 = 10$ case)	0.2853	0.0512	0.5452
Benjamini-Hochberg ($p_1 = 500$ case)	1.0000	0.0451	0.9037

Simulation indicates that the Bonferroni correction controls $FWER$ on level $\alpha = 0.05$. Indeed, we have

$$FWER = P(V > 0) = 1 - P(V = 0) = 1 - (1 - \frac{\alpha}{p})^{p-p_1} \approx \begin{cases} 0.0487, & p_1 = 10 \\ 0.044, & p_1 = 500 \end{cases}.$$

Since $FDR \leqslant FWER$, the Bonferroni procedure also controls FDR. It comes at a cost of unsatisfactory power:

$$P(H_{0i} \text{ rejected} | H_{1i} \text{ true}) = P_{H_{1i}}(|X_i| > \phi^{-1}(1 - \frac{\alpha}{2p})) \approx 0.386.$$

The Benjamini-Hochberg method does not control $FWER$ (for larger number of true alternatives false discoveries were almost certain). It does however, control FDR - we know that for independent test statistics, and such is the case at hand,

$$FDR = \frac{p - p_1}{p}\alpha = \begin{cases} 0.0499, & p_1 = 10 \\ 0.045, & p_1 = 500 \end{cases}.$$

We also obtain significantly better test power which, unlike the Bonferroni case, seems to increase along with the number of true alternatives.

Appendix (to the confidence regions)

```
x_bar <- apply(BankGenuine, 2, mean)
S <- cov(BankGenuine)
n <- dim(BankGenuine)[1]
p <- dim(BankGenuine)[2]
alpha <- 0.05

# does mu fall inside Hotelling's/Bonferroni region
isItIn <- function(mu, n, x_bar, S, p, alpha){
  return(c("Hotelling's:"=
            n*t(x_bar-mu)%*%solve(S)%*%(x_bar-mu)<(n-1)*p/(n-p)*qf(1-alpha, p, n-p),
           "Bonferroni's:"=
            sum(apply(cbind(mu, x_bar, diag(S)), 1, function(v){return(
                   v[1]<v[2]+qt(1-alpha/(2*p), n-1)*sqrt(v[3]/n) &
                   v[1]>v[2]-qt(1-alpha/(2*p), n-1)*sqrt(v[3]/n))})
               )==p
        ))
}

# the means reported
mus <- rbind(c(214.97, 130, 129.67, 8.3, 10.16, 141.52),
             c(214.99, 129.95, 129.73, 8.51, 9.96, 141.55),
             c(214.9473, 129.9243, 129.6709, 8.3254, 10.0389, 141.4954))

# results
apply(mus, 1, isItIn, n=n, x_bar=x_bar, S=S, p=p, alpha=alpha)
```

```
##                 [,1]  [,2] [,3]
## Hotelling's:  FALSE  TRUE TRUE
## Bonferroni's:  TRUE FALSE TRUE
```