

暴露測定數據統計分析探討

Exploration of Statistical Analysis for Exposure Measurement Data

長榮大學職業安全與衛生學系
110學年度第2學期 專題實作
學生:謝幸宜
指導老師:吳俊德助理教授

研究目的

執行作業環境勞工個人有害物暴露測定為職業暴露評估的重要工作，對於暴露測定數據進行適當的統計分析，為瞭解勞工有害物暴露實態，推估勞工有害物暴露引發健康危害效應風險的重要依據。本研究主要目的是以免費統計軟體 R 執行對暴露測定數據的統計分析，熟悉各種相關指令。

暴露測定數據

本研究的暴露測定數據為在 6 個工廠(以A01~A06表示)針對 6 個有機化合物，包括：丙酮(acetone)、正己烷(n-hexane)、苯(benzene)、甲苯(toluene)、乙酸正丁酯(butyl acetate)和二甲苯(xylene)，在區分勞工相似暴露群組(similar exposure group, SEG)下(以編號1~6表示)，對勞工個人暴露空氣採樣測定的測定濃度。

R 軟體統計分析相關指令

為瞭解每一個 SEG 有害物暴露的暴露實態，需對群組暴露測定數據計算相關統計值，如：算術平均值(arithmetic mean, AM)、標準差(standard deviation, SD)、幾何平均值(geometric mean, GM)、幾何標準差(geometric standard deviation, GSD)和第95百分位數(95% percentile, X_{95%})。表 1 為各工廠勞工有害物暴露相似群組暴露實態相關統計值呈現，以 R 計算這些統計值的相關指令如下：

```
stagent01 <- function (factdata01, agent01, seg01) {
  temp01 <- matrix(0, length(seg01), length(agent01))
  for(i in 1:length(seg01)) {
    segdata01 <- factdata01[factdata01$SEG==seg01[i],]
    data01 <- segdata01[colnames(segdata01)==agent01]
    data02 <- data01[agent01]
    AM = round(mean(data02, na.rm=T), 3)
    SD = round(sd(data02, na.rm=T), 3)
    Median = round(median(data02, na.rm=T), 3)
    GM = round(exp(mean(log(data02), na.rm=T)), 3)
    GSD = round(exp(sd(log(data02), na.rm=T)), 3)
    X95 = round(qlnorm(0.95, log(GM), log(GSD)), 3)
    temp01[i, 1] <- AM
    temp01[i, 2] <- SD
    temp01[i, 3] <- Median
    temp01[i, 4] <- GM
    temp01[i, 5] <- GSD
    temp01[i, 6] <- X95
  }
  dimnames(temp01)[[1]] <- c("SEG01", "SEG02", "SEG03", "SEG04", "SEG05", "SEG06")
  dimnames(temp01)[[2]] <- c("AM", "SD", "Median", "GM", "GSD", "X95")
  return(data.frame(temp01))
}

agent01 <- c("Acetone", "n-Hexane", "Benzene", "Toluene", "Butyl acetate", "Xylene")
expostatpar01 <- function (data01, agent01) {
  fact01 <- levels(factor(data01$Factory))
  seg01 <- levels(factor(data01$SEG))
  stat01 <- array(0, dim=c(length(seg01), 6, length(agent01)), length(agent01))
  n01 <- length(agent01)
  for (i in 1:length(fact01)) {
    factdat01 <- data01[data01$Factory==fact01[i],]
    for (j in 1:length(agent01)) {
      temp01 <- stagent01(factdat01, agent01[j], seg01)
      stat01[i, j, ] <- as.matrix(temp01)
    }
  }
  stat02 <- array(stat01, dim=c(length(seg01), 6, length(agent01)), length(fact01))
  dimnames(stat02) <- list(c("SEG01", "SEG02", "SEG03", "SEG04", "SEG05", "SEG06"),
    c("AM", "SD", "Median", "GM", "GSD", "X95"),
    c("Acetone", "n-Hexane", "Benzene", "Toluene", "Butyl acetate", "Xylene"),
    c("A01", "A02", "A03", "A04", "A05", "A06"))
  return(stat02)
}
expostatpar01(shieh01a, agent01)
```

對於每一個 SEG 有害物暴露的暴露平均值和變異可以統計誤差長條圖表示，圖 1 為以丙酮暴露測定濃度在 6 個工廠所區分的 SEG 的誤差長條圖比較，以 R 描繪此圖的相關指令如下：

```
library(scales)
library(ggplot2)
library(ggpubr)

p01 <- ggbarplot(shieh01a, combine=TRUE, x="SEG", y=c("Acetone", "n-Hexane", "Benzene", "Toluene", "Butyl acetate", "Xylene"), ylim=c(0,75), fill="SEG", add=c("mean_se", "jitter"), error.plot = "errorbar", add.params = list(size=0.75, jitter = 0.25)) +
  theme_pubr(legend=c("none")) +
  xlab("勞工相似暴露群組") +
  ylab("有害物暴露 平均值+/-標準誤 (ppb)")
p01
```

對於每一個 SEG 有害物暴露的暴露分佈和變異情形，也可以統計箱型圖表示，圖 2 為針對在 6 個工廠 6 個有害物暴露所區分的 SEG 的箱型圖比較，以 R 描繪此圖的相關指令如下：

```
ggboxplot(shieh01b, combine=TRUE, x="Factory", y=c("Acetone", "n-Hexane", "Benzene", "Toluene", "Butyl acetate", "Xylene"), fill="SEG", error.plot="errorbar", ylim=c(0, 50),
  add.params = list(size=0.75, jitter = 0.25)) +
  labs(title=" ", x="工廠編號", y="有害物質暴露濃度") +
  scale_fill_discrete(name="勞工相似暴露群組")
```

對於每一個 SEG 有害物暴露的暴露分佈描繪，可以統計機率分佈圖表示，圖 3 為以 A01 廠為例，針對 6 個有害物暴露在所區分的 SEG 的暴露機率分佈圖描繪，以 R 描繪此圖的相關指令如下：

```
library(plotly)
packageVersion("plotly")
library(gridExtra)
library(grid)
library(ggplot2)
library(lattice)

factA01a <- data.frame(shieh01b[shieh01b$Factory=="A01",])
pp01a <- ggplot(factA01a, aes(Acetone, y=..scaled.., color=SEG)) + xlim(c(0, 100)) + ylab("Density") + xlab("Acetone") +
  geom_density(kernel="gaussian", geom="line", size=0.75) + theme(legend.background=element_rect(colour=factA01a$SEG),
  legend.justification="left", legend.position=c(0.85,0.85))
pp02a <- ggplot(factA01a, aes(n-Hexane, y=..scaled.., color=SEG)) + xlim(c(0, 20)) + ylab("Density") + xlab("n-Hexane") +
  geom_density(kernel="gaussian", geom="line", size=0.75) + theme(legend.background=element_rect(colour=factA01a$SEG),
  legend.justification="left", legend.position=c(0.85,0.85))
pp03a <- ggplot(factA01a, aes(Benzene, y=..scaled.., color=SEG)) + xlim(c(0, 40)) + ylab("Density") + xlab("Benzene") +
  geom_density(kernel="gaussian", geom="line", size=0.75) + theme(legend.background=element_rect(colour=factA01a$SEG),
  legend.justification="left", legend.position=c(0.85,0.85))
pp04a <- ggplot(factA01a, aes(Toluene, y=..scaled.., color=SEG)) + xlim(c(0, 100)) + ylab("Density") + xlab("Toluene") +
  geom_density(kernel="gaussian", geom="line", size=0.75) + theme(legend.background=element_rect(colour=factA01a$SEG),
  legend.justification="left", legend.position=c(0.85,0.85))
pp05a <- ggplot(factA01a, aes(Butyl acetate, y=..scaled.., color=SEG)) + xlim(c(0, 100)) + ylab("Density") + xlab("Butyl acetate") +
  geom_density(kernel="gaussian", geom="line", size=0.75) + theme(legend.background=element_rect(colour=factA01a$SEG),
  legend.justification="left", legend.position=c(0.85,0.85))
pp06a <- ggplot(factA01a, aes(Xylene, y=..scaled.., color=SEG)) + xlim(c(0, 100)) + ylab("Density") + xlab("Xylene") +
  geom_density(kernel="gaussian", geom="line", size=0.75) + theme(legend.background=element_rect(colour=factA01a$SEG),
  legend.justification="left", legend.position=c(0.85,0.85))
grid.arrange(pp01a+theme(legend.position=c(0.83,0.78), legend.key.size=unit(0.25, "cm")), pp02a+theme(legend.position=c(0.83,0.78),
  legend.key.size=unit(0.25, "cm")), pp03a+theme(legend.position=c(0.83,0.78), legend.key.size=unit(0.25, "cm")),
  pp04a+theme(legend.position=c(0.83,0.78), legend.key.size=unit(0.25, "cm")), pp05a+theme(legend.position=c(0.83,0.78),
  legend.key.size=unit(0.25, "cm")), pp06a+theme(legend.position=c(0.83,0.78), legend.key.size=unit(0.25, "cm")), ncol=3, nrow=2)
```

結果與討論

使用 R 軟體執行暴露測定數據的暴露實態相關統計值計算、暴露變異推估和暴露分佈描繪有準確、快速、簡潔、清晰等優點，然對於相關指令的運用，需要一些時間的訓練和嘗試，方能熟悉且得心應手的操作。有興趣從事職業安全衛生資料統計分析、模式建立、模擬推估等工作的同學，應努力學習 R 軟體的使用，遇見自己有興趣探討的問題，就能多多尋找 R 軟體的相關套件，嘗試寫出適合自己數據分析的程式，如此將有助於利用資料分析結果擬定相關管理與控制策略，有效管理與控制勞工於職場所面臨的危害。

表1 各工廠勞工有害物暴露相似群組暴露實態

SEG	Acetone						n-Hexane						Benzene						Toluene						Butyl acetate						Xylene																																								
	AM	SD	GM	GSD	X _{95%}	SEG	AM	SD	GM	GSD	X _{95%}	SEG	AM	SD	GM	GSD	X _{95%}	SEG	AM	SD	GM	GSD	X _{95%}	SEG	AM	SD	GM	GSD	X _{95%}	SEG	AM	SD	GM	GSD	X _{95%}	SEG	AM	SD	GM	GSD	X _{95%}																														
A01	1.22	1.73	0.82	2.08	2.76	1	0.64	0.32	0.46	3.03	2.83	1	0.76	0.37	0.59	2.59	2.81	1	0.87	0.45	0.61	3.27	4.25	1	0.99	0.51	0.71	3.05	4.45	1	0.99	0.52	0.65	3.68	5.53	1	1.22	1.73	0.82	2.08	2.76	1	0.64	0.32	0.46	3.03	2.83	1	0.76	0.37	0.59	2.59	2.81	1	0.87	0.45	0.61	3.27	4.25	1	0.99	0.51	0.71	3.05	4.45	1	0.99	0.52	0.65	3.68	5.53
A02	9.95	36.4	1.53	4.48	17.99	2	0.91	0.21	0.57	2.95	3.37	2	1.88	4.78	0.8	3.16	5.31	2	1.22	1.38	0.77	3.13	5.01	2	4.65	9.14	1.47	4.46	17.2	2	3.23	69.92	2.97	11.76	171.05	2	1.88	4.78	0.8	3.16	5.31	2	1.22	1.38	0.77	3.13	5.01	2	4.65	9.14	1.47	4.46	17.2	2	3.23	69.92	2.97	11.76	171.05												
A03	1.57	3.33	0.82	2.28	3.16	3	0.74	0.2	0.66	2.01	2.07	3	0.89	0.23	0.81	1.82	2.15	3	12.07	19.79	3.81	4.73	49.03	3	1.16	0.32	1.03	2.02	3.26	3	59.36	191.3	3.47	9.01	129.21	3	0.74	0.2	0.66	2.01	2.07	3	0.89	0.23	0.81	1.82	2.15	3	12.07	19.79	3.81	4.73	49.03	3	1.16	0.32	1.03	2.02	3.26	3	59.36	191.3	3.47	9.01	129.21						
A04	0.64	0.01	0.64	1.02	0.66	4	0.52	0.37	0.31	3.51	2.46	4	0.64	0.42	0.42	2.94	2.49	4	0.72	0.51	0.4	3.83	3.66	4	0.92	0.76	0.51	3.74	4.46	4	1.12	1.18	0.47	5.1	6.89	4	0.52	0.37	0.31	3.51	2.46	4	0.64	0.42	0.42	2.94	2.49	4	0.72	0.51	0.4	3.83	3.66	4	0.92	0.76	0.51	3.74	4.46	4	1.12	1.18	0.47	5.1	6.89						
A05	4.95	12.49	1.5	3.69	12.82	5	0.77	0.42	0.6	2.53	2.73	5	1.94	4.75	0.87	3	5.28	5	1.42	1.86	0.88	3.05	5.49	5	4.41	11.55	1.45	3.64	12.19	5	39.79	144.16	1.97	5.05	73.78	5	0.77	0.42	0.6	2.53	2.73	5	1.94	4.75	0.87	3	5.28	5	1.42	1.86	0.88	3.05	5.49	5	4.41	11.55	1.45	3.64	12.19	5	39.79	144.16	1.97	5.05	73.78						
A06	4.12	7.92	1.53	3.69	13.08	6	0.67	0.3	0.51	2.76	2.7	6	0.8	0.34	0.64	2.39	2.7	6	0.91	0.42	0.68	2.96	4.03	6	1.73	1.44	1.25	2.67	6.27	6	1.63	1.27	0.96	4.08	9.67	6	0.67	0.3	0.51	2.76	2.7	6	0.8	0.34	0.64	2.39	2.7	6	0.91	0.42	0.68	2.96	4.03	6	1.73	1.44	1.25	2.67	6.27	6	1.63	1.27	0.96	4.08	9.67						

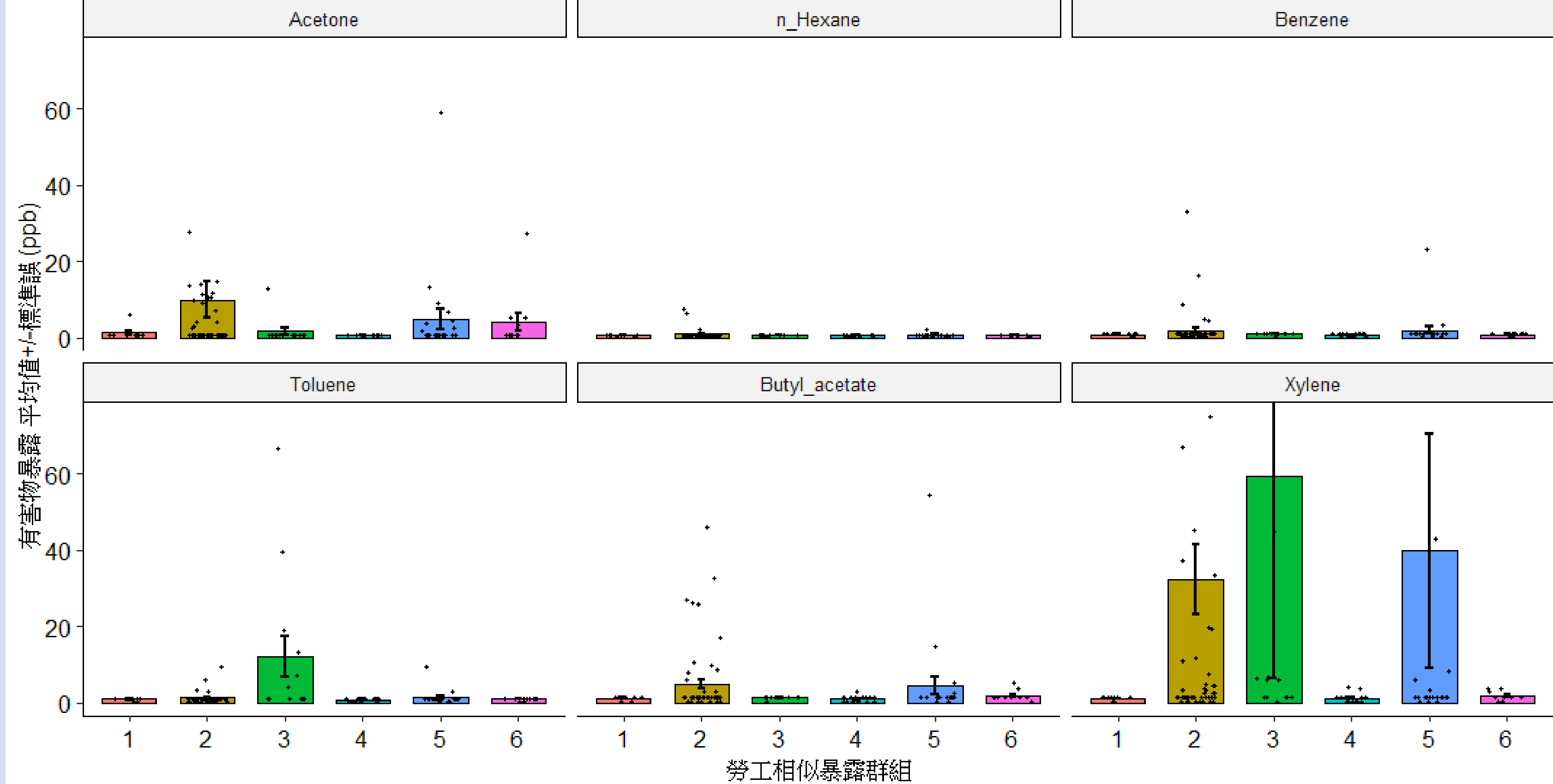


圖1 各工廠個有害物暴露所區分的 SEG 的誤差長條圖比較

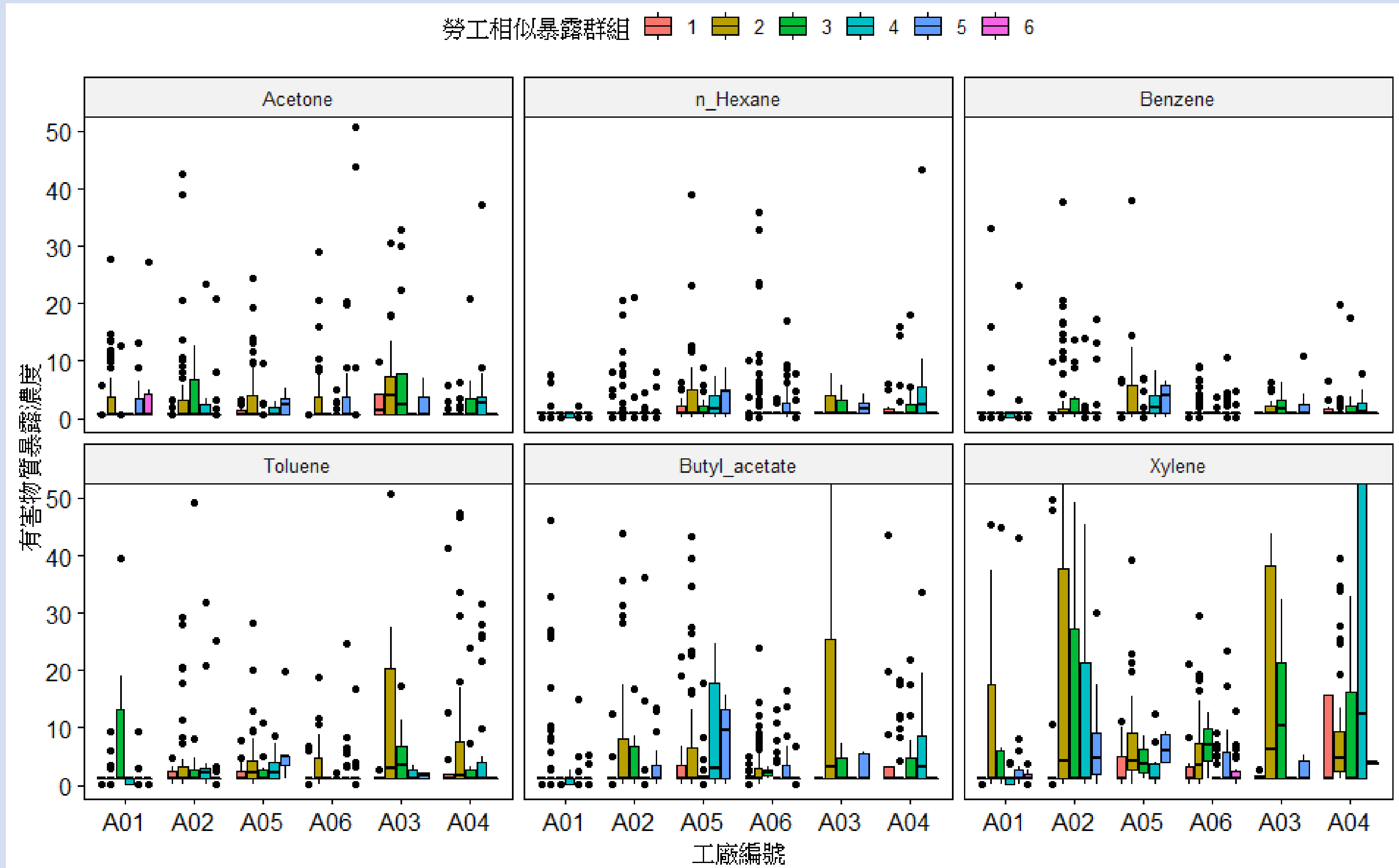


圖2 各工廠有害物暴露所區分的 SEG 的箱型圖比較

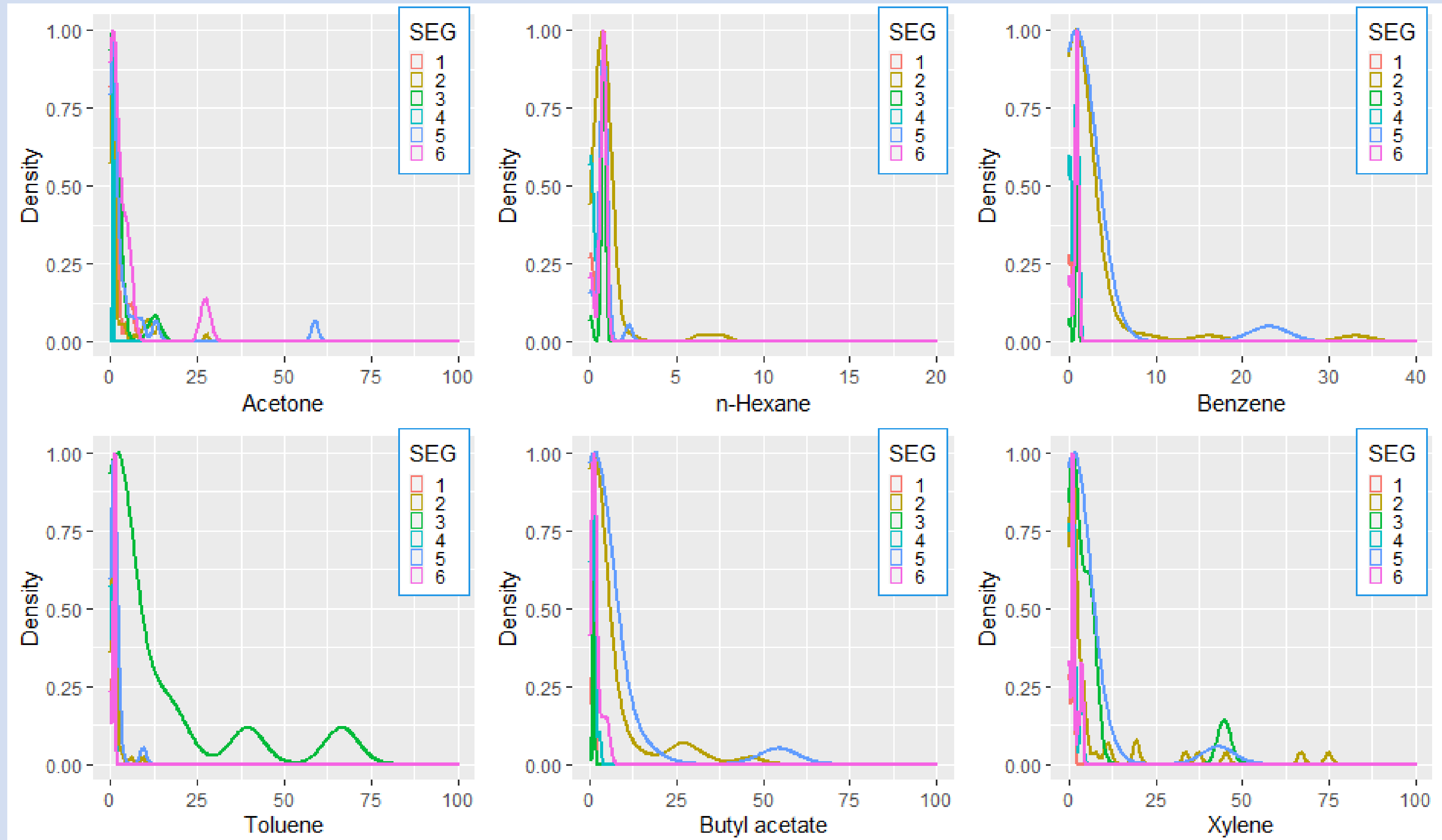


圖3 6 個有害物暴露在所區分 SEG 的暴露機率分佈圖