

Mann-Whitney Test in the Evaluation of Conditioning and Loosing Technological Innovation*

Pengkun Gong¹, Yongkang Xu², Peilin Zhao², Hu Xiang², Junjie Wang², Xingxu Li^{1#}

¹School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming

²Qujing Cigarette Factory, Hongyun Honghe Group, Qujing

Email: #xingxu1967@sina.com

Received: Nov. 11th, 2013; revised: Dec. 5th, 2013; accepted: Dec. 12th, 2013

Copyright © 2013 Pengkun Gong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2013 are reserved for Hans and the owner of the intellectual property Pengkun Gong et al. All Copyright © 2013 are guarded by law and by Hans as a guardian.

Abstract: Modern factories engage in a large number of quality control activities, in which statistical methodologies are indispensable for evaluating the effect of technological innovation. To evaluate the effect of innovation of moisture percentage control technology in the process of conditioning and losing, statistical significance of the difference between after and before the innovation needs to be proved. The fact that the data are non-normally distributed is found after the test of normality, thus traditional T test loses its validity. When employing the Mann-Whitney test in the data, the effect of innovation of moisture percentage control technology was to be found significant ($P < 0.05$). The effect of technological innovation was evaluated effectively and scientifically with this method, which can be popularized in the quality control activities which produce non-normally distributed data.

Keywords: Evaluation of Technical Improvements; Non-Normally Distribution; T Test; Mann-Whitney Test

Mann-Whitney 检验在松散回潮技术改进评价中的应用*

巩鹏堃¹, 徐永康², 赵培林², 向 虎², 王俊杰², 李兴绪^{1#}

¹云南财经大学统计与数学学院, 昆明

²红云红河集团曲靖卷烟厂, 曲靖

Email: #xingxu1967@sina.com

收稿日期: 2013 年 11 月 11 日; 修回日期: 2013 年 12 月 5 日; 录用日期: 2013 年 12 月 12 日

摘 要: 现代企业质量控制活动中, 评价技术改进前后的效果, 已经离不开统计方法的支持。为了评价松散回潮中烟叶含水率控制技术改进效果, 需要对比技术改进前后含水率变异系数是否有显著差异。对数据进行正态性检验后发觉其不服从正态分布, 因此传统的 T 检验失效。用 Mann-Whitney 检验这一非参数检验对数据进行检验, 发现松散回潮含水率控制技术改进效果显著($P < 0.05$)。此方法科学评价了技术改进的效果, 可以对企业质量控制活动中的非正态分布数据应用推广。

关键词: 技术改进效果评价; 非正态分布; T 检验; Mann-Whitney 检验

*资助信息: 本文的研究得到了红云红河烟草(集团)有限责任公司科技计划项目(HYHH2012GY03, HYHH2012SB01)基金资助。

#通讯作者。

1. 引言

松散回潮是烟草生产过程中一项重要工艺,以增加叶片的含水率和温度,提高叶片的耐加工性,使叶片松散。松散回潮是制丝线第一道加工工序,松散回潮含水率是制丝线的重点控制指标,其含水率的均匀性直接影响整个制丝线的后续加工质量。卷烟厂在技术改进之前对松散回潮含水率均采用后馈控制,即根据出口水分仪反馈信号调整加水量。

后馈控制是根据加工结果来调整加工参数,若来料含水率均匀,采用后馈控制方式也能使出口含水率均匀。事实上在松散回潮工序,来料含水率波动较大,出口含水率相应波动也较大,后馈控制存在很大的局限性,因此,需要在松散回潮工序入口加装含水率仪,检测叶片回潮前含水率,将含水率仪前、后馈信号作为调整加水量的依据,实现前、后馈控制相结合,提高出口含水率的均匀性。

含水率标准偏差可衡量松散回潮出口含水率均匀性,标准偏差越小,说明均匀性越好。考虑到批间加工标准的变化会导致含水率均值出现变化,为消除均值的影响,因此选择含水率变异系数来衡量含水率的均匀性。含水率变异数由含水率标准偏差除以含水率均值所得,反应了平均每个含水率单位的标准偏差。

烟厂车间收集到了松散回潮不同控制方式下的批内含水率标准偏差,计算出其批内变异系数,并得出不同控制方式下的变异系数估计值,从而比较前、后馈控制方式与单一的后馈控制方式的优劣。如果技术改进后的前、后馈控制方式具有较小的变异系数,说明技术改进有效。

在企业开展质量控制的活动中,越来越多得需要统计方法来评价质量控制效果。得到变异系数后,所有变异系数都应该视为样本观测值,显然直接用变异系数的样本均值来比较其来自总体的均值大小是不科学的。这时,一般情况下会用 T 检验来判断两组样本观测值所来自总体的均值。

T 检验属于传统的参数检验,在参数检验的方法下,需要假设样本所来自的总体的分布是已知的,所不确定的是参数值。例如,T 检验要求样本所来自的总体是正态分布的。吴喜之(1999)认为,现实中如果没有根据得对样本所来自的总体进行服从某一分布

的假设,结果可能是不可靠的^[1]。这时应该用何种方法来进行判断,是需要研究的问题。

2. Mann-Whitney 检验

2.1. Mann-Whitney 检验与 T 检验的比较

非参数检验就是在总体未知的情况下进行统计推断的方法,只需要做一些简单的假设,具有很好的稳健性。Mann-Whitney 检验就是一种用来比较两组独立样本观测值所来自的总体大小的非参数检验。Sawilowsky (1992)研究得到,在违背正态分布假定的情况下, Mann-Whitney 检验的势是 T 检验的 3 到 4 倍^[2]。这意味着非正态条件下, Mann-Whitney 检验不犯第二类错误概率是 T 检验的 3 到 4 倍。Tana J. Bridge (2007)在其博士学位论文中用蒙特卡洛模拟方法证明了在正态性条件下,与 Mann-Whitney 检验相比, T 检验具有较好的稳健性和较高的势;但是在非正态性条件下, Mann-Whitney 检验具有较高的势^[2]。事实上,假如所得变异系数并不服从正态分布,应用 T 检验就会出现检验统计量和 P-值有偏误的情况,偏误的方向和大小取决于数据的偏度和峰度偏离正态的程度^[3],其结果不可信。因此对于非正态数据,没有必要再做 T 检验。

2.2. Mann-Whitney 检验原理

秩是一个数据在与其比较大小的数据中,按照升幂排列之后,该数据的位置。Mann-Whitney 检验采用统计量为 U 统计量,它的分布在原假设下是已知的。对两组要比较的样本排序,加总来自于样本 1 的观测值的秩和,样本 2 的观测值的秩和就确定了。对于大样本数据来说,计算公式如下:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \quad (1)$$

n_1 表示第一组样本的样本容量, R_1 是样本 1 的所有观测值的秩和。

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2} \quad (2)$$

n_2 表示第二组样本的样本容量, R_2 是样本 2 的所有观测值的秩和。其中 SPSS 输出结果的 U 统计量是 U_1 和 U_2 中较小的值。Mann-Whitney 检验的是两组数

据所来自总体的中位数是否相等, U 统计量大于某一显著性水平下的临界值时, 说明在该显著性水平下拒绝两组数据来自总体中位数相等的原假设。在 SPSS 输出结果中, 根据其 P 值即可以判断是否拒绝原假设。

3. 数据来源说明及描述性分析

3.1. 数据来源说明

因不同牌号其来料含水率均匀性存在较大差异, 导致出口含水率均匀性也存在一定差异, 因此数据分析需对同牌号同模组的两种水分控制方式下的数据进行对比。数据采集时, 每个牌号或模组先在后馈控制方式下每隔一段时间观测一次, 作为一个样本观测值, 观测多次形成一个对照数据组; 然后采用前后馈控制方式, 再进行观测, 形成一个实验数据组。观测时记录含水率标偏和均值, 整理后对每个对照组和实验组数据计算出变异系数, 形成新的对照组和实验组。共有 4 个模组或牌号, 分别是 B 线云烟(软珍品)模组 1、B 线云烟(软珍品)模组 2、B 线云烟(软紫)、B 线红山茶(软)。每个模组或牌号都有一个对照组和实验组。文中为了方便, 依次用序号 t1 至 t8 表示 8 组数据, 序号和每个牌号、模组的对照组和实验组的对应关系如表 1 所示。也就是 t1、t3、t5、t7 表示技术改进前的对照组, 而 t2、t4、t6、t8 依次分别表示技术改进后的、与其对应的实验组。

3.2. 数据描述性分析

为了对样本数据有一定了解, 在做分析之前, 用

Table 1. Objects symbolized by each serial number
表 1. 各个序号的表示对象

序号	模块和牌号
t1	B 线云烟(软珍品)模组 1 后馈控制
t2	B 线云烟(软珍品)模组 1 前、后馈控制
t3	B 线云烟(软珍品)模组 2 后馈控制
t4	B 线云烟(软珍品)模组 2 前、后馈控制
t5	B 线云烟(软紫)后馈控制
t6	B 线云烟(软紫)前、后馈控制
t7	B 线红山茶(软)后馈控制
t8	B 线红山茶(软)前、后馈控制

SPSS 对每组数据进行描述性统计分析, 其结果如表 2 所示。

由表 2 可以看出, 各个模组样本观测值数量不同, 样本观测值最少的是 t4, 为 24, 样本观测值最多的是 t1, 为 162。从均值这一统计量来看, 各个牌号、模组的对照组和实验组的变异系数均值水平明显有差异, 例如 t1 均值为 0.070232, t2 均值为 0.032833, 差值大约为 0.04, 其他三个牌号、模组的对照组和实验组之间的变异系数均值差值大约也都在 0.04 左右。即技术改进前的含水率变异系数均值要比技术改进后的含水率变异系数均值大 0.04 左右。从中位数这一统计量上来看, 技术改进前的各个对照组的中位数也都大于技术改进后的实验组的变异系数中位数, 差值大约为 0.016, 并没有均值的差值那么大。各个对比组变异系数的标准偏差也要远大于其对应实验组的标准偏差, 这说明技术改进前的对比组变异系数变化范围要大于技术改进后的实验组。各个对照组的极小值、极大值、变异系数全距(极大值与极小值之差), 大于各个实验组, 各个对照组三个四分位数也都大于实验组。而且从四分之一分位数到四分之三分位数, 同牌号、模组的对照组与实验组同分位数之间的差值越来越大。从偏度统计量上来看, 各个对照组和实验组的数据均属于不对称分布, 具有偏态分布的性质。因此均值代表性较差, 而中位数与四分位数等位置代表值不受数据偏态分布的影响因而更具有代表性。直观上从中位数和四分位数来判断, 各个牌号、模组技术改进后的实验组样本观测值来自总体的中位数小于技术改进前的对比组来自总体的可能性很大。

4. 正态性检验

虽然从偏度和峰度统计量上直观上看各个数据组的样本观测值来自总体为正态分布的可能性不大, 但是为了从统计意义上检验这些总体是否服从正态分布, 必须进行正态性检验。正态性检验的方法有 Shapiro 正态性检验 (Shapiro-Wilk normality test) 和 Kolmogorov-Smirnov 检验。在检验正态性方面, Shapiro 正态性检验要优于 Kolmogorov-Smirnov 检验。两种检验的原假设都是样本来自总体服从正态分布。各个牌号、模组的变异系数正态性检验结果如表 3 所示。

Table 2. Result of descriptive statistics analysis of 8 data sets
表 2. 8 组数据的描述性统计分析结果数据

序号	N	均值	标准差	偏度	峰度	全距	极小值	极大值	百分位数		
									25	50	75
t1	162	0.070232	0.052271	1.545	3.1	0.309676	0.017477	0.327153	0.03141	0.040104	0.103946
t2	120	0.032833	0.011216	3.902	22.121	0.105917	0	0.105917	0.028306	0.031391	0.034445
t3	33	0.059447	0.048013	1.269	0.388	0.15583	0.022065	0.177895	0.027044	0.029815	0.088979
t4	24	0.028071	0.006587	0.219	-0.935	0.023962	0.017845	0.041807	0.021472	0.027512	0.034355
t5	122	0.079019	0.044732	0.845	-0.356	0.187873	0.031001	0.218874	0.041507	0.062523	0.116238
t6	129	0.036814	0.012887	3.239	14.831	0.106485	0	0.106485	0.031087	0.034633	0.038791
t7	68	0.086374	0.050833	0.804	-0.326	0.209038	0.032405	0.241443	0.042737	0.069004	0.129268
t8	95	0.045269	0.017631	7.756	69.871	0.182821	0.020306	0.203127	0.039663	0.043314	0.048219

Table 3. Result of normality test of 8 data sets
表 3. 8 组数据的正态性检验结果

序号	Kolmogorov-Smirnova			Shapiro-Wilk		
	统计量	df	Sig.	统计量	df	Sig.
t1	0.236	162	0.000	0.806	162	0.000
t2	0.240	120	0.000	0.587	120	0.000
t3	0.324	33	0.000	0.755	33	0.000
t4	0.106	24	0.200	0.955	24	0.341
t5	0.209	122	0.000	0.864	122	0.000
t6	0.206	129	0.000	0.645	129	0.000
t7	0.233	68	0.000	0.857	68	0.000
t8	0.269	95	0.000	0.375	95	0.000

由表 3 可以看出,除了 t4,其余各组都在 0.01 的显著性水平下拒绝了正态分布的原假设($P < 0.01$)。即使 t4 服从正态分布,其对照组 t3 也不服从正态分布,所以比较 t3 和 t4 时,仍然是 Mann-Whitney 检验要更好。正态性检验结果显示,各个牌号、模组变异系数数据的比较都需要用 Mann-Whitney 检验。

5. 各牌号、模组数据的 Mann-Whitney 检验

因为 Mann-Whitney 检验的是两组数据的总体中位数是否相等,所以设 t1、t2、t3、t4、t5、t6、t7、t8 所来自总体的中位数依次为 m1、m2、m3、m4、m5、m6、m7、m8。对于四个牌号、模组的 Mann-Whitney 检验的原假设是 $H_0: m1 = m2, m3 = m4, m5 = m6, m7 = m8$,对应的双侧备择假设应该是 $H_1: m1 \neq m2, m3 \neq m4, m5 \neq m6, m7 \neq m8$ 。由描述性分析可知,

四个牌号、模组的技术改进前的对照组的样本中位数大于技术改进后的实验组样本中位数,因此又可设置单侧备择假设为 $H_2: m1 > m2, m3 > m4, m5 > m6, m7 > m8$ 。表 4 即为 SPSS 输出的 Mann-Whitney 检验结果。

表 4 的前两行是 Mann-Whitney U 统计量和 Wilcoxon W 统计量的值。对于双尾检验 $H_0: m1 = m2, m3 = m4, m5 = m6, m7 = m8$ 对 $H_1: m1 \neq m2, m3 \neq m4, m5 \neq m6, m7 \neq m8$,我们需要的结果是精确显著性(双侧)^[4],显然除了 t3、t4 所对应的 $P < 0.05$ 外,其他各组数据所对应 $P < 0.01$;对于单尾检验 $H_0: m1 = m2, m3 = m4, m5 = m6, m7 = m8$ 对 $H_2: m1 > m2, m3 > m4, m5 > m6, m7 > m8$,我们需要的结果是精确显著性(单侧)^[4],仍然是除了 t3、t4 所对应的 $P < 0.05$ 外,其他各组数据所对应 $P < 0.01$ 。此结果

Table 4. Result of Mann-Whitney test of 8 data sets
表 4. 8 组数据的 Mann-Whitney 检验结果

	t1, t2	t3, t4	t5, t6	t7, t8
Mann-Whitney U	5011	259	1784	1723
Wilcoxon W	12271	559	10169	6283
Z	-6.955	-2.214	-10.585	-5.072
渐近显著性(双侧)	0.000	0.027	0.000	0.000
精确显著性(双侧)	0.000	0.027	0.000	0.000
精确显著性(单侧)	0.000	0.013	0.000	0.000
点概率	0.000	0.001	0.000	0.000

表明, 对于 B 线云烟(软珍品)模组 1、B 线云烟(软珍品)模组 2、B 线云烟(软紫)、B 线红山茶(软)四个牌号、模组, 除了 B 线云烟(软紫)在 0.05 的显著性水平外, 其他模组、牌号的检验都可以在 0.01 的显著性水平下拒绝双尾检验和单尾检验的原假设。有充分理由认为单侧备择假设 $H_2: m_1 > m_2, m_3 > m_4, m_5 > m_6, m_7 > m_8$ 成立。意味着对于四个牌号、模组来说, 技术改进后的变异系数显著小于技术改进前的变异系数, 可以认为技术改进后的含水率均匀性显著优于技术改进前。

6. 结论

本文通过对于 B 线云烟(软珍品)模组 1、B 线云烟(软珍品)模组 2、B 线云烟(软紫)、B 线红山茶(软)这三个牌号四组对比数据的描述性分析, 正态性检验

以及 Mann-Whitney 检验, 可以得出对于这三个牌号, 前、后馈控制方式相对于后馈控制方式在提高出松散回潮工艺中出口含水率的均匀性时有明显改进的结论。有充分证据将前、后馈控制方式这种水分控制技术继续使用并推广。

更进一步, 本文完善了企业技术改进效果评价的方法。即在样本数据服从正态分布时, 采用传统的 T 检验; 而在样本数据不服从正态分布时, 采用 Mann-Whitney 检验或者其他对分布无要求的非参数检验。本文的结果对于企业技术改进评价过程中所得不服从正态分布的数据的检验具有推广价值。

7. 致谢

本文感谢曲靖卷烟厂项目团队(HYHH2012GY03, HYHH2012SB01)提供的数据支持和具体指导!

参考文献 (References)

- [1] 吴喜之 (1999) 非参数统计. 第二版, 中国统计出版社, 北京, 1.
- [2] Bridge, T.J. (2007) Deconstructing the comparative power of the independent samples and the Wilcoxon Mann-Whitney for shift and slight heteroscedasticity. *Wayne State University*, 17-19,33.
- [3] Justine, R., Gondan, M. and Kieser, M. (2012) To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC Medical Research Methodology*, **12**, 2.
- [4] 吴喜之 (2009) 统计学: 从数据到结论. 第三版, 中国统计出版社, 北京, 266-268.