STAT383 HW 6

Lewis Collum

Updated: April 5, 2020

---

5 - Paper Tensile Strength

Hypotheses

$$H_0 : \mu_5 = \mu_{10} = \mu_{15} = \mu_{20}$$

$$H_A : \mu_i \neq \mu_j \text{ for some } i \neq j$$

Data

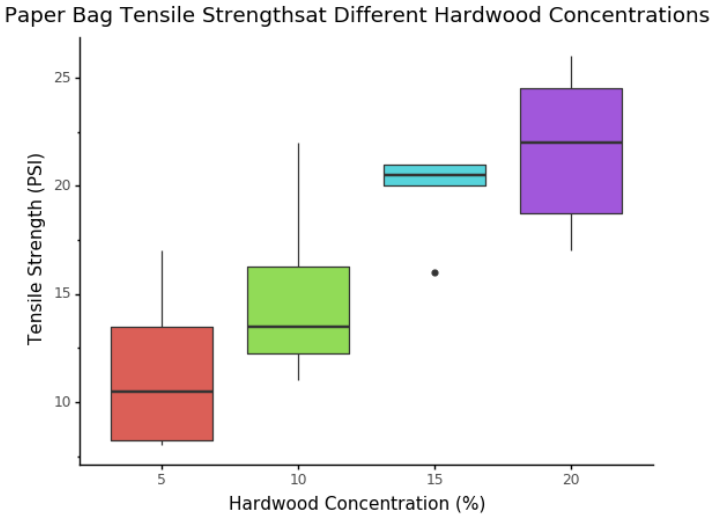
```
import pandas
import plotnine as p9
import matplotlib.pyplot as pyplot

data = pandas.read_excel('q5.xlsx')
data.rename({
    'Hardwood Concentration (%)': 'concentration',
    'Tensile Strength (PSI)': 'psi'}, axis='columns', inplace=True)

data['concentration'] = data['concentration'].astype('category')

g = (p9.ggplot(data)
     + p9.aes(y='psi', x='concentration')
     + p9.geom_boxplot(p9.aes(fill='concentration'))
     + p9.theme_classic()
     + p9.theme(legend_position='none')
     + p9.ggtitle(("Paper Bag Tensile Strengths"
                   "at Different Hardwood Concentrations"))
     + p9.ylab("Tensile Strength (PSI)")
     + p9.xlab("Hardwood Concentration (%)"))

g.draw()
pyplot.savefig('q5_boxplot.png')
```



ANOVA

```
from scipy import stats

f, p = stats.f_oneway(data['psi'], data['concentration'])
print(f'\[\\text{{p-value}} = {round(p, 5)}\]')
```

$$\text{p-value} = 0.0077$$

Tukey HSD

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd
print(tukeyhsd(data['psi'], data['concentration']))
```

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower   upper  reject
---------------------------------------------------
     5     10      3.5 0.3193 -2.0552  9.0552  False
     5     15      8.5 0.0019  2.9448 14.0552   True
     5     20  10.3333  0.001  4.7781 15.8885   True
    10     15      5.0 0.0873 -0.5552 10.5552  False
    10     20   6.8333 0.0126  1.2781 12.3885   True
    15     20   1.8333 0.7726 -3.7219  7.3885  False
---------------------------------------------------
```

Conclusion

From ANOVA:

Since the p-value is less than 0.05, we reject the null hypothesis. The data suggests that the hardwood concentration has some effect on the tensile strength of the bag.

From Tukey HSD:

The difference in tensile strength between bags with hardwood concentrations of 15% and 20% is not significant. However, the difference of those between 10% and 20% is significant.

Analogously, the difference in tensile strength between bags with hardwood concentrations of 5% and 10% is not significant. However, the difference of those between 5% and 15% is significant.

At the edge-cases, bags with a hardwood concentration between 5% and 20% have significantly different tensile strengths.

9 - Paint Weathering

Hypotheses

Difference in Avergage Deterioration between Different Paint Types:

$$H_0 : \alpha_A = \alpha_B = \alpha_C$$

$$H_A : \alpha_i \neq \alpha_j \text{ for some } i \neq j$$

Difference in Avergage Deterioration between Different Environments:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_A : \beta_i \neq \beta_j \text{ for some } i \neq j$$

Interaction between Paint Types and Environments:

$$H_0 : (\alpha\beta)_{ij} = 0$$

$$H_A : (\alpha\beta)_{ij} \neq 0 \text{ for some } i \neq j$$

ANOVA

```
import pandas
import statsmodels.api as sm
from statsmodels.formula.api import ols
from statsmodels.stats import multicomp as mc

def dataFrameToOrg(dataFrame):
    return [x.split(",") for x in dataFrame.to_csv().split("\n")[:-1]]

data = pandas.read_excel('q9.xlsx')

data = data.rename({
    'Paint Type': 'paintType',
    'Environment': 'environment'
}, axis='columns')

model = ols('Deterioration ~ C(paintType) * C(environment)', data).fit()
anova = sm.stats.anova_lm(model, typ = 2)
anova = anova.round(5)
print(dataFrameToOrg(anova))
```

	sum_sq	df	F	PR(>F)
C(paintType)	30.57451	2.0	20.35496	0.00014
C(environment)	26.8365	3.0	11.91092	0.00066
C(paintType):C(environment)	6.02552	6.0	1.33716	0.31396
Residual	9.0124	12.0		

Tukey

```
environmentTukey = mc.pairwise_tukeyhsd(
    data['Deterioration'],
    data['environment'])

paintTypeTukey = mc.pairwise_tukeyhsd(
    data['Deterioration'],
    data['paintType'])

print(f'Environment Tukey:\n{environmentTukey}\n')
print(f'Paint Type Tukey:\n{paintTypeTukey}')
```

```
Environment Tukey:
Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
     1      2    1.155 0.5539 -1.2855 3.5955  False
     1      3  -1.2217 0.5122 -3.6622 1.2188  False
     1      4     1.46  0.363 -0.9805 3.9005  False
     2      3  -2.3767  0.058 -4.8172 0.0638  False
     2      4    0.305    0.9 -2.1355 2.7455  False
     3      4   2.6817 0.0281  0.2412 5.1222   True
---------------------------------------------------

Paint Type Tukey:
Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
     A      B   2.7638 0.0022  0.9846 4.5429   True
     A      C   1.3188 0.1729 -0.4604 3.0979  False
     B      C   -1.445 0.1256 -3.2242 0.3342  False
---------------------------------------------------
```

Conclusion

From ANOVA:

The data suggests that the paint type and the environment effects have significant impact on deterioration.

From Tukey (Paint Type):

The difference in deterioration between surfaces with paint type A and B is significant.

From Tukey (Environment):

The difference in deterioration between surfaces in environments 3 and 4 is significant.