Alexander Caines
MATH-310
3-23-2021

HMWK 3 *With help from Stephanie.*

(1) Assume that the differences between the variances of the data are not statistically significant. A two-way ANOV is necessary here because there are multiple independent factors being considered.

(a) Consider the following null hypotheses: That fiber, sand, or sandfiber have no statistically significant effect on wetmold strength. The corresponding alternative hypotheses being that fiber, sand, or sand-fiber had statistically significant effects on wetmold strength. The results of the ANOVA showed that, at the 95% significance level, the only factor that contributed a statistically significant effect on wetmold strength was fiber. So, we can reject the null that fiber had no effect on wetmold strength. Because sand and sand-fiber did not have a statistically significant effect on wetmold strength, we fail to reject the null that either of them did.

(b) Consider the following null hypotheses: That fiber, sand, or sandfiber have no statistically significant effect on casting hardness. The corresponding alternative hypotheses are that: fiber, sand, or sandfiber have a statistically significant effect on casting hardness. The results of the ANOVA showed that fiber and sand have a statistically significant effect on casting hardness at the 95% confidence level. So, we can reject the null hypothesis that either of the two do not. Since sand-fiber has a statistically significant effect on casting hardness, however, we fail to reject the null that it doesn't.

(c) I tried to make a plot but it honestly didn't show any of the results Stephanie and I got for a and b. So, I am confused.

(2) Because none of the elements in the F-Stat column are less than 0.05, it can be conclulded that there is no significance in effects between the fixed factors listed in the problem on the strength of paper. Find the ANOVA table below.

	SS	dof	msq	f-stat
A	6.94	1	6.94	19.83
B	5.61	3	1.87	5.34
C	12.33	2	6.17	17.61
AB	14.40	3	4.8	13.71
AC	27.32	2	3.66	10.46
BC	15.80	6	2.63	7.52
SSE	8.42	24	0.35	
SST	70.82	32	2.21	

(3) The ANOVA showed that each of the factors did not exhibit a statistically significant effect on the life of a tool at the 95% significance level.

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
Cut_spd      1  850.8   850.8 1815.00 1.81e-06 ***
Cut_dpth     1    2.5     2.5    5.40  0.08080 .  
Feed         1   11.3    11.3   24.07  0.00801 ** 
Residuals    4    1.9     0.5                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

Figure 1: ANOVA results

(4) Given $y = 4000 + 10x$, we know that $Y = 6000 + \varepsilon$, where $\varepsilon = 500$. The reason it can't be the case that both probablities can't hold is because the epsilon must be constant between the two as the variance is constant between elements in the dataset.[1]

(5) Consider the following deductions.

$$\begin{aligned} nb_0 + (\sum x_i)b_1 &= n(\frac{\sum y_i - \beta^{hat}\sum x_i}{n}) + \sum x_i \cdot \left(\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\right) \\ &= \sum y_i + \sum x_i \cdot \left(\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} - \hat{\beta}_1\right) \\ &= \sum y_i + \sum x_i(\hat{\beta}_1 - \hat{\beta}_1) \\ &= \sum y_i \end{aligned}$$

$$Ididn't getaroundtoprovingthis.$$

(6) Let $\bar{x}$ and $\bar{y}$ be the respective means of points (x, y) modeled by a linear regression of the form $Y = \beta_0 + \beta_1 x + \varepsilon$. Note, that for n samples, $\bar{x} = \frac{1}{n}\sum(\hat{x}_i + \varepsilon_i)$ and $\bar{y} = \frac{1}{n}\sum(\hat{y}_i + \varepsilon_i)$. Note, that the epsilons eventually sum to zero[2]. So, $\bar{x} = \hat{x}$ and $\bar{y} = \hat{y}$. So, the approximation given by the regression $(\hat{x}, \hat{y})$ passes through the point average $(\bar{x}, \bar{y})$.

[1] I don't think this response makes any sense.

[2] Somehow. This has to work for the approximation to equal the true mean