Lagarde H/W 5 3/15

1) The only conclusions we can draw from this study are associations. It is an observational study, not a controlled one. There are several socio-economic factors that could matter here, such as education level, ZIP code, family wealth, etc.

2)

a) $H_0: \mu_0 = \mu_1 = \mu_2 = \mu_3$

$H_A: \mu_n \neq \mu_K$, WLOG

$$S_w^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_2^3 + (n_4-1)s_4^2}{n-k}$$

$$= \frac{72(1.61)^2 + 104(1.43)^2 + 239(1.24)^2 + 1079(1.31)^2}{1498-4}$$

$$= \frac{2618.46}{1494} = 1.75$$

$$S_B^2 = \frac{n_1(\bar{x}_1-\bar{x})^2 + n_2(\bar{x}_2-\bar{x})^2 + n_3(\bar{x}_3-\bar{x})^2 + n_4(\bar{x}_4-\bar{x})^2}{k-1}$$

$= 19.06$, given that

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + n_4\bar{x}_4}{n} = \frac{(73)6.22 + 105(5.81) + 240(5.77) + 1080(5.47)}{1498}$$

$= 5.58$

Then, to calculate out the test itself, we have $\rightarrow$

Lagarde

2) $F = \frac{S_B^2}{S_W^2} = \frac{19.06}{1.75} = 10.89$, w/ $df_1 = 3$, $df_2 = 1494$,

b) Thus $df(10.89, 3, 1494)$ gives $p = 6.36 \times 10^{-7} < \alpha = .05$, so we reject H_0 and find that there are differences between the mean cholesterol levels of the groups.

c) that our samples are random, our samples $\sim N(\mu, \sigma^2)$, and that they are independent within the population.

d) from R, we have that

$H_0: \mu_1 + \mu_2 + \mu_3 - 3\mu_4 = 0$

$H_A: \mu_1 + \mu_2 + \mu_3 - 3\mu_4 \neq 0$

has: estimate, se, t test, p value

1.39 .250 5.55 3.32×10^{-8}.

Since $3.32 \times 10^{-8} < .05$, reject H_0. Under Bonferri; $\alpha^* = \frac{.05}{\binom{4}{2}} = \frac{.05}{6} = .0083$. Hence

$\alpha^* \not> p$, we still reject. Using Scheffé,

per lecture notes, we can consider: $t_{n-k, \frac{\alpha}{2}}$ as a cutoff point. Hence, $t_{1494, .025} = -1.96$ and since $5.55 > -1.96$, we reject.

Lagarde

e) again, per R we have

estimate	se	t test	p value
5.36	.164	3.27	.0012

because $.0012 < .05$, we reject H_0. Under Bonferri, we have, again that $\alpha^* = .0083$. Because $\alpha^* > p$, we again reject. Under Scheffé, we see that

$t_{1494, .025} = -1.61$, hence as $-1.61 < 3.27$ we still reject.

3) see pdf - typed

Lagarde

4)

a) with tox as blocking:

	Df	Sum Sq	Mean Sq	F-val	Pr(>F)
sex	1	48	48.25	.367	.546
tox	1	67	66.76	.508	.478
residuals	97	12758	131.53		

w/o blocking:

sex	1	48	48.25	.369	.545
residuals	98	12825	130.87		

b) p-value w/ blocking: .546
w/o blocking: .545

c) yes, in both cases, if the + samples have the same variance as what they are being compared to.

5) Lagarde H/w

a)

	Df	Sum-Sq	Mean-Sq	F-Val	Pr(>F)
size	3	.383	.1275	4.705	.00716
residuals	36	.976	.0271		

because $p < \alpha = .05$, we ftr, concluding that the size of the box affects how long I take to click on it.

b) HSD:

	diff	lwr	upr	p adj
Medium-Large	.0580	-.140	.256	.859
Small-Large	.1367	-.061	.335	.266
XLarge-Large	-.1312	-.329	.061	.298
Small-Medium	.0787	-.119	.277	.710
XL - Medium	-.1892	-.387	.009	.066
XL - Small	-.2679	-.466	-.069	.004

Pairwise-T:

	Large	Medium	Small
Medium	.43599	—	—
Small	.07155	.29222	
Large	.08319	.01446	.00085

HW #5, Problem 3 — Paul Lagarde

Pairwise comparisons using t tests with pooled SD

```
  5      6       7      8
6 1.0000 -       -      -
7 0.0015 4.4e-06 -      -
8 0.0011 2.9e-06 1.0000 -
9 1.0000 0.0625  0.1399 0.1088
```

P value adjustment method: bonferroni

```
>
> pairwise.t.test(Ozone, Month, p.adjust.method = "fdr")
```

Pairwise comparisons using t tests with pooled SD

data: Ozone and Month

```
  5       6       7       8
6 0.19069 -       -       -
7 0.00037 2.2e-06 -       -
8 0.00035 2.2e-06 0.92620 -
9 0.19069 0.01251 0.01998 0.01813
```

P value adjustment method: fdr

```
>
> airquality$Month<-as.factor(airquality$Month)
> p<-aov(Ozone~Month,data=airquality)
> TukeyHSD(p,conf.level = 0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Ozone ~ Month, data = airquality)

$Month
          diff         lwr       upr     p adj
6-5 -10.9731183 -32.27095900 10.324722 0.6139469
7-5  29.7741935   8.65164668 50.896740 0.0013894
8-5  30.4838710   9.36132410 51.606418 0.0009868
9-5  10.5935484 -10.70429233 31.891389 0.6454439
7-6  40.7473118  19.44947111 62.045153 0.0000044
8-6  41.4569892  20.15914853 62.754830 0.0000029
9-6  21.5666667   0.09496314 43.038370 0.0484120
8-7   0.7096774 -20.41286945 21.832224 0.9999830
9-7 -19.1806452 -40.47848588  2.117196 0.0990957
9-8 -19.8903226 -41.18816330  1.407518 0.0795001
```

So we see that for

Bonferri: May and July, May and August, July and September, and August and September have different levels of mean Ozone

Tukey: May and July, May and August, July and September, and August and September

FDR: May and July, May and August, July and September, August and Setpembter, June and July, and June and August.