Homework #9

1 a) $S^2 = \frac{SSE}{n-(k+1)} = \frac{1107.01}{11} = 100.637$

$n = 15$, $\hat{\beta}_2 = -0.92$

$df = 15 - 4 = 11$

$V(\hat{\beta}_2) = c_{22}\sigma^2 \approx \frac{8.1}{10000}(100.637) = 0.081516$

$\sqrt{V(\hat{\beta}_2)} \approx 0.2855$

$H_0: \beta_2 = 0$

$H_a: \beta_2 < 0$

$t = \frac{\hat{\beta}_2 - 0}{\sqrt{V(\hat{\beta}_2)}} = \frac{-0.92}{0.2855} = -3.22$

$p = Prob(t < -3.22 \mid df = 11)$

Using table, $0.005 > p > 0.0025$

Conclusion: small p-value, reject H_0.
likelihood is that $\beta_1 < 0$.
i.e. β_1 is important to the model

b) $\hat{y}(914, 65, 6) = 39.9812$

$t_{0.025}(11) = 2.201$

Let $\vec{a} = \langle 1, 914, 65, 6 \rangle$

By above $S = \sqrt{100.637}$ and

$$\hat{y} \pm t_{0.025}(11)\, S \sqrt{a^T (X^T X)^{-1} a}$$

$$39.9812 \pm 2.201 \sqrt{100.637} \sqrt{\frac{927662}{10000}}$$

$$39.9812 \pm 212.664$$

$$(-172.683,\ 252.645)$$

2. $k=3$

a)
$n_1 = 3$, $\bar{\mu}_1 = 2$
$n_2 = 2$, $\bar{\mu}_2 = 5$
$n_3 = 3$, $\bar{\mu}_3 = 6$

$$\bar{\mu}_0 = \frac{1}{8}(3\cdot 2 + 2\cdot 5 + 3\cdot 6) = 4.25$$

$$SSA = \sum n_i(\bar{\mu}_i - \bar{\mu}_0)^2$$
$$= 3(-2.25)^2 + 2(0.75)^2 + 3(1.75)^2$$
$$= 25.5$$

$$SSW = \sum_{i=1}^{3} \sum_{j=1}^{n_i} (x_{ij} - \bar{\mu}_i)^2$$
$$= (1-2)^2 + (3-2)^2 + (2-2)^2 + (6-5)^2 + (4-5)^2$$
$$+ (6-6)^2 + (8-6)^2 + (4-6)^2$$
$$= 12.$$

b)

	SS	df	MS	F_{obs}	p-value
among	25.5	2	12.75	5.3125	0.0579 by R
within	12	5	2.4		
total	37.5	7			

c) Yes. $p = 0.0579 < 0.10$
At the significance level, we reject H_0.

d) This is silly Q because I forgot to say "Use Table 7"
By Table 7, $0.100 > P(F > 5.3125) > 0.050$

3. a) $\hat{\mu}_0 = \frac{n_1 \hat{\mu}_1 + n_2 \hat{\mu}_2}{n_1 + n_2} = \frac{50 \cdot 26 + 80 \cdot 25}{130} = 25.3846$

$$SSA = \sum_1^2 n_i (\hat{\mu}_i - \hat{\mu}_0)^2$$
$$= 50(26 - 25.38)^2 + \cancel{(80)(25} \; 80(25 - 25.38)^2$$
$$= 30.7692$$

$$SSW = \sum_1^2 (n_i - 1) S_i^2$$
$$= 49(3.5)^2 + (79)3^2$$
$$= 1311.25$$

b)

	SS	df	MS	F_{obs}	$P(>F)$
among	30.7692	1	30.7692	3.00359	0.08548
within	1311.25	128	10.2441		

Note $n = 130$, $k = 2$

At the 5% level, we can not reject the null hypothesis. That is, there is no statistical evidence that the prep course helped student outcomes on the SAT.

c) $df = 128$ $\gg 30$ We can use z-scores.

$$(\bar{\mu}_1 - \bar{\mu}_2) \pm 1.960 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$1 \pm 1.960 \sqrt{\frac{(3.5)^2}{50} + \frac{3^2}{80}}$$

$$1 \pm 1.17191 \quad \text{or} \quad (-0.1719, 2.1719)$$

The true difference in means between the two cohorts is between -0.1719 and 2.1719 points with 95% likelihood.

d) $\mu_1 - \mu_2 = 0$ lies within this interval. This is equivalent to rejecting the 2-sided test at the 5% level.

4a)

```
> scores=c(45,47,67,61,56,72,39,39,43,46,59,27,60,42,55,44,30,18,56,61,60,47,69,42)
> treat=c(rep("A",6),rep("B",6),rep("C",6),rep("D",6))
> table=data.frame(scores,treat)
> results=aov(scores~treat, data=table)
> summary(results)
          Df Sum Sq Mean Sq F value Pr(>F)
treat      3   1380   460.2   3.251 0.0434 *
Residuals 20   2831   141.6
```

4b) At the 5% significance level, the p-value of 0.0434 would indicate that the mean starch level in the crop is not the same throughout all four plots. (At the 1% level, we would conclude the opposite.)

5. $L(\text{all } Y_{ij} \mid \hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_k, S_a^2)$

$$= \prod_{i=1}^{k} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi S_a^2}} \exp\left(-\frac{(Y_{ij}-\hat{\mu}_i)^2}{2S_a^2}\right)$$

$$= \left(\frac{1}{2\pi S_a^2}\right)^{n/2} \exp\left(\frac{-1}{2S_a^2} \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_k}(Y_{ij}-\hat{\mu}_i)^2}_{\text{define this as the stat } T.}\right)$$

$$= \left(\frac{1}{2\pi S_a^2}\right)^{n/2} \exp\left(\frac{-T}{2S_a^2}\right).$$

easier to differentiate $\ln L$.

$$\ln L = -\frac{n}{2}\ln(2\pi S_a^2) - \frac{T}{2S_a^2}$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln S_a^2 - \frac{T}{2S_a^2}.$$

Then $\dfrac{d\ln L}{d(S_a^2)} = \dfrac{-n}{2S_a^2} + \dfrac{T}{2(S_a^2)^2}$

$$\frac{d\ln L}{d(S_a^2)} = 0 \longleftrightarrow \frac{-n}{2S_a^2} + \frac{T}{2(S_a^2)^2} = 0$$

$$-nS_a^2 + T = 0$$

$$\Rightarrow S_a^2 = \frac{T}{n} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_k}(Y_{ij}-\hat{\mu}_i)^2,$$

as stated in class.

6. a) Recall from 325, $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\,dx$.

We have $\int_0^\infty x^{\frac{n}{2}-2} e^{-x/2}\,dx = I$

Let $y = x/2 \rightarrow 2y = x$ and $2dy = dx$.

So $I = \int_0^\infty (2y)^{\frac{n}{2}-2} e^{-y} \cdot 2\,dy$

$= 2^{\frac{n}{2}-1} \int_0^\infty y^{\frac{n}{2}-2} e^{-y}\,dy$

$= 2^{\frac{n}{2}-1} \int_0^\infty y^{z-1} e^{-y}\,dy$ when $z = \frac{n}{2}-1$.

$= 2^{\frac{n}{2}-1}\,\Gamma\left(\frac{n}{2}-1\right)$.

b) For $X \sim \chi^2(n) = \Gamma\left(\frac{n}{2}, 2\right)$, the associated PDF is

$f(x) = \dfrac{x^{\frac{n}{2}-1} e^{-x/2}}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}}$.

Then $E\left[\frac{1}{X}\right] = \int_0^\infty \frac{1}{x} f(x)\,dx$

$= \int_0^\infty \dfrac{x^{\frac{n}{2}-2} e^{-x/2}}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}}\,dx$

$= \dfrac{1}{\Gamma(n/2)\,2^{n/2}} \int_0^\infty x^{\frac{n}{2}-2} e^{-x/2}\,dx$

by (a) $= \dfrac{2^{n/2-1}\Gamma(\frac{n}{2}-1)}{\Gamma(n/2)\,2^{n/2}} = \dfrac{1}{2} \cdot \dfrac{(\frac{n}{2}-1-1)!}{(\frac{n}{2}-1)!}$

$= \dfrac{1}{2} \cdot \dfrac{1}{\frac{n}{2}-1} = \dfrac{1}{n-2}.$

c) $E\left[\dfrac{X_1/m}{X_2/n}\right] = E\left[\dfrac{n}{m} X_1 \dfrac{1}{X_2}\right]$

$= \dfrac{n}{m} E(X_1) E\left(\dfrac{1}{X_2}\right)$ by independence

$= \dfrac{n}{m}\ m \cdot \dfrac{1}{n-2}$ by (b)

$\uparrow$ by properties of $\chi^2(m)$

$= \dfrac{n}{n-2}.$