STAT 510
FINAL EXAM SOLUTIONS
SPRING 2020

1. a) 7
   b) 5
   c) 7
   d) 5

2. 10

3. 11

4. a) 6
   b) 6
   c) 6
   d) 6

5. a) 10
   b) 8
   c) 8
   d) 5

1 a) LET $\overline{y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_{i\cdot})^2$$

FROM OUR NOTES, WE KNOW $(n_i - 1)\frac{S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}$.

THUS, $\frac{S_i^2}{\sigma_i^2} = \frac{(n_i - 1) S_i^2}{\sigma_i^2} \Big/ (n_i - 1) \sim \frac{\chi^2_{n_i - 1}}{n_i - 1}$

BECAUSE $y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}$ ARE ALL INDEPENDENT,

$\frac{S_1^2}{\sigma_1^2}$ INDEPENDENT OF $\frac{S_2^2}{\sigma_2^2}$.

THUS, $\frac{S_1^2}{\sigma_1^2} \Big/ \frac{S_2^2}{\sigma_2^2} \sim F_{n_1 - 1, n_2 - 1}$.

UNDER $H_0: \sigma_1^2 = \sigma_2^2$, $S_1^2 / S_2^2 \sim F_{n_1 - 1, n_2 - 1}$.

THUS, TEST STATISTIC IS S_1^2 / S_2^2.

1 b) UNDER $H_A: \sigma_1^2 = 5\sigma_2^2$,

$$S_1^2 / S_2^2 = \frac{S_1^2/\sigma_1^2}{S_2^2/5\sigma_2^2}$$

$$= 5 \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

$$\sim 5 \, F_{n_1-1,\, n_2-1}$$

THIS IS A SCALED F DISTRIBUTION.
(NOT A NON-CENTRAL F).

1 c) BECAUSE σ_1^2 IS NOT NECESSARILY EQUAL TO σ_2^2, WE NEED TO ACCOUNT FOR THE POTENTIAL OF UNEQUAL VARIANCE. THIS IS ESPECIALLY IMPORTANT WHEN SAMPLE SIZES n_1 AND n_2 MIGHT DIFFER.

$$t = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

THIS IS THE DEFAULT TEST STATISTIC USED BY t.test IN R.

1 d) $\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$ IS A LINEAR COMBINATION OF THE MEAN SQUARES S_1^2 AND S_2^2.

SATTERTHWAITE'S APPROXIMATION GIVES

$$DF = \frac{\left(\frac{1}{n_1} S_1^2 + \frac{1}{n_2} S_2^2\right)^2}{\frac{S_1^4}{n_1^2} / (n_1 - 1) + \frac{S_2^4}{n_2^2} / (n_2 - 1)}$$

2. AS WE HAVE SEEN IN NOTES, THE MLE OF λ IS $\hat{\lambda} = \overline{y}.$, WHICH IS GIVEN AS 10.2 IN THIS CASE.

THE LOG LIKELIHOOD IS

$$\ell(\lambda) = \log(\lambda) \sum_{i=1}^{100} y_i - 100\lambda - \sum_{i=1}^{100} \log(y_i!)$$

$$= \log(\lambda) * 100 * 10.2 - 100\lambda - \sum_{i=1}^{100} \log(y_i!)$$

$$= 1020 * \log(\lambda) - 100 * \lambda - \sum_{i=1}^{100} \log(y_i!)$$

THE INTERVAL IS

$$\{\lambda : \ell(\hat{\lambda}) - \ell(\lambda) \leq \tfrac{1}{2} \chi^2_{1,.95}\}$$

NOTE THE $\sum_{i=1}^{100} \log(y!)$ WILL CANCEL IN $\ell(\hat{\lambda}) - \ell(\lambda)$. WE CAN WRITE A FUNCTION IN R, $f = \text{function}(x)\{1020 * \log(x) - 100 * x\}$, AND FIND VALUES $x_1 < 10.2 < x_2$ SUCH THAT

$$f(10.2) - f(x_1) = f(10.2) - f(x_2) = 0.5 * \text{qchisq}(.95, 1)$$

2. (CONTINUED)

WHILE THIS CAN BE DONE IN SOPHISTICATED WAYS, IT TAKES LITTLE TIME BY TRIAL AND ERROR BECAUSE WE KNOW WHERE TO LOOK. A WALD INTERVAL IS

$$10.2 \pm 1.96\sqrt{10.2/100} \iff (9.57, 10.83)$$

THUS, x_1 SHOULDN'T TOO FAR FROM 9.57 AND x_2 SHOULDN'T TOO FAR FROM 10.83.

$$x_1 \cong 9.587 \quad \& \quad x_2 \cong 10.838$$

NOTE THAT THIS PROBLEM IS SIMILAR TO PROBLEM 5 b) ON HW 10. CODE YOU WROTE FOR THIS HOMEWORK PROBLEM COULD HAVE BEEN USEFUL FOR THIS PROBLEM.

3.
$$\left.\begin{array}{l} t \cong -0.63 \\ \text{P-VALUE} \cong 0.55 \end{array}\right\} \text{SEE CODE}$$

THE MEAN YIELD FOR GENOTYPE 3 IS NOT SIGNIFICANTLY DIFFERENT FROM THE AVERAGE OF THE MEAN YIELDS FOR GENOTYPES 1 + 2.

4. a)
$$\text{VAR}(\bar{y}_{\cdot 11}) = \text{VAR}(\bar{d}_{\cdot} + \bar{e}_{\cdot 11})$$
$$= \sigma_d^2/5 + \sigma_e^2/5$$

b)
$$\text{VAR}(\bar{y}_{\cdot 1 \cdot}) = \sigma_d^2/5 + \text{VAR}(\bar{e}_{\cdot 1 \cdot})$$
$$= \sigma_d^2/5 + \text{VAR}\left(\frac{1}{20} \sum_i \sum_k e_{ik}\right)$$
$$= \sigma_d^2/5 + \frac{1}{400}\, 5\sigma_e^2\, \underset{\sim}{1}' \Sigma_e \underset{\sim}{1}$$
$$= \sigma_d^2/5 + \frac{1}{80}\, \sigma_e^2 \left[4 + 6\phi + 4\phi^2 + 2\phi^3\right]$$
$$= \sigma_d^2/5 + \sigma_e^2\, \frac{2 + 3\phi + 2\phi^2 + \phi^3}{40}$$

4c) $\mathrm{VAR}(\bar{y}_{\cdot 11} - \bar{y}_{\cdot 21}) = \mathrm{VAR}(\bar{e}_{\cdot 11} - \bar{e}_{\cdot 21})$

$$= \mathrm{VAR}(\bar{e}_{\cdot 11}) + \mathrm{VAR}(\bar{e}_{\cdot 21})$$

$$= \frac{\sigma_e^2}{5} + \frac{\sigma_e^2}{5} = 2\sigma_e^2/5$$

4d) $\mathrm{VAR}(\bar{y}_{\cdot 11} - \bar{y}_{\cdot 14}) = \mathrm{VAR}(\bar{e}_{\cdot 11} - \bar{e}_{\cdot 14})$

$$= \mathrm{VAR}\left[\frac{1}{5}\sum_{i=1}^{5}(e_{i11} - e_{i14})\right]$$

$$= \frac{\sigma_e^2 + \sigma_e^2 - 2\,\mathrm{cov}(e_{i11}, e_{i14})}{5}$$

$$= \frac{2\sigma_e^2 - 2\sigma_e^2\phi^3}{5}$$

$$= \frac{2\sigma_e^2(1-\phi^3)}{5}$$

5 a) THIS PROBLEM IS ASKING WHETHER THE SLOPE COEFFICIENT ON AMOUNT OF DRUG IS ZERO. THE REDUCED MODEL CORRESPONDING TO THIS NULL HYPOTHESIS IS m0. THUS, A LIKELIHOOD RATIO TEST TO COMPARE MODELS m0 AND m1 CAN BE CONDUCTED.

$$-2 \log \Lambda = 2(-194.4344 + 197.3743)$$

$$= 2(2.9399) = 5.8798$$

$$\text{p-VALUE} = 1 - \text{pchisq}(5.8798, 1)$$

$$= 0.015$$

THE ESTIMATED SLOPE COEFFICIENT ON AMOUNT OF DRUG IS SIGNIFICANTLY DIFFERENT FROM ZERO.

5 b) HOLDING THE AMOUNT OF DRUG CONSTANT, PIGS FED WITH DIET 2 HAVE MEAN LITTER SIZE $\exp(0.15235) \simeq 1.165$ TIMES GREATER THAN THE MEAN LITTER SIZE OF PIGS FED DIET 1.

WHETHER PIGS ARE FED DIET 1 OR DIET 2, AN INCREASE IN 1 UNIT OF THE THE DRUG IS ASSOCIATED WITH A 2.06% INCREASE IN MEAN LITTER SIZE.

($2.06\% = 100 \times [\exp(0.02038) - 1]$)

5c) SEE SLIDE 9 OF SLIDE SET 29.

$$\mathrm{VAR}(y_{213}) = E(y_{213}) + \left[\exp(\sigma_p^2) - 1\right]\left[E(y_{213})\right]^2$$

$$E(y_{213}) = \exp(\mu + \alpha_2 + \beta_3 + \gamma_{23} + \sigma_p^2/2)$$

$$\hat{E}(y_{213}) = \exp(2.49811 + 0.07063 + -0.04652 + 0.12786 + 0.09924^2/2)$$

$$\cong 14.225$$

$$\hat{\mathrm{VAR}}(y_{213}) \cong 16.23$$

5d)

MODEL	BIC
m 0	$394.7486 + 3\log(72) = 407.58$
m 1	$388.8687 + 4\log(72) = 405.98$
m 2	$385.8227 + 9\log(72) = 424.31$

MODEL m1 IS PREFERRED.

```
####Problem 2 Code
> f = function(lambda){
+        1020 * log(lambda) - 100 * lambda
+ }
>
> x1candidates = seq(9.57, 9.59, by = 0.001)
> x2candidates = seq(10.82, 10.84, by = 0.001)
> cbind(x1candidates, f(10.2) - f(x1candidates)- 0.5*qchisq(.95, 1))
      x1candidates
 [1,]        9.570  0.108875712
 [2,]        9.571  0.102298208
 [3,]        9.572  0.095731839
 [4,]        9.573  0.089176602
 [5,]        9.574  0.082632496
 [6,]        9.575  0.076099518
 [7,]        9.576  0.069577665
 [8,]        9.577  0.063066935
 [9,]        9.578  0.056567327
[10,]        9.579  0.050078837
[11,]        9.580  0.043601463
[12,]        9.581  0.037135204
[13,]        9.582  0.030680056
[14,]        9.583  0.024236017
[15,]        9.584  0.017803085
[16,]        9.585  0.011381258
[17,]        9.586  0.004970533
[18,]        9.587 -0.001429091
[19,]        9.588 -0.007817618
[20,]        9.589 -0.014195049
[21,]        9.590 -0.020561388
> cbind(x2candidates, f(10.2) - f(x2candidates)- 0.5*qchisq(.95, 1))
      x2candidates
 [1,]       10.820 -0.1094536010
 [2,]       10.821 -0.1037191156
 [3,]       10.822 -0.0979759193
 [4,]       10.823 -0.0922240136
 [5,]       10.824 -0.0864634002
 [6,]       10.825 -0.0806940806
 [7,]       10.826 -0.0749160566
 [8,]       10.827 -0.0691293297
 [9,]       10.828 -0.0633339014
[10,]       10.829 -0.0575297735
[11,]       10.830 -0.0517169475
[12,]       10.831 -0.0458954250
[13,]       10.832 -0.0400652076
[14,]       10.833 -0.0342262970
[15,]       10.834 -0.0283786947
[16,]       10.835 -0.0225224024
```

```
[17,]       10.836 -0.0166574216
[18,]       10.837 -0.0107837540
[19,]       10.838 -0.0049014011
[20,]       10.839  0.0009896354
[21,]       10.840  0.0068893540


####Problem 3 Code
> y=c(16.3, 25.6, 22.9,
+     19.3, 19.8, 20.3,
+     22.3, 32.2, 28.1,
+     11.1, 16.7, 14.4,
+     17.0, 27.8, 22.1)
> block = factor(rep(1:5, each = 3))
> geno = factor(rep(1:3, 5))
> o=lm(y ~ block + geno)
> summary(o)

Call:
lm(formula = y ~ block + geno)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3600 -1.0500  0.0000  0.8467  3.3600 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   17.740      1.485  11.947 2.22e-06 ***
block2        -1.800      1.775  -1.014 0.340166    
block3         5.933      1.775   3.343 0.010183 *  
block4        -7.533      1.775  -4.245 0.002819 ** 
block5         0.700      1.775   0.394 0.703578    
geno2          7.220      1.375   5.252 0.000772 ***
geno3          4.360      1.375   3.172 0.013164 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.174 on 8 degrees of freedom
Multiple R-squared:  0.9165,	Adjusted R-squared:  0.8539 
F-statistic: 14.64 on 6 and 8 DF,  p-value: 0.0006351

> b = coef(o)
> v = vcov(o)
> cc = c(0,0, 0, 0, 0, 0.5, -1)
> tstat = t(cc) %*% b / sqrt(t(cc) %*% v %*% cc)
> tstat
          [,1]
[1,] -0.629963
```

```
> pvalue = 2 * (1 - pt(abs(tstat),  8))
> pvalue
          [,1]
[1,] 0.5462863
>
```