

DIRECTIONS: Write your solutions in a single `.tex` file, including R code. Your `.tar` package will consist of that file, its output `.pdf` file, and a separate file for each problem requiring R code, each such file named in the form `x.R`, where `x` is the problem number. Name your `.tar` file as you did in the homework. Your submission must be in the `132groupquiz` directory in `handin`; it must be **timestamped on or before 11:50 a.m. NO LATE SUBMISSIONS; keep submitting the work you have, as you go along, so that you at least have something turned in.** You are not necessarily expected to solve all the problems.

1. (30) This problem involves the built-in dataset `ToothGrowth` in R. Type

```
> ToothGroth # print data frame
> ?ToothGroth # learn more about it
```

to get acquainted.

Fit the model

$$\text{mean length} = \beta_0 + \beta_1 J + \beta_2 \text{dosage} \quad (1)$$

where J is a dummy variable for OJ.

Do NOT include your code in a `.R` file, but DO show ALL your code, including data set up, and the output in your `.tex` file.

Interpret the results.

2. (25) I flip a coin twice, getting X heads. I then flip it X more times, with Y heads among these flips. Define $T = X + Y$. Find $p_T(2)$, with a clean, clear derivation.

3. (25) Write a function with “declaration”

```
pval <- function(x, p0)
```

that will return the approximate p-value for a significance test, as follows.

Here \mathbf{x} is a vector of 0s and 1s, where for instance 1 could mean, Yes, plan to vote for Smith, with 0 meaning No, won't vote for her. We will test the hypothesis

$$H_0 : p = p_0 \quad (2)$$

where p is the true (but unknown) population proportion of 1s. The alternate hypothesis is

$$H_A : p \neq p_0 \quad (3)$$

Example:

```
> set.seed(168)
> x <- sample(0:1, 500, prob=c(0.2, 0.8),
             replace=TRUE)
> pval(x, 0.75)
[1] 0.009823275
```

4. (20) Write a function with “declaration”

```
qqqinv <- function(lmout)
```

where `lmout` is an object of class `lm` (i.e. the value returned from a call to `lm()`), which will return the matrix $(Q'Q)^{-1}$.

Hint: Here you'll need to explore an actual R `lm` object.

Solutions:

1.a

```
> summary(lm(len ~ ., data=tg))

Call:
lm(formula = len ~ ., data = tg)

Residuals:
    Min       1Q   Median       3Q      Max
-6.600 -3.700  0.373  2.116  8.800

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.5725     1.2824   4.345 5.79e-05 ***
supp         3.7000     1.0936   3.383  0.0013 **
dose         9.7636     0.8768  11.135 6.31e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.236 on 57 degrees of freedom
Multiple R-squared:  0.7038,    Adjusted R-squared:  0.6934
F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

Each increase of one unit of the supplement is estimated to result in an increase of 3.7 in mean tooth length. Orange juice has an estimated impact of 9.76. Statements should NOT be made along the lines of “Dose is more significant than OJ.”

2.

$$p_T(2) = P(T = 2) \quad (\text{def. of } p_Y) \quad (4)$$

$$= \sum_{i=0}^2 P(X = i, Y = 2 - i) \quad (2.5) \quad (5)$$

$$= \sum_{i=0}^2 P(X = i) P(Y = 2 - i) \quad (2.5) \quad (6)$$

$$= \sum_{i=0}^2 C_{2,i} 0.5^2 \cdot C_{i,2-i} 0.5^i \quad (\text{binom. pmf}) \quad (7)$$

3.

```
pval <- function(x,p0) {
  estp <- mean(x)
  n <- length(x)
  z <- (estp - p0) / sqrt(p0 * (1-p0) / n)
  2 * (1 - pnorm(abs(z)))
}
```

4. Lots of ways to do this, such as

```
qqqinv <- function(lmout) {
  n <- length(lmout$residuals)
  p1 <- lmout$rank
  s2 <- sum(lmout$residuals^2) / (n-p1)
  vcov(lmou) / s2
}
```