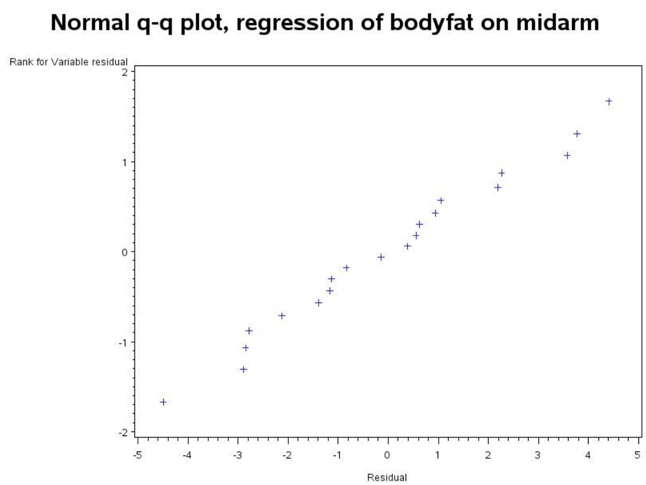
ST 512 Homework assignment #2
Dr. Jason A. Osborne

1. Obtain the residuals from a simple linear regression of bodyfat percentage on midarm circumference ~~(x_2)~~ x_3 using the dataset "`bodyfat.dat`". Obtain a plot of the empirical quantiles of the residuals (sorted residuals) against the corresponding quantiles from the standard normal distribution. That is, obtain a normal q-q plot of the residuals from the simple linear regression of bodyfat on midarm. In computing the normal quantiles, indicate how you avoided having to compute either the 0^{th} or the 100^{th} quantile of the standard normal distribution.



PROC RANK with the option NORMAL=VW is one way to obtain the corresponding normal quantiles for any sample of residuals. This option finds the $\{i/(n+1)\}$ normal quantiles for $i = 1, \ldots, n$.

```
proc reg data=bodyfat;
   model y=x2;
   output out=predz p=yhat r=residual;
run;
proc rank data=predz normal=vw out=bodyfat2;
   var residual;
   ranks normalscores;
run;
goptions device=jpeg;
run;
proc gplot data=bodyfat2;
   title "Normal q-q plot, regression of bodyfat on midarm";
   plot normalscores*residual;
run;
```

2. Rao 8.3 and Rao 8.9

   (a)

$$\begin{aligned} y_{1+} &= 8.93 \\ y_{2+} &= 13.75 \\ y_{3+} &= 18 \end{aligned}$$

   (b)

$$y_{++} = 40.68$$

   (c)

$$\begin{aligned} \bar{y}_{1+} &= 2.23 \\ \bar{y}_{2+} &= 3.44 \\ \bar{y}_{3+} &= 4.5 \end{aligned}$$

$$\begin{aligned} s_1^2 &= 0.91 \\ s_2^2 &= 0.21 \\ s_3^2 &= 0.54 \end{aligned}$$

3. Rao 8.8

   (a) $df_{trt} = 3, MS(trt) = 42, SS(E) = 320, SS(TOT) = 446, F_{obs} = 2.625, F(.05, 3, 20) = 3.089$

   (b) $t = 4$

   (c) $n_i \equiv 6$, since balanced.

   (d) No evidence to reject the (null) hypothesis that mean is constant across four treatments.

4. (Rao 12.3a)

```
                         The GLM Procedure

                                       Sum of
Source                     DF         Squares    Mean Square   F Value   Pr > F
Model                       2     1752.266667     876.133333      6.71   0.0043
Error                      27     3527.600000     130.651852
Corrected Total            29     5279.866667
```

5. (a) (Rao 12.5a)

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + E_i \text{ for } i = 1, \ldots, 20$$

   where the indicator variables x_{ij} are defined by

$$x_{ij} = \begin{cases} 1 & \text{seedling } i \text{ receives treatment } j \\ 0 & \text{else} \end{cases}$$

(b) (Rao 12.5b) for $i = 1, 2, 3, 4$ we have $\beta_0 + \beta_i$ as the mean plant height for treatment i and β_0 as the mean for treatment 5.

(c) (Rao 12.5) - estimated regression parameters:

$$\begin{aligned}
\hat{\beta}_0 + \hat{\beta}_1 &= 34.02 \\
\hat{\beta}_0 + \hat{\beta}_2 &= 31.90 \\
\hat{\beta}_0 + \hat{\beta}_3 &= 30.84 \\
\hat{\beta}_0 + \hat{\beta}_4 &= 34.31 \\
\hat{\beta}_0 &= 34.29
\end{aligned}$$

Successively substituting means leads to

$$\hat{\beta} = \begin{pmatrix} 34.29 \\ -0.27 \\ -2.39 \\ -3.45 \\ 0.02 \end{pmatrix}$$

6. Referring to Rao 12.6b, the contrast described in part (i) is just $\theta_i = \frac{\mu_2+\mu_3}{2} - \frac{\mu_4+\mu_5}{2}$ the contrast described in part (ii) is just $\theta_{ii} = (\mu_2 - \mu_4) - (\mu_3 - \mu_5)$ Substitution of treatment means yields the following estimates:

Parameter	Estimate	Standard Error	t Value	Pr > |t|
i: A-B	-2.93000000	0.52243899	-5.61	<.0001
ii: source x intensity	1.04500000	1.04487798	1.00	0.3331

7. Rao 12.10

(a) Rao 12.10b

$$\mu(x_1, x_2, z) = \beta_0 + \beta_5 z$$

(b) Rao 12.10d - First $H_{0\tau}$ postulates all three adjusted treatment means equal, second $H_{0\tau}$ postulates first two adjusted treatment means equal.

(c) Rao 12.10e

$$\mu(x_1, x_2, z) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \beta_3 z$$

(d) Rao 12.10f

$$F = \frac{SS[R]_f - SS[R]_r}{MS[E]_f} = \frac{4899 - 3283}{14.6} = 110.4$$

Wow, that's big. Conclude treatments 1 and 2 different.

8. Rao 12.11d - use the expression below, with $x' = (1, 1, 0, 30)$

$$\hat{\beta}'x \pm t(0.025, 26)\sqrt{x'\hat{\Sigma}x}$$

(where $\hat{\Sigma} = (X'X)^{-1}MS[E]$) to get $52.8 \pm 2.06(1.21)$ or $(50.3, 55.3)$

9. Rao 12.18

   (a) Rao 12.18a

   $$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 z_i + E_i$$

   where x_{ij} is defined by

   $$x_{ij} = \begin{cases} 1 & \text{bag } i \text{ at location } j \\ 0 & \text{else} \end{cases}$$

   and $\{z_i\}$ denote initial weights and $\{E_i\}$ denote independent normally distributed error terms with common variance σ^2 for bags $i = 1, \ldots, 20$. For $i = 1, 2, 3, 4, \beta_i$ denotes the mean difference between location i and location 5, after controlling for initial weight z. β_5 denotes the mean increase in final weight per unit increase in z for fixed location. β_0 is the intercept of the (population) line describing the association between mean final weight and initial weight in location 5.

   (b) Rao 12.18b - let $\mu_j(z)$ denote mean final weight for initial weight z at location j:

   $$\begin{aligned} \hat{\mu}_1(z) &= 2.25 + 1.08z \\ \hat{\mu}_2(z) &= 2.21 + 1.08z \\ \hat{\mu}_3(z) &= 4.15 + 1.08z \\ \hat{\mu}_4(z) &= 3.60 + 1.08z \\ \hat{\mu}_5(z) &= 2.49 + 1.08z \end{aligned}$$

   (c) Rao 12.18c - by evaluating the expressions above at $\bar{z} = 27.7$, we get

   $$\begin{aligned} \hat{\mu}_1(\bar{z}) &= 30.15 \\ \hat{\mu}_2(\bar{z}) &= 30.12 \\ \hat{\mu}_3(\bar{z}) &= 32.05 \\ \hat{\mu}_4(\bar{z}) &= 31.50 \\ \hat{\mu}_5(\bar{z}) &= 30.40 \end{aligned}$$

   (d) Rao 12.18d - 90% c.i. for β_1 is $\hat{\beta}_1 \pm t(0.05, 14)SE(\hat{\beta}_1)$ ore $-0.24 \pm 1.76(0.5766)$ -0.24 ± 1.02 or $(-1.26, 0.77)$.

(e) Rao 12.18f - conclusion about $H_0 : \beta_5 = 0$ based on $t_{\beta_5} = 22.75$ which is significantly different from 0 ($p < 0.0001$). Therefore, final weight associated w/ initial weight.

(f) Rao 12.18g - 95% c.i. given by $\hat{\beta}_5 \pm t(0.025, 14)\sqrt{0.0075MS[E]}$ or $1.08 \pm 2.14(0.0476)$ or $(0.98, 1.19)$.

(g) Rao 12.18h - 95% p.i. for Y given $x' = (1, 0, 0, 1, 0, 25)$ given by $x'\hat{\beta} \pm t(0.025, 14)\sqrt{MS[E] + x'\hat{\Sigma}x}$ or $31.23 \pm 2.14\sqrt{0.30 + 0.08}$ or 31.23 ± 1.31 or $(29.9, 32.5)$.