58] For $Y = \beta_1 + \beta_2 x + e$

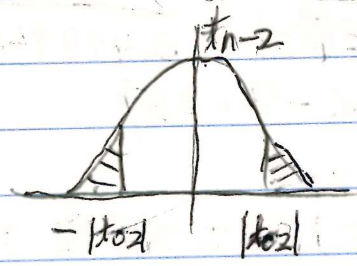
the Anova F test and t test for $H_0\ \beta_k = 0$ are equivalent since

$$(t_{n-2})^2 = F_{1,n-2}$$

and $$t_{02}^2 = \left[\frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}\right]^2 = F_0 .\ \text{So } 2P(t_{n-2} < -|t_{02}|)$$

$$= 2\,P(t_{n-2} > |t_{02}|) = P[t_{n-2}^2 > |t_0|^2]$$

$$= P(F_{1,n-2} > F_0) = pval.$$




$$pval = P(t_{n-2} > |t_0|) + P(t_{n-2} < |t_0|)$$
$$= 2\,P(t_{n-2} < -|t_0|)$$
$$= 2\,P(t_{n-2} > |t_0|).$$

ex] mussel data $Y = \log M$, $n = 82$

| | coef | SE | tvalue | pvalue |
|---|---|---|---|---|
| constant | -5.07459 | 1.85124 | -2.741 | 0.0076 |
| $x_2 = \log(S)$ | 0.573167 | 0.11646 | 4.922 | 0.0000 |
| $x_3 = \log(H)$ | 1.12399 | 0.49894 | 2.253 | 0.0270 |

a) Test $\beta_2 = 0$ $\quad H_0\ \beta_2 = 0 \cdot H_A\ \beta_2 \neq 0$

$t_{02} = 4.922$

$pval = 0.0$

(since pval < δ) reject $H_0$, log(S) is needed in the MLR model for log(M) given that log(H) is in the model.

b) Test $\beta_3 = 0$ using $\delta = 0.01$ (not 0.05)

$H_0\ \beta_3 = 0 \quad H_A\ \beta_3 \neq 0$

$t_{03} = 2.253$

pval $= 0.027$

Fail to reject $H_0$ since $.027 > \delta = .01$.
log(H) is not needed in the MLR model for log(M) given that log(S) is in the model.

Note that $t_{03} = \frac{\hat{\beta}_3}{SE(\hat{\beta}_3)} = \frac{1.12399}{0.49894} = 2.2528$

and pval $= 2P(\underbrace{t_{n-3}}_{79>30} < -2.25)$

$\approx 2P(Z < -2.25) = 2P(Z > 2.25)$

$= 2[1 - P(Z < 2.25)] = 2(1 - .9878) = .0244$

| | .05 |
|---|---|
| 2.2 | .9878 |

pvalue from output is more accurate

59] Usually use 90% 95% or 99% CI

| df>30 | | | |
|---|---|---|---|
| $\infty$ | 1.645 | 1.96 | 2.576 |
| percentile | .95 | .975 | .995 |
| CI | 90% | 95% | 99% |

ex] A 99% CI for $\beta_3$ is

$\hat{\beta}_3 \pm \underbrace{t_{n-p,\,1-\frac{.01}{2}}}_{\text{use } z_{0.995} \text{ since } n-3=79>30} SE(\hat{\beta}_3)$

$= 1.12399 \pm 2.576(.49894) = 1.12399 \pm 1.2853$

$= [-0.1613 \quad , 2.4093]$.

Since the 99% CI contains 0, fail to reject $H_0\ \beta_3 = 0$ if $\delta = 0.01$

ex] 

| df = $\nu = n-p$ | 90% | 95% | 99% |
|---|---|---|---|
| | .95 | .975 | .995 |
| 5 | 2.015 | 2.571 | 4.032 |
| 7 | 1.895 | 2.365 | 3.499 |
| 9 | 1.833 | 2.262 | 3.250 |
| $n-p > 30$ $\infty$ | 1.645 | 1.96 | 2.576 |
| | $t_{n-p,\,.95}$ | $t_{n-p,\,.975}$ | $t_{n-p,\,.995}$ |

§2.6 60] p45

Full model $Y = \beta_1 + \beta_2 x_2 + \dots + \beta_p x_p + e$

Reduced model $Y = \beta_1 + \beta_2 x_{i_2} + \dots + \beta_q x_{i_q} + e$

$\{i_2, i_3, \dots, i_q\} \subset \{2, 3, \dots, p\}$

61] p46 Know for final Partial F test

i) $H_0$ reduced model is good $H_A$ use the full model

ii) $F_R = \left[ \dfrac{SSE(R) - SSE(F)}{df_R - df_F} \right] / MSE(F)$

iii) pval $= P(F_{df_R - df_F,\, df_F} > F_R)$

iv) If pval $< \delta$, reject $H_0$, so use full model
If pval $\geq \delta$, fail to reject $H_0$, so reduced model is good

62] p46 full model

12.5

| source | df | SS | MS | $F_0$ | pval for |
|---|---|---|---|---|---|
| model or reg | $p-1$ | SSR | MSR | $F_0 = \frac{MSR}{MSE}$ | $H_0\ \beta_2 = \dots = \beta_p = 0$ |
| error or residual | $n-p = df_F$ | SSE(F) | MSE(F) | | |

reduced model

| source | df | SS | MS | $F_0$ | pval for |
|---|---|---|---|---|---|
| model or reg | $q-1$ | SSR | MSR | $F_0 = \frac{MSR}{MSE}$ | $H_0\ \beta_2 = \dots = \beta_q = 0$ |
| error or residual | $n-q = df_R$ | SSE(R) | MSE(R) | | |

63] R has useful output for the partial F test

```
full ← lm(Y~ x2 + x3 + x4 + x5 + x6 + x7)
red ← lm(Y~ x2 + x5 + x7)
anova(red, full)
```

output

model 1 $Y \sim x_2 + x_5 + x_7$

model 2 $Y \sim x_2 + x_3 + x_4 + x_5 + x_6 + x_7$

| | Res. Df | RSS | df | Sum of sq | F | Pr > F |
|---|---|---|---|---|---|---|
| 1 | $df_R$ | SSE(R) | | | | |
| 2 | $df_F$ | SSE(F) | $df_R - df_F$ | SSE(R) − SSE(F) | $F_R$ | pval |

$\underbrace{F_R \quad \text{pval}}_{\text{needed}}$

see HW3 D, ex 2.10, 2.11

ex] $Y$ = # women married to civilians in district

$x_2$ = pop of district in 1843

$x_3$ = # married civilian men

$x_4$ = # married military men

$x_5$ = # women married to military men

reduced Y and $X_3$

| full source | df | SS | MS | F | pval |
|---|---|---|---|---|---|
| reg | 4 | | | | |
| resid | $21 = df_F$ | 3241299 | 154348 | | |

| reduced source | df | SS | MS | F | pval |
|---|---|---|---|---|---|
| reg | 1 | | | | |
| resid | 24 | 3268017 | 136167 | | |

```
z <- as.data.frame(marry)
full <- lm (mwmn~pop+mmen+mmilmen+milwmn, data=z)
red <- lm (mwmn ~ mmen)
```

| | Res.Df | RSS | df | Sum of sq | F | Pr>F |
|---|---|---|---|---|---|---|
| 1 | 24 | 3268017 | | | | |
| 2 | 21 | 3241299 | 3 | 26718 | 0.0577 | 0.9813 |

$H_0$ reduced model is good $H_A$ use full model

$F_R = 0.0577$

pval = 0.9813

fail to reject $H_0$, the reduced model is good

see 53): F table If n-p > 60 use den df = ∞ otherwise use closer df.

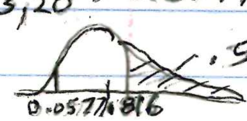
without R output

$$F_R = \left[\frac{SSE(R) - SSE(F)}{df_R - df_F}\right] \Big/ MSE(F) = \frac{\left[\frac{3268017 - 3241299}{24-21}\right]}{154348}$$

$$= \left(\frac{26718}{3}\right) \Big/ 154348 = 0.0577$$

$$pval = P(F_{3,21} > 0.0577) \approx P(F_{3,20} > 0.0577)$$

20 .5 | .816 so pval > 0.5



ex} 502 Final 3-5) B.5

| | predictor | coef | SE | T | pval |
|---|---|---|---|---|---|
| $x_1$ | constant | 57.4237 | 8.49058 | 6.76 | 0.000 |
| $x_2$ | chem 2 | 0.78912 | 0.16839 | ? ?? | 0.0007 |

Final Used $Y = \beta_1 + \beta_2 x_2 + e$

(So now 3)) want 95% CI for $\beta_2$

$n = 49$

| $df = n-p = \nu$ | t table |
|---|---|
| $= 49-2 = 47 > 30 \to \infty$ | 1.96 |
| | 95% |

desired work

$$\hat{\beta}_2 \pm z_{.975} SE(\hat{\beta}_2) = .78912 \pm 1.96(.16839)$$

$$= \boxed{[.459, 1.119]}$$

Suppose $n = 23$

| $df = n-p = \nu$ | t table |
|---|---|
| $= 23-2 = 21$ | 2.080 |
| | 95% |

work

$$\hat{\beta}_2 \pm t_{21,.975} SE(\hat{\beta}_2) = .78912 \pm 2.080(.16839)$$

$$\doteq \boxed{[.4389, 1.1394]}$$

t table: if $n-p > 30$ use $df = \nu = \infty$ line.

SP08 5) $H_0: \beta_2 = 0$ $\quad H_A: \beta_2 \neq 0$

$$t_{02} = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} = \frac{.78912}{.16839} = 4.69$$

pval = .0007

reject $H_0$ chem 2 is needed in the MLR model for heat

(Leave out "given that the constant is in the model")

S02 Final 6)

| Full | df | SS | MS | F | pval |
|---|---|---|---|---|---|
| reg | 6 | | | | |
| resid | 67 ($df_F$) | 17240.9 (SSE(F)) | 257.327 (MSE(F)) | | |

| red | df | SS | MS | F | pval |
|---|---|---|---|---|---|
| reg | 1 | | | | |
| resid | 72 ($df_R$) | 18403.8 (SSE(R)) | 255.608 | | |

full
reg $p-1$
resid $n-p$
$n-1$

red
reg $8-1$
resid $n-8$
$n-1$

i) $H_0$ the reduced model is good $\quad H_A$ use the full model

ii) $$F_R = \frac{\left(\frac{SSE(R) - SSE(F)}{df_R - df_F}\right)}{MSE(F)} = \frac{\left[\frac{18403.8 - 17240.9}{72 - 67}\right]}{257.327}$$

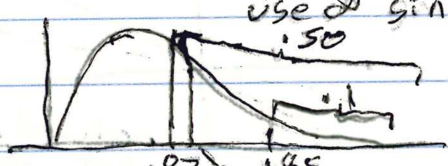
$$= \frac{232.58}{257.327} = 0.904$$

iii) pval $= P(F_{df_R - df_F,\, df_F} > F_R)$

$= P(F_{5,67} > .904)$

use $\infty$ since $67 > 60$

| | | 5 |
|---|---|---|
| $\infty$ | .5 | .870 |
| | .9 | 1.85 |




$1 - .9 = .1 < \text{pval} < .5 = 1 - .5$

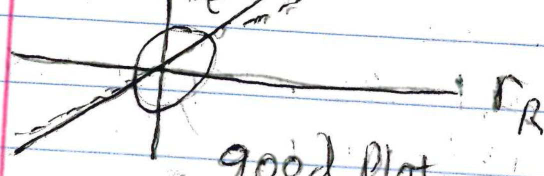
iv) fail to reject $H_0$
the reduced model is good

64] * For the partial F test, $df_R - df_F = n-q-(n-p)$ $= p-q$ = # of parameters set to 0 in reduced model = # variables deleted from reduced model = # variables in full model − # " "reduced" "

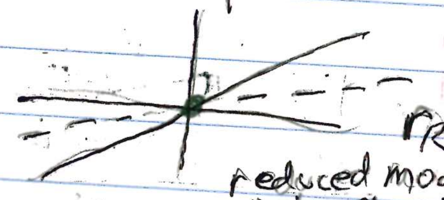
65] p34 & 47 : an **RR plot** is a plot of residuals from the reduced model vs those of the full model. An **FF plot** is a plot of the fitted values from the reduced model vs those of the full model.

66] If the full and reduced model are good and $n \geq 5p$, the plotted points in the RR and FF plots should scatter about the identity line with correlation $\geq 0.95$. In the RR plot, it should be hard to see that the OLS line and identity line intersect at the origin if the reduced model is good.

not clear where the 2 lines intersect




good plot




reduced model not good