Midterm Exam

David M. Rocke

February 18, 2025

1. The dataframe `uts` contains the data from an experiment in which the tensile strength of 24 samples of engineered cartilage was measured. Some samples were treated with condroitinase ABC and some were controls. At 2 weeks, 6 controls and 5 treated samples were measured. At 4 weeks, 7 controls and 6 treated samples were measured.

   The coefficient labeled `time_4` is the estimated difference in UTS between 2 weeks and 4 weeks. A positive value means an increase between 2 weeks and 4 weeks.

   The coefficient labeled `treat_Yes` is the estimated difference in UTS between untreated and treated specimens. A positive value means an increase in the treated samples vs. the control samples.

Use the normal percentage point.

a) Find a 95% confidence interval for the amount by which the tensile strength is changed by the treatment.
b) Find a 95% confidence interval for the amount by which the tensile strength is changed by waiting from 2 weeks to 4 weeks.
c) What are the plots that you would like to look at to see if the assumptions of the linear model are valid?

```
> uts.lm <- lm(uts~time+treat,data=uts.data)
> summary(uts.lm)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   216.39      39.04   5.543 1.68e-05 ***
time4 week     50.96      45.13   1.129    0.272
treatYes      226.02      45.13   5.009 5.87e-05 ***
```

a)

$$226.04 \pm (1.96)(45.13)$$
$$226.04 \pm 88.45$$
$$(137.56, 314.47)$$

b)

$$50.96 \pm (1.96)(45.13)$$
$$50.96 \pm 88.45$$
$$(-37.49, 139.42)$$

c) residual vs. predicted, normal qq, etc.

2. An analytical chemist wants to use a calibration line which predicts `log(Peak.Area)` from `log(Concentration)` because this may stabilize the variance of the peak area so that it is the same at every concentration.

```
> lpa <- log(zinc$Peak.Area)}
> lc <- log(zinc$Concentration+10)}
> lzinc.lm <- lm(lpa~lc)
> summary(lzinc.lm)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.26017    0.07903   41.25   <2e-16 ***
lc           0.84881    0.01169   72.64   <2e-16 ***
> predobj <- predict.lm(lzinc.lm,newdata=data.frame(lc=log(10)),
                  interval="predict",level=0.99,type="response")
> predobj
      fit      lwr      upr
  5.21463 4.463724 5.965536
```

With the new variables `lc` and `lpa` (10 is added to the concentrations because `log(0)` is not defined and perform a regression to use as a calibration curve. Suppose we want to compute the critical level CL as the peak area that is large enough that it is not consistent with 0 concentration. We can compute the CL for this situation with 0.995 one-sided CI by using `predict` as above.

- Find the CL as a a log peak area (`lpa`): $\mathrm{CL\text{-}LPA} = 5.9655$.
- Find the started log concentration (`lc`) corresponding to that. Now $\mathrm{LPA} = 3.2602 + 0.8488\mathrm{LC}$ so

  $$\mathrm{CL\text{-}LC} = (\mathrm{CL\text{-}LPA} - 3.2602)/0.8488 = 3.1872$$

- Find the actual concentration corresponding to that. $\mathrm{LC} = \log(C + 10)$ so $\mathrm{CL\text{-}C} = \exp(CL - LC) - 10 = 14.22$

3. A standard cell culture medium is specified to have 30 g/L of calcium. Two competing vendors are selected, and 100 samples of each one are analyzed. The samples from vendor 1 (V1) had a mean of 29.5 g/L with a standard deviation of 3.3 g/L. The samples from vendor 2 (V2) had a mean of 30.6 g/L and a standard deviation of 2.8 g/L. Use the normal critical values instead of the t critical values.

    a) Test the hypothesis $H_0 : \mu_1 = 30$ for V1 at the 5% level. What is your conclusion?
    b) Test the hypothesis $H_0 : \mu_1 = 30$ for V2 at the 5% level. What is your conclusion?
    c) Test the hypothesis $H_0 : \mu_1 = \mu_2$ at the 5% level. What is your conclusion?
    d) Find a 95% confidence interval for the difference of the means.

a) Test the hypothesis $H_0 : \mu_1 = 30$ for V1 at the 5% level. What is your conclusion?

$$z_1 = \frac{29.5 - 30}{3.3/\sqrt{100}} = -1.51 \text{ Do not reject.}$$

b) Test the hypothesis $H_0 : \mu_1 = 30$ for V2 at the 5% level. What is your conclusion?

$$z_1 = \frac{30.6 - 30}{2.8/\sqrt{100}} = 2.14 \text{ Reject.}$$

c) Test the hypothesis $H_0 : \mu_1 = \mu_2$ at the 5% level. What is your conclusion?

$$z = \frac{29.5 - 30.6}{\sqrt{3.3^2/100 + 2.8^2/100}} = \frac{-1.1}{0.432} = 2.54, \text{Reject.}$$

d) Find a 95% confidence interval for the difference of the means.

$$-1.1 \pm (1.96)(0.432)$$

$$-1.1 \pm 0.846$$

$$(-1.95, -0.25)$$

4. Consider the data from the highway paint problem in BHH.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	59.042	2.019	29.237	1.22e-14	***
factor(site)2	8.750	2.332	3.752	0.001922	**
factor(site)3	4.500	2.332	1.930	0.072765	.
factor(site)4	-4.750	2.332	-2.037	0.059696	.
factor(site)5	9.250	2.332	3.967	0.001240	**
factor(site)6	3.000	2.332	1.287	0.217759	
supplierGS	9.833	1.904	5.165	0.000115	***
supplierL	-2.000	1.904	-1.050	0.310141	
supplierZK	9.000	1.904	4.727	0.000270	***

	supplierGS	supplierL	supplierZK
supplierGS	3.6250	1.8125	1.8125
supplierL	1.8125	3.6250	1.8125
supplierZK	1.8125	1.8125	3.6250

4. Consider the data from the highway paint problem in BHH.

```
             Estimate Std. Error t value Pr(>|t|)
supplierGS      9.833      1.904   5.165 0.000115 ***
supplierL      -2.000      1.904  -1.050 0.310141
supplierZK      9.000      1.904   4.727 0.000270 ***

           supplierGS supplierL supplierZK
supplierGS     3.6250    1.8125     1.8125
supplierL      1.8125    3.6250     1.8125
supplierZK     1.8125    1.8125     3.6250
```

Find a 95% confidence interval for the difference of the wear between GS and ZK. Use the normal percentage point.

$$(9.833 - 9.000) \pm 1.960\sqrt{3.6250 + 3.6250 - 2(1.8125)}$$

$$(9.833 - 9.000) \pm 1.960\sqrt{3.6250}$$

$$0.833 \pm (1.960)(1.904)$$

$$0.833 \pm 3.732$$

$$(-2.899, 4.565)$$