HW SOLUTIONS

MELIKAMP

1. Problems due by August 4

All drawings are omitted. All equalities in computations should be assumed to be approximate, even though = is used.

Problem (6.6.5).

Solution. We are assuming that population variances are equal.

(a) Use the formula from p. 239:

$$\overline{X}_1 - \overline{X}_2 \pm Z_{1-\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where S_p is the pooled standard deviation:

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}. \qquad (\dagger)$$

From the given data we know that $\overline{X}_1 = 80.2$, $\overline{X}_2 = 75.4$, $s_1 = 5.7$, $s_2 = 6.1$, $n_1 = n_2 = 100$, so we can write

$$S_p = \sqrt{\frac{(100-1)5.7^2 + (100-1)6.1^2}{100+100-2}} = 5.9$$

and the CI is

$$80.2 - 75.4 \pm 1.96 \cdot 5.9 \sqrt{\frac{1}{100} + \frac{1}{100}} = 4.8 \pm 1.64,$$

or

$$(3.16, 6.44).$$

(b) Finding a 95% confidence interval is, in a way, equivalent to running a two-sided test with $\alpha = 0.05$. The test will result in rejecting $H_0 : \mu_1 - \mu_2 = 0$ iff the confidence interval does not contain 0, as is the case here. In other words, there is enough evidence to suggest that $\mu_1 \neq \mu_2$. □

Date: August 3, 2009.

Problem (6.6.7).

Solution. From the given data we know that $\overline{X}_1 = 70.5$, $\overline{X}_2 = 76.3$, $s_1 = 24.3$, $s_2 = 21.6$, $n_1 = n_2 = 25$. We are assuming that population variances are unequal, so we use formulas from p. 247:

$$\overline{X}_1 - \overline{X}_2 \pm t_{1-\alpha/2}\sqrt{\frac{s_2^2}{n_1} + \frac{s_2^2}{n_2}},$$

where

$$\mathrm{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \approx 48,$$

and so the CI is

$$70.5 - 76.3 \pm 2.011\sqrt{\frac{24.3^2}{25} + \frac{21.6^2}{25}} = -5.8 \pm 13.08$$

or

$$(-18.88, 7.28).$$

□

Problem (6.6.10).

Solution. From the given data we know that $\overline{X}_1 = 5.88$, $\overline{X}_2 = 6.29$, $s_1 = 4.7$, $s_2 = 4.89$, $n_1 = 8$, $n_2 = 7$. We are assuming that population variances are unequal, so we use formulas from p. 247.

(i) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$, $\alpha = 0.05$.

(ii)
$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_2^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

(iii) Reject H_0 iff $|t| \geq t_{1-\alpha/2}$. Here we need to compute df, of course (just as in the previous problem):

$$\begin{aligned} \mathrm{df} &= 13, \\ t_{1-\alpha/2} &= 2.16. \end{aligned}$$

(iv) $|t| = \left|\frac{5.88 - 6.29}{\sqrt{\frac{4.7^2}{8} + \frac{4.89^2}{7}}}\right| = |-0.16| < 2.16$, so we will fail to reject H_0.

(v) There is not enough evidence to reject H_0; $\alpha = 0.05$.

□

Problem (6.6.15).

Solution. From the given data we know that $\overline{X}_1 = 390$ (males), $\overline{X}_2 = 425$ (females), $s_1 = 57$, $s_2 = 65$, $n_1 = 25$, $n_2 = 40$. We are assuming that population variances are equal, so we use formulas from p. 239. We will also use Z statistic, although arguments can be made for using t here.

(i) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$, $\alpha = 0.05$.

(ii)
$$Z = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$
where S_p is defined just as in (†).

(iii) Reject H_0 iff $|Z| \geq Z_{1-\alpha/2} = 1.96$.

(iv) $S_p = 62.07$,
$$|Z| = \left|\frac{390 - 425}{62.07\sqrt{\frac{1}{25} + \frac{1}{40}}}\right| = |-2.21| > 1.96,$$
so we will reject H_0.

(v) The data provides sufficient evidence to conclude that population means are significantly different; $\alpha = 0.05$.

□

Problem (7.10.1).

Solution. From what is given,
$n_1 = 120$,
$n_2 = 150$,
$\hat{p}_1 = \frac{X_1}{n_1} = \frac{30}{120} = 0.25$ and
$\hat{p}_2 = \frac{62}{150} = 0.413$.

The difference between population proportions can be estimated by

$$\begin{aligned} & \hat{p}_1 - \hat{p}_2 \pm Z_{1-\alpha/2}\sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\right) + \left(\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)} \\ = \quad & 0.25 - 0.413 \pm 1.96\sqrt{\frac{0.25 \cdot 0.75}{120} + \frac{0.413 \cdot 0.587}{150}} \\ = \quad & -0.16 \pm 0.11 \end{aligned}$$

or

$$(-0.27, -0.05).$$

□

Problem (7.10.2).

Solution. It is given that $X = 423$, $n = 1200$, so...

(a) $\hat{p} = 423/1200 = 0.3525$.

(b) The standard error is found from

$$\text{s.e.}(p) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.3525(1-0.3525)}{1200}} = 0.014.$$

(c) A 95% confidence interval for p is

$$\hat{p} \pm Z_{1-\alpha/2} \cdot \text{s.e.}(p) = 0.3525 \pm 0.0274$$

or

$$(0.3251, 0.3799).$$

□

Problem (10.8.1).

Solution. From the given data,
$\bar{X} = 6.25$, $s_X = 6.96$,
$\bar{Y} = 40.38$, $s_Y = 31.2$,
$\text{Cov}(X,Y) = 197.89$, so

(a) $r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X\,\text{Var}Y}} = \frac{\text{Cov}(X,Y)}{s_X \cdot s_Y} = \frac{197.89}{6.96 \cdot 31.2} = 0.9113.$

(b) We run a two-sided test with $\alpha = 0.05$:
  (i) $H_0 : \rho = 0$, $H_0 : \rho \neq 0$, $\alpha = 0.05$.
  (ii) $t = r\sqrt{\frac{n-2}{1-r^2}}$, $\text{df} = n-2$.
  (iii) Reject H_0 iff $|t| \geq t_{1-\alpha/2}$.
  (iv) $t = 0.9113\sqrt{\frac{8-2}{1-0.9113^2}} = 5.42$, while $\text{df} = 6$ and $t_{1-\alpha/2} = 2.447$.
  (v) There is sufficient evidence to reject H_0 and conclude that there is significant correlation between X and Y.

(c) $\hat{\beta}_1 = r\sqrt{\frac{\text{Var}Y}{\text{Var}X}} = r\frac{s_X}{s_Y} = 0.9113 \cdot \frac{31.2}{6.96} = 4.085,$
$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 40.38 - 4.085 \cdot 6.25 = 14.85$, so

$$\hat{Y} = 14.85 + 4.085 \cdot X.$$

□

Problem (10.8.2).

Solution. From the given data,
$\bar{X} = 9.67$, $s_X = 3.56$,
$\bar{Y} = 7.25$, $s_Y = 1.56$,
$\text{Cov}(X,Y) = 4.12$, so

(a) Chart omitted.

(b) $r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}X\,\text{Var}Y}} = 0.742.$

(c) $\hat{\beta}_1 = r\sqrt{\frac{\operatorname{Var} Y}{\operatorname{Var} X}} = 0.33$,
$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 4$, so

$$\hat{Y} = 4 + 0.33 \cdot X.$$

(d) $\hat{Y} = 4 + 0.33 \cdot 11 = 7.63$.
(e) The difference is $3 \cdot \beta_1 = 0.99$.

□