

PartA

$$Q1: \text{样本均值 } \bar{x} = \frac{\sum_{i=1}^{25} x_i}{N} = \frac{201}{25} = 8.04, \quad \bar{y} = \frac{\sum_{i=1}^{25} y_i}{N} = \frac{94400}{25} = 3776$$

标准差

$$s_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} = \sqrt{\frac{1}{N-1} \left( \sum_{i=1}^N x_i^2 - N\bar{x}^2 \right)}$$
$$= \sqrt{\frac{1}{25-1} \times (1768.5 - 25 \times 8.04^2)} = 2.52$$

$$s_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2} = \sqrt{\frac{1}{N-1} \left( \sum_{i=1}^N y_i^2 - N\bar{y}^2 \right)}$$
$$= \sqrt{\frac{1}{25-1} \times (361365000 - 25 \times 3776^2)} = 452.34$$

样本协方差

$$s_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) = E(x, y) - E(x)E(y) = \frac{1}{N} \sum_{i=1}^{25} x_i y_i - \bar{x}\bar{y}$$
$$= \frac{1}{25} \times 785850 - 8.04 \times 3776 = 1074.96$$

Q2

定义符号：总体平均月薪  $\mu$

原假设  $H_0: \mu \neq 4000$  (ABC 大学的工人平均月薪不等于 4000 元)

备择假设  $H_1: \mu = 4000$  (ABC 大学的工人平均月薪等于 4000 元)

由于样本量  $N=25$ ，属于小样本，总体方差未知，使用 t 检验

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{N}} = \frac{3776 - 4000}{452.34/\sqrt{25}} = \frac{-224}{90.468} = -2.48$$

在置信水平为 1% ( $\alpha = 0.01$ )，自由度  $N-1=24$  的条件下，查表得  $t_{\alpha/2}(24) = 2.797$ ，

因为  $|t| < |t_{\alpha/2}(24)|$ ，样本统计量未落入拒绝域，因此接受  $H_0$ ，说明大学的工人平均月薪不等于 4000 元。

在置信水平为 5% ( $\alpha = 0.05$ )，自由度  $N-1=24$  的条件下，查表得  $t_{\alpha/2}(24) = 2.064$ ，

因为  $|t| > |t_{\alpha/2}(24)|$ ，样本统计量落入拒绝域，因此拒绝  $H_0$ ，接受  $H_1$ ，说明大学的工人平均月薪等于 4000 元

Q3

定义符号：总体平均教育水平 $\mu$

原假设 $H_0: \mu = 9$ （ABC 大学的工人平均教育水平是 9 年）

备择假设 $H_1: \mu \neq 9$ （ABC 大学的工人平均教育水平不是 9 年）

由于样本量  $N=25$ ，属于小样本，总体方差未知，使用 t 检验

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{N}} = \frac{8.04 - 9}{2.52/\sqrt{25}} = \frac{-0.96}{0.504} = -1.90$$

在置信水平为 5% ( $\alpha = 0.05$ )，自由度  $N-1=24$  的条件下，查表得  $t_{\alpha/2}(24) = 2.064$ ，

因为  $|t| < |t_{\alpha/2}(24)|$ ，样本统计量未落入拒绝域，因此接受  $H_0$ ，说明大学的工人平均教育水平是 9 年。

PartB

Q4

$$\begin{aligned}\beta_1 &= \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{n \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} = \frac{25 \times 785850 - 201 \times 94400}{25 \times 1768.5 - (201)^2} \\ &= \frac{19646250 - 18974400}{44212.5 - 40401} = \frac{671850}{3811.5} = 176.27\end{aligned}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 3776 - 176.27 \times 8.04 = 3776 - 1417.21 = 2358.79$$

$\beta_0$  代表当受教育水平为 0 时，工人的月薪为 2358.79 元

$\beta_1$  代表受教育水平每增加一个单位（1 年），工人的月薪增加 176.27 元

方程为： $\hat{y}_i = \beta_0 + \beta_1 x_i = 2358.79 + 176.27 x_i$

Q5

判定系数

$$\begin{aligned}R^2 &= \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} = \frac{\sum (\hat{y}_i^2 - 2\hat{y}_i \bar{y}_i + \bar{y}_i^2)}{\sum (y_i^2 - 2y_i \bar{y}_i + \bar{y}_i^2)} = \frac{\sum \hat{y}_i^2 - \sum 2\hat{y}_i \bar{y}_i + \sum \bar{y}_i^2}{\sum y_i^2 - \sum 2y_i \bar{y}_i + \sum \bar{y}_i^2} \\ &= \frac{\sum \hat{y}_i^2 - 2\bar{y}_i \sum \hat{y}_i + \sum \bar{y}_i^2}{\sum y_i^2 - 2\bar{y}_i \sum y_i + \sum \bar{y}_i^2} \\ &= \frac{(N \times \beta_0^2 + 2\beta_0 \beta_1 \sum x_i + \beta_1^2 \sum x_i^2) - 2\bar{y}_i (N \times \beta_0 + \beta_1 \sum x_i) + \sum \bar{y}_i^2}{\sum y_i^2 - 2\bar{y}_i \sum y_i + \sum \bar{y}_i^2} \\ &= \frac{(25 \times 2358.79^2 + 2 \times 2358.79 \times 176.27 \times 201 + 176.27^2 \times 1768.5) - 2 \times 3776 \times (25 \times 2358.79)}{361365000 - 2 \times 3776 \times 94400 + 25 \times 3776^2} \\ &= 0.96\end{aligned}$$

判定系数是测度了回归直线对观测数据的拟合程度，若所有观测点都落在直线上，判定系数为 1，拟合是完全的；如果 y 的变化与 x 无关，x 完全无助于解释 y 的变差，则判定系数为 0。

回归标准误差

$$\hat{\sigma} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N-2}} = \sqrt{\frac{SSE}{N-2}} = \sqrt{\frac{SST - SSR}{N-2}} = \sqrt{\frac{173498.13}{25-2}} = 86.85$$

回归标准误差是度量各实际观测点在直线周围的散布情况的一个统计量，是对误差项的标准差的估计，可以看作在排除了 x 对 y 的线性影响后，y 随机波动大小的一个估计量，反映了用估计的回归方程预测因变量 y 时预测误差的大小。

Q6

回归系数的标准差

$$\sigma_{\hat{\beta}_1} = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}} = \frac{86.85}{\sqrt{1768.5 - \frac{1}{25} \times 201^2}} = 7.03$$

这个标准误差是大的。

$$\begin{aligned}\sigma_{\hat{\beta}_0} &= \hat{\sigma} \sqrt{\frac{1}{N} + \frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \hat{\sigma} \times \sqrt{\frac{1}{N} + \frac{\bar{x}}{\sum x_i^2 - 2\bar{x}_i \sum x_i + \sum \bar{x}_i^2}} \\ &= 86.85 \times \sqrt{\frac{1}{25} + \frac{8.04}{1768.5 - 2 \times 8.04 \times 201 + 25 \times 8.04^2}} = 19.99\end{aligned}$$

Q7

工人#3 的学习水平为 5，即  $x=5$ ，将其带入 Q4 所得的回归方程

$$\hat{y}_3 = \beta_0 + \beta_1 x_3 = 2358.79 + 176.27 \times 5 = 3240.14$$

得到按他的学习水平，预期月薪为 3240.14 元，而实际上为 3300 元，相差 59.86 元是回归线未作出解释的部分。

Q8

因为月薪的单位由元放大变成千元，是原来的 1000 倍，因此新的回归方程  $(1000\hat{y}_i)$ ，且每个工人的学习等级多两年，因此新的回归方程  $(x_i - 2)$ 。新的回归方程为

$$(1000\hat{y}_i) = 2358.79 + 176.27(x_i - 2)$$

$$\hat{y}_i = 2.35879 + 0.17627(x_i - 2)$$

$$\hat{y}_i = 2.35879 + 0.17627x_i - 0.35254$$

$$\hat{y}_i = 2.00625 + 0.17627x_i$$

新的  $\beta_0 = 2.00625$ ， $\beta_1 = 0.17627$

新的

$$R^2 = \frac{SSR}{SST} = \frac{(N \times \beta_0^2 + 2\beta_0\beta_1 \sum x_i + \beta_1^2 \sum x_i^2) - 2\bar{y}_i(N \times \beta_0 + \beta_1 \sum x_i) + \sum \bar{y}_i^2}{\sum y_i^2 - 2\bar{y}_i \sum y_i + \sum \bar{y}_i^2}$$

$$= \frac{(25 \times 2.00625^2 + 2 \times 2.00625 \times 0.17627 \times (201 + 25 \times 2) + 0.17627^2 \times 1768.5) - 2 \times \frac{3776}{1000} \times (25 \times 2.00625 + 0.17627 \times (201 + 25 \times 2)) + 25 \times \left(\frac{3776}{1000}\right)^2}{\left(\frac{361365000}{1000000}\right) - 2 \times \left(\frac{3776}{1000}\right) \times \left(\frac{94400}{1000}\right) + 25 \times \left(\frac{3776}{1000}\right)^2}$$

$$= 0.962$$