University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Homework 9

Answer the following questions:

a. Suppose that two independent random samples of n_1 and n_2 observations are selected from normal populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively. Find a confidence interval for the variance ratio $\frac{\sigma_1^2}{\sigma_2^2}$ with confidence level $1-\alpha$.

b. The sample mean $\bar{X}$ is a good estimator of the population mean μ. It can also be used to predict a future value of X independently selected from the population. Assume that you have a sample mean $\bar{x}$ and a sample variance s^2, based on a random sample of n measurements from a normal population. Construct a prediction interval for a new observation x, say x_p. Use $1-\alpha$ confidence level. Hint: Start with $X_p - \bar{X}$ and then use t distribution.

c. (from *Mathematical Statistics and Data Analysis*), by J. Rice, 2nd Edition.
In a study done at the National Institute of Science and Technology (Steel et al. 1980), asbestos fibers on filters were counted as part of a project to develop measurement standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and punches of 3-mm diameter were taken from the filter and mounted on a transmission electron microscope. An operator counted the number of fibers in each of 23 grid squares, yielding the following counts:

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

Assume that the Poisson distribution with unknown parameter λ would be a plausible model for describing the variability from grid square to grid square in this situation.

a. Use the method of maximum likelihood to estimate the parameter λ.
b. Use the asymptotic properties of the maximum likelihood estimates to construct a 95% confidence interval for λ. As a reminder, for large samples the distribution of $\frac{\hat{\theta}-\theta}{\sqrt{\frac{1}{nI(\theta)}}}$ is approximately standard normal, where $I(\theta)$ is the Fisher information.

d. Let $X_1, X_2, \cdots, X_9$ and $Y_1, Y_2, \cdots, Y_{12}$ represent two independent random samples from the respective normal distributions $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$. It is given that $\sigma_1^2 = 3\sigma_2^2$, but σ_2^2 is unknown. Define a random variable which has a t distribution and use it to find a 95% confidence interval for $\mu_1 - \mu_2$.

e. Suppose that a simple linear regression of miles per gallon (Y) on car weight (x) has been performed on 32 observations. The least squares estimates are $\hat{\beta}_0 = 68.17$ and $\hat{\beta}_1 = -1.112$, with $s_e = 4.281$. Other useful information: $\bar{x} = 30.91$ and $\sum_{i=1}^{32}(x_i - \bar{x})^2 = 2054.8$. Answer the following questions:

a. Construct a 95% confidence interval for β_1.
b. Construct a 95% confidence interval for σ^2.
c. Construct a confidence interval for $3\beta_0 - 2\beta_1 - 50$.

f. Consider the simple regression model $y_i = \beta_0 + \beta_1 x_1 + \epsilon_i$. The Gauss-Markov conditions hold and also $\epsilon_i \sim N(0, \sigma)$. Construct a prediction interval for the average of m new observations of Y for a given new $x = x_0$.

Feng Shiwei UID:3035256428

a. $\left.\begin{array}{l}\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1} \\ \dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}\end{array}\right\} \dfrac{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)}{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)} = \dfrac{\sigma_1^2}{\sigma_2^2}\cdot\dfrac{S_2^2}{S_1^2} \sim F_{n_2-1,n_1-1}$

$$P\left(F_{\alpha/2;n_2-1,n_1-1} < \frac{\sigma_1^2}{\sigma_2^2}\cdot\frac{S_2^2}{S_1^2} < F_{1-\alpha/2;n_2-1,n_1-1}\right) = 1-\alpha$$

$$\therefore P\left(\frac{S_1^2}{S_2^2}F_{\alpha/2;n_2-1,n_1-1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{1-\alpha/2;n_2-1,n_1-1}\right) = 1-\alpha.$$

$\therefore$ The confident interval for $\frac{\sigma_1^2}{\sigma_2^2}$ is

$$\left(\frac{S_1^2}{S_2^2}F_{\alpha/2;n_2-1,n_1-1},\ \frac{S_1^2}{S_2^2}F_{1-\alpha/2;n_2-1,n_1-1}\right)$$

b. $\left.\begin{array}{l}X_p \sim N(\mu,\sigma) \\ \bar{X} \sim N(\mu,\frac{\sigma}{\sqrt{n}})\end{array}\right\} \Rightarrow X_p-\bar{X} \sim N(0,\sigma\sqrt{1+\frac{1}{n}})$

$\left.\begin{array}{l}X_p-\bar{X} \sim N(0,\sigma\sqrt{1+\frac{1}{n}}) \\ \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}\end{array}\right\} \dfrac{X_p-\bar{X}}{S\sqrt{1+\frac{1}{n}}} \sim t_{n-1}.$

$$\therefore P\left(-t_{\alpha/2;n-1} < \frac{X_p-\bar{X}}{S\sqrt{1+\frac{1}{n}}} < t_{\alpha/2;n-1}\right) = 1-\alpha$$

$$\therefore P\left(\bar{X}-S\sqrt{1+\tfrac{1}{n}}\,t_{\alpha/2;n-1} < X_p < \bar{X}+S\sqrt{1+\tfrac{1}{n}}\,t_{\alpha/2;n-1}\right) = 1-\alpha$$

$\therefore$ The prediction interval for X_p is

$$\left(\bar{X}-S\sqrt{1+\tfrac{1}{n}}\,t_{\alpha/2;n-1},\ \bar{X}+S\sqrt{1+\tfrac{1}{n}}\,t_{\alpha/2;n-1}\right)$$

c. (a) $$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$L = \frac{\lambda^{\sum x_i} \cdot e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

$$\therefore \ln L = \sum_{i=1}^{n} x_i \ln\lambda - n\lambda - \sum_{i=1}^{n} \ln x_i!$$

$$\therefore \frac{\partial \ln L}{\partial \lambda} = \sum_{i=1}^{n} x_i \frac{1}{\lambda} - n = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\therefore \hat{\lambda} = \bar{X} = 24.913$$

(b) $$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i$$

$$\therefore \frac{\hat{\lambda}-\lambda}{\sqrt{\frac{1}{I_n(\lambda)}}} = \frac{\hat{\lambda}-\lambda}{\sqrt{\frac{1}{-E(-\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i)}}} = \frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \longrightarrow N(0,1)$$

$$\therefore \hat{\lambda} \longrightarrow N(\lambda, \frac{\lambda}{n})$$

$$P\left(-z_{\alpha/2} < \frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\therefore P\left(-z_{\alpha/2} < \frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}/n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\therefore P\left(\hat{\lambda} - z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}} < \lambda < \hat{\lambda} + z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}\right) = 1-\alpha.$$

$\because \hat{\lambda} = \bar{x} = 24.913$, $n = 23$, $z_{\alpha/2} = z_{0.025} = 1.96$

$\therefore$ The confident interval is $(22.87, 26.95)$

d. $\bar{X} \sim N(\mu_1, \frac{\sigma_1}{3})$. $\bar{Y} \sim N(\mu_2, \frac{\sigma_2}{2\sqrt{3}})$

$$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \sqrt{\tfrac{5}{12}}\sigma_2)$$

$$\therefore \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{5}{12}}\sigma_2} \sim N(0,1)$$

$$\because \frac{8S_x^2}{\sigma_1^2} \sim \chi_8^2, \quad \frac{11S_y^2}{\sigma_2^2} \sim \chi_{11}^2$$

$$\therefore \frac{8S_x^2}{\sigma_1^2} + \frac{11S_y^2}{\sigma_2^2} \sim \chi_{19}^2$$

$$\therefore \frac{\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{5/12}\,\sigma_2}}{\sqrt{\left(\dfrac{8S_x^2}{\sigma_1^2}+\dfrac{11S_y^2}{\sigma_2^2}\right)/19}} = \sqrt{\frac{228}{5}}\,\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{8}{3}S_x^2+11S_y^2}} \sim t_{19}$$

$$\therefore P\left(-t_{\alpha/2;19} < \sqrt{\frac{228}{5}}\,\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{8}{3}S_x^2+11S_y^2}} < t_{\alpha/2;19}\right) = 1-\alpha$$

$\therefore$ The confident interval is

$$\left(\bar{X}-\bar{Y}-\sqrt{\frac{5}{228}}\sqrt{\frac{8}{3}S_x^2+11S_y^2}\;t_{0.025;19},\ \bar{X}-\bar{Y}+\sqrt{\frac{5}{228}}\sqrt{\frac{8}{3}S_x^2+11S_y^2}\;t_{0.025;19}\right)$$

e. (a) $\because \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum(X_i-\bar{X})^2}}\right)$

$$\left.\begin{array}{l}\therefore \dfrac{\hat{\beta}_1-\beta_1}{\sigma/\sqrt{\sum(X_i-\bar{X})^2}} \sim N(0,1) \\ \because \dfrac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}\end{array}\right\} \Longrightarrow \frac{\hat{\beta}_1-\beta_1}{S_e/\sqrt{\sum(X_i-\bar{X})^2}} \sim t_{n-2}$$

$$\therefore P\left(-t_{\alpha/2;n-2} < \frac{\hat{\beta}_1-\beta_1}{S_e/\sqrt{\sum(X_i-\bar{X})^2}} < t_{\alpha/2;n-2}\right) = 1-\alpha$$

$$\therefore P\left(\hat{\beta}_1 - t_{\alpha/2;n-2} S_e/\sqrt{\sum(X_i-\bar{X})^2} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2;n-2} S_e/\sqrt{\sum(X_i-\bar{X})^2}\right) = 1-\alpha$$

$\because \hat{\beta}_1 = -1.112$, $t_{0.025;30} = 2.042$, $S_e = 4.281$, $\sum(X_i-\bar{X})^2 = 2054.8$

$\therefore$ The 95% confidence interval for β_1 is $(-1.305, -0.919)$

(b) $\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$ $\quad \therefore P\left(\chi^2_{\alpha/2;n-2} < \frac{(n-2)S_e^2}{\sigma^2} < \chi^2_{1-\alpha/2;n-2}\right) = 1-\alpha$

$$\therefore P\left(\frac{(n-2)S_e^2}{\chi^2_{1-\alpha/2;n-2}} < \sigma^2 < \frac{(n-2)S_e^2}{\chi^2_{\alpha/2;n-2}}\right) = 1-\alpha$$

$\because S_e = 4.281$. $\chi^2_{1-\alpha/2;n-2} = \chi^2_{0.975;30} = 46.98$, $\chi^2_{\alpha/2;n-2} = \chi^2_{0.025;30} = 16.79$

$\therefore$ The confident interval for σ^2 is $(11.703, 32.746)$

(C). $\hat{\beta}_0 \sim N\left(\beta_0, \sigma\sqrt{\frac{1}{n}+\frac{\bar{X}^2}{\Sigma(X_i-\bar{X})^2}}\right)$

$\hat{\beta}_1 \sim N\left(\beta_1, \sigma/\sqrt{\Sigma(X_i-\bar{X})^2}\right)$

$$\operatorname{cov}(\hat{\beta}_0, \hat{\beta}_1) = \operatorname{cov}(\bar{y}-\hat{\beta}_1\bar{X}, \hat{\beta}_1) = \operatorname{cov}(\bar{y}, \hat{\beta}_1) - \bar{X}\operatorname{var}(\hat{\beta}_1)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\operatorname{cov}(y_i, \hat{\beta}_1) - \bar{X}\operatorname{var}(\hat{\beta}_1)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\operatorname{cov}\left(y_i, \frac{(X_i-\bar{X})y_i}{\Sigma(X_i-\bar{X})^2}\right) - \bar{X}\operatorname{var}(\hat{\beta}_1)$$

$$= 0 - \bar{X}\cdot\frac{\sigma^2}{\Sigma(X_i-\bar{X})^2}$$

$$= \frac{-\sigma^2\bar{X}}{\Sigma(X_i-\bar{X})^2}$$

$\therefore 3\hat{\beta}_0-2\hat{\beta}_1-50 \sim N\left(3\beta_0-2\beta_1-50, \left[9\sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{\Sigma(X_i-\bar{X})^2}\right)+4\sigma^2\cdot\frac{1}{\Sigma(X_i-\bar{X})^2}+\frac{12\sigma^2\bar{X}}{\Sigma(X_i-\bar{X})^2}\right]^{\frac{1}{2}}\right)$

$3\hat{\beta}_0-2\hat{\beta}_1-50 \sim N(3\beta_0-2\beta_1-50, \sigma\cdot A)$ where A is known.

$$\therefore \frac{(3\hat{\beta}_0-2\hat{\beta}_1-50)-(3\beta_0-2\beta_1-50)}{A\sigma} \sim N(0,1)$$

$$\because \frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}$$

$$\therefore \frac{(3\hat{\beta}_0-2\hat{\beta}_1-50)-(3\beta_0-2\beta_1-50)}{A\cdot S_e} \sim t_{n-2}$$

Let $\alpha=0.05$. $3\beta_0-2\beta_1-50 \in 3\hat{\beta}_0-2\hat{\beta}_1-50 \pm A\cdot S_e\cdot t_{\alpha/2;n-2}$.

$A=2.156$, $S_e=4.281$, $t_{0.025;30}=2.042$

$\therefore$ the confident interval is $(137.887, 175.581)$

f. $\bar{Y}_m$ is the average of m new observations.

$$\bar{Y}_m = \frac{1}{m}\sum_{i=1}^{m}(\beta_0+\beta_1 x_0+\varepsilon_i)$$

$$E(\bar{Y}_m) = \beta_0+\beta_1 x_0 . \quad var(\bar{Y}_m) = \frac{\sigma^2}{m}.$$

$\hat{Y}_0$ is the predictor of the $\bar{Y}_m$

$$\hat{Y}_0 = \hat{\beta}_0+\hat{\beta}_1 x_0$$

$$E(\hat{Y}_0) = \beta_0+\beta_1 x_0$$

$$var(\hat{Y}_0) = var(\bar{y}+\beta_1(x_0-\bar{x})) = \frac{\sigma^2}{n}+\frac{\sigma^2(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\therefore \bar{Y}_m - \hat{Y}_0 \sim N\left(0, \sigma\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right)$$

$$\because \frac{(n-2)s_e^2}{\sigma^2} \sim \chi^2_{n-2}$$

$$\therefore \frac{\bar{Y}_m-\hat{Y}_0}{s_e\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

the pivot should contains both θ ($\bar{Y}_m$) and $\hat{\theta}$ ($\hat{Y}_0$)

$\therefore$ The prediction interval for $\bar{Y}_m$ is

$$\left(\hat{Y}_0 - t_{\alpha/2;n-2}\, s_e\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}},\ \hat{Y}_0 + t_{\alpha/2;n-2}\, s_e\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right)$$