Formulas Exam II

SUMMARY OF HYPOTHESIS TESTS FOR μ

Large Sample ($n \geq 30$)

To test

$H_0 : \mu = \mu_0$

versus

$$H_a : \begin{array}{l} \mu > \mu_0, \text{ upper tail test} \\ \mu < \mu_0, \text{ lower tail test} \\ \mu \neq \mu_0, \text{ two-tailed test} \end{array}$$

Test statistic: $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

Replace σ by S, if σ is unknown.

$$\text{Rejection region}: \begin{cases} z > z_{\alpha}, & \text{upper tail RR} \\ z < -z_{\alpha}, & \text{lower tail RR} \\ |z| > z_{\alpha/2}, & \text{two tail RR} \end{cases}$$

Small Sample ($n < 30$)

To test

$H_0 : \mu = \mu_0$

versus

$$H_a : \begin{array}{l} \mu > \mu_0, \text{ upper tail test} \\ \mu < \mu_0, \text{ lower tail test} \\ \mu \neq \mu_0, \text{ two-tailed test} \end{array}$$

Test statistic: $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

$$\text{RR}: \begin{cases} t > t_{\alpha,n-1}, & \text{upper tail RR} \\ t < -t_{\alpha,n-1}, & \text{lower tail RR} \\ |t| > t_{\alpha/2,n-1}, & \text{two tail RR} \end{cases}$$

Assumption: $n \geq 30$

Assumption: Random sample comes from a normal population

Decision: Reject H_0, if the observed test statistic falls in the RR and conclude that H_a is true with $(1 - \alpha)100\%$ confidence. Otherwise, keep H_0 so that there is not enough evidence to conclude that H_a is true for the given α and more experiments may be needed.

PROCEDURE FOR FITTING A LEAST-SQUARES LINE

1. Form the n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, and compute the following quantities: $\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} y_i^2$, and $\sum_{i=1}^{n} x_i y_i$. Also compute the sample means, $\overline{x} = (1/n)\sum_{i=1}^{n} x_i$ and $\overline{y} = (1/n)\sum_{i=1}^{n} y_i$.
2. Compute

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

and

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$$

3. Compute $\hat{\beta}_0$ and $\hat{\beta}_1$ by substituting the computed quantities from step 1 into the equations

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

and

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

4. The fitted least-squares line is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

5. For a graphical representation, in the xy-plane, plot all the data points and draw the least-squares line obtained in step 4.

Theorem 8.2.1 *Let $Y = \beta_0 + \beta_1 x + \varepsilon$ be a simple linear regression model with $\varepsilon \sim N(0, \sigma^2)$, and let the errors ε_i associated with different observations $y_i (i = 1, \ldots, N)$ be independent. Then*

(a) *$\hat{\beta}_0$ and $\hat{\beta}_1$ have normal distributions.*
(b) *The mean and variance are given by*

$$E\left(\hat{\beta}_0\right) = \beta_0, \quad Var\left(\hat{\beta}_0\right) = \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)\sigma^2,$$

and

$$E\left(\hat{\beta}_1\right) = \beta_1, \quad Var\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{xx}},$$

where $S_{xx} = \sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum\limits_{i=1}^{n} x_i\right)^2$. In particular, the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively.

$$E(SSE) = (n-2)\sigma^2.$$

Thus, an unbiased estimator of the error variance, σ^2, is $\hat{\sigma}^2 = (SSE)/(n-2)$. We will denote $(SSE)/(n-2)$ by MSE (Mean Square Error).

The t distributions below have n-2 degrees of freedom:

$$Z_1 = \frac{\hat{\beta}_1 - \beta_1}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim N(0,1). \qquad t_{\beta_1} = \frac{Z}{\sqrt{\frac{\left(\frac{SSE}{\sigma^2}\right)}{n-2}}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{S_{xx}}}}$$

$$Z_0 = \frac{\hat{\beta}_0 - \beta_0}{\sigma\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{yy}}\right)} \sim N(0,1). \qquad t_{\beta_0} = \frac{z_0}{\sqrt{\frac{\frac{SSE}{\sigma^2}}{n-2}}} = \frac{\hat{\beta}_0 - \beta_0}{\left[MSE\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)\right]^{1/2}}$$

PROCEDURE FOR OBTAINING CONFIDENCE INTERVALS FOR β_0 AND β_1

1. Compute S_{xx}, S_{xy}, S_{xy}, $\overline{y}$, and $\overline{x}$ as in the procedure for fitting a least-squares line.
2. Compute $\hat{\beta}_1$, $\hat{\beta}_0$ using equations $\hat{\beta}_1 = (S_{xy})/(S_{xx})$ and $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$, respectively.
3. Compute SSE by $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$.
4. Define MSE (mean square error) to be

$$MSE = \frac{SSE}{n-2},$$

where n = Number of pairs of observations $(x_1, y_1), \ldots, (x_n, y_n)$.

5. A $(1-\alpha)100\%$ confidence interval for β_1 is given by

$$\left(\hat{\beta}_1 - t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}}, \hat{\beta}_1 + t_{\alpha/2,n-2}\sqrt{\frac{MSE}{S_{xx}}}\right)$$

where $t_{\alpha/2}$ is the upper tail $\alpha/2$-point based on a t-distribution with $(n-2)$ degrees of freedom.

6. A $(1-\alpha)100\%$ confidence interval for β_0 is given by

$$\left(\hat{\beta}_0 - t_{\alpha/2,n-2}\left[MSE\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)\right]^2, \hat{\beta}_0 + t_{\alpha/2,n-2}\left[MSE\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)\right]^{1/2}\right).$$

The *coefficient of determination* is the proportion of the sum of squares of deviations of the y values that can be credited to a linear relationship between x and y. This is defined by

$$r^2 = \frac{S_{yy} - SSE}{S_{yy}}$$