Chapter 9

For testing hypotheses about μ:

1. For Large sample or known σ, use z values.
2. For small sample and unknown σ, use t values.

p-value approach

Hypothesis type	*p*-value	*p*-value
Lower tail	$P(Z < z)$	$P(T < t)$
Upper tail	$P(Z > z)$	$P(T > t)$
2-tailed	$2P(Z > \|z\|)$	$2P(T > \|t\|)$

Critical values in observed mean metric

$$\bar{x}_{\alpha L} = \mu - z_\alpha \frac{\sigma}{\sqrt{n}}, \quad \bar{x}_{\alpha U} = \mu + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\bar{x}_{\frac{\alpha}{2}} = \mu \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Test statistics

$z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ or $t_{n-1} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ will be compared with critical values

1. z_α or t_α for upper tailed hypotheses.
2. $-z_\alpha$ or $-t_\alpha$ for lower tailed hypotheses.
3. $\pm z_{\alpha/2}$ or $\pm t_{\alpha/2}$ for two tailed hypotheses.

For testing hypotheses about π:

Test statistic

$z = \dfrac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$ where $p = \dfrac{x}{n}$ under the assumptions: $n\pi \geq 5$ and $n(1-\pi) \geq 5$.

Chapter 10

The i^th^ **paired difference** $d_i = x_{1i} - x_{2i}$.

Test statistic is $t_{n-1} = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$ $\quad \bar{d} = \dfrac{\sum_{i=1}^{n} d_i}{n}$ and

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} d_i\right)^2}{n-1}}.$$

p-value approach

Hypothesis type	*p*-value	*p*-value
Lower tail	$P(Z < z)$	$P(T < t)$
Upper tail	$P(Z > z)$	$P(T > t)$
2-tailed	$2P(Z > \|z\|)$	$2P(T > \|t\|)$

For testing difference between two independent population means:

If σ_1 and σ_2 known $\quad z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

If σ_1 and σ_2 unknown **AND** n_1 and $n_2 \geq 30$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

If σ_1 and σ_2 unknown **AND** n_1 or $n_2 < 30$

(assuming equal σ's) $\quad t_{n_1+n_2-2} = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

where $s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

If σ_1 and σ_2 unknown **AND** n_1 or $n_2 < 30$

(assuming unequal σ's) $\quad t_{df} = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

where $df = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left(\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}\right)}$

To test hypotheses about $\pi_1 - \pi_2$

If $\bar{p} = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{x_1 + x_2}{n_1 + n_2}$ then the test statistic is $z = \dfrac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}(1 - \bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Assumptions:

1. $n_1\pi_1 \geq 5$ and $n_1(1-\pi_1) \geq 5$.
2. $n_2\pi_2 \geq 5$ and $n_2(1-\pi_2) \geq 5$.

For testing hypotheses about σ_1^2 and σ_2^2

Test statistic $F_0 = \dfrac{s_1^2}{s_2^2}$ $\quad$ with $df_1 = n_1 - 1$ $\; df_2 = n_2 - 1$.

1. F_0 is compared with F_{α, n_1-1, n_2-1} (One-tailed).

2. F_0 is compared with $F_{\frac{\alpha}{2}, n_1-1, n_2-1}$ (Two-tailed).

Chapter 12

For testing hypotheses about *σ*

H_0: $\sigma^2 = \sigma_0^2$. Test statistic $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$ has a χ^2 distribution with $df = n-1$.

1. χ^2 is compared with χ^2_α (Upper-tailed).
2. χ^2 compared with $\chi^2_{1-\alpha}$ (Lower-tailed).
3. χ^2 is compared with $\chi^2_{\frac{\alpha}{2}}$ or $\chi^2_{1-\frac{\alpha}{2}}$ (2-tailed).

Chi-Square Test for *c* Independent Proportions

H_0: $\pi_1 = \pi_2 = \ldots = \pi_c$ versus H_A: Not all π_j are equal

If $\chi^2 = \sum_{j=1}^{c} \dfrac{(o_j - e_j)^2}{e_j} > \chi^2_\alpha$ [with df $=c-1$], Reject H_0.

o_j = Observed cell frequency e_j =Expected cell frequency

Marascuillo's Test for Pair-wise Proportions

H_0: $\pi_i = \pi_j$ versus H_A: $\pi_i \neq \pi_j$.

If $|p_i - p_j| > \sqrt{\chi^2_\alpha}\sqrt{\dfrac{p_i(1-p_i)}{n_i} + \dfrac{p_j(1-p_j)}{n_j}}$

[with df $=(c-1)$] then Reject H_0.

Test of Independence

H_0: The two characteristics are independent.

H_A: The two characteristics are **NOT** independent.

Test statistic $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \dfrac{(o_{ij} - e_{ij})^2}{e_{ij}}$

$$e_{ij} = \frac{(i^{th}\text{ Row total})(j^{th}\text{ Column total})}{\text{Total sample size}}$$

If $\chi^2 > \chi^2_\alpha$ [with df = $(r-1)(c-1)$] then Reject H_0.

McNemar's Test of 2 Related Proportions

H_0: $\pi_1 = \pi_2$ versus H_A: $\pi_1 \neq \pi_2$.

Test statistic $z = \dfrac{B - C}{\sqrt{B + C}}$

If $z > z_{\alpha/2}$ then Reject H_0.

Goodness-of-fit test

H_0: The distribution is as assumed.

H_A: The distribution is **NOT** as assumed.

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

If $\chi^2 > \chi^2_\alpha$ (With df = $k-1$) then Reject H_0

Chapter 13

Sample correlation coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

$$= \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{\left[\sum x^2 - (\sum x)^2/n\right]\left[\sum y^2 - (\sum y)^2/n\right]}} = \frac{Sxy}{\sqrt{Sxx\,Syy}}$$

For testing H_0: ρ=0 vs. H_A: ρ≠0

Test statistic $t_{n-2} = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$ has df = $n-2$

If $t_{n-2} > t_{\alpha/2, n-2}$ then reject H_0.

Estimated regression model $\hat{y}_i = b_0 + b_1 x$

The Least Square Estimates are $b_1 = \dfrac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$

$$b_1 = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{(\sum x)^2}{n}} = \frac{Sxy}{Sxx}$$

And $b_0 = \bar{y} - b_1\bar{x}$

Error $e_i = Y_i - \hat{Y}_i$

Total Sum of Squares

$$SST = Syy = \sum(y - \bar{y})^2$$

$$= \sum_{1}^{n} y_i^2 - n\bar{y}^2$$

Regression Sum of Squares $SSR = \sum(\hat{y} - \bar{y})^2$

Sum of Squares Error $SSE = SST - SSR$

$$SSE = \frac{Sxx\,Syy - (Sxy)^2}{Sxx}$$

(ALSO $SSE = \sum(y - \hat{y})^2$)

$$Sxy = \sum xy - (\sum x)(\sum y)/n$$

Coefficient of Determination

R- Squared = $R^2 = \frac{SSR}{SST}$

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = r^2$$

Standard Error of the Estimate

$$s_\varepsilon = \sqrt{\frac{SSE}{n-1-1}}$$

Standard Error of the Slope

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum(x-\bar{x})^2}} = \frac{s_\varepsilon}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} = \frac{s_\varepsilon}{\sqrt{Sxx}}$$

For testing H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$
The test statistic & **C.I. for the slope**

$$t_{n-2} = \frac{b_1 - \beta_1}{s_{b_1}} \qquad \& \qquad b_1 \pm t_{\alpha/2} s_{b_1}$$

C.I. for the **mean** of y given a particular x_p

$$\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

C.I. estimate for an **Individual value of y** given a particular x_p

$$\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x-\bar{x})^2}}$$

Chapter 13, 14, 15 & 16

For testing H_0: ρ=0 vs. H_A: ρ≠0
Durbin-Watson Test statistic

$$d = \frac{\sum_{t=1}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$, If $d < d_L$ reject H_0.

If $d > d_U$ don't reject H_0. Inconclusive otherwise.

Chapter 14 & 15

Estimated multiple regression model

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$$

Two variable model is

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 \qquad \& \qquad e_i = y_i - \hat{y}_i$$

is Errors (residuals) from regression model

Proportion of variation in *y* explained by all ***x*** variables adjusted for sample size and the number of ***x*** variables.

$$R_A^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k-1}\right)$$

For testing
H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
H_A: at least one $\beta_i \neq 0$

Test statistic $F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE}$

with $df_1 = k$ and $df_2 = n-k-1$

Contribution of X_j given other X variables

$$SSR(X_j \mid \text{All except } X_j) = SSR(\text{All}) - SSR(\text{All except } X_j)$$

Coefficient of Partial Determination X_j given other X variables already included

$$r^2_{YX_j \bullet (\text{All except } X_j)} = \frac{SSR(X_j \mid \text{All except } X_j)}{SST - SSR(\text{All}) + SSR(X_j \mid \text{All except } X_j)}$$

For testing H_0: $\beta_i = 0$ vs. H_A: $\beta_i \neq 0$
The test statistic & C.I. for the slope β_i are

$$t_{n-k-1} = \frac{b_i - 0}{s_{b_i}} \quad \text{or} \quad F_{1,n-k-1} = \frac{SSR(X_j \mid \text{All except } X_j)}{MSE}$$

$$t^2_{n-k-1} = F_{1,n-k-1}$$

C.I. for the slope β_i is $b_i \pm t_{\alpha/2} s_{b_i}$,.

The estimate of the standard error of the regression model

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}$$

Variance Inflationary Factor $VIF_j = \frac{1}{1-R_j^2}$

Contribution of X_j given other X variables

$$SSR(X_j \mid \text{All except } X_j) = SSR(\text{All}) - SSR(\text{All except } X_j)$$

C_p statistic: $C_p = \frac{(1-R_k^2)(n-T)}{1-R_T^2} - [n - 2(k+1)]$

For testing H_0: $\beta_c = \beta_d = \ldots\ \beta_m = 0$
against H_A: at least one $\beta_i \neq 0$
Test statistic

$$F = \frac{\dfrac{SSR_{Full} - SSR_{Reduced}}{m}}{\dfrac{SSE}{n-k-1}} = \frac{\dfrac{SSE_{Reduced} - SSE_{Full}}{m}}{MSE}$$

with $df_1 = m$ and $df_2 = n - k - 1$

Chapter 16

Simple Index number formula & Unweighted aggregate price index formula (respectively)

$$I_t = \frac{y_t}{y_0}100 \ \ \& \ \ I_t = \frac{\sum p_t}{\sum p_0}100$$

Paasche Weighted Aggregate Price Indexes

$$I_t = \frac{\sum q_t p_t}{\sum q_t p_0}100$$

& Laspeyres $I_t = \dfrac{\sum q_0 p_t}{\sum q_0 p_0}100$

Deflation formula $y_{adj_t} = \dfrac{y_t}{I_t}100$

Forecasting formula & Residual formula are

$$F_t = \hat{y} = b_0 + b_1 t \ \ \& \ \ e_t = y_t - F_t$$

respectively.

Mean Square Error & Mean Absolute Deviation are (respectively)

$$MSE = \frac{\sum (y_t - F_t)^2}{n} \ \ \& \ \ MAD = \frac{\sum |y_t - F_t|}{n}$$

Multiplicative Time-Series Model

$$y_t = T_t \times S_t \times C_t \times I_t$$

T_t = Trend value $\quad$ S_t = Seasonal value
C_t = Cyclical value $\quad$ I_t = Irregular (random) value

Ratio-to-Moving Average formula & Deseasonalizing formula (respectively)

$$S_t \times I_t = \frac{y_t}{T_t \times C_t} \qquad \& \qquad T_t \times C_t \times I_t = \frac{y_t}{S_t}$$

Single Exponential Smoothing Model

$F_{t+1} = F_t + \alpha(y_t - F_t) = \alpha y_t + (1-\alpha)F_t$ where
α: smoothing constant.

Double Exponential Smoothing Model

$F_{t+1} = C_t + T_t$ where $C_t = \alpha y_t + (1-\alpha)(C_{t-1} + T_{t-1})$
& $T_t = \beta(C_t - C_{t-1}) + (1-\beta)T_{t-1}$

α: Constant-process smoothing constant
β: Trend-smoothing constant

Exponential Trend Model $Y_t = \beta_o \beta_1^{X_t} \varepsilon_t$

Transformed Exponential Trend Model

$$\log(Y_t) = \log(\beta_o) + X_t \log(\beta_1) + \log(\varepsilon_t)$$

Exponential Model for Quarterly data

$$Y_t = \beta_o \beta_1^{X_t} \beta_2^{Q_1} \beta_3^{Q_2} \beta_4^{Q_3} \varepsilon_t$$

Transformed Exponential Model for Quarterly data

$$\log(Y_t) = \log(\beta_o) + X_t \log(\beta_1) + Q_1 \log(\beta_2) + Q_2 \log(\beta_3) + Q_3 \log(\beta_4) + \log(\varepsilon_t)$$

***p*th-order Autoregressive Model**

$$Y_t = A_o + A_1 Y_{t-1} + A_2 Y_{t-2} \ldots + A_p Y_{t-p} + \varepsilon_t$$

where
A_p: the *p*th autoregressive parameter

Some Useful STAT211 Formulas

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \text{ or } \frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}{n-1}$$

Binomial $P(x) = \dfrac{n!}{x!(n-x)!} p^x q^{n-x}$,

$\mu_X = E[X] = np$, $\sigma_X = \sqrt{npq}$

Poisson $P(x) = \dfrac{(\lambda t)^x}{x!} e^{-\lambda t}$,

$\mu_X = E[X] = \lambda t$, $\sigma_X = \sqrt{\lambda t}$

Hypergeometric $P(x) = \dfrac{C_{n-x}^{N-X} C_x^X}{C_n^N}$,

Discrete Uniform $P(x) = \dfrac{1}{k}$, $\quad$ for k discrete points

<u>**Continuous**</u>

Uniform function $f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$

$\mu_X = E[X] = \dfrac{a+b}{2}$, $\sigma_X = \sqrt{\dfrac{(b-a)^2}{12}}$

Exponential function $f(x) = \lambda e^{-\lambda x}$,

$P(0 \le x \le a) = 1 - e^{-\lambda a}$, $\mu_X = E[X] = \dfrac{1}{\lambda}$, $\sigma_X = \dfrac{1}{\lambda}$