Point Estimation

$\mu_{\hat{\theta}} - \theta = \textit{bias}$ of $\hat{\theta}$ as an estimate of θ
$\text{MSE}_{\hat{\theta}} = \mu_{(\hat{\theta}-\theta)^2} = (\mu_{\hat{\theta}} - \theta)^2 + \sigma^2_{\hat{\theta}}$

Inference Patterns

Confidence interval for θ :

$\hat{\theta} \pm$ (table value for confidence) $\times\, \sigma_{\hat{\theta}}$

Test of $H_0 : \theta = \theta_0$:

Test statistic: $\dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

P-value $= P$(a result at least as extreme as the test statistic $\mid H_0$), which depends on H_1:

$H_1 : \theta > \theta_0 \implies P\text{-value} = \text{right tail area}$
$H_1 : \theta < \theta_0 \implies P\text{-value} = \text{left tail area}$
$H_1 : \theta \neq \theta_0 \implies P\text{-value} = \text{both tails}$

Test-Interval Relationship

Level-α test of $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ retains H_0 $\iff$ μ_0 is inside a $1-\alpha$ confidence interval for μ.

χ^2 Tests

Goodness-of-Fit :

$E_i = np_i$, where n = sample size, p_i = expected proportion in category i

$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1}$,
where k = #category values

Independence / Homogeneity :

$$E_{ij} = \frac{O_{i.}O_{.j}}{O_{..}} = \frac{\text{(row total)(column total)}}{\text{table total}}$$

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{\nu},$$

where ν = (#rows - 1)(#columns - 1)

One Mean, μ, $n > 30$

$\bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$; sample size $n = \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2$
$Z = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim N(0,1)$

Small Samples with Normal X

$\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$
$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$

Difference of Two Means, $\mu_X - \mu_Y$

Large Samples

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}$$

$$Z = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \sim N(0,1)$$

Small Samples with Normal X and Y

$$\nu = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X-1} + \frac{(s_Y^2/n_Y)^2}{n_Y-1}}, \text{ rounded down}$$

$$(\bar{X} - \bar{Y}) \pm t_{\nu,\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

$$T = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} \sim t_{\nu}$$

One Proportion, p

($np > 10$ and $n(1-p) > 10$)
Plus-four: $\tilde{n} = n + 4$, $\tilde{p} = \frac{X+2}{\tilde{n}}$: $\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$
sample size $n = -4 + \left(\frac{z_{\alpha/2}}{m}\right)^2 \tilde{p}(1-\tilde{p})$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$$

Difference of Two Proportions, $p_X - p_Y$

(#successes, #failures all > 10)
Plus-four: $\tilde{n}_X = n_X + 2, \tilde{p}_X = \frac{X+1}{\tilde{n}_X}, \tilde{n}_Y = \cdots$

$$(\tilde{p}_X - \tilde{p}_Y) \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}_X(1-\tilde{p}_X)}{\tilde{n}_X} + \frac{\tilde{p}_Y(1-\tilde{p}_Y)}{\tilde{n}_Y}}$$

$$Z = \frac{(\hat{p}_X - \hat{p}_Y) - 0}{\sqrt{\hat{p}(1-\hat{p})(1/n_X + 1/n_Y)}} \sim N(0,1),$$

where pooled $\hat{p} = \frac{X+Y}{n_X+n_Y}$

Multiple Tests

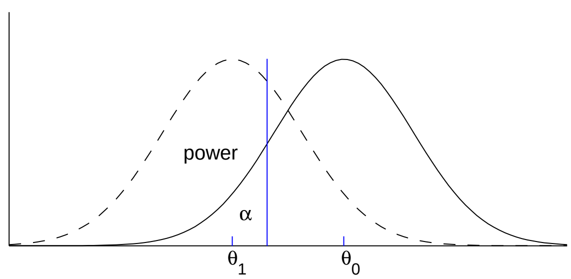
Problem: When running N tests at level α, we expect to reject H_0 when it's true for about $N\alpha$ of them. Bonferroni multiplies each P-value by N and compares NP to level α. Guidelines:

- $NP < \alpha \implies$ reject H_0 conclusively
- $P < \alpha$ but $NP \nless \alpha \implies$ retest the hypothesis with small original P

Level α and Power

For $H_0 : \theta = \theta_0$,

- A level α test rejects $H_0 \iff P$-value $\leq \alpha$: level $\alpha = P(\text{reject } H_0 \,|\, H_0 \text{ is true })$
- The point $\hat{\theta}_{\text{critical}}$ gives a P-value $= \alpha$. For $H_1 : \theta < \theta_0$, it's indicated by the vertical line, with area α left of the line under the solid curve
- The power of the test is $P(\text{reject } H_0 \,|\, H_0 \text{ is false because } \theta = \theta_1) =$ area left of line under dotted curve



Matched Pairs

Choose subjects in pairs to minimize differences within a pair. Study $\{D_i = X_i - Y_i\}$ as one sample from a population of differences.

F Test for Equality of Variances

For normal X and Y, to test $H_0 : \sigma_X^2 = \sigma_Y^2$, use $f = s_X^2 / s_Y^2$, swapping X and Y if necessary so $f \geq 1$.

$H_1 : \sigma_X^2 > \sigma_Y^2 \implies P$-value $= P(F_{n_X-1,n_Y-1} > f)$, area right of f

$H_1 : \sigma_X^2 \neq \sigma_Y^2 \implies P$-value = twice the P-value from one-sided test

[Regression is not on exam 2.]

Simple Regression

Assume $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, with $\varepsilon_1, \cdots, \varepsilon_n$ random, independent, and $\sim N(0, \sigma^2)$
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_1 = r\frac{s_y}{s_x}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
Residual plot of $\{(\hat{y}_i, e_i)\}$ should show no pattern or varying vertical spread

Inference on Coefficients

$\sigma^2 \approx s^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2 = (1-r^2)\frac{n-1}{n-2}s_y^2$
$\hat{\beta}_0 \pm t_{n-2,\alpha/2} s_{\hat{\beta}_0}$, where $s_{\hat{\beta}_0} = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}}$
$\hat{\beta}_1 \pm t_{n-2,\alpha/2} s_{\hat{\beta}_1}$, where $s_{\hat{\beta}_1} = \frac{s}{s_x\sqrt{n-1}}$
$t = \frac{\hat{\beta}_0 - \beta_{0_0}}{s_{\hat{\beta}_0}} \sim t_{n-2}$ tests $H_0 : \beta_0 = \beta_{0_0}$
$t = \frac{\hat{\beta}_1 - \beta_{1_0}}{s_{\hat{\beta}_1}} \sim t_{n-2}$ tests $H_0 : \beta_1 = \beta_{1_0}$

Inference on Mean Response, $\mu_y = \beta_0 + \beta_1 x$

$(\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{n-2,\alpha/2} s_{\hat{y}}$, where $s_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}}$
$\frac{\hat{y} - \mu_0}{s_{\hat{y}}} \sim t_{n-2}$ tests $H_0 : \mu_y = \mu_0$