Formulario Segundo Parcial CO3321

1. **Pruebas de hipótesis:**

   a) Para la media (todos los casos):

   Bilateral: $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$
   Unilateral derecha: $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$
   Unilateral izquierda: $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$

   1) Para la media con varianza poblacional conocida:

      Estadístico de prueba bajo H_0: $Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} \sim N(0,1)$
      Bilateral: $RR = (-\infty\, ,\, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}\, ,\, \infty)$
      $pvalor = 2P(Z > |z_{obs}|)$

      Unilateral derecha: $RR = (z_\alpha\, ,\, \infty)$
      $pvalor = P(Z > z_{obs})$

      Unilateral izquierda: $RR = (-\infty\, ,\, -z_\alpha)$
      $pvalor = 1 - P(Z > z_{obs})$

   2) Para la media con varianza poblacional desconocida y $n \geq 30$:

      Estadístico de prueba bajo H_0: $Z = \frac{\bar{X}-\mu_0}{S_{\bar{X}}} \sim N(0,1)$
      Bilateral: $RR = (-\infty\, ,\, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}\, ,\, \infty)$
      $pvalor = 2P(Z > |z_{obs}|)$

      Unilateral derecha: $RR = (z_\alpha\, ,\, \infty)$
      $pvalor = P(Z > z_{obs})$

      Unilateral izquierda: $RR = (-\infty\, ,\, -z_\alpha)$
      $pvalor = 1 - P(Z > z_{obs})$

   3) Para la media con varianza poblacional desconocida y $n < 30$:

      Estadístico de prueba bajo H_0: $T = \frac{\bar{X}-\mu_0}{S_{\bar{X}}} \sim t_{n-1}$
      Bilateral: $RR = (-\infty\, ,\, -t_{n-1,\frac{\alpha}{2}}) \cup (t_{n-1,\frac{\alpha}{2}}\, ,\, \infty)$
      $pvalor = 2P(T > |t_{obs}|)$

      Unilateral derecha: $RR = (t_{n-1,\alpha}\, ,\, \infty)$
      $pvalor = P(T > t_{obs})$

      Unilateral izquierda: $RR = (-\infty\, ,\, -t_{n-1,\alpha})$
      $pvalor = 1 - P(T > t_{obs})$

   b) Pruebas de hipótesis para proporciones:

Estadístico de prueba bajo H_0: $Z = \frac{\hat{p}-p_0}{\sigma_{p_0}} \sim N(0,1)$

Bilateral: $H_0 : p = p_0$ contra $H_1 : p \neq p_0$
$RR = (-\infty\,,\,-z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}\,,\,\infty)$
$pvalor = 2P(Z > |z_{obs}|)$

Unilateral derecha: $H_0 : p \leq p_0$ contra $H_1 : p > p_0$
$RR = (z_\alpha\,,\,\infty)$
$pvalor = P(Z > z_{obs})$

Unilateral izquierda: $H_0 : p \geq p_0$ contra $H_1 : p < p_0$
$RR = (-\infty\,,\,-z_\alpha)$
$pvalor = 1 - P(Z > z_{obs})$

c) Pruebas de hipótesis para diferencias de proporciones:

Estadístico de prueba bajo H_0:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1)$$

Bilateral: $H_0 : p_1 - p_2 = p_0$ contra $H_1 : p_1 - p_2 \neq p_0$
$RR = (-\infty\,,\,z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}\,,\,\infty)$
$pvalor = 2P(Z > |z_{obs}|)$

Unilateral derecha: $H_0 : p_1 - p_2 \leq p_0$ contra $H_1 : p_1 - p_2 > p_0$
$RR = (z_\alpha\,,\,\infty)$
$pvalor = P(Z > z_{obs})$

Unilateral izquierda: $H_0 : p_1 - p_2 \geq p_0$ contra $H_1 : p_1 - p_2 < p_0$
$RR = (-\infty\,,\,z_\alpha)$
$pvalor = 1 - P(Z > z_{obs})$

d) Pruebas de hipótesis para la varianza (o desviación estándar):

Estadístico de prueba bajo H_0:

$$\chi^2 = \frac{(n-1)S_1^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

Bilateral: $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$
$RR = (0,a) \cup (b,\infty)$
donde $a = \chi^2_{n-1,1-\frac{\alpha}{2}}$ y $b = \chi^2_{n-1,\frac{\alpha}{2}}$
$pvalor = 2 * min(P(\chi^2 > \chi^2_{obs}), P(\chi^2 < \chi^2_{obs}))$

Unilateral derecha: $H_0 : \sigma^2 \leq \sigma_0^2$ contra $H_1 : \sigma^2 > \sigma_0^2$
$RR = (b,\infty)$
donde $b = \chi^2_{n-1,\alpha}$
$pvalor = P(\chi^2 > \chi^2_{obs})$

Unilateral izquierda: $H_0 : \sigma^2 \geq \sigma_0^2$ contra $H_1 : \sigma^2 < \sigma_0^2$
$RR = (0, a)$
donde $a = \chi^2_{n-1,1-\alpha}$
$pvalor = 1 - P(\chi^2 > \chi^2_{obs})$

e) Pruebas de hipótesis para la media (datos apareados):
Igual que en las pruebas de hipótesis para la media reemplazando X por D, donde $D = X_i - Y_i$

f) Pruebas de hipótesis para la diferencia de medias (Todos los casos):

Bilateral: $H_0 : \mu_1 - \mu_2 = \mu_0$ contra $H_1 : \mu_1 - \mu_2 \neq \mu_0$
Unilateral derecha: $H_0 : \mu_1 - \mu_2 \leq \mu_0$ contra $H_1 : \mu_1 - \mu_2 > \mu_0$
Unilateral izquierda: $H_0 : \mu_1 - \mu_2 \geq \mu_0$ contra $H_1 : \mu_1 - \mu_2 < \mu_0$

1) Para la diferencia de medias con varianzas poblacionales conocidas:

Estadístico de prueba bajo H_0:

$$Z = \frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Bilateral: $RR = (-\infty - z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty)$
$pvalor = 2P(Z > |z_{obs}|)$

Unilateral derecha: $RR = (z_\alpha, \infty)$
$pvalor = P(Z > z_{obs})$

Unilateral izquierda: $RR = (-\infty - z_\alpha)$
$pvalor = 1 - P(Z > z_{obs})$

2) Para la diferencia de medias con varianzas poblacionales desconocidas y $n_1 \geq 30$ y $n_2 \geq 30$:

Estadístico de prueba bajo H_0:

$$Z = \frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

Bilateral: $RR = (-\infty \ , \ -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}} \ , \ \infty)$
$pvalor = 2P(Z > |z_{obs}|)$

Unilateral derecha: $RR = (z_\alpha \ , \ \infty)$
$pvalor = P(Z > z_{obs})$

Unilateral izquierda: $RR = (-\infty \ , \ -z_\alpha)$
$pvalor = 1 - P(Z > z_{obs})$

3) Para la media con varianzas poblacionales desconocidas e iguales y $n_1 < 30$ o $n_2 < 30$:

Estadístico de prueba bajo H_0:

$$T = \frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \sim t_{n_1+n_2-2}$$

donde $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

Estadístico de prueba bajo H_0:
Bilateral: $RR = (-\infty\ ,\ -t_{n_1+n_2-2\ ,\ \frac{\alpha}{2}}) \cup (t_{n_1+n_2-2, \frac{\alpha}{2}}\ ,\ \infty)$
$pvalor = 2P(T > |t_{obs}|)$

Unilateral derecha: $RR = (t_{n_1+n_2-2,\alpha}\ ,\ \infty)$
$pvalor = P(T > t_{obs})$

Unilateral izquierda: $RR = (-\infty\ ,\ -t_{n_1+n_2-2,\alpha})$
$pvalor = 1 - P(T > t_{obs})$

4) Para la media con varianzas poblacionales desconocidas y distintas y $n < 30$:

Estadístico de prueba bajo H_0:

$$T = \frac{\bar{X} - \bar{Y} - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\nu$$

donde $\nu = \left[\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} \right]$

Bilateral: $RR = (-\infty\ ,\ -t_{\nu,\frac{\alpha}{2}}) \cup (t_{\nu,\frac{\alpha}{2}}\ ,\ \infty)$
$pvalor = 2P(T > |t_{obs}|)$

Unilateral derecha: $RR = (t_{\nu,\alpha}\ ,\ \infty)$
$pvalor = P(T > t_{obs})$

Unilateral izquierda: $RR = (-\infty\ ,\ -t_{\nu,\alpha})$
$pvalor = 1 - P(T > t_{obs})$

g) Pruebas de hipótesis para el cociente de varianzas:

Estadístico de prueba bajo H_0: $F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1\,,\,n_2-1}$

Bilateral: $H_0 : \sigma_1 = \sigma_2$ contra $H_1 : \sigma_1 \neq \sigma_2$
$RR = (0, a) \cup (b, \infty)$

donde $b = f_{n_1-1\,,\,n_2-1\,,\,\alpha/2}$ y $a = \frac{1}{f_{n_2-1\,,\,n_1-1\,,\,\alpha/2}}$

$pvalor = 2 * min(P(F > f_{obs}), P(F < f_{obs}))$

Unilateral derecha: $H_0 : \sigma_1 \leq \sigma_2$ contra $H_1 : \sigma_1 > \sigma_2$
$RR = (b, \infty)$
$pvalor = P(F > f_{obs})$

donde $b = f_{n_1-1\,,\,n_2-1\,,\,\alpha}$

Unilateral izquierda: $H_0 : \sigma_1 \geq \sigma_2$ contra $H_1 : \sigma_1 < \sigma_2$
$RR = (0, a)$

donde $a = \frac{1}{f_{n_2-1\,,\,n_1-1\,,\,\alpha}}$
$pvalor = 1 - P(F > f_{obs})$

2. **Bondad de ajuste:**

   Estadístico de prueba bajo H_0: $\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \sim \chi^2_{k-1-r}$

   $RR = (\chi^2_{k-1-r,\alpha}, \infty)$

3. **Regresión Lineal Simple**:

   $SS_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \qquad SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$
   $SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad S^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

   - X variable independiente:
     $y = \hat{\beta}_0 + \hat{\beta}_1 x$, donde $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$ y $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$
   - Y variable independiente:
     $x = \hat{\beta}_0 + \hat{\beta}_1 y$, donde $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{yy}}$ y $\hat{\beta}_0 = \bar{x} - \hat{\beta}_1\bar{y}$

   a) Error estándar de la estimación: $S = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$

   b) Coeficiente de correlación muestral: $R = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$

   c) Coeficiente de determinación: $R^2 = \frac{(SS_{xy})^2}{SS_{xx}SS_{yy}}$

   d) Intervalo de confianza para β_0:
   El estadístico es

   $$T = \frac{\hat{\beta}_0 - \beta_0}{S\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{SS_{xx}}}} \sim t_{n-2}$$

   El intervalo de confianza es $I = \hat{\beta}_0 \pm t_{n-2;\frac{\alpha}{2}} S\sqrt{\frac{1}{n} + \frac{\hat{X}^2}{SS_{xx}}}$

   e) Intervalo de confianza para β_1:

El estadístico es

$$T = \frac{(\hat{\beta}_1 - \beta_1)\sqrt{SS_{xx}}}{S} \sim t_{n-2}$$

El intervalo de confianza es $I = \hat{\beta}_1 \pm \frac{t_{n-2;\frac{\alpha}{2}} S}{\sqrt{SS_{xx}}}$

f) Pruebas de hipótesis para β_0:
El estadístico de prueba bajo H_0 es:

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{S\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{SS_{xx}}}} \sim t_{n-2}$$

1) Si la prueba es bilateral:
$H_0 : \beta_0 = \beta_{00}$ contra $H_1 : \beta_0 \neq \beta_{00}$
$RR = (-\infty, -t_{n-2;\alpha/2}) \cup (t_{n-2;\alpha/2}, \infty)$
$p - valor = 2P(T > |t_{obs}|)$

2) Si la prueba es unilateral derecha:
$H_0 : \beta_0 \leq \beta_{00}$ contra $H_1 : \beta_0 > \beta_{00}$
$RR = (t_{n-2;\alpha}, \infty)$
$p - valor = P(T > t_{obs})$

3) Si la prueba es unilateral izquierda:
$H_0 : \beta_0 \geq \beta_{00}$ contra $H_1 : \beta_0 < \beta_{00}$
$RR = (-\infty, -t_{n-2;\alpha})$
$p - valor = 1 - P(T > t_{obs})$

g) Pruebas de hipótesis para β_1:
El estadístico de prueba bajo H_0 es:

$$T = \frac{(\hat{\beta}_1 - \beta_{10})\sqrt{SS_{xx}}}{S} \sim t_{n-2}$$

1) Si la prueba es bilateral:
$H_0 : \beta_1 = \beta_{10}$ contra $H_1 : \beta_1 \neq \beta_{10}$
$RR = (-\infty, -t_{n-2;\alpha/2}) \cup (t_{n-2;\alpha/2}, \infty)$
$p - valor = 2P(T > |t_{obs}|)$

2) Si la prueba es unilateral derecha:
$H_0 : \beta_1 \leq \beta_{10}$ contra $H_1 : \beta_1 > \beta_{10}$
$RR = (t_{n-2;\alpha}, \infty)$
$p - valor = P(T > t_{obs})$

3) Si la prueba es unilateral izquierda:
$H_0 : \beta_1 \geq \beta_{10}$ contra $H_1 : \beta_1 < \beta_{10}$
$RR = (-\infty, -t_{n-2;\alpha})$
$p - valor = 1 - P(T > t_{obs})$

h) Intervalo de predicción:
El estadístico es

$$T = \frac{\hat{y} - y}{S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_{xx}}}} \sim t_{n-2}$$

El intervalo de predicción es:

$$I = \hat{y} \pm t_{n-2,\frac{\alpha}{2}} S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_{xx}}}$$

donde $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_o$ es el valor de predicción cuando $x = x_o$

i) Pruebas de hipótesis para la predicción cuando $x = x_0$:
El estadístico de prueba bajo H_0 es:

$$T = \frac{\hat{y} - y_0}{S\sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{SS_{xx}}}} \sim t_{n-2}$$

1) Si la prueba es bilateral:
$H_0 : y = y_0$ contra $H_1 : y \neq y_0$
$RR = (-\infty, -t_{n-2;\alpha/2}) \cup (t_{n-2;\alpha/2}, \infty)$
$p - valor = 2P(T > |t_{obs}|)$

2) Si la prueba es unilateral derecha:
$H_0 : y \leq y_0$ contra $H_1 : y > y_0$
$RR = (t_{n-2;\alpha}, \infty)$
$p - valor = P(T > t_{obs})$

3) Si la prueba es unilateral izquierda:
$H_0 : y \geq y_0$ contra $H_1 : y < y_0$
$RR = (-\infty, -t_{n-2;\alpha})$
$p - valor = 1 - P(T > t_{obs})$

4. **Regresión Lineal Múltiple:**

Tabla ANOVA de Regresión Múltiple:

	Sum of Sq	**df**	**Mean Square**	**F**	$Pr(>F)$
Regression	SSR	p	$MSR = \frac{SSR}{p}$	$\frac{MSR}{MSE}$	$P(F > fobs)$
Residual	SSE	$n-p-1$	$MSE = s^2 = \frac{SSE}{n-p-1}$		
Total	$SSR + SSE$	$n-1$	$MSR + MSE$		

Tabla de Coeficientes de Regresión Múltiple:

	Beta estimate	**Std. error**	**t**	$Pr(>T)$
Intercept	$\hat{\beta}_0$	$s\sqrt{c_{00}}$	$\frac{\hat{\beta}_0}{S\sqrt{c_{00}}}$	$2P(T>\lvert tobs\rvert)$
x_1	$\hat{\beta}_1$	$s\sqrt{c_{11}}$	$\frac{\hat{\beta}_1}{S\sqrt{c_{11}}}$	$2P(T>\lvert tobs\rvert)$
x_2	$\hat{\beta}_2$	$s\sqrt{c_{22}}$	$\frac{\hat{\beta}_2}{S\sqrt{c_{22}}}$	$2P(T>\lvert tobs\rvert)$
...	...	...	...	...
x_p	$\hat{\beta}_p$	$s\sqrt{c_{pp}}$	$\frac{\hat{\beta}_p}{S\sqrt{c_{pp}}}$	$2P(T>\lvert tobs\rvert)$

donde los C_{ii} se obtienen a partir de:

$$(X^TX)^{-1}=\begin{pmatrix} c_{00} & c_{01} & ... & c_{0p} \\ c_{10} & c_{11} & ... & c_{1p} \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ c_{p0} & c_{p1} & ... & c_{pp} \end{pmatrix}$$

Coeficiente de determinación múltiple:

$$R^2 = 1 - \frac{SSE}{SS_{yy}}$$

Coeficiente de determinación múltiple ajustado:

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

Error típico de estimación:

$$S_e = \sqrt{\sum_{i=1}^{n} \frac{(y_i-\hat{y})^2}{n-p-1}}$$

Estadístico de prueba para modelos reducido y completo:

$H_0 : \beta_{g+1} = \beta g + 2 = ... = \beta p = 0$ contra H_1 : Alguno de estos es distinto de cero.

El estadístico de prueba bajo H_0 es:

$$F = \frac{(SSE_1 - SSE_2)(n-p-1)}{SSE_2(p-g)} \sim F_{p-g,n-p-1}$$

donde:

$SSE_1 = Y^TY - \hat{\beta}^T X^T Y = \sum_{i=1}^{n} y_i^2 - \hat{\beta}^T X^T Y$ es la suma de los cuadrados de las desviaciones de los valores observados de y, donde $y = \beta_0 + \beta_1 x_1 + ... + \beta_g x_g + \varepsilon$ es el modelo reducido.

$SSE_2 = Y^TY - \hat{\beta}^T X^T Y = \sum_{i=1}^{n} y_i^2 - \hat{\beta}^T X^T Y$ es la suma de los cuadrados de las desviaciones de los valores observados de y, donde $y = \beta_0 + \beta_1 x_1 + ... + \beta_g x_g + \beta_{g+1} x_{g+1} + ... + \beta_p x_p + \varepsilon$ es el modelo completo.