StatCrunch Lab 11 for Statistics 301

Topics: tests and CIs of (independent) $p_1 - p_2$, (dependent or paired) μ_d , independent $\mu_1 - \mu_2$, tests of $\frac{\sigma_1}{\sigma_2}$, F distribution probability and percentiles

Differences in independent proportions $p_1 - p_2$.
Consider number of male doctors in military and civilian hospitals. Test claim there is a *smaller* proportion of male doctors in military than in civilian life at $\alpha = 0.05$. Calculate 95% CI of difference in proportions.

	military (1)	civilian (2)
male doctors	358	6786
total doctors	407	7363

1. *Hypothesis test.*
   (a) *Statement.*
   $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 < 0$
   (b) *Test.*

$$P(\hat{p}_1 - \hat{p}_2 \leq -0.042) = P\left(\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \leq -3.03\right) \approx P\left(Z \leq -3.03\right) \approx 0.001$$

   Stat, Proportions, Two samples, with summary, Sample 1: Number of successes: 358, Number of observations: 407, Sample 2: Number of successes: 6786, Number of observations: 7363, Next, Null: prop. = 0 Alternative: < Calculate.
   Level of significance $\alpha = 0.05$
   (c) *Conclusion.*
   Since p–value = $0.001 < \alpha = 0.050$, reject null guess: $H_0 : p_1 - p_2 = 0$.

2. *Confidence interval.*

$$\begin{aligned}\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} &= \left(\frac{358}{407} - \frac{6786}{7363}\right) \\ &\pm 1.96 \cdot \sqrt{\frac{\frac{358}{407}\left(1 - \frac{358}{407}\right)}{407} + \frac{\frac{6786}{7363}\left(1 - \frac{6786}{7363}\right)}{7363}} \\ &= (-0.074, -0.010)\end{aligned}$$

   Options, choose Confidence Interval 0.95 Calculate.

Differences in dependent means $\mu_d = \mu_1 - \mu_2$.
A study is conducted to determine cellular response to progesterone in females. Blood cells from female 1 are broken into two groups. One group of these blood cells are injected with progesterone; the other group, the *control*, is, for comparison purposes, left untreated. Blood cells of females 2, 3 and 4 are handled in same way. Assume normality with no outliers. Test if mean progesterone response *greater* than mean control response at 5%. Calculate 95% CI.

female	progesterone (1)	control (2)
1	5.85	5.23
2	2.28	1.21
3	1.51	1.40
4	2.12	1.38

Relabel var1 progesterone, var2 control. Type data into these two columns. Data, Data expression, Expression: progesterone - control, New column name: difference, Compute.

1. *Hypothesis test.*
   (a) *Statement.*
   $H_0 : \mu_d = 0$ versus $H_1 : \mu_d > 0$
   (b) *Test.*

   $$\text{p–value} = P(\bar{d} \geq 0.635) = P\left(\frac{\bar{d} - \mu_0}{\frac{s}{\sqrt{n}}} \geq \frac{0.635 - 0}{\frac{0.398}{\sqrt{4}}}\right) \approx P\left(t \geq 3.19\right) \approx 0.025$$

   Stat, T statistics, Paired, Sample 1 in: progesterone, Sample 2 in: control, check Save differences, Next, Null: prop. = 0 Alternative: > Calculate. Level of significance $\alpha = 0.05$
   (c) *Conclusion.*
   Since p–value = 0.025 < $\alpha = 0.050$, reject null guess: $H_0 : \mu_d = 0$.

2. *Confidence interval.*

   $$\bar{d} \pm t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}} = 0.635 \pm 3.19 \times \frac{0.398}{\sqrt{4}} = (0.0011, 1.2689)$$

   Options, Edit, choose Confidence Interval 0.95, Calculate.

Differences in independent means $\mu_1 - \mu_2$.
A study is conducted to determine cellular response to progesterone in females. Blood cells from four females are injected with progesterone; blood cells from four *different* females are, for comparison purposes, left untreated. Test if average progesterone response is *greater* than average control response at 5%. Calculate 95% CI. Assume normality with no outliers.

female	progesterone (1)	female	control (2)
1	5.85	5	5.23
2	2.28	6	1.21
3	1.51	7	1.40
4	2.12	8	1.38

1. *Hypothesis test.*

   (a) *Statement.*
   $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 > 0$

   (b) *Test.*

$$P(\bar{x}_1 - \bar{x}_2 \geq 0.635) = P\left(\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \geq \frac{(2.94 - 2.305) - 0}{\sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}}}\right) \approx P\left(t \geq 0.458\right) = 0.33$$

   Stat, T statistics, Two sample, with data, Sample 1 in: progesterone, Sample 2 in: control, so *not* check Pool variances, Next, Null: prop. = 0 Alternative: > Calculate.
   Level of significance $\alpha = 0.05$

   (c) *Conclusion.*
   Since p–value = 0.33 > $\alpha = 0.050$, do not reject null guess: $H_0 : \mu_1 - \mu_2 = 0$.

2. *Confidence interval.*

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (2.94 - 2.305) \pm 2.45 \cdot \sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}} = (-2.75, 4.03)$$

Test of $\frac{\sigma_1}{\sigma_2}$.
Consider statistics on plasma levels for males and females given in table. Test if $\sigma_1 > \sigma_2$ at $\alpha = 0.05$. Assume both simple random samples normal with no outliers and collected independently of one another.

	males (1)	females (2)
$\bar{x}$	3.259	1.413
s	0.16	0.09
n	9	6

1. *Statement.*
   $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 > \sigma_2$

2. *Test.*

$$\text{p–value} = P\left(F \geq \frac{s_1^2}{s_2^2}\right) = P\left(F \geq \frac{0.16^2}{0.09^2}\right) \approx P\left(F \geq 3.16\right) \approx 0.11$$

   Stat, Variance, Two sample, with summary, Sample 1 Variance: 0.0256 (which is 0.16^2), Size: 9, Sample 2 Variance: 0.0081 (which is 0.09^2), Size: 6, Next, Null: variance ratio = 1 Alternative: > Calculate.
   Level of significance $\alpha = 0.05$

3. *Conclusion.*
   Since p–value $= 0.11 > \alpha = 0.05$, do not reject null guess: $H_0 : \sigma_1 = \sigma_2$.

Probability and Percentile For F Distribution.
At McDonalds in Westville, waiting time to order (in minutes) follows an F distribution. Consider following figure with two F distributions, each with different shaded areas (probabilities).

- For a F with (2,2) df, probability of waiting less than 3.9 minutes $P(F < 3.9) = 0.80$

  Stat, Calculators, F, Num. DF: 2, Den. DF: 2, Prob(X <=) 3.9 = $\boxed{?}$

- The 72nd percentile for F with (2,2) degrees of freedom, is 2.6.

  Stat Calculators, F, Num. DF: 2, Den. DF: 2, Prob(X <=) $\boxed{?}$ = 0.72