Statistical Significance and Univariate/Bivariate Tests

ECO 307: Introductory Econometrics

1

1.1 Goals

Goals

- Specific goals:
  - Re-familiarize ourselves with basic statistics ideas: sampling distributions, hypothesis tests, p-values.
  - Be able to distinguish different types of data and prescribe appropriate statistical methods.
  - Conduct a number of hypothesis tests using methods appropriate for questions involving only one or two variables.

2 Statistical Significance

2.1 Sampling Distribution

Probability Distribution

- **Probability distribution:** summary of all possible values a variable can take along with the probabilities in which they occur.
- Usually displayed as:

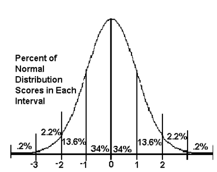
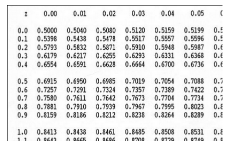
Picture 

Table 

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749

Formula

$$f(x|\mu,\sigma) =$$

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

- **Normal distribution:** often used "bell shaped curve", reveals probabilities based on how many standard deviations away an event is from the mean.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, proportion, difference between two sample means, standard deviation, variance, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
- Is this the same thing as the probability distribution of the population? NO! They may coincidentally have the same shape though.

Example

- Sampling Distribution Simulator
- In reality, you only do an experiment once, so the sampling distribution is a hypothetical distribution.
- Why are we interested in this?

Desirable qualities

What are some qualities you would like to see in a sampling distribution?

- The average of the sample statistics is equal to the true population parameter.
- Want the variance *of the sampling distribution* to be as small as possible. Why?
- Want the *sampling distribution* to be normal, regardless of the distribution of the population.

2.2 Central Limit Theorem

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a *sample mean* ($\bar{x}$) is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (unbiased):

    $$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

    $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem: Small samples

If n is small (rule of thumb for a single variable: $n < 30$)

- The sample mean is still unbiased.
- The formula for the standard deviation of the sampling distribution still holds ($\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$), but with a small n, the sampling distribution may be wide.
- Sampling distribution will be normal *only if* the distribution of the population is normal, so using the central limit theorem requires this additional assumption.

Example 1

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 1.5lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.5lbs$ or greater?

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{7.5 - 7}{1.5/\sqrt{30}} = 1.826$$

In R, we can compute $P(z > 1.826)$ using the formula
`=1 - pnorm(1.826)`. The probability the sample mean is greater than $7.5lbs$ is:

$$P(\bar{x} > 7.5) = P(z > 1.826) = 0.0339$$

Example 2

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 1.5lbs$.

What is the probability that a randomly selected baby will have a weight of 7.5lbs or more? What do you need to assume to answer this question?

Must assume the population is normally distributed. Why?

$$z = \frac{x - \mu}{\sigma} = \frac{7.5 - 7}{1.5} = 0.33$$

In R, we can compute $P(z > 0.33)$ using the formula
`=1 - pnorm(0.33)`. The probability that a baby is greater than 7.5lbs is:

$$P(x > 7.5) = P(z > 0.33) = 0.3707$$

Example 3

- Suppose average birth weight of all babies is $\mu = 7lbs$, and the standard deviation is $\sigma = 1.5lbs$.
- Suppose you collect a sample of 30 newborn babies whose mothers smoked during pregnancy.
- Suppose you obtained a sample mean $\bar{x} = 5\ lbs$. If you assume the mean birth weight of babies whose mothers smoked during pregnancy has the same sampling distribution as the rest of the population, what is the probability of getting a sample mean this low?

Example 3 continued

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{5 - 7}{1.5/\sqrt{30}} = -7.30$$

In R, we can compute $P(z < -7.30)$ using the formula
`=pnorm(-7.30)`.

The probability the sample mean is less than or equal to $5lbs$ is:

$$P(\bar{x} < 5) = P(z < -7.30) = 0.00000000000014$$

That is, if smoking during pregnancy actually truly led to an average birth weight of 7 pounds (we began with this assumption), there was only a 0.00000000000014 (or 0.000000000014%) chance of getting a sample mean as low as six or lower.

This is an extremely unlikely event if the assumption is true. Therefore it is likely the assumption is not true.

2.3 Hypotheses Tests

Statistical Hypotheses

- A **hypothesis** is a claim or statement about a property of a population.
  - Example: The population mean for income per household in the United States is $45,000.
- A **hypothesis test** (or **test of significance**) is a standard procedure for testing a claim about a property of a population.
- Recall the example about birth weights with mothers who smoke during pregnancy.
  - Hypothesis: Smoking during pregnancy leads to an average birth weight of 7 pounds (same average as with mothers who do not smoke during pregnancy).

Null and Alternative Hypotheses

- The **null hypothesis** is a statement that the value of a population parameter (such as the population mean) **is equal to** some value.
  - H_0: $\mu = 7$.
- The **alternative hypothesis** is an alternative to the null hypothesis; a statement that says a parameter **is different than** the value given in the null hypothesis.
- Pick *only one* of the following for your alternative hypothesis. Which one depends on your research question.
  - H_a: $\mu < 7$.
  - H_a: $\mu > 7$.
  - H_a: $\mu \neq 7$.
- In hypothesis testing, assume the null hypothesis is true until there is strong statistical evidence to suggest the alternative hypothesis.
- Similar to an "innocent until proven guilty" policy.

Hypothesis tests

- (Many) hypothesis tests are all the same:

$$z \text{ or } t = \frac{\text{sample statistic} - \text{null hypothesis value}}{\text{standard deviation of the sampling distribution}}$$

- Example: hypothesis testing about μ:
  - Sample statistic = $\bar{x}$.
  - Standard deviation of the sampling distribution of $\bar{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

P-values

- The P-value of the test statistic, is the area of the sampling distribution from the sample result in the direction of the alternative hypothesis.
- Interpretation: If the null hypothesis is correct, than the p-value is the probability of obtaining a sample that yielded your statistic, or a statistic that provides even stronger evidence against the null hypothesis.
- The p-value is therefore a measure of **statistical significance**.
  - If p-values are very small, there is strong statistical evidence in favor of the alternative hypothesis.
  - If p-values are large, there is insignificant statistical evidence. When large, you fail to reject the null hypothesis.
- **Significance level:** often denoted by α, a threshold p-value for deciding to reject versus fail to reject a null hypothesis.
- Common significance levels: $\alpha = 0.05$, $\alpha = 0.1$, $\alpha = 0.01$.
- Best practice is writing research: report the p-value. Different readers may have different opinions about how small a p-value should be before saying your results are statistically significant.

3 Univariate Tests

3.1 Hypothesis Testing about Mean

Single Mean T-Test

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- T-test is used instead of Z-test for reasons you may have learned in a statistics class. Interpretation is the same.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
  - Null: The population mean is equal to some specified value.
  - Alternative: The population mean is [greater/less/different] than the value in the null.

Example Questions

- Suppose we have a dataset of individual schools and the average pay for teachers in each school, and the level of spending as a ratio of the number of students (spending per pupil).
  - Show some descriptive statistics for teacher pay and expenditure per pupil.
  - Is there statistical evidence that teachers make less than $50,000 per year?
  - Is there statistical evidence that expenditure per pupil is more than $7,500?

3.2 Hypothesis Testing about Proportion

Single Proportion T-Test

- **Proportion:** Percentage of times some characteristic occurs.
- Example: percentage of consumers of soda who prefer Pepsi over Coke.

$$\text{Sample proportion} = \frac{\text{Number of items that has characteristic}}{\text{sample size}}$$

- Example questions:
  - Are more than 50% of potential voters in Wisconsin most likely to vote for Donald Trump in the next presidential election?
  - Suppose typical brand-loyalty turn-over in the mobile phone industry is 15%. Is there statistical evidence that AT&T has brand-loyalty turnover more than 15%?
- You can alternatively just use a single mean test for a proportion, where the variable is binary (0,1) and can be treated as interval/ratio data.

4 Bivariate Tests

4.1 Difference in Population Means (Independent Samples)

Difference in Means (Independent Samples)

- Suppose you want to know whether the mean from one population is larger than the mean for another.
- Independent samples means you have different individuals in your two sample groups.

- Examples:
  - Compare sales volume for stores that advertise versus those that do not.
  - Compare production volume for employees that have completed some type of training versus those who have not.
- Statistic: Difference in the sample means ($\bar{x}_1 - \bar{x}_2$).
- Hypotheses:
  - Null hypothesis: the difference between the two means is zero.
  - Alternative hypothesis: the difference is [above/below/not equal] to zero.

4.2 Difference in Population Means (Paired Samples)

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
  - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
  - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
  - Do students spend more time studying than watching TV?
  - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are not independent samples, because you have the same individuals in each group.

4.3 Correlation

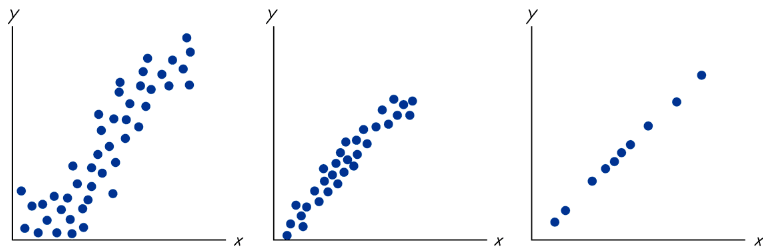
Correlation

- A **correlation** exists between two variables when one of them is related to the other in some way, such that there is **co-movement**.
- The **Pearson linear correlation coefficient** is a measure of the strength of the linear relationship between two variables.

- Null hypothesis: there is zero linear correlation between two variables.
- Alternative hypothesis: there is a linear correlation (positive / negative / either) between two variables.

- **Spearman Rank Test**
  - Non-parametric test.
  - Behind the scenes - replaces actual data with their *rank*, compute the Pearson correlation using ranks.
  - Appropriate also for *ordinal data*.
  - Useful for nonlinear, monotonic relationships
  - Same hypotheses.

Positive linear correlation



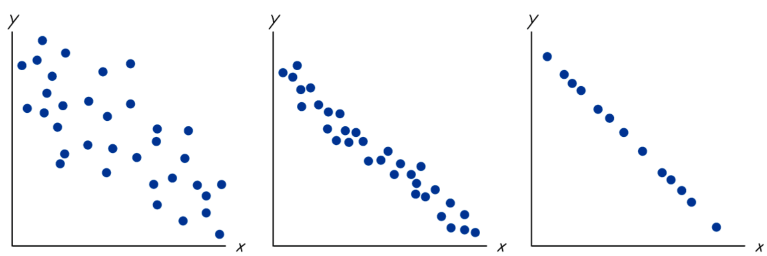
(a) Positive correlation between x and y

(b) Strong positive correlation between x and y

(c) Perfect positive correlation between x and y

- Positive correlation: two variables move in the same direction.
- Stronger the correlation: closer the correlation coefficient is to 1.
- Perfect positive correlation: $\rho = 1$

Negative linear correlation



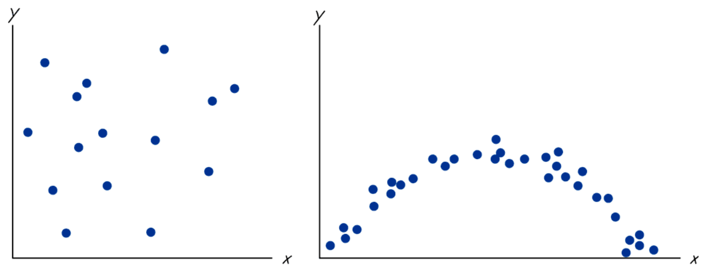
(d) Negative correlation between x and y

(e) Strong negative correlation between x and y

(f) Perfect negative correlation between x and y

- Negative correlation: two variables move in opposite directions.
- Stronger the correlation: closer the correlation coefficient is to -1.
- Perfect negative correlation: $\rho = -1$

No linear correlation



(g) No correlation between x and y

(h) Nonlinear relationship between x and y

- Panel (g): no relationship at all.
- Panel (h): strong relationship, but not a *linear* relationship.
  - Cannot use regular correlation to detect this.