Applied Regression Modeling: A Business Approach
Chapter 1: Foundations
Sections 1.5–1.7

by Iain Pardoe

© Iain Pardoe, 2006

Interval estimation

- Goal: estimate the population mean $\mathrm{E}(Y)$.
- Best point estimate: the sample mean m_Y.
- How far off might we be? Can we quantify our uncertainty?
- Confidence interval: point estimate $\pm$ uncertainty.
- Example: 80% confidence interval for $\mathrm{E}(Y)$ in home prices application is $278.603 \pm 12.893 = (265.710, 291.496)$.
- In other words, based on this dataset, we are 80% confident that the population mean home price is between $\$266{,}000$ and $\$291{,}000$.
- This leaves quite a bit of room for error (20%), so 90% and 95% intervals are more common.
- Question: will a 90% interval be narrower or wider than the 80% interval?

Confidence interval for E(Y)

- Example: 80% confidence interval.
- Pr $(-90^{\text{th}}$ percentile $< t_{n-1} < 90^{\text{th}}$ percentile$) = 0.80$ where the 90^{th} percentile comes from t_{n-1} (t-distribution with $n-1$ df).
- Question: why does an 80% interval require 90^{th} percentiles? (draw a picture)

Confidence interval for E(Y)

- Example: 80% confidence interval.
- $\Pr(-90^{\text{th}}\text{ percentile} < t_{n-1} < 90^{\text{th}}\text{ percentile}) = 0.80$ where the 90^{th} percentile comes from t_{n-1} (t-distribution with $n-1$ df).
- Question: why does an 80% interval require 90^{th} percentiles? (draw a picture)
- Next step: plug in $t_{n-1} = \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}}$.
- Algebra ...

Confidence interval for E(Y)

- Example: 80% confidence interval.
- Pr $(-90^{\text{th}}$ percentile $< t_{n-1} < 90^{\text{th}}$ percentile$) = 0.80$ where the 90^{th} percentile comes from t_{n-1} (t-distribution with $n-1$ df).
- Question: why does an 80% interval require 90^{th} percentiles? (draw a picture)
- Next step: plug in $t_{n-1} = \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}}$.
- Algebra . . .
- Pr $(m_Y - 90^{\text{th}}\text{percentile}(s_Y/\sqrt{n}) < \mathrm{E}(Y) < m_Y + 90^{\text{th}}\text{percentile}(s_Y/\sqrt{n})) = 0.80$.
- In other words, the 80% confidence interval can be written $m_Y \pm 90^{\text{th}}$ percentile $(s_Y/\sqrt{n})$.
- Question: what is the formula for a 90% interval?

© Iain Pardoe, 2006

Calculating confidence intervals

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% confidence interval for $\mathrm{E}(Y)$.

© Iain Pardoe, 2006

Calculating confidence intervals

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% confidence interval for $\mathrm{E}(Y)$.
- 90^{th} percentile of t_{29} is 1.311.

Calculating confidence intervals

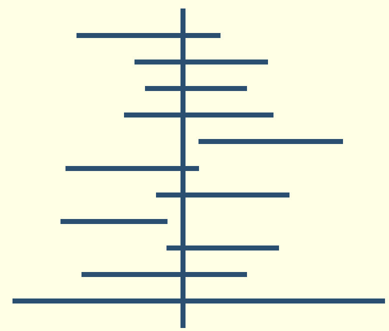
- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% confidence interval for $\mathrm{E}(Y)$.
- 90^{th} percentile of t_{29} is 1.311.
- $m_Y \pm 90^{\text{th}}$ percentile $(s_Y/\sqrt{n}) =$
  $278.603 \pm 1.311\,(53.8656/\sqrt{30}) =$
  $278.603 \pm 12.893 = (265.710, 291.496)$.

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% confidence interval for $\mathrm{E}(Y)$.
- 90^{th} percentile of t_{29} is 1.311.
- $m_Y \pm 90^{\text{th}}\text{ percentile}\,(s_Y/\sqrt{n}) =$
  $278.603 \pm 1.311\,(53.8656/\sqrt{30}) =$
  $278.603 \pm 12.893 = (265.710, 291.496)$.
- Calculate a 90% confidence interval for $\mathrm{E}(Y)$.

© Iain Pardoe, 2006

Confidence interval interpretation

- Loosely speaking: based on this dataset, we are 80% confident that the population mean home price is between $\$266{,}000$ and $\$291{,}000$.
- More precisely: If we were to take a large number of random samples of size 30 from a population of sale prices and calculate an 80% confidence interval for each, then 80% of those confidence intervals would contain the (unknown) population mean.
- E.g., 10 confidence intervals for samples from a population with $\mathrm{E}(Y)$ marked by the vertical line:



- 8 of the intervals contain $\mathrm{E}(Y)$, while 2 don't.

© Iain Pardoe, 2006

Hypothesis testing

- Confidence intervals tell us a range of plausible values for $\mathrm{E}(Y)$ with a specified confidence level.
- By contrast, hypothesis tests ask whether a particular value is plausible or not.
- Example: does a population mean of $\$255{,}000$ seem plausible given our sample of 30 home prices?

- Confidence intervals tell us a range of plausible values for $\mathrm{E}(Y)$ with a specified confidence level.
- By contrast, hypothesis tests ask whether a particular value is plausible or not.
- Example: does a population mean of $\$255{,}000$ seem plausible given our sample of 30 home prices?
  - Upper-tail test: can we reject the possibility that $\mathrm{E}(Y) = 255$ in favor of $\mathrm{E}(Y) > 255$?

© Iain Pardoe, 2006

Hypothesis testing

- Confidence intervals tell us a range of plausible values for $\mathrm{E}(Y)$ with a specified confidence level.
- By contrast, hypothesis tests ask whether a particular value is plausible or not.
- Example: does a population mean of $\$255{,}000$ seem plausible given our sample of 30 home prices?
  - Upper-tail test: can we reject the possibility that $\mathrm{E}(Y) = 255$ in favor of $\mathrm{E}(Y) > 255$?
  - Lower-tail test: can we reject the possibility that $\mathrm{E}(Y) = 255$ in favor of $\mathrm{E}(Y) < 255$?

© Iain Pardoe, 2006

Hypothesis testing

- Confidence intervals tell us a range of plausible values for $\mathrm{E}(Y)$ with a specified confidence level.
- By contrast, hypothesis tests ask whether a particular value is plausible or not.
- Example: does a population mean of $\$255{,}000$ seem plausible given our sample of 30 home prices?
  - Upper-tail test: can we reject the possibility that $\mathrm{E}(Y) = 255$ in favor of $\mathrm{E}(Y) > 255$?
  - Lower-tail test: can we reject the possibility that $\mathrm{E}(Y) = 255$ in favor of $\mathrm{E}(Y) < 255$?
  - Two-tail test: can we reject the possibility that $\mathrm{E}(Y) = 255$ in favor of $\mathrm{E}(Y) \neq 255$?

© Iain Pardoe, 2006

The rejection region method

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- If NH is true, then the sampling distribution of the t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}}$ is t_{n-1}.
- Recall t_{n-1} has a bell-shape centered at zero with most of it's area ($\approx 95\%$) between -2 and $+2$.
- So, if the value of the t-statistic is "not too far" from zero, we cannot reject NH.
- Conversely, a t-statistic much larger than zero favors AH (*larger* since this is an *upper*-tail test).
- How large does the t-statistic have to be before we reject NH in favor of AH?

The rejection region method

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- If NH is true, then the sampling distribution of the t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}}$ is t_{n-1}.
- Recall t_{n-1} has a bell-shape centered at zero with most of it's area ($\approx 95\%$) between -2 and $+2$.
- So, if the value of the t-statistic is "not too far" from zero, we cannot reject NH.
- Conversely, a t-statistic much larger than zero favors AH (*larger* since this is an *upper*-tail test).
- How large does the t-statistic have to be before we reject NH in favor of AH?
- *Significance level* (e.g., 5%) determines a *rejection region* beyond a *critical value* (e.g., 95^{th} percentile of t_{n-1}).

© Iain Pardoe, 2006

Rejection region example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.

© Iain Pardoe, 2006

Rejection region example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \dfrac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40.$

© Iain Pardoe, 2006

Rejection region example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \frac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \frac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.

Rejection region example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \frac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \frac{278.603-255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.
- Critical value is the 95^{th} percentile of t_{29} which is 1.699.

© Iain Pardoe, 2006

Rejection region example

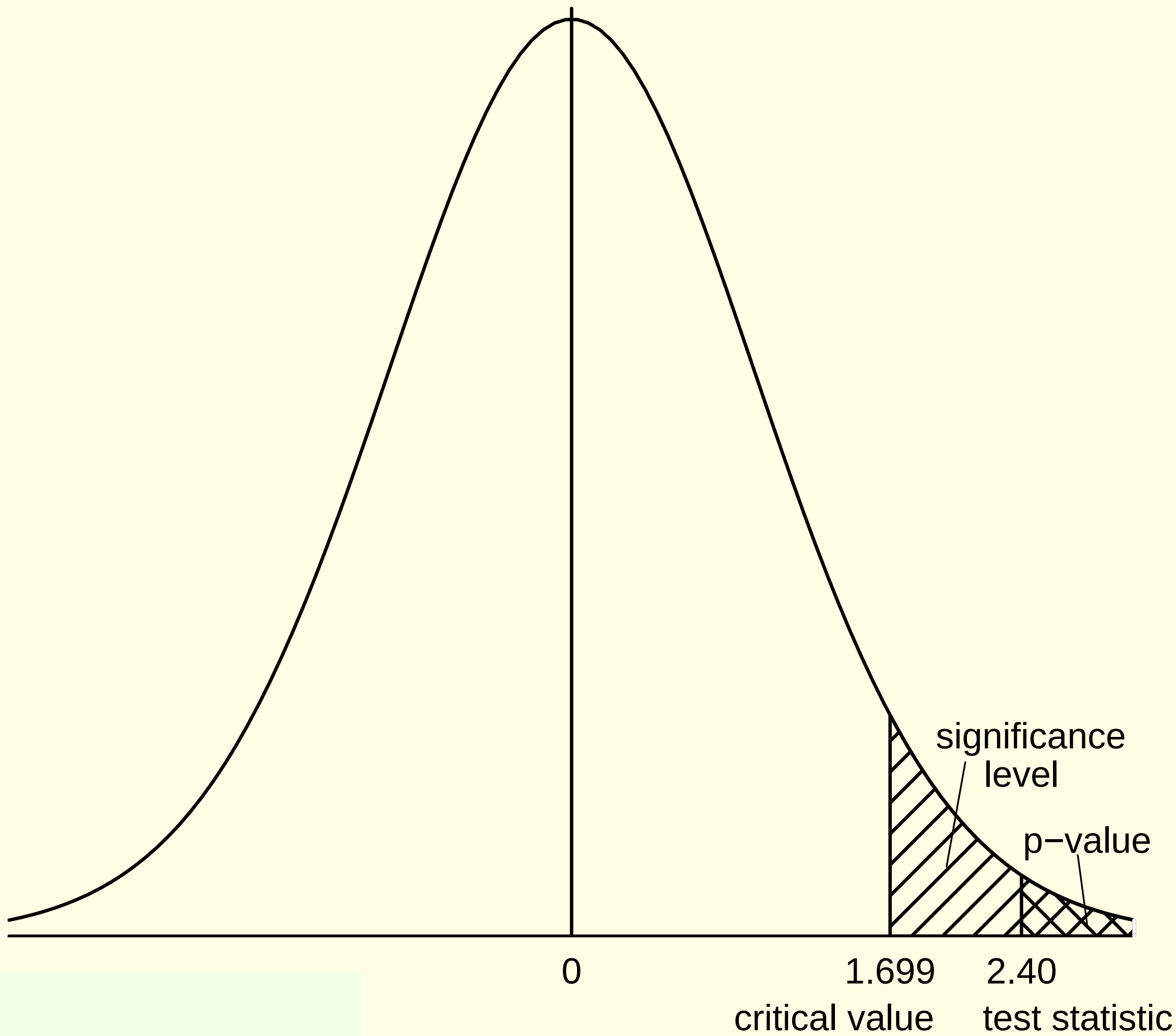
- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \dfrac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40.$
- Significance level $= 5\%$.
- Critical value is the 95^{th} percentile of t_{29} which is 1.699.
- Since t-statistic $(2.40) >$ critical value (1.699), we reject NH in favor of AH.

Rejection region example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \dfrac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.
- Critical value is the 95^{th} percentile of t_{29} which is 1.699.
- Since t-statistic $(2.40) >$ critical value (1.699), we reject NH in favor of AH.
- In other words, the sample data suggest that the population mean is greater than $\$255{,}000$ (at a 5% significance level).

Hypothesis test for home prices example

Test stat. is in rejection region, p-value $<$ signif. level:



The p-value method

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- If NH is true, then the sampling distribution of the t-statistic $= \frac{m_Y - \mathrm{E}(Y)}{s_Y / \sqrt{n}}$ is t_{n-1}.
- Recall t_{n-1} has a bell-shape centered at zero with most of it's area ($\approx 95\%$) between -2 and $+2$.
- So, if the upper-tail area beyond the t-statistic is "not too small," we cannot reject NH.
- Conversely, a very small upper tail-area favors AH.
- How small does the upper-tail area, called the *p-value*, have to be before we reject NH in favor of AH?

The p-value method

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- If NH is true, then the sampling distribution of the t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}}$ is t_{n-1}.
- Recall t_{n-1} has a bell-shape centered at zero with most of it's area ($\approx 95\%$) between -2 and $+2$.
- So, if the upper-tail area beyond the t-statistic is "not too small," we cannot reject NH.
- Conversely, a very small upper tail-area favors AH.
- How small does the upper-tail area, called the *p-value*, have to be before we reject NH in favor of AH?
- Smaller than the significance level (e.g., 5%).

© Iain Pardoe, 2006

A p-value example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.

A p-value example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \dfrac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40.$

© Iain Pardoe, 2006

A p-value example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \dfrac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.

A p-value example

- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \frac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \frac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.
- Since the t-statistic (2.40) is between 2.045 and 2.462, the p-value must be between 0.01 and 0.025.

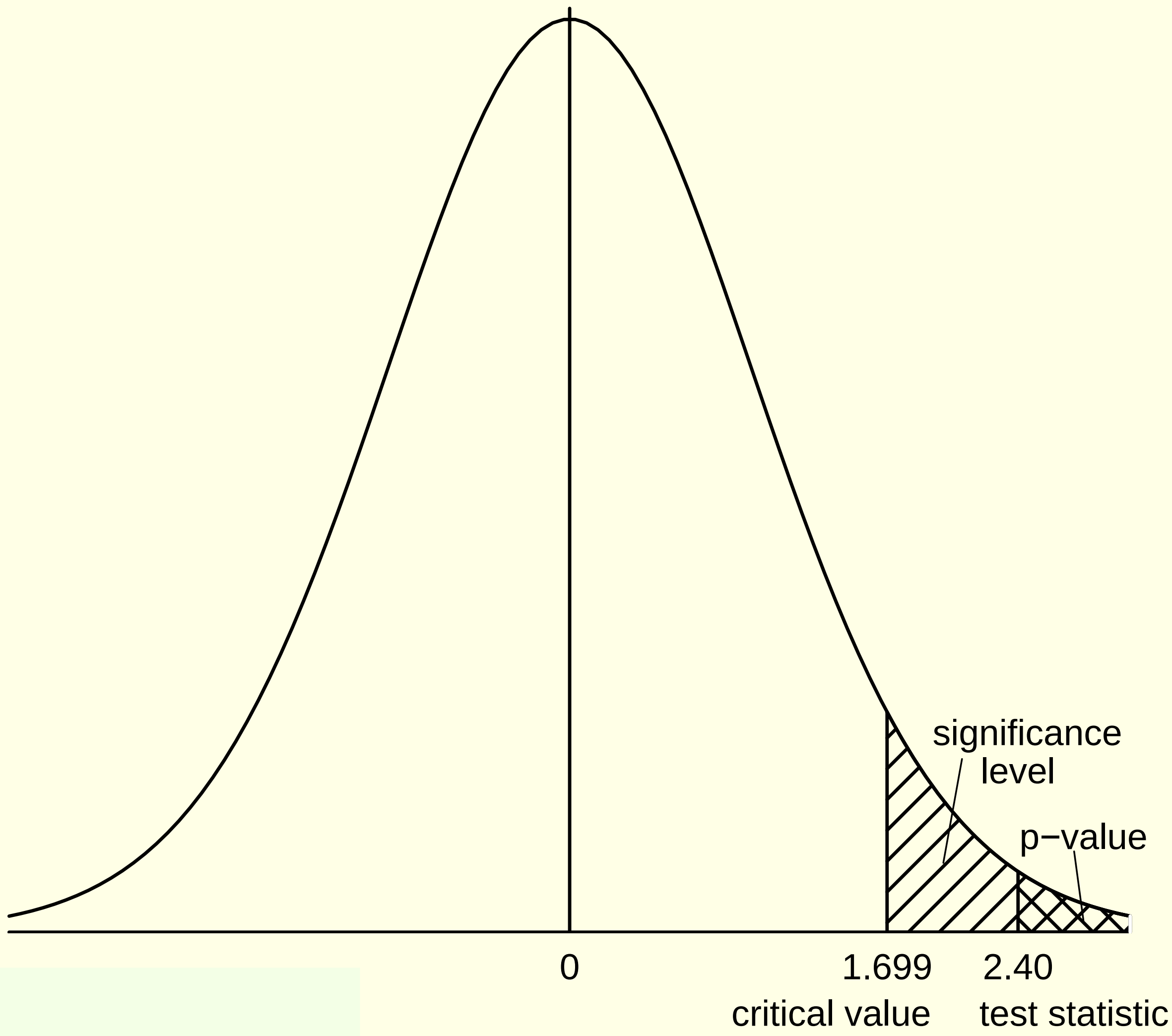
- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \frac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \frac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.
- Since the t-statistic (2.40) is between 2.045 and 2.462, the p-value must be between 0.01 and 0.025.
- Since p-value $<$ significance level, we reject NH in favor of AH.

© Iain Pardoe, 2006

A p-value example

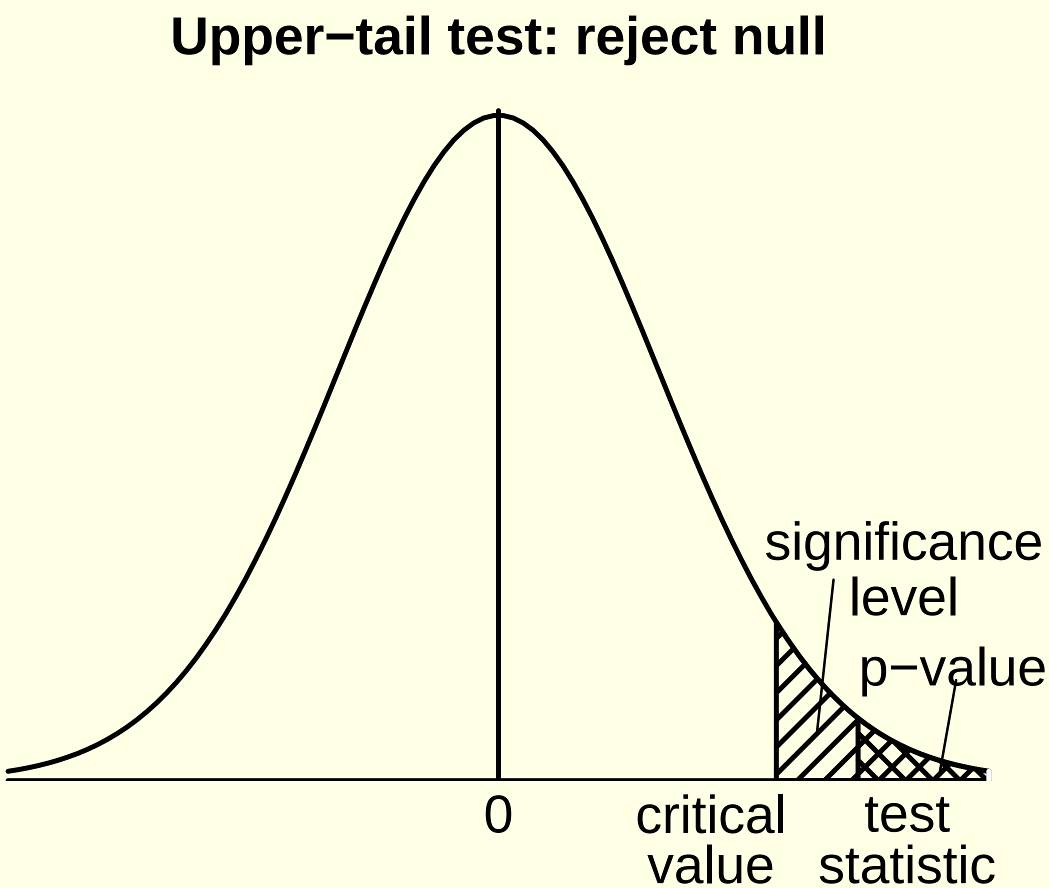
- Upper-tail test: *null hypothesis* NH: $\mathrm{E}(Y) = 255$ versus *alternative hypothesis* AH: $\mathrm{E}(Y) > 255$.
- t-statistic $= \dfrac{m_Y - \mathrm{E}(Y)}{s_Y/\sqrt{n}} = \dfrac{278.603 - 255}{53.8656/\sqrt{30}} = 2.40$.
- Significance level $= 5\%$.
- Since the t-statistic (2.40) is between 2.045 and 2.462, the p-value must be between 0.01 and 0.025.
- Since p-value $<$ significance level, we reject NH in favor of AH.
- In other words, the sample data suggest that the population mean is greater than $\$255{,}000$ (at a 5% significance level).

Test stat. is in rejection region, p-value < signif. level:

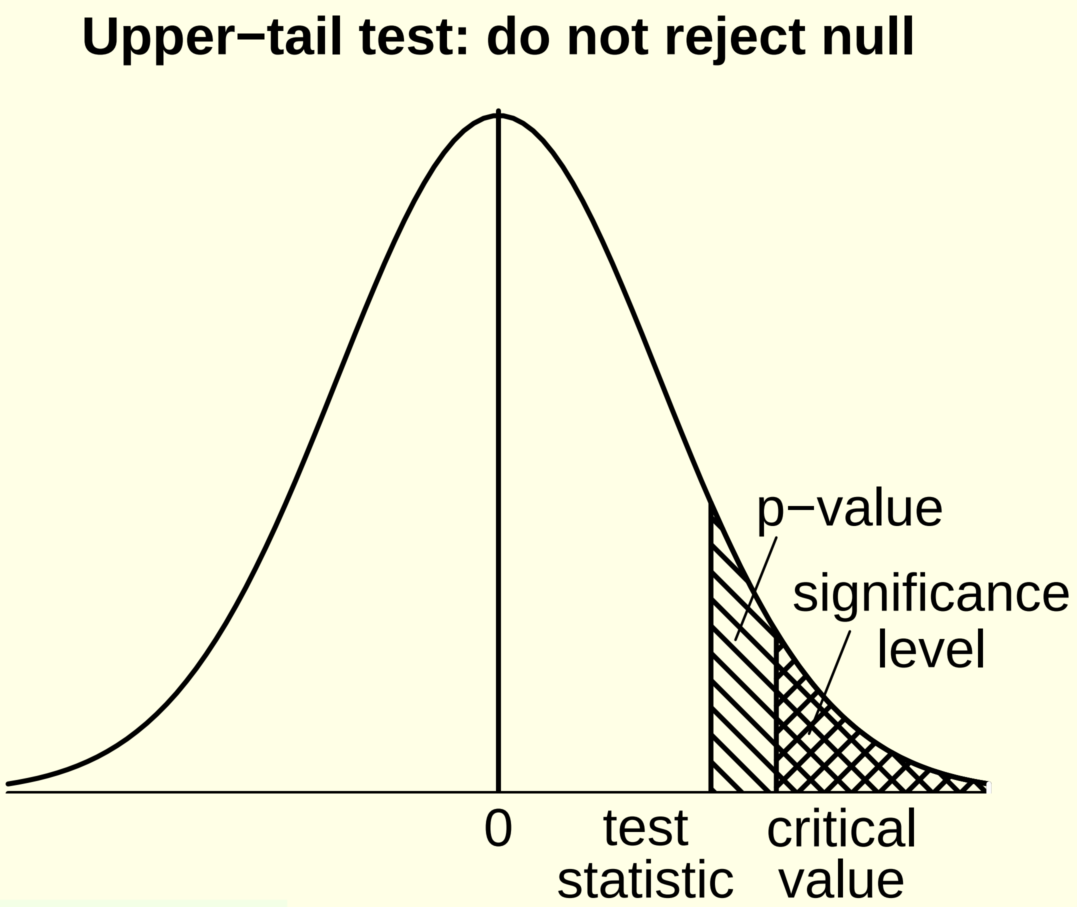


One-tail hypothesis tests

E.g., NH: $\mathrm{E}(Y)=255$ vs. AH: $\mathrm{E}(Y)>255$ (@5%).
Test stat. is in rejection region, p-value $<$ signif. level:



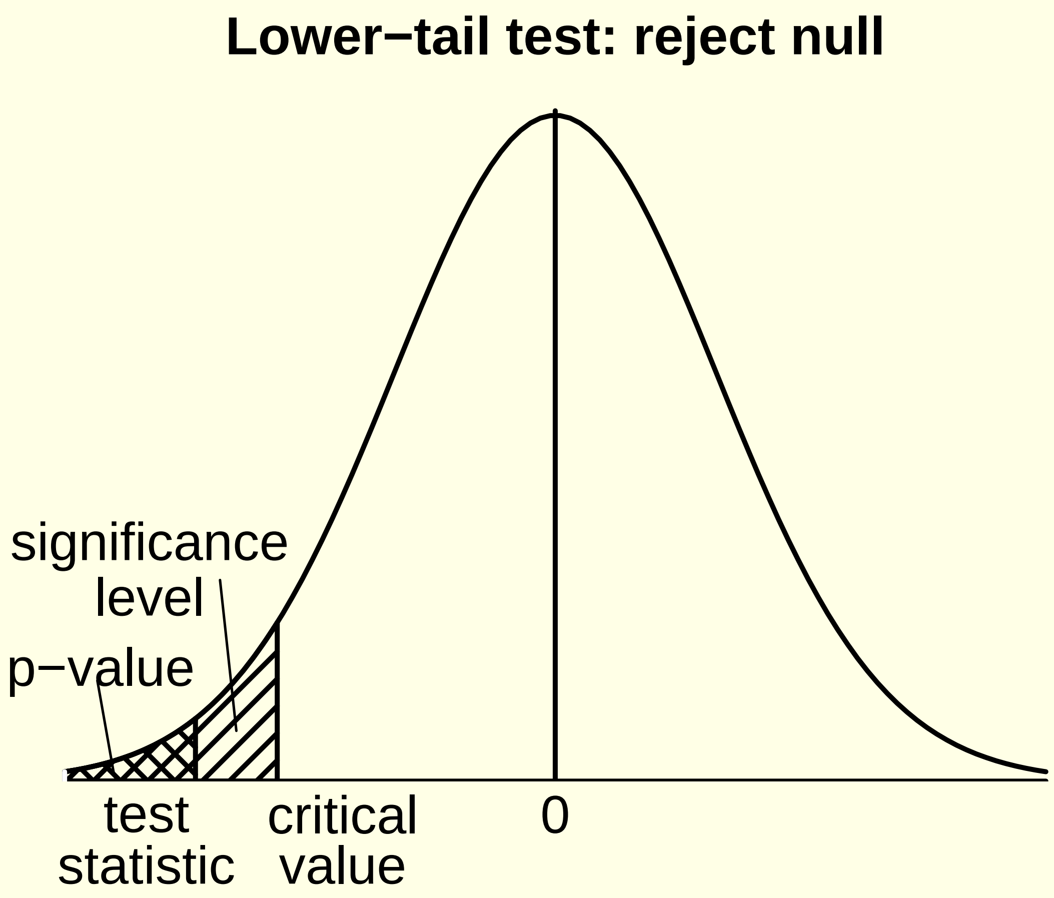
E.g., NH: $\mathrm{E}(Y) = 265$ vs. AH: $\mathrm{E}(Y) > 265$ (@5%).
Test stat. not in rejection region, p-value $>$ signif. level:



© Iain Pardoe, 2006

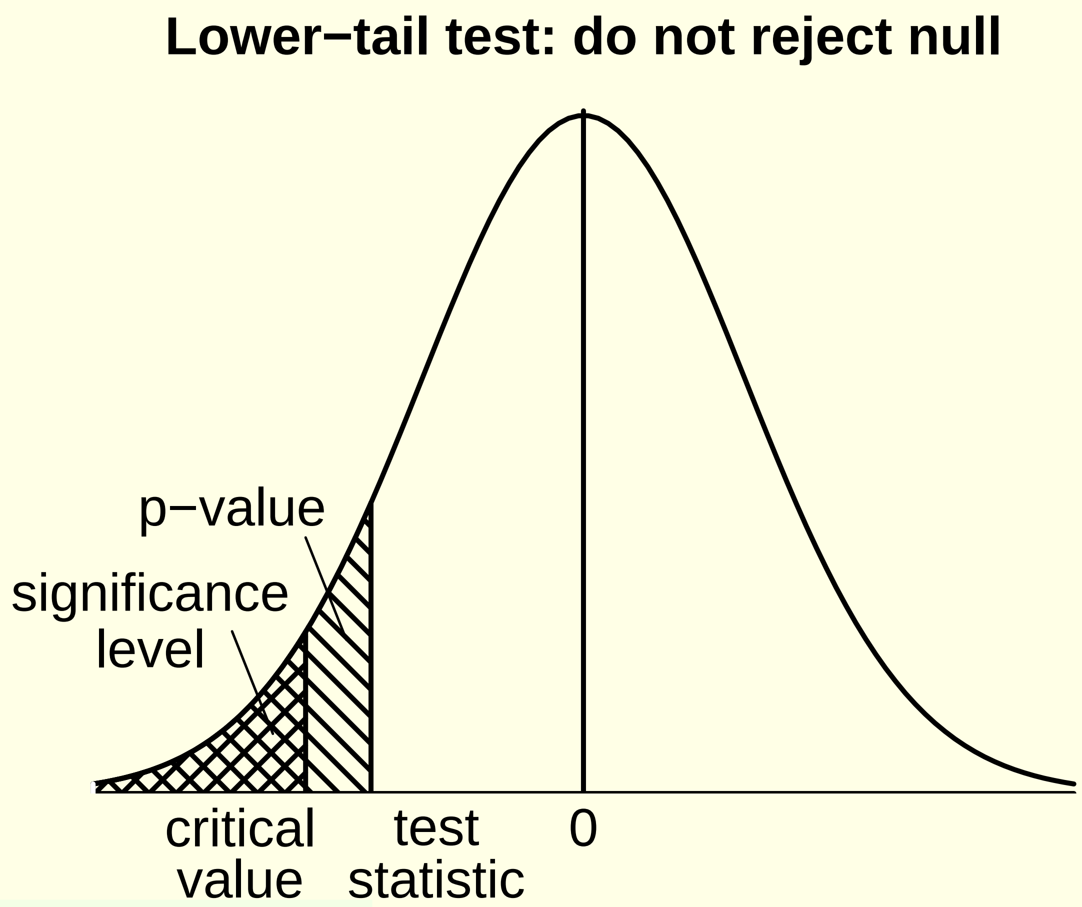
One-tail hypothesis tests

E.g., NH: $\mathrm{E}(Y)=300$ vs. AH: $\mathrm{E}(Y)<300$ (@5%).
Test stat. is in rejection region, p-value $<$ signif. level:



One-tail hypothesis tests

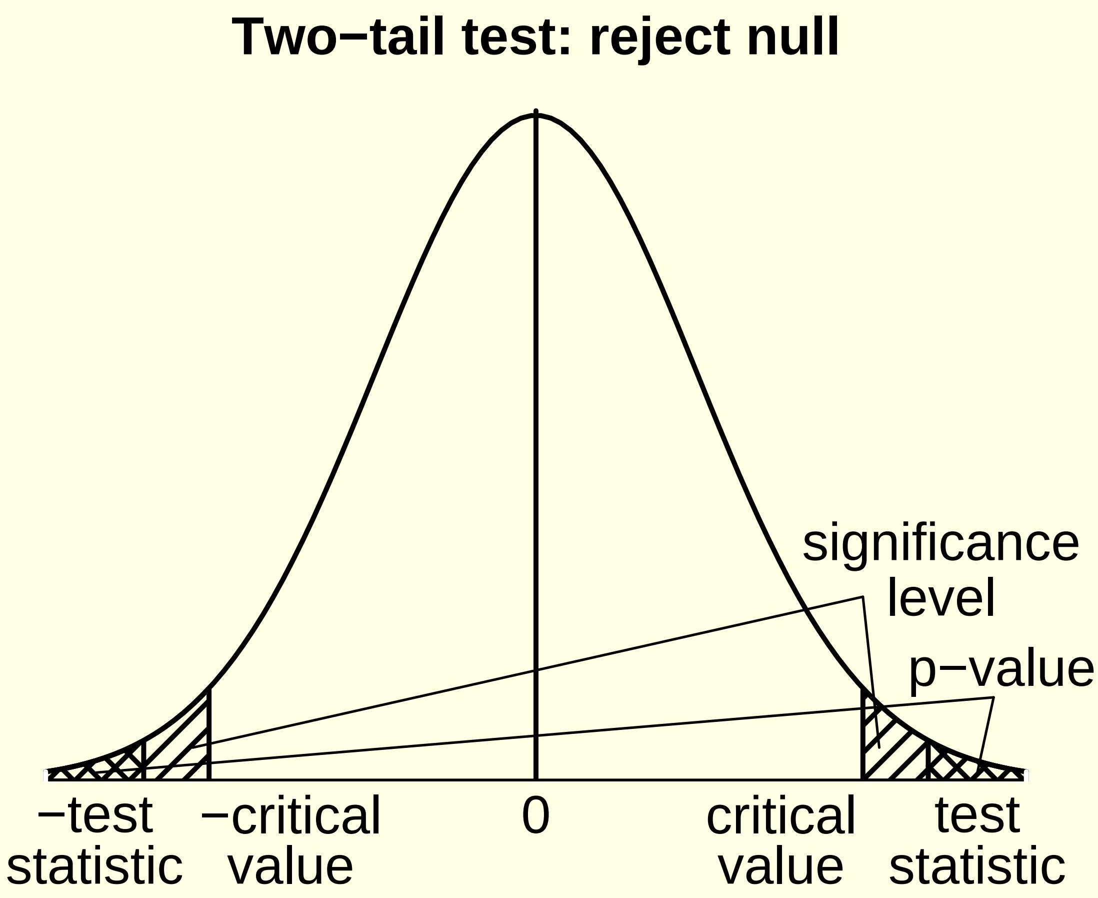
E.g., NH: $\mathrm{E}(Y) = 290$ vs. AH: $\mathrm{E}(Y) < 290$ (@5%).
Test stat. not in rejection region, p-value $>$ signif. level:



© Iain Pardoe, 2006

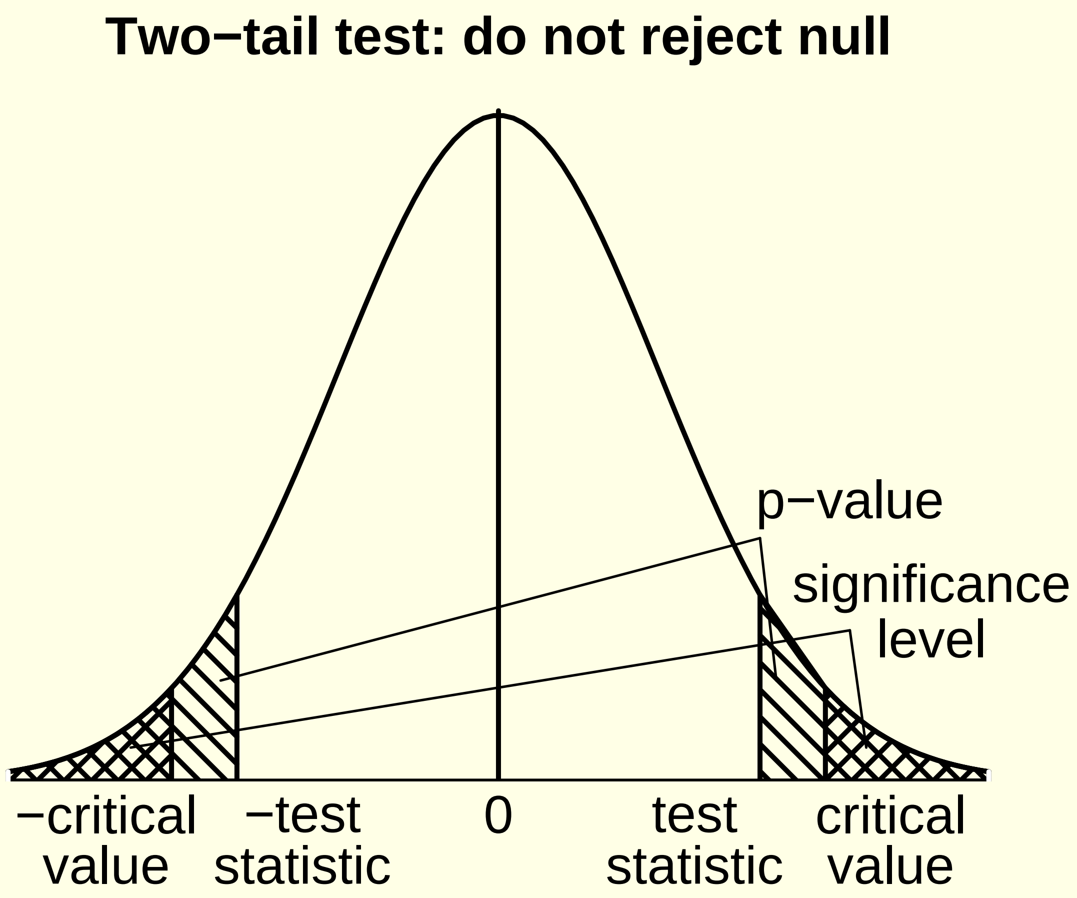
Two-tail hypothesis tests

E.g., NH: $\mathrm{E}(Y) = 255$ vs. AH: $\mathrm{E}(Y) \neq 255$ (@5%).
Test stat. is in rejection region, p-value $<$ signif. level:



Two-tail hypothesis tests

E.g., NH: $\mathrm{E}(Y) = 265$ vs. AH: $\mathrm{E}(Y) \neq 265$ (@5%).
Test stat. not in rejection region, p-value $>$ signif. level:



Hypothesis test errors

- Four possible hypothesis test outcomes:

		Decision	
		Do not reject NH in favor of AH	Reject NH in favor of AH
Reality	NH true	correct decision	type 1 error
	NH false	type 2 error	correct decision

- Pr(type 1 error) $=$ signif. level; analyst selects this.
- But, setting it too low can increase the chance of a type 2 error occuring.
- Trade-off: set signif. level at 5% (sometimes 1% or 10%); reduce chance of type 2 error by having n as large as possible, using sound statistical methods.
- Also, use hypothesis tests judiciously and always keep in mind the possibility of making these errors.

- New problem: predict an individual Y-value picked at random from the population.
- Is this easier or more difficult than estimating the population mean?

© Iain Pardoe, 2006

Prediction intervals

- New problem: predict an individual Y-value picked at random from the population.
- Is this easier or more difficult than estimating the population mean?
- More difficult: imagine predicting the sale price of a new home on the market (versus estimating the average sale price of homes in this market)—which answer would you be less certain about?

Prediction intervals

- New problem: predict an individual Y-value picked at random from the population.
- Is this easier or more difficult than estimating the population mean?
- More difficult: imagine predicting the sale price of a new home on the market (versus estimating the average sale price of homes in this market)—which answer would you be less certain about?
- Approach: calculate a *prediction interval*—like a confidence interval but with a larger range of uncertainty.
- Confidence interval: point estimate $\pm$ estimation uncertainty.
- Prediction interval: point estimate $\pm$ prediction uncertainty.

© Iain Pardoe, 2006

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*?

Prediction error

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.

© Iain Pardoe, 2006

Prediction error

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.
- Variance of estimation error $(\mathrm{E}(Y) - m_Y)$: s_Y^2/n.

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.
- Variance of estimation error $(\mathrm{E}(Y) - m_Y)$: s_Y^2/n.
- Var. of random error (e^*): s_Y^2.

© Iain Pardoe, 2006

Prediction error

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.
- Variance of estimation error $(\mathrm{E}(Y) - m_Y)$: s_Y^2/n.
- Var. of random error (e^*): s_Y^2.
- Var. of prediction error $(Y^* - m_Y)$: $s_Y^2(1+1/n)$.

© Iain Pardoe, 2006

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.
- Variance of estimation error $(\mathrm{E}(Y) - m_Y)$: s_Y^2/n.
- Var. of random error (e^*): s_Y^2.
- Var. of prediction error $(Y^* - m_Y)$: $s_Y^2(1+1/n)$.
- Confidence interval for $\mathrm{E}(Y)$:
  $m_Y \pm$ t-percentile$(s_Y/\sqrt{n})$.

Prediction error

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.
- Variance of estimation error $(\mathrm{E}(Y) - m_Y)$: s_Y^2/n.
- Var. of random error (e^*): s_Y^2.
- Var. of prediction error $(Y^* - m_Y)$: $s_Y^2(1+1/n)$.
- Confidence interval for $\mathrm{E}(Y)$:
  $m_Y \pm$ t-percentile$(s_Y/\sqrt{n})$.
- Prediction interval for Y^*:
  $m_Y \pm$ t-percentile $\left(s_Y\sqrt{1+1/n}\right)$.

© Iain Pardoe, 2006

Prediction error

- Model: $Y_i = \mathrm{E}(Y) + e_i \quad (i = 1, \ldots, n)$.
- Y-value to be predicted: $Y^* = \mathrm{E}(Y) + e^*$.
- Point estimate of Y^*? Sample mean, m_Y.
- Prediction error: $Y^* - m_Y = (\mathrm{E}(Y) - m_Y) + e^*$.
- Variance of estimation error $(\mathrm{E}(Y) - m_Y)$: s_Y^2/n.
- Var. of random error (e^*): s_Y^2.
- Var. of prediction error $(Y^* - m_Y)$: $s_Y^2(1+1/n)$.
- Confidence interval for $\mathrm{E}(Y)$:
  $m_Y \pm$ t-percentile$(s_Y/\sqrt{n})$.
- Prediction interval for Y^*:
  $m_Y \pm$ t-percentile $\left(s_Y\sqrt{1+1/n}\right)$.
- Which is wider?

© Iain Pardoe, 2006

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% prediction interval for Y.

© Iain Pardoe, 2006

Calculating prediction intervals

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% prediction interval for Y.
- 90^{th} percentile of t_{29} is 1.311.

Calculating prediction intervals

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% prediction interval for Y.
- 90^{th} percentile of t_{29} is 1.311.
- $m_Y \pm 90^{\text{th}}$ percentile $\left(s_Y\sqrt{1+1/n}\right) =$
  $278.603 \pm 1.311 \left(53.8656\sqrt{1+1/30}\right) =$
  $278.603 \pm 71.785 = (206.818, 350.388)$.

Calculating prediction intervals

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% prediction interval for Y.
- 90^{th} percentile of t_{29} is 1.311.
- $m_Y \pm 90^{\text{th}}$ percentile $\left(s_Y \sqrt{1+1/n}\right) =$ $278.603 \pm 1.311 \left(53.8656\sqrt{1+1/30}\right) =$ $278.603 \pm 71.785 = (206.818, 350.388)$.
- We're 80% confident the sale price of an individual, randomly selected home in this market will be between $\$207{,}000$ and $\$350{,}000$.

Calculating prediction intervals

- Example: home prices $Y_1, \ldots, Y_{30}$.
- Sample mean, m_Y, is 278.603.
- Sample standard deviation, s_Y, is 53.8656.
- Calculate an 80% prediction interval for Y.
- 90^{th} percentile of t_{29} is 1.311.
- $m_Y \pm 90^{\text{th}}$ percentile $\left(s_Y\sqrt{1+1/n}\right) =$
  $278.603 \pm 1.311 \left(53.8656\sqrt{1+1/30}\right) =$
  $278.603 \pm 71.785 = (206.818, 350.388)$.
- We're 80% confident the sale price of an individual, randomly selected home in this market will be between $\$207{,}000$ and $\$350{,}000$.
- Calculate a 90% prediction interval for Y.